

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Wirtschaftswissenschaften

Lehrstuhl für Wirtschaftspädagogik

**Feedback zu Concept Maps im
betriebswirtschaftlichen Planspielunterricht
– eine empirische Untersuchung**

Dissertationsschrift

zur Erlangung des akademischen Grades

„Doctor rerum politicarum“

Vorgelegt von: Dipl.-Hdl. Jeannine Ryssel, geb. am 19.03.1979 in Görlitz

Vorgelegt am: 17.09.2018

Verteidigt am: 12.03.2019

Gutachter: Prof. Dr. Bärbel Fürstenau

Prof. Dr. Susanne Strahinger

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abkürzungsverzeichnis	VII
Abbildungsverzeichnis	IX
Tabellenverzeichnis	XIII
1 Einleitung.....	1
2 Feedback in Lehr-Lernprozessen.....	9
2.1 Theoretische Grundlagen zum Feedback in Lehr-Lernprozessen	9
2.1.1 Bestimmung des Begriffs Feedback in Lehr-Lernprozessen.....	9
2.1.2 Funktionen und Ebenen von Feedback in Lehr-Lernprozessen	12
2.1.3 Klassifikation von Feedback in Lehr-Lernprozessen	16
2.1.3.1 Klassifikation von Feedback in Lehr-Lernprozessen anhand ausgewählter Kriterien.....	16
2.1.3.2 Klassifikation von Feedback in Lehr-Lernprozessen anhand des Kriteriums der Komplexität der Feedbacknachricht nach einfachem und elaboriertem Feedback.....	19
2.1.4 Feedbackmodelle zur Feedbackrezeption und/oder -gestaltung.....	22
2.1.4.1 Makromodelle von Kulhavy und Stock (1989) sowie von Bangert- Drowns u. a. (1991).....	22
2.1.4.2 Mikromodell von Narciss (2006)	25
2.1.4.3 Rahmenmodell von Strijbos und Müller (2014).....	30
2.1.4.4 Erweitertes Rahmenmodell zur Gestaltung und Verarbeitung von Feedback.....	36
2.2 Empirische Befunde zur Lernwirksamkeit des Feedbacks	40
2.2.1 Überblick über Metanalysen zur Lernwirksamkeit von Feedback	40
2.2.2 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht	46
2.2.3 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Feedback-Form	57

2.2.4	Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von (inter-) personalen Faktoren.....	61
2.2.5	Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von situativen Faktoren.....	69
2.2.6	Zusammenfassung der empirischen Befunde zur Lernwirksamkeit des Feedbacks	75
3	Concept Maps in Lehr-Lernprozessen	83
3.1	Theoretische Grundlagen zu Concept Maps in Lehr-Lernprozessen	83
3.1.1	Bestimmung des Begriffs und Klassifikation von Concept Maps.....	83
3.1.2	Anwendungsgebiete von Concept Maps in Lehr-Lernprozessen	89
3.1.3	Theoretische Ansätze zur Lernwirksamkeit von Concept Maps	94
3.2	Empirische Befunde zur Lernwirksamkeit von Concept Maps	102
3.2.1	Überblick über Metanalysen zur Lernwirksamkeit von Concept Maps	102
3.2.2	Studien zum Einfluss von Vorwissen, verbalen Voraussetzungen, kognitiven Fähigkeiten und dem Training der Methode auf die Lernwirksamkeit von Concept Maps.....	105
3.2.3	Studien zur Lernwirksamkeit von Feedback zu Concept Maps	117
4	Empirische Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps – Konzeption, Erhebung und Auswertung	125
4.1	Gestaltung der Untersuchung	125
4.1.1	Fragestellungen und Hypothesen.....	125
4.1.2	Stichprobe und Untersuchungsdesign.....	140
4.2	Gestaltung der Lernmaterialien und Interventionsmaßnahmen.....	144
4.2.1	Planspielunterricht mit Easy Business TM	144
4.2.2	Gestaltung des Concept Mapping.....	149
4.2.3	Gestaltung des Feedbacks.....	151
4.3	Erhebungsinstrumente	158
4.3.1	Wissenstest als Paralleltest	158
4.3.2	Sprachfähigkeitstest.....	163
4.3.3	Fragebogen zur Einschätzung des Umgangs mit Concept Mapping, der Nützlichkeit des Feedbacks sowie des Umgangs mit Fehlern.....	164
4.4	Auswertung der Wissenstests und des Sprachfähigkeitstests	169

4.4.1	Auswertung der Wissenstests	169
4.4.2	Auswertung des Sprachfähigkeitstests	172
4.5	Auswertung der Concept Maps	174
4.5.1	Kategoriale und strukturelle Inhaltsanalyse.....	174
4.5.2	Qualitative Bewertung der Concept Maps.....	184
4.6	Auswertung des Fragebogens zur Einschätzung des Umgangs mit Concept Mapping, des Umgangs mit Fehlern sowie der Nützlichkeit des Feedbacks... 193	
4.6.1	Auswertung des Teil-Fragebogens zur Einschätzung des Concept Mapping.....	193
4.6.2	Auswertung des Teil-Fragebogens zur Einschätzung der Nützlichkeit des Feedbacks	200
4.6.3	Auswertung des Teil-Fragebogens zur Einschätzung des Umgangs mit Fehlern	203
5	Empirische Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps – Analysen und Ergebnisse.....	205
5.1	Analysen zum Einfluss der Art des Feedbacks.....	205
5.1.1	Analysen zum Einfluss der Art des Feedbacks auf den Wissenserwerb ...	205
5.1.1.1	Analyse der Lernwirksamkeit des Concept Mapping	205
5.1.1.2	Gruppenunterschiede hinsichtlich des Wissenserwerbs	208
5.1.1.3	Einfluss von personalen und weiteren Faktoren auf den Wissenserwerb.....	213
5.1.1.3.1	Einfluss von Vorwissen, Sprachfähigkeit, der Einschätzung zum Umgang mit Concept Mapping und zum Umgang mit Fehlern auf den Wissenserwerb	213
5.1.1.3.2	Einfluss von niedrigem bzw. hohem Vorwissen auf den Wissenserwerb.....	218
5.1.1.3.3	Einfluss von niedriger bzw. hoher Sprachfähigkeit auf den Wissenserwerb.....	224
5.1.1.4	Einfluss des Lernzielniveaus auf den Wissenserwerb.....	230
5.1.1.4.1	Gruppenunterschiede hinsichtlich des Wissenserwerbs für einfache Aufgaben.....	230

5.1.1.4.2	Gruppenunterschiede hinsichtlich des Wissenserwerbs für komplexe Aufgaben.....	235
5.1.1.5	Einfluss von personalen und weiteren Faktoren auf den Wissenserwerb einfacher bzw. komplexer Aufgaben	239
5.1.1.6	Zusammenfassung der Ergebnisse und Bewertung der Hypothesen hinsichtlich des Wissenserwerbs	241
5.1.2	Einfluss der Art des Feedbacks auf die Qualität der Maps.....	251
5.1.2.1	Gruppenunterschiede hinsichtlich der Map-Qualität.....	251
5.1.2.2	Einfluss des Vorwissens, der Sprachfähigkeit, der Einschätzung zum Umgang mit Concept Mapping und zum Umgang mit Fehlern auf die Qualität der Maps	256
5.1.2.3	Zusammenfassung der Ergebnisse und Bewertung der Hypothesen hinsichtlich der Qualität der Maps.....	263
5.1.3	Einfluss der Qualität der Maps auf den Wissenserwerb.....	268
5.1.3.1	Einfluss der Qualität des 1., 2. und 3. Maps auf den 1. bzw. 2. Nachtest	268
5.1.3.2	Einfluss von niedriger bzw. hoher Qualität des 2. Maps auf den Wissenserwerb.....	269
5.1.3.3	Einfluss von niedriger bzw. hoher Qualität des 3. Maps auf den Wissenserwerb.....	276
5.1.3.4	Einfluss der Qualität des 1., 2. und 3. Maps auf den Wissenserwerb einfacher und komplexer Aufgaben	282
5.1.3.5	Zusammenfassung der Ergebnisse und Bewertung der Hypothesen zum Einfluss der Qualität der Maps auf den Wissenserwerb.....	286
5.2	Analysen zum Umgang mit dem Feedback sowie Analyse der Fehler in den Concept Maps.....	292
5.2.1	Analysen zum Umgang mit dem Feedback.....	292
5.2.2	Analysen der Fehler in den Concept Maps.....	298
5.2.3	Zusammenfassung der Ergebnisse zum Umgang mit dem Feedback und der Fehleranalyse	304

5.3 Interpretation der Ergebnisse	307
6 Schlussbetrachtung	323
6.1 Zusammenfassung	323
6.2 Kritische Würdigung, Optimierungsvorschläge und Ausblick.....	337
Literaturverzeichnis	349

Abkürzungsverzeichnis

ANCOVA	Kovarianzanalyse
ANOVA	Univariate Varianzanalyse
AUC	Answer-until-correct
AV	Abhängige Variable
CM	Concept Mapping (Bezeichnung für Versuchsgruppe)
EG	Experimentalgruppe
EiFB	Einfaches Feedback (Bezeichnung für Versuchsgruppe)
EiFBN	Einschätzung der Nützlichkeit des einfachen Feedbacks
ElFB	Elaboriertes Feedback (Bezeichnung für Versuchsgruppe)
ElFBgew	Gewünschtes elaboriertes Feedback
ElFBN	Einschätzung der Nützlichkeit des elaborierten Feedbacks
FB	Feedback (Bezeichnung für Versuchsgruppe)
FBgew	Gewünschtes Feedback
GuV	Gewinn- und Verlustrechnung
H	Eingeschätzte Handhabbarkeit hinsichtlich des Concept Mapping
H_nh	Niedrige bzw. hohe eingeschätzte Handhabbarkeit hinsichtlich des Concept Mapping (Variablenbezeichnung)
IN	Eingeschätzte Interessantheit und Nützlichkeit hinsichtlich des Concept Mapping
IN_nh	Niedrige bzw. hohe eingeschätzte Interessantheit und Nützlichkeit hinsichtlich des Concept Mapping (Variablenbezeichnung)
ITF	Informatives tutorielles Feedback
ITFL	Interactive-Two-Feedback-Loop
Kein FB	Kein Feedback (Bezeichnung für Versuchsgruppe)
KC	Knowledge about concepts
KCR	Knowledge of correct response
KG	Kontrollgruppe
KH	Knowledge on how to proceed
KM	Knowledge about mistakes
KMO	Kaiser-Meyer-Olkin
KMC	Knowledge on meta-cognition
KP	Knowledge of performance
KR	Knowledge of result
KTC	Knowledge on task constraints

MANOVA	Multivariate Varianzanalyse
MSA	Measure of sampling adequacy (KMO-Test für einzelne Variablen)
MTF	Multiple-try-feedback
NT	Nachtest
OHP	Overheadprojektor
ScorM	Scoringwerte des Mappings
SF	Sprachfähigkeit
SF_nh	Niedrige bzw. hohe Sprachfähigkeit
STF	Single-try-feedback
ÜG	Übereinstimmungsgrad
ÜG _{diff}	Differenzierter Übereinstimmungsgrad/Scoringwert
UV	Unabhängige Variable
UF	Umgang mit Fehlern
VT	Vortest
VW	Vorwissen
VW_nh	niedriges bzw. hohes Vorwissen (Variablenbezeichnung)
WIT-2	Wilde-Intelligenz-Test-2

Abbildungsverzeichnis

Abbildung 1: Zusammenhang der Klassifikationsvorschläge zum elaborierten Feedback.....	21
Abbildung 2: Mindful-Processing-Modell	25
Abbildung 3: Interactive-Two-Feedback-Loop-Model	28
Abbildung 4: Übersicht über die Gestaltungsprinzipien des ITFL-Modells	30
Abbildung 5: Interaktionales Rahmenmodell zur Gestaltung und Verarbeitung von Feedback.....	33
Abbildung 6: Erweitertes Rahmenmodell zur Gestaltung und Verarbeitung von Feedback.....	39
Abbildung 7: Concept Map über Concept Maps	89
Abbildung 8: Aufbau und Ablauf der empirischen Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps.....	143
Abbildung 9: Handlungsorientiertes Lernen	145
Abbildung 10: Boxplots für den Vortest, 1. Nachtest und 2. Nachtest (ohne Aufgaben 4b und 10b).....	172
Abbildung 11: Histogramm zur Verteilung der Mittelwerte zum „Sprachlichen Denken“ entsprechend ihrer Häufigkeiten	174
Abbildung 12: Relative Häufigkeiten der Verarbeitungsstrategien des einfachen Feedbacks	184
Abbildung 13: Ein Beispiel zur Bestimmung des absoluten Distanzmaßes.....	186
Abbildung 14: Boxplots der Scoringwerte für das Mapping1, Mapping 2 und Mapping 3.....	192
Abbildung 15: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die KG und die Gruppe CM über drei Messzeitunkte.....	208
Abbildung 16: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen kein FB, EiFB und EIFB über drei Messzeitunkte	212
Abbildung 17: Graphische Darstellung der (adjustierten) Mittelwerte Standardabweichungen für den 2. Nachtest	217
Abbildung 18: Graphische Darstellung der Gruppenmittelwerte des 1. Nachtests in Abhängigkeit vom Vorwissensniveau	220
Abbildung 19: Graphische Darstellung der Gruppenmittelwerte des 2. Nachtests in Abhängigkeit vom Vorwissen	221

Abbildung 20: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitunkte in Abhängigkeit vom <i>VW</i> -Niveau.....	223
Abbildung 21: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit von der <i>SF</i>	226
Abbildung 22: Graphische Darstellung der Gruppenmittelwerte des 2. Nachttests in Abhängigkeit von der <i>SF</i>	227
Abbildung 23: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitpunkte in Abhängigkeit vom Niveau der <i>SF</i>	230
Abbildung 24: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitunkte für <i>einfache Aufgaben</i>	233
Abbildung 25: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitunkte für <i>komplexe Aufgaben</i>	237
Abbildung 26: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitunkte für <i>komplexe Aufgaben</i> in Abhängigkeit vom Niveau der <i>SF</i>	241
Abbildung 27: Graphische Darstellung der Scoringwerte für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitpunkte.....	256
Abbildung 28: Graphische Darstellung der Werte für <i>ScorM2</i> in Abhängigkeit vom Niveau für <i>IN_nh</i>	261
Abbildung 29: Graphische Darstellung der Werte für <i>ScorM2</i> in Abhängigkeit vom Niveau für <i>H_nh</i>	263
Abbildung 30: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit vom Niveau für <i>ScorM2</i>	272
Abbildung 31: Graphische Darstellung der Gruppenmittelwerte des 2. Nachttests in Abhängigkeit vom Niveau für <i>ScorM2</i>	273
Abbildung 32: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>ElFB</i> über drei Messzeitunkte in Abhängigkeit vom Niveau für <i>ScorM2</i>	275
Abbildung 33: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit vom Niveau für <i>ScorM3</i>	278

Abbildung 34: Graphische Darstellung der Gruppenmittelwerte des 2. Nachtests in Abhängigkeit vom Niveau für ScorM3	279
Abbildung 35: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen <i>kein FB</i> , <i>EiFB</i> und <i>EIFB</i> über drei Messzeitpunkte in Abhängigkeit vom Niveau für ScorM3	282
Abbildung 36: Graphische Darstellung der Scoringwerte für die Gruppen <i>EiFB</i> und <i>EIFB</i> über sechs Messzeitpunkte.....	295
Abbildung 37: Relative Häufigkeiten der Verarbeitungsstrategien für Map 1 für das einfache und elaborierte Feedback	296
Abbildung 38: Relative Häufigkeiten der Verarbeitungsstrategien für Map 2 für das einfache und elaborierte Feedback	297
Abbildung 39: Relative Häufigkeiten der Verarbeitungsstrategien für Map 3 für das einfache und elaborierte Feedback	297
Abbildung 40: Ergebnisse der Analysen zur Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs	328
Abbildung 41: Ergebnisse der Analysen zur Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps.....	331
Abbildung 42: Ergebnisse der Analysen zum Einfluss der Mapqualität auf den Wissenserwerb.....	334

Tabellenverzeichnis

Tabelle 1:	Differenzierung der kognitiven Feedback-Funktionen nach der Antwortrichtigkeit.....	14
Tabelle 2:	Klassifikation von Feedback in Lehr-Lernprozessen anhand ausgewählter Kriterien.....	18
Tabelle 3:	Klassifikationsvorschläge für einfaches und elaboriertes Feedback	19
Tabelle 4:	Mittlere Effektstärken verschiedener Moderatorvariablen auf die Feedbackwirkung.....	43
Tabelle 5:	Mittlere Effektstärken signifikanter Moderatorvariablen auf die Feedbackwirkung.....	45
Tabelle 6:	Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht	56
Tabelle 7:	Kognitive und Metakognitive Prompts.....	112
Tabelle 8:	Zuordnung der Hypothesen zu den Forschungsfragen	139
Tabelle 9:	Einordnung der Wissenstestaufgaben in die Lernzieltaxonomie nach Anderson und Krathwohl (2001) für den Vortest und 1. Nachtest.....	161
Tabelle 10:	Aufbau des WIT-2	163
Tabelle 11:	Teil-Fragebogen zur Einschätzung des Umgangs mit Concept Mapping	165
Tabelle 12:	Fragen zur Einschätzung der Nützlichkeit des einfachen Feedbacks	166
Tabelle 13:	Frage zum gewünschten elaborierten Feedback (Gruppe einfaches Feedback).....	166
Tabelle 14:	Fragen zur Einschätzung des elaborierten Feedbacks	167
Tabelle 15:	Frage zum gewünschten Feedback (Gruppe kein Feedback)	167
Tabelle 16:	Teil-Fragebogen zur Einschätzung des Umgangs mit Fehlern.....	169
Tabelle 17:	Mittelwerte und Standardabweichungen für die Ergebnisse des Vortests, 1. Nachtests und 2. Nachtests	172
Tabelle 18:	Anzahl der Konzepte und Relationen der Basislexika.....	175
Tabelle 19:	Anzahl der genannten Propositionen und durchschnittlicher Netzumfang	180
Tabelle 20:	Kumulierte Anzahl der Propositionen zur Ermittlung der Modalnetze..	181
Tabelle 21:	Bewertung des ersten Maps des Probanden Ei04 anhand des entwickelten Scoringverfahrens.....	190

Tabelle 22:	Mittelwerte und Standardabweichungen für die Scoringwerte der drei Mapping-Aktivitäten.....	192
Tabelle 23:	KMO-Kriterium und Bartlett-Test für die Zwei-Faktoren-Lösung mit neun Variablen.....	197
Tabelle 24:	Kommunalitäten für die Zwei-Faktoren-Lösung mit neun Variablen....	197
Tabelle 25:	Rotierte Faktorladungen für die Zwei-Faktoren-Lösung mit neun Variablen.....	198
Tabelle 26:	Interne Konsistenz für den Faktor 1 für die Zwei-Faktorenlösung mit neun Variablen.....	198
Tabelle 27:	Kriterien zur Prüfung des Datenmaterials und Gütekriterien für die Zwei-Faktoren-Lösung mit neun Variablen.....	199
Tabelle 28:	Mittelwerte und Standardabweichungen für die Faktoren Interessantheit-Nützlichkeit und Handhabbarkeit.....	200
Tabelle 29:	Mittelwerte und Standardabweichungen für EIFBN.....	202
Tabelle 30:	Interne Konsistenz für EIFBN.....	202
Tabelle 31:	Standardabweichungen und Mittelwerte für EIFBN.....	203
Tabelle 32:	Mittelwerte und Standardabweichungen zum gewünschten Feedback..	203
Tabelle 33:	Interne Konsistenz für den Umgang mit Fehlern.....	204
Tabelle 34:	Mittelwerte und Standardabweichungen für den Umgang mit Fehlern..	204
Tabelle 35:	Mittelwerte und Standardabweichungen für die Gruppe CM und die KG für den Vortest, 1. Nachtest und 2. Nachtest.....	206
Tabelle 36:	Ergebnisse der ANOVAs mit Messwiederholung für die Lernwirksamkeit des Concept Mapping.....	207
Tabelle 37:	Ergebnisse der einfaktoriellen ANOVAs mit der unabhängigen Variable [UV] <i>Gruppe</i> und den AVs <i>1. Nachtest</i> bzw. <i>2. Nachtest</i>	209
Tabelle 38:	Effektstärken für den Haupteffekt <i>Gruppe</i> der einfaktoriellen ANOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i>	209
Tabelle 39:	Mittelwerte und Standardabweichungen für die Gruppen kein FB, EiFB und EIFB für den Vortest, 1. und 2. Nachtest.....	210
Tabelle 40:	Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i>	211

Tabelle 41: Effektstärken für den Interaktionseffekt <i>Zeit x Gruppe</i> der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i> über zwei bzw. drei Messzeitpunkte	213
Tabelle 42: Ergebnisse der einfaktoriellen ANCOVA mit der UV <i>Gruppe</i> , der AV <i>1. Nachtest</i> sowie den Kovariaten <i>VW, SF, H, IN</i> und <i>UF</i>	215
Tabelle 43: Ergebnisse der einfaktoriellen ANCOVA mit der UV <i>Gruppe</i> , der AV <i>2. Nachtest</i> sowie den Kovariaten <i>VW, SF, H, IN</i> und <i>UF</i>	216
Tabelle 44: Effektstärken für die einfaktoriellen ANCOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i> sowie den Kovariaten <i>VW, SF, H, IN</i> und <i>UF</i>	216
Tabelle 45: Ergebnisse der zweifaktoriellen ANOVAs mit der UV <i>Gruppe</i> und <i>VW_nh</i> sowie der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i>	218
Tabelle 46: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Gruppe</i> und <i>VW_nh</i> sowie der AV <i>Gruppenmittelwerte Wissenstest</i>	221
Tabelle 47: Ergebnisse der zweifaktoriellen ANOVAs mit der UV <i>Gruppe</i> und <i>SF_nh</i> sowie der AV <i>1. Nachtest</i> sowie <i>2. Nachtest</i>	224
Tabelle 48: Effektstärken für den Haupteffekt <i>Gruppe</i> der zweifaktoriellen ANOVA mit den UVs <i>Gruppe</i> und <i>SF_nh</i> sowie der AV <i>1. Nachtest</i>	225
Tabelle 49: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Gruppe</i> und <i>SF_nh</i> sowie der AV <i>Gruppenmittelwerte Wissenstest</i>	228
Tabelle 50: Ergebnisse der einfaktoriellen ANOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i> für einfache Aufgaben.....	231
Tabelle 51: Effektstärken für den Haupteffekt <i>Gruppe</i> der einfaktoriellen ANOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i> für einfache Aufgaben.....	231
Tabelle 52: Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i> für einfache Aufgaben	232

Tabelle 53:	Effektstärken für den Interaktionseffekt <i>Zeit x Gruppe</i> der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i> über zwei bzw. drei Messzeitpunkte für einfache Aufgaben	234
Tabelle 54:	Ergebnisse der einfaktoriellen ANOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i> für komplexe Aufgaben.....	235
Tabelle 55:	Effektstärken für den Haupteffekt <i>Gruppe</i> der einfaktoriellen ANOVAs mit der UV <i>Gruppe</i> und der AV <i>1. Nachtest</i> bzw. <i>2. Nachtest</i> für komplexe Aufgaben	236
Tabelle 56:	Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i> für komplexe Aufgaben	237
Tabelle 57:	Effektstärken für den Interaktionseffekt <i>Zeit x Gruppe</i> der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Gruppenmittelwerte Wissenstest</i> über zwei bzw. drei Messzeitpunkte für <i>komplexe Aufgaben</i>	238
Tabelle 58:	Ergebnisse der Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs - Prüfung der Hypothesen	250
Tabelle 59:	Mittelwerte und Standardabweichungen für die Gruppen kein FB, EiFB und EIFB ScorM1, ScorM2 und ScorM3	251
Tabelle 60:	Ergebnisse der MANOVA mit der UV <i>Gruppe</i> und den AVs ScorM1, ScorM2 und ScorM3.....	252
Tabelle 61:	Ergebnisse der MANCOVA mit der UV <i>Gruppe</i> , der Kovariaten ScorM1 und den AVs ScorM2 sowie ScorM3	252
Tabelle 62:	Effektstärken für die MANCOVA mit der UV <i>Gruppe</i> , der Kovariaten ScorM1 und den AVs ScorM2 sowie ScorM3	253
Tabelle 63:	Ergebnisse der zweifaktoriellen ANOVA mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Scoringwerte</i>	255
Tabelle 64:	Effektstärken für die zweifaktorielle ANOVA mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Scoringwerte</i>	255
Tabelle 65:	Ergebnisse der einfaktoriellen MANCOVA mit der UV <i>Gruppe</i> , den Kovariaten ScorM1, VW, SF, H, IN, UF und der AV <i>Scoringwerte</i>	257
Tabelle 66:	Ergebnisse der MANCOVA mit der UV <i>Gruppe</i> , den Kovariaten ScorM1, VW, SF, H, IN, UF und der AV ScorM2 bzw. ScorM3.....	257

Tabelle 67:	Effektstärken für die einfaktorielle MANCOVA mit der UV Gruppe, den Kovariaten ScorM1, VW, SF, H, IN, UF und der AV ScorM2 bzw. ScorM3.....	258
Tabelle 68:	Ergebnisse der zweifaktoriellen ANCOVA mit den UVs Gruppe und IN_nh, den Kovariaten ScorM1 sowie der AV ScorM2.....	259
Tabelle 69:	Ergebnisse der zweifaktoriellen ANCOVA mit den UVs Gruppe und H_nh, den Kovariaten ScorM1 sowie der AV ScorM2.....	261
Tabelle 70:	Ergebnisse der Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps - Prüfung der Hypothesen	267
Tabelle 71:	Ergebnisse der einfaktoriellen ANCOVA mit der UV Gruppe, den Kovariaten ScorM1, ScorM2 und ScorM3 sowie der AV 1. Nachttest ...	268
Tabelle 72:	Ergebnisse der einfaktoriellen ANCOVA mit der UV Gruppe, den Kovariaten ScorM1, ScorM2 und ScorM3 sowie der AV 2. Nachttest ...	269
Tabelle 73:	Ergebnisse der zweifaktoriellen ANOVAs mit der UV Gruppe und ScorM2_nh sowie der AV 1. Nachttest bzw. 2. Nachttest.....	270
Tabelle 74:	Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs Gruppe und ScorM2_nh sowie der AV Gruppenmittelwerte Wissenstest	273
Tabelle 75:	Ergebnisse der zweifaktoriellen ANOVAs mit der UV Gruppe und ScorM3_nh sowie der AV 1. Nachttest bzw. 2. Nachttest.....	276
Tabelle 76:	Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs Gruppe und ScorM3_nh sowie der AV Gruppenmittelwerte Wissenstest.....	279
Tabelle 77:	Ergebnisse der einfaktoriellen ANCOVA mit der UV Gruppe, den Kovariaten ScorM1, ScorM2 und ScorM3 sowie der AV 1. Nachttest für einfache Aufgaben.....	283
Tabelle 78:	Ergebnisse der einfaktoriellen ANCOVA mit der UV Gruppe, den Kovariaten ScorM1, ScorM2 und ScorM3 sowie der AV 2. Nachttest für einfache Aufgaben.....	283
Tabelle 79:	Ergebnisse der einfaktoriellen ANCOVA mit der UV Gruppe, den Kovariaten ScorM1, ScorM2 und ScorM3 sowie der AV 1. Nachttest für komplexe Aufgaben	284

Tabelle 80:	Ergebnisse der einfaktoriellen ANCOVA mit der UV <i>Gruppe</i> , den Kovariaten <i>ScorM1</i> , <i>ScorM2</i> und <i>ScorM3</i> sowie der AV <i>2. Nachtest für komplexe Aufgaben</i>	285
Tabelle 81:	Ergebnisse des Einflusses der Mapqualität auf den Wissenserwerb - Prüfung der Hypothesen	291
Tabelle 82:	Ergebnisse der einfaktoriellen ANOVA mit der UV <i>Gruppe</i> und der AV <i>gewünschtes FB allgemein</i>	292
Tabelle 83:	Ergebnisse der einfaktoriellen ANOVA mit der UV <i>Gruppe</i> und der AV <i>Mittelwert EiFBN</i>	293
Tabelle 84:	Ergebnisse der zweifaktoriellen ANOVA mit Messwiederholung mit den UVs <i>Zeit</i> und <i>Gruppe</i> und der AV <i>Scoringwerte</i>	295
Tabelle 85:	Falsche und fehlende Propositionen des 1. Mapping	300
Tabelle 86:	Falsche und fehlende Propositionen des 2. Mapping	301
Tabelle 87:	Falsche und fehlende Propositionen des 3. Mapping	303
Tabelle 88:	Ergebnisse der Analyse zum Umgang mit dem Feedback - Prüfung der Hypothesen	304

1 Einleitung

Neue technische, ökonomische und gesellschaftliche Entwicklungstendenzen haben ein grundlegendes Überdenken der beruflichen Erstausbildung und Weiterbildung ausgelöst. Anforderungen an junge Berufsanfänger¹ sind durch eine zunehmende Komplexität gekennzeichnet (Achtenhagen, 1995). Um diese zu bewältigen, ist ein aktiver und selbstgesteuerter Umgang mit komplexen Sachverhalten erforderlich (Mandl & Fischer, 2000, S. 3). Um Lernende schon frühzeitig auf die komplexen Problemsituationen der Arbeitswelt vorzubereiten, muss die Ausbildung daher zukunftsfähig gestaltet werden. Vor diesem Hintergrund werden verstärkt Komplexe Lehr-Lern-Arrangements, z. B. Planspiele, eingesetzt. „Ein Planspiel ist eine besondere Form eines dynamischen Modells, in dem sich die Spielteilnehmer mit konfliktären und problemhaltigen Situationen“ unter Beachtung von Regeln und Einhaltung einer bestimmten Zielsetzung kollektiv und individuell handelnd auseinandersetzen (Rebmann, 2001, S. 10). Hierbei gilt es jedoch zu beachten, dass Planspiele selbst bei einer angemessenen Konstruktion nicht automatisch lernwirksam sind. Umfangreiche Forschung hat gezeigt, dass Planspiele keine Selbstläufer sind und es einer instruktionalen Unterstützung, z. B. durch den Einsatz von Lernstrategien, bedarf (Achtenhagen, Tramm & Preiss, 1992, S. 76; Dörner, Göbel, Effelsberg & Wiemeyer, 2016). Eine mögliche Lernstrategie stellt dabei die Anwendung von Concept Maps dar.

Concept Maps sind zweidimensionale Strukturdarstellungen, welche Wissen oder Informationen in Form von Netzwerken graphisch darstellen und organisieren können (Novak & Cañas, 2008, S. 1; Ruiz-Primo & Shavelson, 1996, S. 570). Die Anwendung von Concept Maps ermöglicht es, komplexe Sachverhalte angemessen zu visualisieren. Ein theoretischer Erklärungsansatz für die Lernwirksamkeit von Concept Maps stellt dabei die Theorie der semantischen Netzwerke dar, welche von einer netzwerkartigen Struktur des Gedächtnisses ausgeht (Collins & Quillian, 1969). Concept Maps entsprechen somit der angenommenen Repräsentation kognitiver Strukturen. Diese Strukturgleichheit führt dazu, dass das Erstellen von Concept Maps kein zusätzliches Umkodieren von der Netzwerkform erfordert. Beim Erstellen eines Textes sind dagegen aufwändige Umkodierungsprozesse von der Netzwerkform des Gedächtnisses in die lineare Form vonnöten

¹ Aus Gründen der besseren Lesbarkeit wird in diesem Beitrag auf eine geschlechtsspezifische Schreibweise verzichtet. Mit der männlichen Schreibweise sind alle Geschlechter gemeint. Beispielsweise schließt der Begriff „Berufsanfänger“ auch „Berufsanfängerinnen“ ein.

(Fürstenau, 2011, S. 46). Concept Maps werden im Lehr-Lernkontext einerseits zur Wissensdiagnose verwendet, um vernetztes Wissen und dessen Veränderung zu messen. Andererseits dienen Concept Maps der instruktionalen Unterstützung, um das Verstehen von Zusammenhängen zu fördern (Fürstenau, 2011, S. 56; Mandl & Fischer, 2000, S. 6 ff.). Beide Anwendungsgebiete können dabei auch verknüpft werden.

Bisher durchgeführte Metaanalysen (Nesbit & Adesope, 2006; Schroeder, Nesbit, Anguiano, & Adesope, 2017) zeigen die hohe Wirksamkeit des Lernens mit Concept Maps auf, wobei die Befundlage nicht eindeutig ist. Die Lernwirksamkeit ist dabei von verschiedenen Faktoren abhängig. Nesbit und Adesope (2006) heben den Einfluss von individuellen Voraussetzungen wie den verbalen Fähigkeiten sowie dem Vorwissen hervor. Des Weiteren spielt die Einstellung der Lernenden zur Methode (2004) sowie eine Vertrautheit mit dem Concept Mapping eine bedeutende Rolle (Chang, Sung, & Chen, 2002; Hilbert & Renkl, 2008). Im Kontext des Planspielunterrichts wurde bisher wenig geforscht. Es zeigte sich auch hier ein Einfluss des Vorwissens auf die Lernwirksamkeit (Ryssel & Fürstenau, 2011; Sommer, Fürstenau, Ryssel, & Kunath, 2009). Insgesamt besteht hierbei weiterer Forschungsbedarf.

Ergänzend zu den möglichen Faktoren, die die Wirksamkeit des Lernens mit Concept Maps beeinflussen, stellt sich weiterhin die Frage, inwieweit das Erstellen von Concept Maps allein eine Tiefenverarbeitung der erlernten Inhalte fördert oder ob es eines zusätzlichen Feedbacks zu den in den Concept Maps dargestellten Inhalten bedarf. Hierbei spielt eine formale Rückmeldung zur Verbesserung des Umgangs mit der Lernstrategie eine untergeordnete Rolle. Vielmehr geht es darum, den Lernenden z. B. Hinweise zur Richtigkeit bzw. Falschheit der Propositionen zu geben und Hilfestellungen zur inhaltlichen Verbesserung der Maps anzubieten.

Feedback stellt einen der größten Einflussfaktoren auf das Lernen dar (Hattie, 2015). Zur Erforschung der Lernwirksamkeit von Feedback existieren einige Metaanalysen, wobei die Metanalysen von Bangert-Drowns u. a. (1991) sowie Kluger und DeNisi (1996) die am häufigsten zitierten sind. Beide Metaanalysen zeigen, dass Feedback nicht generell lernwirksam ist, sondern von verschiedenen Faktoren moderiert wird. Feedback lässt sich anhand verschiedener Kriterien gestalten. Zum einen sollte das Feedback dabei in Abhängigkeit von der Feedbackform (Feedbackquelle, Feedbackempfänger, Zeitpunkt der Feedbackgabe, Transparenz für den Feedbackempfänger, Präsentationsformat, Anzahl erneuter Lösungsversuche) auch auf unterschiedliche Weise gegeben werden (Strijbos &

Müller, 2014; Narciss, 2006). Zum anderen hat die Komplexität der Feedbacknachricht einen Einfluss auf die Gestaltung des Feedbacks. Dabei kann zwischen einfachem Feedback, bei dem der Anteil der richtig gelösten Aufgaben, die richtige Lösung oder die Information, ob die Antwort richtig oder falsch ist, rückgemeldet wird und elaboriertem Feedback unterschieden werden. Elaboriertes Feedback beinhaltet umfassende und differenzierte Informationen, z. B. zur Anzahl oder den Ursachen der Fehler, Hinweise auf Lösungsstrategien, ergänzende Erklärungen oder Beispiele zu Begriffen. Die Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht wurde bisher am meisten erforscht (Hattie & Wollenschläger, 2014, S. 138 f.; Shute, 2008, S. 292; Strijbos, Narciss, & Dünnebier, 2010, S. 293). Es zeigt sich kein genereller Vorteil für das elaborierte gegenüber dem einfachen Feedback. Die uneinheitliche Befundlage lässt auf den Einfluss weiterer (inter-)personaler und situativer Faktoren schließen. Dabei konnte z. B. ein großer Einfluss der Kompetenz des Feedbackempfängers (Vorwissen, Sprachfähigkeit, Antwortsicherheit) nachgewiesen werden (z. B. Hattie & Timperley, 2007; Butler & Winne, 1995). In Bezug auf den Einfluss des Vorwissens kann elaboriertes Feedback zu einer Angleichung des Wissensniveaus von vorwissensschwachen an das von vorwissensstarken Lernenden führen, wobei vorwissensstarke Probanden diese zusätzliche externe instruktionale Unterstützung nicht benötigen (z. B. Krause, 2007). Darüber hinaus kann auch der Umgang mit Fehlern von Bedeutung sein, d. h. inwieweit eine Fehlerbereitschaft besteht bzw. wie bedeutsam Fehler von Lernenden eingeschätzt werden. Des Weiteren können situative Faktoren, wie die Aufgabenanforderungen der Lernaufgabe (z. B. Kluger & DeNisi, 1996), oder die wahrgenommene Nützlichkeit des Feedbacks (z. B. Harks, Rakoczy, Hattie, Besser, & Klieme, 2014) beeinflussend wirken.

Im Rahmen des Lernens mit Concept Maps existiert bisher wenig Forschung, die die Wirkung von ergänzendem Feedback untersucht. Mehrheitlich verwenden die Untersuchungen computergestütztes Concept Mapping und verbinden dieses mit automatisiertem Feedback (Hwang, Wu, & Ke, 2011; Ifenthaler, 2010; Chang u. a., 2002). Einen Überblick über computergestützte Concept-Mapping-Tools, welche automatisiertes Feedback verwenden, geben Álvarez-Montero, Jacobo-García und Rocha-Ruiz (2015). Sie stellten in ihrem Literaturreview heraus, dass das Feedback im Rahmen von computergestützten Tools oft nur intuitiv gestaltet wird, ohne die Ergebnisse der Feedbackliteratur zu berücksichtigen.

sichtigen. Darüber hinaus existieren zahlreiche Concept Mapping Tools, welche Feedback implementieren, jedoch die Wirksamkeit nicht evaluieren. Concept Maps lassen sich auch im Paper & Pencil-Verfahren per Hand erstellen. Ergänzendes Feedback kann hierbei ebenfalls papierhaft und/oder mündlich gegeben werden. Zu dieser Art der medialen Unterstützung wurde bisher wenig geforscht. Es existieren Studien zum computerbasierten Mapping, welche papierbasiertes mit computerbasiertem Feedback vergleichen (z. B. Wu, Hwang, Milrad, Ke, & Huang, 2012). Weitere Studien vergleichen papierbasiertes Feedback zu Maps, welche per Hand erstellt wurden mit computergestütztem Feedback zu am Computer erstellten Maps (Yang, 2015). Die Ergebnisse lassen keinen eindeutigen Schluss zu, der für einen Vorteil des computerbasierten Feedbacks spricht.

Den Forschungsdefiziten soll mit dieser Studie Rechnung getragen werden. Ziel der Studie ist es zu untersuchen, inwieweit das Erstellen von Concept Maps im Paper & Pencil-Verfahren allein eine Tiefenverarbeitung der erlernten Inhalte fördert oder ob es eines zusätzlichen Feedbacks zu den in den Concept Maps dargestellten Inhalten bedarf. Dabei stellt sich die Frage, welches Feedback lernwirksamer ist. Das Erstellen von Concept Maps beansprucht die kognitiven Ressourcen der Lernenden. Das Anwenden dieser Lernstrategie kann die Lernenden somit möglicherweise überfordern. Für eine tiefgehende Verarbeitung der Inhalte kann es daher unter Umständen sinnvoll sein, den Lernenden eine Rückmeldung zu den in den Maps dargestellten Zusammenhängen zu geben, um die kognitive Belastung zu reduzieren. Dies kann von verschiedenen Faktoren abhängig sein, deren Einfluss ebenfalls überprüft werden soll.

Insgesamt ergeben sich drei Forschungsfragen:

- F1) Inwieweit ist ergänzendes Feedback zu Concept Maps lernwirksamer als kein ergänzendes Feedback?
- F2) Ist elaboriertes Feedback lernwirksamer als einfaches Feedback?
- F3) Inwieweit beeinflussen das Vorwissen, die Sprachfähigkeit, die Einschätzung des Umgangs mit Concept Mapping, die Einschätzung der Nützlichkeit des Feedbacks, die Einschätzung des Umgangs mit Fehlern sowie die Aufgabenanforderungen der Lernaufgabe die Lernwirksamkeit des Feedbacks?

Die Untersuchung wird im Rahmen des betriebswirtschaftlichen Planspielunterrichts durchgeführt, um den eingangs beschriebenen Anforderungen an junge Berufsanfänger hinsichtlich des Umgangs mit komplexen Problemsituationen gerecht zu werden. Ausgewählt wird das von der Business Training International GmbH entwickelte Planspiel Easy

BusinessTM, welches spielerisch den Wertschöpfungsprozess eines Produktionsunternehmens vermittelt und die Folgen des Handelns in der jährlichen Rechnungslegung erleben lässt. Das Planspiel wurde ursprünglich für Schüler in Abschlussklassen allgemeinbildender Schulen oder für Schüler in Eingangsklassen beruflicher Schulen entwickelt. Als Zielgruppe für diese Studie werden Schüler der achten und neunten Klasse zweier sächsischer Mittelschulen² ausgewählt. Hierfür wird die Komplexität des Planspiels etwas reduziert, z. B. wird auf das Vermitteln von Kennzahlen verzichtet. Die Inhalte des Planspiels stellen dabei ergänzende und vertiefende Inhalte zum Lehrplan des Fachs Wirtschaft-Technik-Haushalt/Soziales dar (Sächsisches Staatsministerium für Kultus, 2009). Diese Zielgruppe wurde bewusst ausgewählt, um die Schüler schon frühzeitig auf die komplexen Anforderungen des Berufslebens vorzubereiten. Des Weiteren wird vermutet, dass sich der Einfluss verschiedener Faktoren, wie Vorwissen oder Sprachfähigkeit, auf die Lernwirksamkeit des Feedbacks bei dieser Zielgruppe möglicherweise besonders auswirkt.

Um die Lernwirksamkeit zu operationalisieren, wird der Wissenserwerb anhand der Ergebnisse eines Paralleltests (Vor- und Nachtest) sowie anhand der Qualität der erstellten Maps analysiert. Auf diese Weise dienen die Concept Maps nicht nur der instruktionalen Unterstützung, sondern werden auch zur Wissensdiagnose verwendet.

Kapitel 2 gibt einen Überblick über theoretische Grundlagen zum Feedback in Lehr-Lernprozessen (Kapitel 2.1) und empirische Befunde zur Lernwirksamkeit des Feedbacks (Kapitel 2.2). Im Kapitel 2.1.1 wird das für die Arbeit relevante Verständnis von Feedback erläutert. Die Kapitel 2.1.2 und 2.1.3 verdeutlichen anschließend, dass Feedback ein sehr vielschichtiges Konzept im Lernprozess ist. Kapitel 2.1.2 thematisiert die verschiedenen Funktionen von Feedback. Im Kapitel 2.1.3 wird Feedback in Lehr-Lernprozessen anhand der Komplexität der Feedbacknachricht sowie weiterer Kriterien klassifiziert. In Kapitel 2.1.4 werden verschiedene Feedbackmodelle zur Feedbackrezeption und/ oder –gestaltung vorgestellt, um insbesondere Einflussfaktoren auf die Lernwirksamkeit des Feedbacks zu identifizieren. Hierbei werden zunächst die Makromodelle von Kulhavy und Stock (1989) und Bangert-Drowns u. a. (1991) und das Mikromodell von Narciss (2006) dargestellt. Des Weiteren wird das Rahmenmodell von Strijbos und Müller (2014) vorgestellt. Abschließend wird ein erweitertes Rahmenmodell zur Gestaltung und Verar-

² Die Mittelschulen in Sachsen tragen seit dem Jahr 2013 die Bezeichnung Oberschulen. Da die Untersuchung vorher durchgeführt wurde, wird die Bezeichnung Mittelschule verwendet.

beitung von Feedback basierend auf dem Modell von Strijbos und Müller (2014) entwickelt sowie um weitere Erkenntnisse aus dem Kapitel 2.1 ergänzt. Einen Überblick über den Stand der empirischen Forschung zur Lernwirksamkeit des Feedbacks gibt Kapitel 2.2. Im Kapitel 2.2.1 werden zunächst die Ergebnisse der Studie von Hattie (2015) in Bezug auf den Einfluss des Feedbacks auf das Lernen sowie die Ergebnisse der Metaanalysen von Bangert-Drowns u. a. (1991) sowie Kluger und DeNisi (1996) vorgestellt. Kapitel 2.2.2 und 2.2.3 stellen anschließend Studien vor, die die Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht bzw. der Feedbackform untersuchen. Kapitel 2.2.4 und 2.2.5 betrachten Studien zur Lernwirksamkeit in Abhängigkeit von (inter-)personalen sowie situativen Faktoren. Im Kapitel 2.2.6 werden die Ergebnisse der empirischen Befunde zur Lernwirksamkeit des Feedbacks zusammengefasst.

Kapitel 3 gibt einen Überblick über theoretische Grundlagen zu Concept Maps in Lehr-Lernprozessen (Kapitel 3.1) und empirische Befunde zur Lernwirksamkeit von Concept Maps (Kapitel 3.2). Im Kapitel 3.1.1 wird der Begriff Concept Map bestimmt sowie klassifiziert. Kapitel 3.1.2 stellt Anwendungsgebiete von Concept Maps vor. Im Kapitel 3.1.3 werden theoretische Ansätze zur Lernwirksamkeit von Concept Maps erläutert. Im Kapitel 3.2.1 werden zunächst die Ergebnisse der Studie von Hattie (2015) in Bezug auf den Einfluss des Concept Mapping auf das Lernen sowie die Ergebnisse der Metaanalysen von Nesbit und Adesope (2006) sowie Schroeder u. a. (2017) vorgestellt. Kapitel 3.2.2 betrachtet Studien zu Einflussfaktoren auf die Lernwirksamkeit von Concept Maps. Im Kapitel 3.2.3 werden abschließend Studien vorgestellt, welche Feedback als Ergänzung zum Concept Mapping einsetzen.

Im Kapitel 4 wird die Konzeption, Erhebung und Auswertung der empirischen Untersuchung zur Lernwirksamkeit des Feedbacks zu Concept Maps, die dieser Arbeit zugrunde liegt, vorgestellt. Hierzu wird in Kapitel 4.1 zunächst auf die Gestaltung der Untersuchung eingegangen und die Fragestellungen, die Hypothesen, die Stichprobe sowie das Untersuchungsdesign erläutert. Kapitel 4.2 thematisiert die Gestaltung der Lernmaterialien und Interventionsmaßnahmen. Dabei wird in Kapitel 4.2.1 der Planspielunterricht mit Easy BusinessTM vorgestellt und die Gestaltung des im Rahmen der Studien eingesetzten Concept Mapping (Kapitel 4.2.2) und Feedbacks (Kapitel 4.2.3) erläutert. Die Erhebungsinstrumente werden im Kapitel 4.3 vorgestellt, wobei Kapitel 4.3.1 den Wissens-

test als Paralleltest, Kapitel 4.3.2 den Sprachfähigkeitstest und Kapitel 4.3.3 den Fragebogen zur Einschätzung des Umgangs mit Concept Mapping, der Nützlichkeit des Feedbacks sowie des Umgangs mit Fehlern betrachten. Kapitel 4.4 geht auf die Auswertung des Wissenstests (Kapitel 4.4.1) und die Auswertung des Sprachfähigkeitstests (Kapitel 4.4.2) ein. Im Kapitel 4.5 wird die Auswertung der Concept Maps beschrieben. Hierzu wird zunächst eine kategoriale und strukturelle Inhaltsanalyse durchgeführt sowie eine Reliabilitätsanalyse des Codiervorgehens vorgenommen (Kapitel 4.5.1). Anschließend erfolgt eine qualitative Bewertung der Concept Maps (Kapitel 4.5.2). Zu diesem Zweck wird zunächst der Übereinstimmungsgrad mit dem Referenznetz anhand der Galanter-Metrik berechnet und anschließend ein differenziertes Scoringverfahren auf Basis der Galanter-Metrik entwickelt. Dieses Scoringverfahren wird des Weiteren einer Reliabilitätsanalyse unterzogen und evaluiert. Im Kapitel 4.6 wird der Fragebogen zur Einschätzung des Umgangs mit Concept Mapping (Kapitel 4.6.1), des Umgangs mit Fehlern (Kapitel 4.6.2) sowie der Nützlichkeit des Feedbacks (Kapitel 4.6.3) ausgewertet.

Kapitel 5 beinhaltet Analysen und Ergebnisse der empirischen Untersuchung zur Lernwirksamkeit des Feedbacks zu Concept Maps. Im Kapitel 5.1 werden dazu inferenzstatistische und deskriptive Analysen zum Einfluss der Art des Feedbacks vorgestellt. Um die Lernwirksamkeit als abhängige Variable zu operationalisieren, werden der Wissenserwerb sowie die Qualität der Maps berücksichtigt. Die Kapitel 5.1.1 und 5.1.2 analysieren die Einflüsse der Art des Feedbacks auf den Wissenserwerb sowie auf die Qualität der Maps. Dabei werden auch (inter-)personale und situative Einflussfaktoren untersucht. Des Weiteren wird der Einfluss der Qualität der Maps auf den Wissenserwerb erforscht (Kapitel 5.1.3). Die Ergebnisse werden jeweils zusammengefasst und im Hinblick auf die aufgestellten Hypothesen bewertet. Kapitel 5.2 analysiert zunächst den Umgang der Probanden mit dem Feedback (Kapitel 5.2.1). Des Weiteren werden die erstellten Concept Maps in Bezug auf ihre inhaltliche Richtigkeit überprüft, um Fehlkonzepte bzw. fehlende Propositionen zu identifizieren (Kapitel 5.2.2). Die Ergebnisse werden ebenfalls zusammengefasst und im Hinblick auf die Hypothesen bewertet (Kapitel 5.2.3). Kapitel 5.3 interpretiert die Ergebnisse unter Bezug auf die aufgestellten Hypothesen sowie die theoretischen Grundlagen der Kapitel 2 und 3.

Im Rahmen der Schlussbetrachtung im Kapitel 6 wird die Arbeit zunächst zusammengefasst (Kapitel 6.1). Des Weiteren werden die Gütekriterien der Untersuchung überprüft.

Unter Berücksichtigung weiterer kritischer Aspekte werden Optimierungsvorschläge unterbreitet und ein Ausblick für eine zukünftige Forschung zur Lernwirksamkeit von Feedback zu Concept Maps gegeben.

2 Feedback in Lehr-Lernprozessen

Der Begriff Feedback ist sowohl im alltäglichen als auch im wissenschaftlichen Sprachgebrauch weit verbreitet. Dabei variieren jedoch je nach Kontext die Präzision und die Breite des Begriffsverständnisses und somit die Arten von Feedbackmeldungen. Im Rahmen der theoretischen Grundlagen zu Feedback in Lehr-Lernprozessen (Kapitel 2.1) wird daher das für die Arbeit relevante Verständnis von Feedback dargelegt (Kapitel 2.1.1). Darüber hinaus werden die Funktionen des Feedbacks erläutert (Kapitel 2.1.2) sowie Feedback klassifiziert (Kapitel 2.1.3). Darauf aufbauend werden Feedbackmodelle zur Feedbackrezeption und/oder –gestaltung vorgestellt (Kapitel 2.1.4). Kapitel 2.2 betrachtet empirische Befunde zur Lernwirksamkeit des Feedbacks. Hierbei wird ein Überblick über Metaanalysen gegeben (Kapitel 2.2.1) sowie Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht (Kapitel 2.2.2) sowie der Feedback-Form (Kapitel 2.2.3) vorgestellt. Des Weiteren werden Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von (inter-)personalen (Kapitel 2.2.4) und situativen Faktoren (Kapitel 2.2.5) erläutert. Die empirischen Befunde werden abschließend zusammengefasst (Kapitel 2.2.6).

2.1 Theoretische Grundlagen zum Feedback in Lehr-Lernprozessen

2.1.1 Bestimmung des Begriffs Feedback in Lehr-Lernprozessen

Feedback bezeichnet laut Duden online (2017) eine „zielgerichtete Steuerung eines technischen, biologischen oder sozialen Systems durch Rückmeldung der Ergebnisse, wobei die Eingangsgröße durch Änderung der Ausgangsgröße beeinflusst werden kann“. Dies entspricht der Sichtweise der Kybernetik (Lehre von den Regelungswissenschaften), aus der der Begriff Feedback ursprünglich stammt. Das Feedback ist dabei ein Bestandteil des Regelkreismodells. Resultierend aus dem Vergleich des Ist-Zustands mit dem Soll-Zustand werden Informationen rückgemeldet, um Aufschluss darüber zu geben, wie groß die Diskrepanz zwischen angestrebtem und gegebenem Zustand ist (Fengler, 2009, S. 12). Übertragen auf Lehr-Lernprozesse bezieht sich das Feedback direkt auf das Missverhältnis zwischen der aktuellen und der korrekten Antwort und zielt darauf, die Diskrepanz zu beheben. Informationen über das Verhaltensergebnis werden erst dann als Feedback bezeichnet, wenn sie gezielt kontrolliert und manipuliert werden können und implizieren somit weitestgehend standardisiertes und automatisiertes Feedback (Narciss, 2006, S. 17 ff.). Ausgehend von diesem *engen Begriffsverständnis* haben die kognitive Wende sowie

eine rasante Entwicklung computerbasierter Lehr-Lerntechnologien zu einem *erweiterten* Begriffsverständnis im Lehr- Lernkontext geführt (ebd.). Dabei rückt zum einen die korrigierende Funktion des Feedbacks in den Vordergrund, zum anderen lassen sich neue Feedbackarten mittels moderner Informationstechnologien gestalten und präsentieren (ebd.). Unter Bezugnahme auf Narciss (2006, S. 18), Kluger und DeNisi (1996, S. 255) sowie Krause (2007, S. 51) bezeichnet Feedback innerhalb von Lehr-Lernkontexten alle Informationen, die Lernenden von einer *externen* Informationsquelle, wie dem Lehrer, dem Lernprogramm oder dem Versuchsleiter während oder nach der Aufgabebearbeitung angeboten werden:

- a) zur Fehlerkorrektur oder Bestätigung korrekter Antworten, um vorgegebene Lernziele zu erreichen,
- b) um Auskunft über den eigenen Lernfortschritt zu erhalten oder
- c) um die eigene Leistung bzw. eigene Stärken und Schwächen mit anderen Lernenden zu vergleichen.

Dem Lernenden wird folglich ein Feedback zu seiner Leistung gegeben, welche mit einem Lernziel, dem eigenen Lernfortschritt oder mit anderen Lernenden verglichen wird. In diesem Zusammenhang wird zwischen der sachlichen, der individuellen und der sozialen Bezugsnorm³ unterschieden (Rheinberg, 2008). Die Informationen der externen Quelle können sich auch indirekt auf die Diskrepanz zwischen Soll und Ist beziehen, z. B. auf das Beheben dieser Diskrepanz (Narciss, 2006, S. 18).

Feedback ist ein wesentlicher Bestandteil im Prozess der *formativen* Leistungsbeurteilung (Brookhart, 2010, S. 7; Sadler, 1989, S. 120) und sollte hierbei als Teil des Lehr-Lern-Prozesses betrachtet werden, um diesen zu optimieren (Maier, 2010, S. 299). Sowohl die Lehrenden als auch die Lernenden sind Akteure in diesem Prozess (Black & Wiliam, 2009, S. 9). Für eine Stärkung des selbstregulierten Lernens sollten die Lernenden dabei auch angehalten werden, ihre Leistung selbst einzuschätzen (self assessment) oder sich gegenseitig zu bewerten (peer assessment). Das Feedback sollte gezielt genutzt werden, um über den weiteren Lehr-Lern-Prozess zu informieren und diesen an den aktuellen Könnensstand der Lernenden anzupassen (Maier, 2010, S. 299). Die Regulation des Lehr-Lern-Prozesses kann sich dabei auf alle drei Bezugsnormen beziehen, wobei die sachliche und die individuelle Bezugsnorm im Vordergrund stehen sollten. Zum einen kann somit

³ Bezugsnormen bezeichnen „Standards, mit denen man ein Resultat vergleichen muss, wenn man das Resultat als Leistung bewerten will“ (Rheinberg, 2008, S. 178).

eine Annäherung an das Lernziel erreicht werden und zum anderen eröffnet die formative Leistungsbeurteilung neue Optionen für individuelle Lernentwicklungen und Fördermaßnahmen (Maier, 2014, S. 26). Insgesamt sollten dabei Fehler zugelassen werden, als willkommener Anlass, daraus zu lernen und somit den Lehr-Lernprozess zu optimieren (Stern, 2010, S. 33). Inwieweit Fehler als Lernchance verstanden werden, hängt von verschiedenen Faktoren ab. Insgesamt sollte eine fehleroffene Lernumgebung, d. h. eine *positive Fehlerkultur*⁴ vorherrschen. Dabei stellt sich die Frage, wie eine positive Fehlerkultur gestaltet werden sollte (Hascher & Hagenauer, 2010, S. 377). Nach Oser, Hascher und Spychiger (1999) sind hierfür zunächst drei Schritte vonnöten: die Fehlererklärung (was ist falsch), die Fehlerursache (warum ist es falsch) und die Fehlerkorrektur (Möglichkeiten zur Fehlerkorrektur). Spychiger, Kuster und Oser (2006, S. 90) identifizieren anhand von Unterrichtsbeobachtungen zwei Grundbedingungen für eine positive Fehlerkultur: ein *positives Unterrichtsklima* und eine *Lernorientierung*. Ein *positives Unterrichtsklima* beinhaltet dabei das Lehrerverhalten im Sinne einer Fehlerfreundlichkeit bzw. der Normtransparenz. Die Fehlerfreundlichkeit bezieht sich auf grundlegend fehleroptimistische Einstellungen, welche Zielverfehlungen als Ausdruck von Fertigkeiten und fehlerhaftes Handeln als Potential zum Erweitern der Handlungsmöglichkeiten erachtet. Die Normtransparenz sagt aus, inwieweit die Lehrperson soziale Normen und Regeln transparent vermittelt.⁵ Bei gegebener Normtransparenz können die Lernenden z. B. nachvollziehen, warum ihr Verhalten im Unterricht kritisiert wird. Als dritte Kategorie für ein positives Unterrichtsklima sollte die Fehlerangst berücksichtigt werden. Fehlerangst bedeutet dabei, dass ein Fehler Angst, Scham bzw. Schuld auslöst. Hierbei ist eine gänzliche Abwesenheit von Fehlerangst nicht unbedingt lernförderlich. Es zeigte sich, dass eher mittlere Werte fehlerkulturpositiv sind. Zu der *Lernorientierung* zählt der Umgang mit Fehlern, d. h. inwieweit eine Fehlerbereitschaft besteht bzw. wie bedeutsam Fehler vom Lernenden eingeschätzt werden (Spychiger u. a., 2006).⁶

⁴ „Eine positive Fehlerkultur, stellt eine Situation dar, in der dem Fehler und dem Fehlermachen Platz eingeräumt wird. Fehlersituationen werden weder vermieden, noch negativ bewertet. An Fehlern wird gearbeitet und der Umgang mit ihnen geübt. Sie werden als integrativer Bestandteil des Lernprozesses anerkannt, gleichzeitig soll die Hemmung gegenüber des Fehlermachens abgebaut werden.“ (Spychiger, Oser, Hascher, & Mahler, 1999, S. 44).

⁵ Soziale Normen adressieren zukünftig zu erwartendes Verhalten, bestimmten Verhaltensregelmäßigkeiten entsprechendes Verhalten sowie gesolltes, desideratives Verhalten und sind mit einem Sanktionsrisiko bei Abweichungen verbunden (Popitz, 1980).

⁶ Die vier Grunddimensionen der Fehlerkultur Lernorientierung, Fehlerfreundlichkeit, Normtransparenz und Fehlerangst entstanden im Rahmen einer Schülerbefragung. Dabei zeigte sich, dass die Skala Lernorientierung die gewichtigste Dimension des Fragebogens darstellte (höchster Varianzanteil).

Im Rahmen der Leistungsbewertung besteht oftmals die Problematik, dass unverbundenes Faktenwissen abgefragt wird. Dies ist der Tatsache geschuldet, dass Feedback oft summativen Charakter hat, d. h. zur abschließenden Bewertung eines Lehr-Lernprozesses für die Notengebung oder eine Selektionsentscheidung herangezogen wird (*summative Leistungsbeurteilung*). Die Herausforderung besteht darin, Tests bzw. andere Bewertungsverfahren zu entwickeln, welche auf das Abprüfen von konzeptuellem Wissen abzielen, um komplexe Lehr-Lernprozesse zu schaffen. Erst wenn Informationen über das Verständnis grundlegender Begriffszusammenhänge zur Verfügung stehen, ist eine sinnvolle Optimierung des Lehr-Lernprozesses im Rahmen der formativen Leistungsbeurteilung unter der Voraussetzung einer positiven Fehlerkultur möglich. Dabei kann auf eine summative Leistungsbeurteilung nicht verzichtet werden, diese sollte jedoch das Abprüfen von konzeptuellem Wissen fokussieren (Maier, 2014, S. 24). Auf diese Weise lässt sich eine summative Leistungsbeurteilung auch formativ nutzen (Black & Wiliam, 2009, S. 7).

Für eine weitere Begriffsbestimmung muss *Feedback* von *Rückmeldungen* abgegrenzt werden. Feedback bezieht sich eng auf den Adressaten und nimmt direkt Bezug zum gezeigten Verhalten bzw. zu der (Lern-)Leistung des Individuums. Rückmeldungen finden im Gegensatz dazu eher verhaltensfern und zeitfern bezogen auf Systeme und Systemeinheiten statt. Als typisches Beispiel können hier die einmal jährlich stattfindenden, externen Schulvergleichsarbeiten genannt werden. Dabei werden die Lernleistungen der Schüler auf Klassenebene aggregiert und mit zeitlicher Distanz schriftlich an die Lehrkräfte zurückgemeldet, um Qualitätsentwicklungsprozesse initiieren zu können. Feedback und Rückmeldungen sollten jedoch aufgrund der Schnittstellen und Überschneidungen nicht als getrennte Konzepte behandelt werden. Die Übertragung von Ergebnissen der Feedbackforschung auf Rückmeldekontexte ist sinnvoll (Müller & Ditton, 2014, S. 13 ff.). Beide Begriffe werden daher für die Arbeit synonym verwendet.

2.1.2 Funktionen und Ebenen von Feedback in Lehr-Lernprozessen

Feedback kann *kognitive*, *metakognitive* und *motivationale* Funktionen haben (Narciss, 2006, S. 78 ff.; Bangert-Drowns u. a., 1991, S. 215) in Abhängigkeit davon, welche Ebene im Lernprozess angesprochen wird. Hierbei gibt es laut Kluger und DeNisi (1996) drei hierarchisch organisierte Bezugsebenen: die aufgabenbezogene Lernebene (*task learning*), die motivationale Aufgabenebene (*task motivation*) und die selbstbezogene Meta-

Ebene (*meta-task processes*).⁷ Hattie hat diesen Ansatz weiterentwickelt und unterscheidet vier Feedback-Ebenen: die *Aufgabenebene*, die *Prozessebene*, die *Ebene der Selbstregulation* und die *Selbstebene* (Hattie, 2014, S. 135 ff.; Hattie & Timperley, 2007). Feedback auf der Aufgaben- und der Prozessebene hat eine kognitive Funktion, Feedback auf der Ebene der Selbstregulation eine metakognitive Funktion und Feedback auf der Selbstebene hat eine motivationale Funktion.

Feedback auf der *Aufgabenebene* gibt an, wie gut eine Aufgabe ausgeführt wurde und ist daher oftmals korrektiv. Das Feedback ist häufig spezifisch und im Unterricht an die ganze Klasse gerichtet. Der Lehrende informiert darüber, inwieweit die Aufgabe richtig ist. Er fordert weitere Antworten ein und bietet zusätzliche Informationen zur Aufgabebearbeitung an, um das Wissen über die Aufgabe zu vergrößern (Hattie, 2014, S. 135). Der Lernende sollte auf dieser Ebene die Fragen „Wohin gehe ich?“ und „Was sind meine Ziele?“ beantworten können (ebd., S. 132). Dies setzt voraus, dass Lehrpersonen die Ziele des Unterrichts angemessen und transparent kommunizieren. Hierbei sollte den Lehrpersonen insbesondere die formative Komponente des Feedbacks bewusst sein. Ziele sollten daher nicht nur die durch den Lehrplan vorgegebenen Lernziele, sondern auch die individuelle Lernleistung betreffen, z. B. auf das Verstehen des Stoffes oder das Verbessern bestimmter Fähigkeiten gerichtet sein und somit auch individuelles Feedback gegeben werden (ebd., S. 133). Das Feedback auf der Aufgabenebene dient insgesamt als wichtige Basis für das Feedback auf der nächsten Ebene (ebd., S. 135).

Das Feedback auf der *Prozessebene* zielt auf die Prozesse und Strategien ab, welche zur Fertigstellung der Aufgabe benötigt werden. Hierbei können dem Lernenden alternative Lösungswege aufgezeigt werden und er kann bei der Entwicklung von kognitiven Lernstrategien, z. B. um Verknüpfungen zwischen verschiedenen Konzepten zu erstellen, unterstützt werden. Das Feedback kann des Weiteren auf das Aufzeigen von Fehlern und Wissenslücken gerichtet sein, d. h. dem Lernenden werden Strategien angeboten, um Fehler zu identifizieren und aus ihnen zu lernen. Hierbei sollten die drei Schritte der Fehlerklärung, der Fehlerursache und der Fehlerkorrektur nach Oser, Hascher und Spychiger (1999) beachtet werden (vgl. Kapitel 2.1.1). Feedback auf dieser Ebene kann ein tiefergehendes Lernen als auf der Aufgabenebene ermöglichen (Hattie, 2014, S. 135 f.). Es

⁷ Die drei Zielebenen sowie weitere theoretische Annahmen bilden die von Kluger und DeNisi (1996) entwickelte Feedback-Interventions-Theorie (FIT).

ermöglicht das Beantworten der Frage „Welche Fortschritte wurden in Richtung Ziel gemacht?“ (ebd., S. 132).

Auf der Aufgaben- und der Prozessebene können kognitive Feedback-Funktionen in Abhängigkeit von der Antwortrichtigkeit differenziert werden (Narciss, 2006, S. 77). Werden bei *falschen* Antworten die Anzahl, der Ort, die Art und/ oder Fehlerursachen hervorgehoben, ist das Feedback informierend. Des Weiteren können fehlende Wissensselemente ergänzt, falsche Wissensselemente korrigiert, ungenaue Wissensselemente angepasst, d. h. abgegrenzt bzw. präzisiert, sowie falsch verknüpfte Wissensselemente restrukturiert werden (Butler & Winne, 1995). Sofern eine *richtige* Antwort gegeben wurde, betätigt Feedback zunächst und kann ferner darüber informieren, warum die Antwort korrekt ist. Des Weiteren können grundsätzlich richtige Antworten auch Ergänzungen oder Anpassungen im Sinne von Abgrenzungen oder Präzisierungen erfordern. Nachfolgende Tabelle 1 fasst die dargestellten kognitiven Feedback-Funktionen in Anlehnung an Butler und Winne (1995) noch einmal zusammen.

Tabelle 1: Differenzierung der kognitiven Feedback-Funktionen nach der Antwortrichtigkeit

Funktion	Richtige Antwort	Falsche Antwort
• Bestätigen	X	
• Informieren	X	X
• Ergänzen	X	X
• Anpassen (Abgrenzen, Präzisieren)	X	X
• Korrigieren		X
• Restrukturieren		X

Auf der *Ebene der Selbstregulation* sollen die Lernenden dabei unterstützt werden, ihr Lernen zu bewerten, zu kontrollieren und selbst zu regulieren, um somit das Feedback im Hinblick auf das Reduzieren der Diskrepanz zwischen Ist- und Sollzustand effektiver zu nutzen. Die Lernenden sollen selbständig entscheiden können, welcher Schritt im Lernprozess als nächster folgt (Hattie, 2014, S. 134 ff.). Hierbei kann das Feedback über metakognitive Strategien sowie ihre Einsatzmöglichkeiten informieren und falsche Strategien ggf. korrigieren. Darüber hinaus hat das Feedback eine lenkende Funktion, wenn Lernende z. B. mittels reflexiver oder sondierender Fragen dazu animiert werden, die Angemessenheit der Lösungsstrategien zu durchdenken (Butler & Winne, 1995) und somit die Fragen zu beantworten „Wann, wo und warum habe ich die Fehler gemacht?“ (Hattie, 2014, S. 136). Das Feedback wird auf der Ebene der Selbstregulation daher nicht nur von

einer externen Quelle angeboten, sondern kann auch durch Erfahrungen selbst generiert werden und somit intern sein (vgl. Kapitel 2.1.3).

Das Feedback auf der *Selbstebene*, welches sich an das „Selbst“ richtet, enthält eine bewertende sowie eine ermutigende Komponente und hat somit eine motivationale Funktion. Auf dieser Ebene besteht die Herausforderung für die Lehrpersonen, dem Lernenden trotz möglicher Fehler das Gefühl zu geben, dass das Bewältigen der Aufgabe überhaupt einen Wert für ihn hat und somit einen Anreiz für künftige Situationen derselben Art schafft (Narciss, 2006, S. 80). An dieser Stelle spielt folglich eine positive Fehlerkultur eine besondere Rolle (vgl. Kapitel 2.1.1). Dabei ist zu beachten, dass voreiliges und unnötiges Lob die Lernenden verwirren kann und folglich den Lernprozess hemmt (Hyland & Hyland, 2001). Die Leistungen der Lernenden sollten demnach angemessen wertgeschätzt werden. Andernfalls kann Lob die Aufmerksamkeit von der Aufgabe, den Prozessen oder der Selbstregulation weg lenken und die Wirkung des Feedbacks somit verwässern (Hattie, 2014, S. 137 f.).

2.1.3 Klassifikation von Feedback in Lehr-Lernprozessen

2.1.3.1 Klassifikation von Feedback in Lehr-Lernprozessen anhand ausgewählter Kriterien

Feedback kann nach der *Komplexität der Feedbacknachricht* (vgl. Kapitel 2.1.3.2), der *Feedbackquelle* (Niegemann u. a., 2008, S. 232; Butler & Winne, 1995), dem *Feedbackempfänger* (Krause, 2007, S. 47), dem *Zeitpunkt der Feedbackgabe* (Kopp & Mandl, 2014, S. 156 f.; Narciss, 2008, S. 137 ff.; Krause, 2007, S. 55), der *Transparenz für den Feedbackempfänger* (Butler & Winne, 1995, S. 264; Bangert-Drowns u. a., 1991, S. 214 f.; Lehmenkühler, Roscher, & Theis, 1976), dem *Präsentationsformat* (Hattie & Wollenschläger, 2014, S. 138; Kopp & Mandl, 2014, S. 155; Krause, 2007, S. 54; Narciss, 2006, S. 139) und der *Anzahl erneuter Lösungsversuche* (Narciss, 2008, S. 138 ff.; Clariana, 1993) klassifiziert werden.

Hinsichtlich der *Feedbackquelle* lässt sich in *internes* und *externes Feedback* unterteilen. Sobald Feedback von einer externen Informationsquelle gegeben wird, handelt es sich um externes Feedback (vgl. Kapitel 2.1.1). Davon abzugrenzen ist *internes Feedback*, welches der Lernende aufgrund von Erfahrungen selbst generieren kann (Niegemann u. a., 2008, S. 232; Butler & Winne, 1995).

Feedback kann sich in Bezug auf den *Feedbackempfänger* entweder an Einzelpersonen (*individuelles Feedback*) oder an Gruppen (*Gruppenfeedback*) richten (Krause, 2007, S. 47) und thematisiert je nach Aggregationsniveau *individuelle* oder *kollektive Aspekte* (Nadler, 1979, S. 312).

Hinsichtlich des *Zeitpunkts der Feedbackgabe* kann zunächst nach *sofortigem* oder *verzögertem Feedback* unterschieden werden. Feedback kann dem Lernenden direkt nach der Bearbeitung einer Aufgabe (sog. *immediate Feedback*) oder erst zu einem späteren Zeitpunkt (sog. *delayed Feedback*) gegeben werden (Kopp & Mandl, 2014, S. 156 f.; Narciss, 2008, S. 137 f.; Krause, 2007, S. 55). Hierbei gibt es keine einheitliche Definition dafür, wann von sofortigem und wann von verzögertem Feedback gesprochen wird (Smith & Kimball, 2010, S. 91). Der jeweilige Zeithorizont für sofortiges bzw. verzögertes Feedback ergibt sich dabei in Abhängigkeit vom Aufgabentyp (Smits, Boon, Sluijsmans, & van Gog, 2008, S. 184). Eine Rückmeldung nach jeder Antwort oder Lösungsstufe stellt sofortiges Feedback dar, welches eher bei einfachen Aufgabentypen sinnvoll

ist. Bei komplexen Aufgabentypen wird sofortiges Feedback nach vollständiger Beendigung einer Aufgabe gegeben. Eine Rückmeldung nach der Bearbeitung jeder Aufgabe würde bei einem Multiple Choice Test beispielsweise einem sofortigen Feedback entsprechen, bei einer Rückmeldung am Ende des Tests würde verzögertes Feedback gegeben werden. Beim Schreiben eines Aufsatzes hingegen wäre es nicht sinnvoll, eine Rückmeldung nach jedem geschriebenen Satz zu erhalten. Daher kann an dieser Stelle bei einer Rückmeldung nach Fertigstellung des Textes von sofortigem Feedback gesprochen werden, verzögertes Feedback entspräche einer Rückmeldung z. B. nach einem Zeitraum von zwei Tagen. Informationen können nach Narciss (2008, S. 138) *gleichzeitig* oder *sequenziell* präsentiert werden, wobei dies von der Komplexität der Feedbacknachricht abhängig ist (vgl. Kapitel 2.1.3.2). Aufgrund einer begrenzten Kapazität des Arbeitsgedächtnisses⁸ empfiehlt es sich, komplexe Feedbackinformationen nacheinander zu präsentieren.

In Bezug auf die *Transparenz für den Feedbackempfänger* kann Feedback einerseits hinsichtlich der Absichtlichkeit entweder *absichtlich* oder *beiläufig* gegeben werden (Butler & Winne, 1995, S. 264; Bangert-Drowns u. a., 1991, S. 214 f.). Wenn die Lehrperson das Ziel verfolgt, den Lernenden über die Angemessenheit seiner Leistung zu informieren, erfolgt dies absichtlich (Bangert-Drowns u. a., 1991, S. 215). Dabei können die Informationen direkt von einer Person zu einer anderen, z. B. Lehrer – Schüler, oder durch sog. „intervening agents“ (ebd.), z. B. über schriftliche Anmerkungen oder einen Computerbildschirm ausgetauscht werden (Butler & Winne, 1995, S. 264; Bangert-Drowns u. a., 1991, S. 215). Feedback kann aber auch eine beiläufige Konsequenz einer zufälligen Interaktion mit der Umwelt, d. h. unbeabsichtigt sein (ebd.). Des Weiteren kann zwischen *explizitem* und *implizitem* oder auch indirektem Feedback unterschieden werden, wobei der Feedbackgeber den Fehler beim impliziten Feedback nicht direkt benennt. Eine fehlerhafte Äußerung kann z. B. mit Hilfe von sogenannten „recasts“ umformuliert werden. Beim expliziten Feedback erhält der Lernende einen konkreten Hinweis auf seinen Fehler und die richtige Lösung (Yilmaz & Granena, 2016; Lehmenkühler u. a., 1976).

Eine weitere Klassifikationsmöglichkeit stellt das *Präsentationsformat* dar. Kopp und Mandl (2014, S. 155 f.) sowie Krause (2007, S. 54) unterscheiden hierbei zwischen *mündlicher* und *schriftlicher* Feedbackgabe. Bei mündlichem Feedback wird die Interpretation

⁸ Gemäß der Cognitive-Load-Theorie macht es die begrenzte Kapazität des Arbeitsgedächtnisses schwierig, viele gleichzeitig angebotene Informationen zu verarbeiten (z. B. Sweller & Chandler, 1994, S. 185).

erleichtert, da neben verbalen auch nonverbale Signale ausgesendet werden. Darüber hinaus ist es dem Lernenden möglich direkte Rückfragen zu stellen. Bei schriftlichem Feedback wird die Feedbackbotschaft nicht durch den Tonfall, die Mimik oder Gestik ergänzt, Formulierungen fallen stärker ins Gewicht und sollten daher mit Bedacht gewählt werden. Da hier keine direkten Rückfragen möglich sind, sollte auf Verständlichkeit⁹ geachtet werden (Krause, 2007, S. 54). Hattie und Wollenschläger (2014, S. 138) berücksichtigen die Besonderheiten des computergestützten Feedbacks und nehmen dieses Format mit auf. Sie differenzieren zwischen *verbalem*, *textbasiertem* und *computergestütztem* Feedback. An dieser Stelle werden keine weiteren Erläuterungen zu der Unterscheidung vorgenommen. Es kann jedoch davon ausgegangen werden, dass verbales Feedback hier lediglich mündliches Feedback umfasst und sich textbasiertes Feedback auf Rückmeldungen in schriftlicher Form bezieht. Narciss (2008, S. 139) geht auf die Möglichkeit ein, Feedback *unimodal* oder *multimodal* zu geben. Informationen werden entweder über nur ein Sinnesorgan (z. B. visuell) oder mehrere präsentiert (z. B. visuell und auditiv).

Die *Anzahl erneuter Lösungsversuche*, die dem Lernenden nach einer als falsch zurückgemeldeten Antwort ermöglicht werden, kann variieren. Dabei kann dem Lernenden entweder kein Versuch, d. h. single-try feedback [STF] oder eine bestimmte Anzahl an Versuchen, d. h. multiple-try feedback [MTF] gewährt werden, dieselbe Aufgabe richtig zu lösen. Eine dritte Möglichkeit ist answer-until-correct [AUC], bei der so lange geantwortet werden kann, bis die Aufgabe richtig gelöst wurde (Narciss, 2008, S. 138; Clariana, 1993). Diese Unterscheidung wird bei computergestütztem Feedback angewendet. Nachfolgende Tabelle 2 fasst die Klassifikationsvorschläge noch einmal zusammen.

Tabelle 2: Klassifikation von Feedback in Lehr-Lernprozessen anhand ausgewählter Kriterien

Klassifikationsvorschlag	Kategorien
Feedbackquelle	intern vs. extern
Feedbackempfänger	Individuum vs. Gruppe
Zeitpunkt der Feedbackgabe	sofort vs. verzögert gleichzeitig vs. sequenziell
Transparenz für den Feedbackempfänger	absichtlich vs. beiläufig implizit vs. explizit
Präsentationsformat	mündlich vs. schriftlich verbal vs. textbasiert vs. computergestützt unimodal vs. multimodal
Anzahl erneuter Lösungsvorschläge	STF vs. MTF vs. AUC

⁹ Hierbei sei auf das „Hamburger Verständlichkeitskonzept“ verwiesen, das Merkmale verständlicher Texte beschreibt. Diese sind Einfachheit, Gliederung/Ordnung, Kürze/Prägnanz und anregende Zsätze (Langer, Schulz von Thun, & Tausch, 2011).

Die Klassifikation nach der Komplexität der Feedbacknachricht wird in nachfolgendem Abschnitt genauer betrachtet.

2.1.3.2 Klassifikation von Feedback in Lehr-Lernprozessen anhand des Kriteriums der Komplexität der Feedbacknachricht nach einfachem und elaboriertem Feedback

Hinsichtlich des Inhalts kann Feedback in Abhängigkeit von der Komplexität der Feedbacknachricht nach *einfachem* und *elaboriertem Feedback* differenziert werden (z. B. Kopp & Mandl, 2014, S. 157; Narciss, 2006, S. 19 ff.). Im Vordergrund steht dabei, welche und wie viele Informationen dem Lernenden zur Verfügung gestellt werden.

Einfaches Feedback unterscheidet zwischen knowledge of performance [KP], d. h. der Anteil der richtig gelösten Aufgaben wird rückgemeldet, knowledge of result [KR], d. h. eine Information darüber, ob die Antwort richtig oder falsch ist, wird gegeben, und knowledge of correct response [KCR], d. h. die richtige Lösung wird mitgeteilt (ebd.).

Elaboriertes Feedback lässt sich ebenfalls weiter untergliedern. Hier sind die Klassifikationsvorschläge jedoch weniger eindeutig. Diese werden in Anlehnung an Narciss (2006, S. 19 ff.) in nachfolgender Tabelle 3 dargestellt.

Tabelle 3: Klassifikationsvorschläge für einfaches und elaboriertes Feedback

Einfaches Feedback	Elaboriertes Feedback		
	<i>Kulhavy & Stock, 1989</i>	<i>Mason & Bruning, 2001</i>	<i>Narciss, 2006</i>
- KP - KR - KCR	- task-specific feedback ¹⁰ - instruction-based - extra-instructional	- topic-contingent - response-contingent - bug-related - attribute-isolation	- KTC - KC - KM - KH - KMC

Kulhavy und Stock (1989, S. 286) unterscheiden zwischen Feedback, das sich überwiegend auf das Umformulieren der richtigen Aufgabe und somit lediglich KCR bezieht (task-specific elaboration), Feedback, das vorwiegend auf den ursprünglichen Lernmaterialien der Instrukionsphase basiert, z. B. Erklärungen, warum die Aufgabe richtig oder falsch ist (instruction-based elaboration) oder Feedback, das zusätzliche Informationen, wie neue Beispiele oder alternative Präsentationen des Lehrstoffs, zur Verfügung stellt

¹⁰ Diese Feedbackart ist im Grunde dem einfachen Feedback zuzuordnen und wird daher bei der weiteren Betrachtung des elaborierten Feedbacks außer Acht gelassen.

(extra-instructional elaboration). Extra-instructional Feedback setzt dabei nicht notwendigerweise das Präsentieren der Lösung voraus. Mason und Bruning (2001) unterscheiden vier Feedback-Arten. Topic-contingent Feedback bietet umfassende Informationen bezüglich der Zielthematik an, welche dabei helfen sollen, die richtige Antwort zu identifizieren, ohne diese direkt zu geben. Response-contingent bzw. extra-instructional Feedback stellt antwortspezifische Informationen bereit. Neben der richtigen Antwort werden hier Informationen gegeben, warum eine Antwort richtig oder falsch ist. Dies kann dabei auch über die Lernmaterialien der Instruktionsphase hinausgehen. Bug-related Feedback gibt dem Lernenden Informationen, die ihm helfen, Fehler zu identifizieren und selbst zu korrigieren, ohne die richtige Lösung direkt zu geben. Beim attribute-isolation Feedback werden zentrale Komponenten des Zielkonzepts hervorgehoben, um das generelle Verständnis zu verbessern. Bei dieser Form muss dabei die richtige Lösung nicht gegeben werden. Die Informationen können sich beim bug-related und beim attribute-isolation Feedback auf Lernmaterialien der Instruktionsphase beziehen oder darüber hinausgehen. Narciss (2006, S. 23) unterteilt fünf Formen. Knowledge on task constraints [KTC] enthält aufgabenbezogene Informationen zur Art der Aufgabe, zu Bearbeitungsregeln, zu Aufgabenanforderungen und/oder zu Teilaufgaben. Knowledge about concepts [KC] sind Informationen, die sich auf aufgabenrelevantes konzeptuelles Wissen beziehen, d. h. Hinweise auf Fachbegriffe, auf den Begriffskontext, Erklärungen oder Beispiele zu Begriffen. Knowledge about mistakes [KM] beinhaltet fehlerbezogene Informationen, d. h. die Anzahl, den Ort, die Art und die Ursachen der Fehler. Knowledge on how to proceed [KH] bezieht sich auf strategisches Wissen, welches für die Aufgabenlösung benötigt wird, d. h. fehlerspezifische Korrekturhinweise, aufgabenspezifische Lösungshinweise, Hinweise auf Lösungsstrategien, Leitfragen sowie Lösungsbeispiele. Knowledge on meta-cognition [KMC] sind Informationen, die für die Regulation des Lernprozesses notwendig sind, d. h. Hinweise zu metakognitiven Strategien und metakognitive Leitfragen. Bei allen fünf Formen kann sich das Feedback auf Lernmaterialien der Instruktionsphase beziehen oder darüber hinausgehen. Narciss (2006, S. 22) betont, dass diese Feedbackformen unabhängig von der richtigen Lösung gegeben werden sollten.

Nachfolgende Abbildung 1 fasst die Zusammenhänge der verschiedenen Klassifikationsvorschläge noch einmal zusammen.

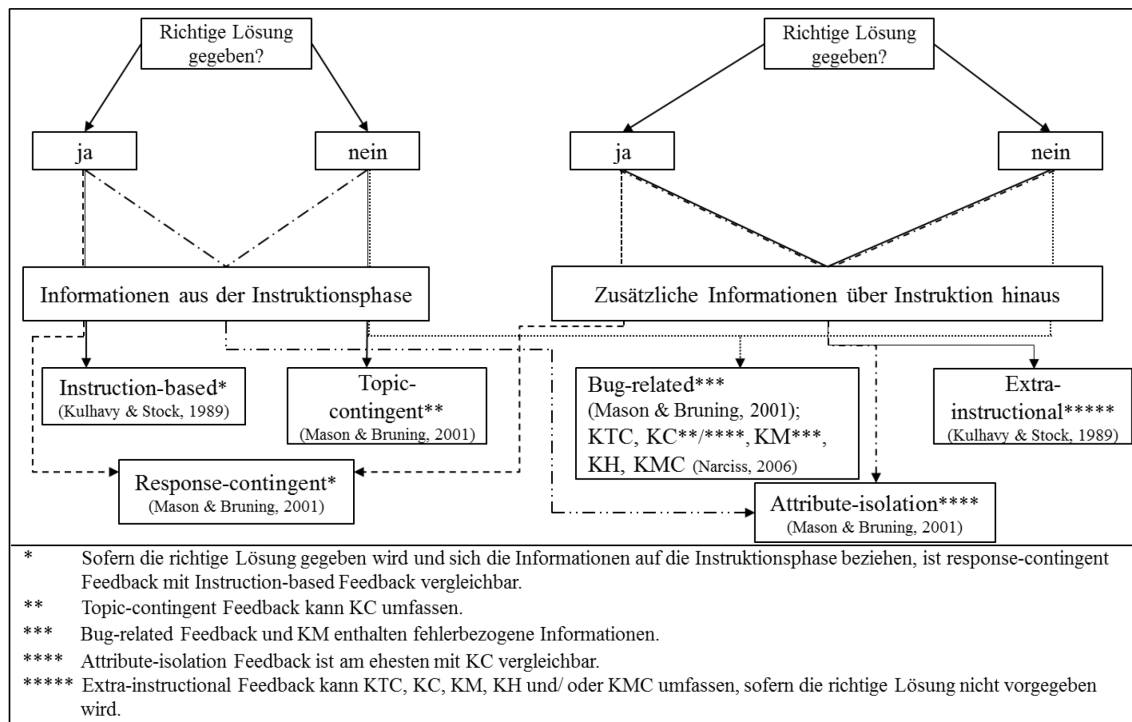


Abbildung 1: Zusammenhang der Klassifikationsvorschläge zum elaborierten Feedback

Der Abbildung 1 ist zu entnehmen, dass sich die verschiedenen Feedbackarten zunächst danach differenzieren lassen, inwieweit dem Lernenden vorab die richtige Lösung gegeben wird. Des Weiteren kann Feedback Informationen aus der Instruktionsphase und/oder zusätzliche Informationen über die Instruktion hinaus bereitstellen. Das instruction-based Feedback sowie das topic-contingent Feedback umfassen Informationen der Instruktionsphase, wobei der Lernende beim topic-contingent Feedback die richtige Lösung selbst finden muss. Zusätzliche Informationen über die Instruktion hinaus werden beim extra-instructional Feedback bereitgestellt. Hierbei kann die richtige Lösung entweder vorgegeben werden oder nicht. Das response-contingent Feedback sowie das attribute-isolation Feedback enthalten entweder Informationen aus der Instruktionsphase oder informieren darüber hinaus, wobei dem Lernenden lediglich beim response-contingent Feedback die richtige Antwort präsentiert wird. Beim attribute-isolation Feedback kann die richtige Lösung vorgegeben werden oder nicht. Bug-related Feedback sowie die fünf Formen von Narciss (2006) KTC, KC, KM, KH und KMC enthalten Informationen der Instruktionsphase oder geben zusätzliche Informationen, ohne den Lernenden unmittelbar mit der richtigen Lösung zu konfrontieren.

Die verschiedenen elaborierten Feedbackarten der genannten Autoren lassen sich teilweise vergleichen. Sofern die richtige Lösung gegeben wird und sich die Informationen

auf die Instruktionsphase beziehen, ist response-contingent Feedback mit instruction-based Feedback vergleichbar. Topic-contingent Feedback kann KC umfassen. Sowohl Bug-related Feedback als auch KM enthalten fehlerbezogene Informationen. Attribute-isolation Feedback ist am ehesten mit KC vergleichbar. Extra-instructional Feedback kann KTC, KC, KM, KH und/ oder KMC umfassen, sofern die richtige Lösung nicht vorgegeben wird.

Insgesamt zeigt sich, dass die Klassifikation von Narciss (2006) zum elaborierten Feedback am umfassendsten und ausdifferenziertesten ist, da sie als einziges strategisches und metakognitives Wissen berücksichtigt. Die Formen lassen sich dabei auch auf Feedback übertragen, bei dem die Präsentation der richtigen Lösung voraus geht.

2.1.4 Feedbackmodelle zur Feedbackrezeption und/oder -gestaltung

2.1.4.1 Makromodelle von Kulhavy und Stock (1989) sowie von Bangert-Drowns u. a. (1991)

Nach dem Auflösungsgrad der Betrachtung lassen sich Feedbackmodelle in Makro- und Mikromodelle unterscheiden. In diesem Abschnitt werden die Modelle von Kulhavy und Stock (1989) und Bangert-Drowns u. a. (1991) vorgestellt, welche Makromodelle darstellen (Strijbos & Müller, 2014). Hierbei handelt es sich um eher abstrakte Rahmenmodelle der Feedbackverarbeitung aus einer Makroperspektive. Sie geben einen Überblick über den Rezeptionsprozess. Personale Variablen werden dabei lediglich auf einer allgemeinen Ebene als individuelle Merkmale zusammengefasst.

Response-Certitude-Modell (Kulhavy & Stock, 1989)

Das Response-Certitude-Modell von Kulhavy und Stock (1989) ist ein Modell, in welchem kognitive Prozesse bei der Bearbeitung von Lernaufgaben und Feedback sowie inhaltliche Merkmale unterschiedlicher Feedbackarten (vgl. Kapitel 2.1.3) berücksichtigt werden. Als Grundlage dieses Modells dienten Studien zur Wirkung des informativen Feedbacks bei Lernen mit Texten in Kombination mit Multiple-Choice-Aufgaben. Nach dem Durcharbeiten eines Lehrtextes werden Multiple-Choice-Aufgaben gelöst, ohne auf den Text zugreifen zu können. Danach erhalten die Lernenden ein Feedback entsprechend experimentell variierten Bedingungen und lösen abschließend erneut Multiple-Choice-Aufgaben. Basierend auf dem systemtheoretischen Prinzip kann beim Bearbeiten von Lernaufgaben mit Feedback eine Diskrepanz zwischen der Antwort der Lernenden und der Feedbackinformation entstehen. Kulhavy und Stock (1989) nehmen dabei an, dass

das Ausmaß dieser Diskrepanz vom Grad der Antwortsicherheit bestimmt wird. Eine hohe Antwortsicherheit kann zu einer hohen Diskrepanz führen, sofern das Feedback die Korrektheit der Antwort nicht bestätigt. Der Lernende sollte daraufhin nach Gründen für die Diskrepanz suchen. Dies führt zu einer intensiven Anstrengung des Lernenden und folglich zu einer hohen Feedbackbearbeitungszeit sowie einer längeren Behaltensdauer. Bei korrekten Antworten kann das Feedback bei einer niedrigen Antwortsicherheit die Antwortsicherheit für zukünftige Aufgaben erhöhen. Der Zusammenhang zwischen Diskrepanz und Feedbackbearbeitungszeit sowie zwischen Antwortsicherheit und Behaltensdauer konnte empirisch bestätigt werden (einen Überblick geben Mory, 2004; Kulhavy & Wager, 1993).

Kulhavy und Stock (1989) gehen des Weiteren davon aus, dass die elaborativen Feedbackkomponenten bei Antworten mit niedriger Antwortsicherheit am effektivsten sind. Dies konnte nicht bestätigt werden. Teilweise zeigte sich sogar elaboriertes Feedback bei falschen Antworten mit hoher und nicht mit niedriger Antwortsicherheit am effektivsten (z. B. Butterfield & Metcalfe, 2001; Swindell, 1992). Die möglichen Wirkungen unterschiedlicher Feedbackarten können folglich nicht mit der Variable Antwortsicherheit hinreichend erklärt werden. Dies könnte u. a. daran liegen, dass es neben der Antwortsicherheit noch weitere Einflussfaktoren gibt. Hierbei könnten kognitive, motivationale und metakognitive Voraussetzungen der Lernenden bzw. deren Informationsverarbeitungsstrategien eine Rolle spielen (Narciss, 2014, S. 47 f.). Ergänzend zu diesen empirisch bestätigten Einschränkungen des Response-Certainty-Modells zeigen sich theoretische und methodische Defizite insbesondere im Hinblick auf die Generalität des Modells. Die Antwortsicherheit beruht auf einer subjektiven Einschätzung und wird mittels fünfstufiger Likert-Skalen erfasst, was die Reliabilität und Validität der damit erhobenen Daten infrage stellt. Das Modell wurde lediglich für Aufgaben mit einer eindeutigen Antwort empirisch bestätigt, eine Übertragung auf Aufgaben mit mehreren Antworten ist fraglich, da die Diskrepanz hierbei nicht eindeutig berechnet werden kann (Mory, 2004).

Mindful-Processing-Modell von Bangert-Drowns u. a. (1991)

Im Modell von Bangert-Drowns u. a. (1991)¹¹ stellt das *mindful processing* von Informationen ein zentrales Konzept dar. Das *mindful processing* von Informationen wurde von Salomin und Globerson im Jahr 1987 eingeführt und beinhaltet eine zielgerichtete, sinnvolle, aktive und intensive Verarbeitung von Informationen. Gemäß Bangert-Drowns u. a. (1991) ist Feedback dann besonders wirksam, wenn den Lernenden ein *mindful processing* der Feedbackinformationen ermöglicht wird.

Das Modell beinhaltet insgesamt *fünf* Stufen. Der individuelle Ausgangszustand des Lernenden stellt die *erste* Stufe dar und ist gemäß dem Modell durch dessen Vorwissen, Interesse, Ziele bzw. Zielorientierung sowie einen bestimmten Grad der Selbstwirksamkeit¹² geprägt (Bangert-Drowns u. a., 1991, S. 217). Auf der *zweiten* Stufe werden mittels Lernaufgaben oder Fragen bei dem Lernenden Informationsverarbeitungsstrategien aktiviert, um diese bewältigen zu können (Bangert-Drowns u. a., 1991, S. 217). Der Grad der Aktivierung ist dabei von den kognitiven Anforderungen der Aufgabe sowie den individuellen Wissensstrukturen abhängig (Narciss, 2006, S. 51). Auf der *dritten* Stufe wählt der Lernende eine Lösung aus bzw. antwortet auf die Fragen, wobei das Ausmaß der Antwortsicherheit die Erwartungen des Lernenden an das Feedback beeinflusst (Bangert-Drowns u. a., 1991, S. 217). Im Anschluss an die Aufgabenbearbeitung erhält der Lernende Feedback. In dem Modell wird dabei dem Problem begegnet, dass sich ein Lernender, der jederzeit (also auch vor Beginn der Aufgabenbearbeitung und Aktivierung von Informationsverarbeitungsstrategien) auf das Feedback zugreifen kann, nicht aufmerksam mit der Lernaufgabe auseinandersetzt und die im Rahmen des Feedbacks bereitgestellten Antworten möglicherweise einfach kopiert. Kulhavy (1977) prägte für das Phänomen des vorzeitigen Bereitstellens von Feedback den Begriff der „presearch availability“ (S. 217). Mit Hilfe der Feedbackinformationen evaluiert der Lernende auf der *vierten* Stufe seine Lösung bzw. Antwort. Hierbei spielt die Antwortsicherheit erneut eine Rolle. Bangert-Drowns u. a. (1991) stützen sich dabei auf die Erkenntnisse von Kulhavy und Stock (1989). Als Resultat dieses Evaluationsprozesses gleicht der Lernende auf der *fünften* Stufe sein Wissen und Interesse sowie seine Ziele an. Darüber hinaus wirkt das

¹¹ Bangert-Drowns u. a. (1991) nutzten als Ausgangspunkt für ihr Modell empirische und theoretische Erkenntnisse aus vorhandenen Reviews (z. B. Kulik & Kulik, 1988; Kulhavy, 1977) und erweiterten das Modell anschließend mit Hilfe einer eigenen umfangreichen Metaanalyse (Näheres hierzu in Kapitel 2.2.1).

¹² Selbstwirksamkeit bezieht sich auf die subjektive Überzeugung eines Lernenden, Ziele aufgrund eigener Kompetenzen erreichen zu können (Woolfolk, 2008, S. 404).

Feedback auch auf die Selbstwirksamkeit. Das Ergebnis dieses Anpassungsprozesses bildet den Ausgangszustand für den nächsten Lernprozess.

Die fünf Stufen werden in nachfolgender Abbildung 2 in Anlehnung an Dempsey, Driscoll und Swindell (1993, S. 40) dargestellt.

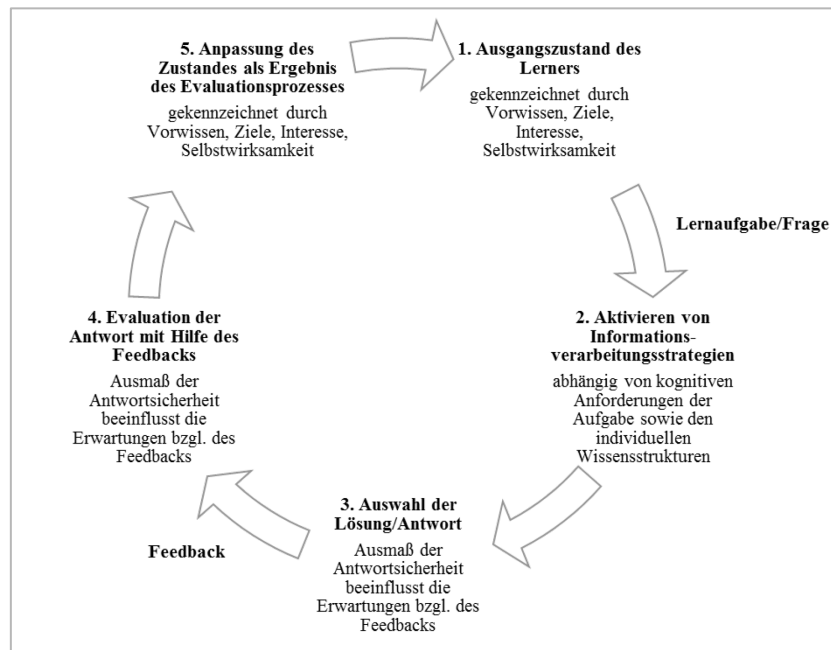


Abbildung 2: Mindful-Processing-Modell

Das Modell berücksichtigt motivationale und kognitive Faktoren, die die Wirksamkeit von Feedback beeinflussen. Es bezieht neben der Antwortsicherheit im Vergleich zum Modell von Kulhavy und Stock (1989) weitere Faktoren mit ein, über deren Zusammenwirken aufgrund der Vielzahl und Komplexität möglicher Lehr-Lernsituationen jedoch keine Aussagen getroffen werden. Das Modell wurde in einer Metanalyse überprüft und erweitert (vgl. Kapitel 2.2.1). Es wird insgesamt als wichtiger theoretischer Ansatz in der Feedbackforschung diskutiert (z. B. Shute, 2008; Narciss, 2006, 2008; Mory, 2004).

2.1.4.2 Mikromodell von Narciss (2006)

Das folgende Kapitel fokussiert das Mikromodell von Narciss (2006). Mikromodelle differenzieren zwischen verschiedenen Charakteristika des Feedback-Empfängers, welche in Prozessverläufe der Feedbackverarbeitung eingeordnet werden. Dabei werden kognitive, metakognitive und motivationale Regulationsmechanismen betrachtet und auf konkrete Entscheidungsheuristiken bei der Aufnahme von Feedbackbotschaften hingewiesen (Strijbos & Müller, 2014).

Neben dem Modell von Narciss (2006) existieren noch weitere Mikromodelle, z. B. von Butler und Winne (1995) oder von Kluger und DeNisi (1996). Auf eine detaillierte Darstellung dieser Modelle wird an dieser Stelle verzichtet, da bei Narciss (2006) die Modelle von Butler und Winne (1995) und Kluger und DeNisi (1996) integriert werden.

Narciss (2006, 2008) entwickelte das *Interactive-Two-Feedback-Loop-Model* [ITFL-Modell] mit dem Ziel, neben der Bedeutung der individuellen Faktoren auch die Bedeutung von Faktoren der inhaltlichen und formalen Qualität des externen Feedbacks zu berücksichtigen. Des Weiteren lassen sich mit Hilfe des Modells Prinzipien zur Gestaltung und Evaluation von informativen tutoriellen Feedbackarten¹³ ableiten. Es verbindet Erkenntnisse aus Lehr-Lernmodellen mit systemtheoretischen Annahmen sowie Einsichten der bisherigen Feedbackforschung (Narciss, 2006, S. 69).

Als Ausgangspunkt dienen die grundlegenden Komponenten eines Regelkreises, die Regelstrecke (der zu regulierende Prozess mit Stellglied, Regelgröße und Messfühler) und die Regeleinrichtung (um die Regelstrecke zu beeinflussen bzw. zu kontrollieren). Narciss (2006) unterscheidet in diesem Modell zwischen einem externen und einem internen Regelkreis. Der externe Regelkreis wird durch das Lehrmedium (z. B. Lehrperson, Mitschüler, computerbasiertes Lehr-Lernsystem) repräsentiert, der interne Regelkreis durch die lernende Person. Die Wirkung des Feedbacks hängt von dem Zusammenwirken beider Regelkreise ab.

Sowohl der externe als auch der interne Regelungsprozess basieren auf der für die Aufgabenbewältigung relevanten Regelgröße. Diese legt genau fest, was reguliert werden soll, z. B. Leistung, Fehlerrate oder Lernfortschritt. Eine Regelgröße kann kognitive, metakognitive und/oder motivationale Leistungskriterien beinhalten. Der aktuelle Wert der Regelgröße (Ist-Wert), z. B. die aktuelle Leistung, wird vom Lehrmedium sowie dem Lernenden selbst erfasst. Die aktuelle Ausprägung der Regelgröße wird dabei vom externen und internen Messfühler gemessen. Der externe Ist-Wert des Lernenden wird im externen Regelkreis weiterverarbeitet. Hierzu vergleicht das Lehrmedium den externen Ist- mit dem extern definierten Soll-Wert. Der externe Soll-Wert beruht dabei auf einer externen Repräsentation der Aufgabenanforderungen und hängt von situativen Merkmalen

¹³ Laut Narciss (2014, 2006) macht es eine moderne Wissensgesellschaft erforderlich, Feedback zu geben, welches eine aktive Konstruktion und Kommunikation von Wissen anregt und nicht nur evaluative, ergebnisorientierte Informationen bereitstellt. Daher verwendet sie hierfür den Begriff „informatives tutorielles Feedback“.

der Lehr-Lernsituation (z. B. Lehr-Lernziele und Lehr-Lerninhalte) ab. Der Vergleich zwischen externem Soll-Wert und externem Ist-Wert wird in der externen Regeleinrichtung vorgenommen. Ausgehend hiervon wird eine externe Stellgröße als externes Feedback gebildet (z. B. Rückmeldung der Fehler) und an den Lernenden weitergegeben. Der interne Ist-Wert wird in der internen Regeleinrichtung der lernenden Person verarbeitet und mit dem internen Soll-Wert verglichen. Der interne Soll-Wert ergibt sich aus der subjektiven Repräsentation der für die Bewältigung von Lernaufgaben notwendigen Anforderungen, welche im Wesentlichen von den individuellen Eingangsvoraussetzungen des Lernenden (Vorwissen, metakognitive und motivationale Strategien, individuelle Lernziele) abhängen. Des Weiteren wird das externe Feedback dem internen Soll-Wert sowie dem internen Ist-Wert gegenübergestellt. Resultierend aus diesen Vergleichsprozeden generiert der Lernende eine interne Stellgröße, die an den Eingang der Regelstrecke (= Stellglied) weiter geleitet wird und dort die Grundlage für die Auswahl und Aktivierung korrigierender Maßnahmen bildet. Diese können sich wiederum auf die externe Regelgröße auswirken. Die interne Stellgröße kann auch zur Veränderung der subjektiven Repräsentation der Aufgabenanforderungen führen. Bei einer großen Diskrepanz zwischen externem und internem Soll-Wert werden die Regelkreise für eine Anpassung mehrere Male durchlaufen (Narciss, 2006, 2008).

Nachfolgende Abbildung 3 stellt die zentralen Komponenten des ITFL-Modells und deren Zusammenhänge noch einmal zusammenfassend dar.

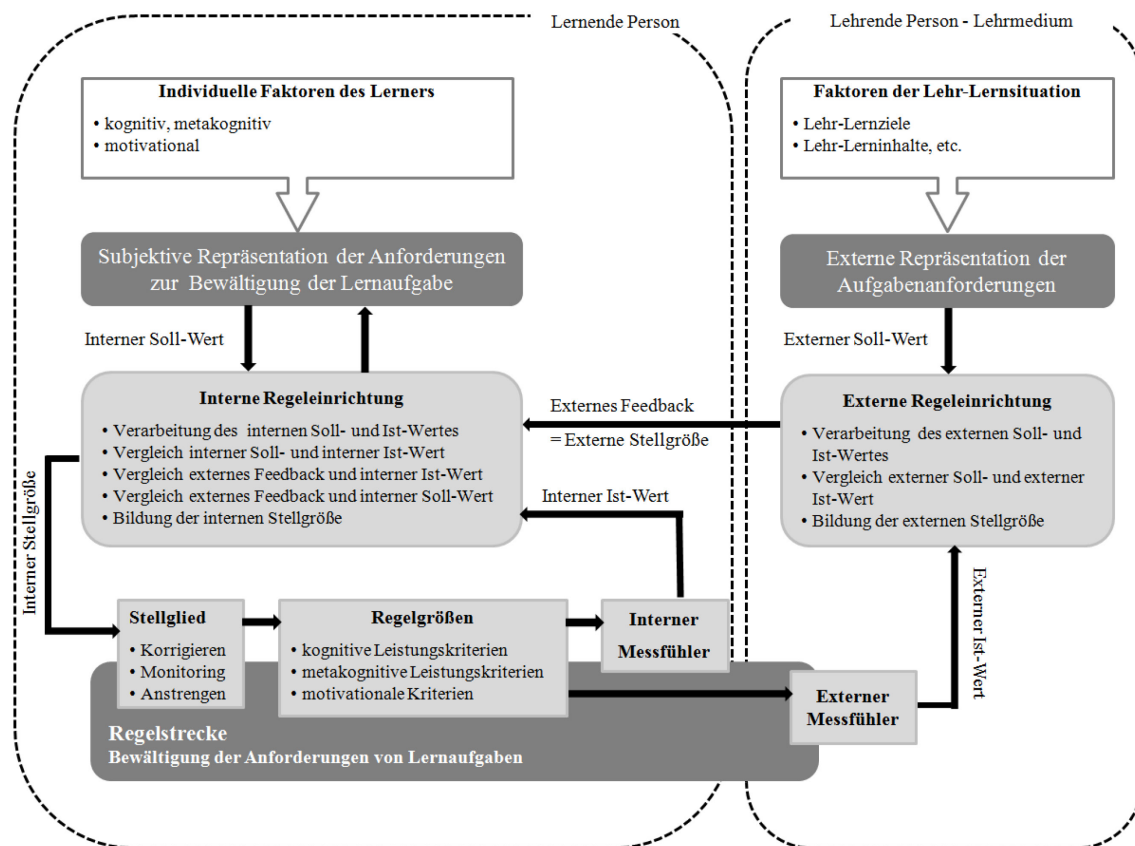


Abbildung 3: Interactive-Two-Feedback-Loop-Model (Narciss, 2006, S. 70)

Um Lehr-Lernprozesse mit Hilfe von externem Feedback optimal zu regulieren, müssen verschiedene Bedingungen erfüllt sein (Narciss, 2006, S. 74 f.). In Lehr-Lernsituationen sollten dabei zunächst die Anforderungen, welche mit den relevanten Lernzielen, den relevanten Lerninhalten und mit den Lernaufgaben verknüpft sind, präzise bestimmt werden. Hierzu ist es notwendig, die Lernziele sorgfältig zu definieren und zu operationalisieren, wobei an dieser Stelle auf Lernzieltaxonomien oder Bildungsstandards zurückgegriffen werden kann (Narciss, 2006, S. 75). Auf diese Weise können valide und reliabel überprüfbare Lernkriterien im Sinne von Soll-Werten abgeleitet werden. Um eine Regelgröße mittels Korrekturmaßnahmen angemessen zu regulieren, ist es des Weiteren erforderlich, typische, systematische Fehler und Probleme beim Bewältigen der Aufgabenanforderungen zu erkennen und Informationen zum Beheben der Fehler bzw. Probleme bereitzustellen. Einen weiteren Einfluss hat die Qualität des Feedbacks, d. h. wie gut es dem Lehrmedium gelingt, den Diskrepanzwert zwischen Soll und Ist in eine Stellgröße zu transformieren. Hierbei kann es z. B. der Fall sein, dass bei komplexen Lernaufgaben mehrere Feedbackinformationen kombiniert werden müssen (Narciss, 2014, S. 68 f.). Des Weiteren spielen Qualitätsfaktoren der internen Feedbackschleife eine Rolle. Die

Wirksamkeit des externen Feedbacks hängt dabei u. a. von der Fähigkeit bzw. der Motivation der lernenden Person ab, ihren Lernprozess zu überwachen bzw. Informationen zu nutzen.

Aus dem ITFL-Modell leitet Narciss (2006) verschiedene Gestaltungs- und Evaluationsprinzipien ab. Feedback ist eine multidimensionale Instruktionsmaßnahme und beinhaltet mindestens *drei Facetten*: eine *funktionale* (vgl. Kapitel 2.1.2), eine *formale* (vgl. Kapitel 2.1.3.1) und eine *inhaltliche* Facette (vgl. Kapitel 2.1.3.2). Hinsichtlich der funktionalen Facette unterscheidet Narciss (2006) zwischen kognitiven, motivationalen und metakognitiven Funktionen. Die formale Facette legt u. a. den Zeitpunkt und den Modus der Feedbackrepräsentation fest. Die inhaltliche Facette besteht aus mindestens zwei Komponenten: der evaluativen und der informativen Komponente. Die evaluative Komponente bezieht sich dabei auf einfaches Feedback und die informative Komponente auf elaboriertes Feedback (Narciss, 2006, S. 82 f.). Welche Funktionen, welcher Inhalt und welche Form bei der Gestaltung des Feedbacks am besten geeignet sind, hängt von *situativen* und *individuellen* Faktoren ab. Zu den situativen Faktoren zählen Lernziele, Lernaufgaben sowie mögliche Quellen für Fehlkonzepte¹⁴ und Fehler. Die individuellen Faktoren umfassen Vorwissen und Fertigkeiten sowie die Lernmotivation (Narciss, 2006, S. 81 ff.). Die Gestaltungsprinzipien werden in nachfolgender Abbildung 4 in Anlehnung an Narciss (2006, S. 87) dargestellt.

¹⁴ Narciss (2006) erläutert an dieser Stelle den Begriff „Fehlkonzept“ nicht näher. Nach Rutke (2007) sind Fehlkonzepte resistente Vorstellungen, welche Schüler von fachlichen Inhalten haben, die trotz der Erarbeitung der Fachinhalte im Unterricht, oft bestehen bleiben. Da nichtwissenschaftliche Konzepte im alltäglichen Kontext für den Lernenden nützlich sein können, wird oft eher von "alternativen Vorstellungen" oder "Alltagsvorstellungen" gesprochen (Hank, 2015, S. 9). Dabei sollten die Schülervorstellungen nicht als Fehlvorstellungen abgetan werden, sondern von den Lehrpersonen gezielt im Unterricht eingesetzt werden.

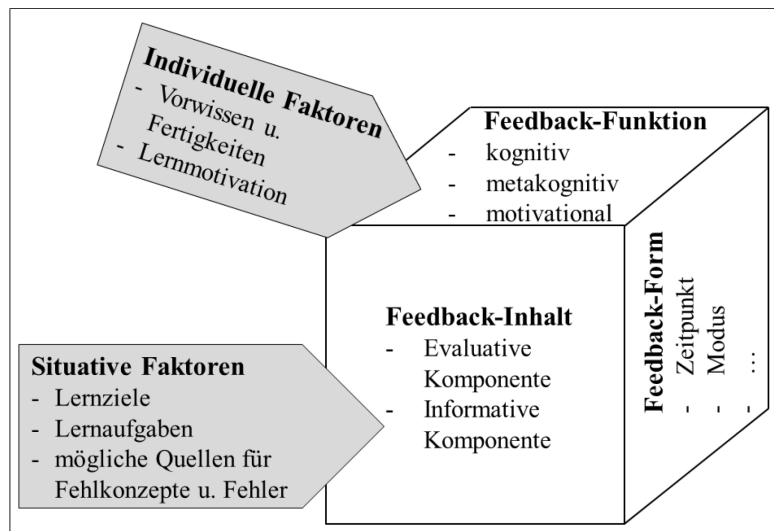


Abbildung 4: Übersicht über die Gestaltungsprinzipien des ITFL-Modells (in Anlehnung an Narciss, 2006, S. 87)

Bei der Evaluation von Feedback in Lehr-Lernsituationen sollte die Multidimensionalität des Feedbacks ebenso beachtet werden. Darüber hinaus sollten neben den kognitiven, auch motivationale und metakognitive Wirkungen untersucht werden. Aufgrund dessen, dass Feedback Wirkungen zu unterschiedlichen Zeitpunkten entfalten kann, sollten darüber hinaus Daten während des gesamten Lernprozesses und nicht nur in einem Posttest erfasst werden. Da die Wirkung des Feedbacks von individuellen und situativen Faktoren beeinflusst wird, gilt es diese Faktoren entsprechend zu kontrollieren (Narciss, 2014, S. 78).

2.1.4.3 Rahmenmodell von Strijbos und Müller (2014)

Mit dem Rahmenmodell von Strijbos und Müller (2014) werden kognitive und situative Aspekte aufgegriffen und bestehende Modelle, die primär die Rezeption von Feedback thematisieren (z. B. Kluger & DeNisi, 1996; Butler & Winne, 1995), um interaktionale und relationale Aspekte, im Hinblick auf die Beziehung zwischen dem Sender und dem Empfänger von Feedback sowie den Kontext, ergänzt. Auf diese Weise soll dem Defizit Rechnung getragen werden, dass bisherige Modelle personale Variablen bei der Verarbeitung von Feedback bisher noch nicht aus interaktionaler Perspektive integriert haben. Das Modell hat einen allgemeingültigen Charakter. Es ist universell einsetzbar und bezieht sich nicht wie viele Modelle der Feedbackforschung explizit auf bestimmte Kontexte (Strijbos und Müller, 2014, S. 86, 92).

Strijbos und Müller (2014) beziehen allgemeine Sender-Nachricht-Empfänger-Modelle der Kommunikation mit ein und fokussieren dabei insbesondere das Vier-Seiten-Modell

von Schulz von Thun (2015). Das Modell wurde zwar bisher nicht empirisch untersucht, kann jedoch als bildhafte Metapher für die Unterschiede zwischen einer gesendeten und einer empfangenen Nachricht verwendet werden. Schulz von Thun (2015) unterscheidet vier Seiten einer Nachricht: Sachinhalt, Appell, Selbstoffenbarung und Beziehung. Der Sachinhalt bezieht sich auf die Sachinformation, welche mit der Nachricht übermittelt wird. Aus einer Nachricht geht im Rahmen des Beziehungsaspekts hervor, was der Sender von seinem Gesprächspartner hält und wie er zu ihm steht. Der Selbstoffenbarungsaspekt beinhaltet Informationen über den Sender, ist also das, was der Sender beim Übermitteln der Nachricht von sich kundgibt. Der Appell ist das, wozu der Sender den Gesprächspartner veranlassen möchte. Eine Nachricht enthält demnach viele Botschaften und beinhaltet gemäß Schulz von Thun (2015) immer die vier Seiten einer Nachricht. Der Empfänger nimmt daher eine Nachricht auch auf vier Ebenen wahr. Sender und Empfänger gewichten die vier Seiten abhängig von individuellen Merkmalen unterschiedlich, weshalb Missverständnisse entstehen können. Der Empfänger muss das Feedback folglich nicht unter derselben Zielperspektive interpretieren, wie es der Sender gegeben hat. Ein Hinweis auf eine inhaltliche Verbesserungsmöglichkeit kann dabei z. B. als Abwertung der Person des Empfängers verstanden werden (Schulz von Thun, 2015).

Individuum und Umwelt bzw. Situation bedingen sich entsprechend aktueller sozialisationstheoretischer Ansätze (Hurrelmann, Grundmann, & Walper, 2008) gegenseitig und werden im Rahmenmodell von Strijbos und Müller (2014) mittels Überschneidungen der jeweiligen Bereiche visualisiert (vgl. Abbildung 5). Die Gestaltung und die Verarbeitung des Feedbacks werden von personalen Variablen, Bedingungen der Situation bzw. des Kontextes sowie Repräsentationen bezüglich des Gegenübers beeinflusst. Das Rahmenmodell hat einen hybriden Charakter: Aus der Makroperspektive betrachtet es den gesamten Prozess von der Feedback-Gestaltung bis zur Feedback-Verarbeitung. Personale Faktoren des Senders und Empfängers von Feedback werden nur überblicksartig eingeordnet. Indem auf einen Abgleich von gesendetem Feedback (GFB) und internem Feedback (IFB) und die diesbezüglichen Selbstregulationsprozesse verwiesen wird, lässt sich auch eine Schnittstelle zu den Modellen auf der Mikroebene erkennen (Strijbos & Müller, 2014, S. 88).

Das Modell geht davon aus, dass die Gestaltung der Feedbacknachricht von personalen Variablen des Senders beeinflusst wird. Darüber hinaus aktiviert der Feedbackgeber gespeicherte mentale Repräsentationen bezüglich des Feedback-Empfängers und richtet das

Feedback danach aus. Der Sender antizipiert dabei eine mögliche Rückkopplung durch den Empfänger nach dem Feedback und nimmt prospektiv eine Anpassung der Nachricht vor. Auf diese Weise können negative Konsequenzen für den Feedbackgeber selbst vermieden werden. Dabei kann angenommen werden, dass zwischen den personalen Variablen des Senders sowie der Repräsentation bezüglich des Feedback-Empfängers eine Interaktion stattfindet (visualisiert mittels einer unterbrochenen Linie in Abbildung 5). Sender mit starkem Mitgefühl könnten z. B. eine negative Nachricht positiv verzerren, sofern Sie dem Empfänger eine Verletzlichkeit unterstellen. Bei der Gestaltung des Feedbacks nehmen des Weiteren Aspekte der Situation Einfluss. Unter Bezugnahme auf verschiedene Autoren (z. B. Hattie & Wollenschläger, 2014; Kopp & Mandl, 2014; Black & Wiliam, 2009) schlagen Strijbos und Müller (2014, S. 89) folgende Faktoren vor: kulturelle Rahmung, Modus der Übermittlung (z. B. schriftlich vs. mündlich), Setting (z. B. Individual vs. Gruppenfeedback), Grad der Verbindlichkeit bzw. Anonymität, Zeitpunkt (z. B. unmittelbar vs. verzögert) und Funktion einer Feedbackintervention (formativ vs. summativ). Feedback kann von Kultur zu Kultur unterschiedlich wahrgenommen werden und sollte daher hinsichtlich der kulturellen Rahmung immer in einem kulturellen Kontext, d. h. in Bezug auf die Merkmale der beteiligten Personen und Gruppen, betrachtet werden

Das gesendete Feedback (GFB) wird durch den Empfänger ebenfalls in einem mehrfach interaktionalen Prozess verarbeitet. Das empfangene Feedback (EFB) stellt die endgültige interne Repräsentation des Feedbacks dar und muss nicht mit dem GFB übereinstimmen (vgl. Vier-Seiten-Modell). Das EFB entsteht mittels der Interaktion folgender Faktoren: der personalen Aspekte des Feedbacksenders, der personalen Aspekte des Feedbackempfängers, der Merkmale des GFB, der Repräsentationen des Feedbackempfängers bezüglich des Feedbacksenders, der Ausprägung des IFB und den kontextualen Bedingungen (z. B. der Aufgabe). Die mentalen Repräsentationen bezüglich des Senders können sich auf stabile individuelle Merkmale (z. B. Geschlecht), Beziehungsaspekte (z. B. formaler Statusunterschied) oder variable Aspekte (z. B. Glaubwürdigkeit) beziehen. Darüber hinaus spielt IFB eine Rolle, welches durch Selbstmonitoring unter Einbezug kognitiver, metakognitiver und motivationaler Prozesse entsteht. Das IFB wird dabei mit externen Stimuli abgeglichen. Insbesondere bei negativem Feedback können so Handlungsimpulse zur Zielerreichung ausgelöst werden.

Selbstverteidigung aktiviert werden (Strijbos & Müller, 2014, S. 94 f.). Die Lehrperson kann an dieser Stelle Belastungen, welche durch eine interne Attribution auftreten kann (z. B. Anstrengung), abschwächen, indem sie z. B. die Aufgabenschwierigkeit variiert. Der Lernende wird bei einer lohnenswerten Anstrengung negatives Feedback eher akzeptieren (Semmer & Jacobshagen, 2010).

Selbstwirksamkeitsüberzeugungen beinhalten den Glauben bestimmte Situationen meistern zu können (Bandura, 1997, 1991). Positive Rückmeldungen führen zu einer Steigerung der Selbstwirksamkeitserwartung, infolgedessen zu einer Erhöhung individueller Ziele und zu einer möglichen Leistungssteigerung. Voraussetzung ist hierbei, dass die Ursachen der Rückmeldeinhalte als kontrollierbar eingeschätzt werden. Bei negativem Feedback wirkt eine ausgeprägte Selbstwirksamkeitserwartung immunisierend gegen die Abwertung des eigenen Selbst. Die Ursachen der Rückmeldeinhalte werden als kontrollierbar eingeschätzt und aktiv nach Lösungsmöglichkeiten zur Zielerreichung gesucht. Kritik kann somit eher akzeptiert werden und wird als weniger bedrohlich oder belastend empfunden. Insgesamt sollten positive Selbstwirksamkeitsüberzeugungen beim Feedbackempfänger gefördert werden (Strijbos und Müller, 2014, S. 99).

Hinsichtlich der *Zielorientierung* beziehen sich Strijbos und Müller (2014, S. 99 ff.) auf die klassische Differenzierung nach Elliot und Dweck (1988), welche zwischen der Performanz- oder Leistungsorientierung und der Lernorientierung unterscheiden. Zielorientierungen beinhalten grundsätzlich Vorstellungen und Überzeugungen, welche der Lernende mit dem Prozess der Zielerreichung verbindet. Lernende beabsichtigen bei der Performanz- oder Leistungsorientierung ihre Fähigkeiten so zu präsentieren, dass sie insbesondere verglichen mit anderen kompetent wirken und Misserfolg vermeiden. Bei der Lernorientierung stehen das Verbessern von Kompetenzen, die Entwicklung von Fähigkeiten und das erfolgreiche Bewältigen einer Aufgabe im Fokus des Lernenden.

In Bezug auf die *Persönlichkeitsdimensionen (Big 5)* werden Neurotizismus bzw. emotionale Stabilität, Extraversion (Geselligkeit, Optimismus), Offenheit für Erfahrungen, Verträglichkeit (Altruismus, Empathie, Nachgiebigkeit) sowie Gewissenhaftigkeit unterschieden (Costa & McCrae, 1992). Eine geringe emotionale Stabilität kann mit negativen Affekten nach dem Feedback einhergehen. Menschen mit starker emotionaler Stabilität sind eher offen für Feedback und auch motiviert, Feedbackbotschaften weiter zu verwenden. Die Extraversion kann sich indirekt über affektive Reaktionen auf die Akzeptanz des Feedbacks auswirken. Optimistische Menschen suchen eventuell eher nach weiteren

Rückmeldungen als pessimistische Menschen. Offenheit gegenüber neuen Erfahrungen kann zur Folge haben, dass diese Menschen Feedback mehr wertschätzen. Personen mit einer hohen Verträglichkeit zeigen eine hohe Akzeptanz gegenüber dem Feedback, suchen nach weiterem Feedback und setzen sich auch eher neue Ziele als Personen mit niedriger Verträglichkeit. Die Gewissenhaftigkeit kann ebenfalls einen positiven Einfluss auf die Verwendung von Feedback haben. Verantwortungsbewusste Menschen fühlen sich verpflichtet, Rückmeldeergebnisse zu verwenden. Sie suchen weiteres Feedback und zeigen eher Verhaltensänderungen, welches auf Feedback zurückzuführen ist (Strijbos & Müller, 2014, S. 101 ff.).

Hinsichtlich des *formalen Status* kann zwischen vertikalen und horizontalen Sender-Empfänger-Konstellationen unterschieden werden. Im Lehr-Lernkontext ist dies die Beziehung zwischen dem Lehrer und dem Lernenden bzw. die Beziehung zwischen mindestens zwei Lernenden. Bei einer vertikalen Konstellation zeigt sich dabei ein Unterschied bei formalem Status und Macht. Die Übermittlung des Feedbacks geschieht überwiegend abwärtsgerichtet und ist am Anfang der Bildungsbiographie von einem Machtungleichgewicht gekennzeichnet. Im Zuge der Schullaufbahn bis hin zum tertiären Bereich neigen die Lernenden mehr und mehr dazu, die Lehrkräfte, die Lerninhalte und das Feedback in Frage zu stellen. Beim Feedback auf horizontaler Ebene existieren keine Machtunterschiede aufgrund des formalen Status. Jedoch schätzen Lernende Rückmeldungen von Peers¹⁵ oft weniger kompetent ein (Strijbos & Müller, 2014, S. 104 ff.). Der *informelle Status* einer Person wird im Wesentlichen von der vorhandenen oder zugeschriebenen Kompetenz bestimmt. Beim Feedbackempfänger beeinflusst das Vorwissen gemäß Strijbos und Müller (2014, S. 108 f.) u. a. den Wissenserwerb. Darüber hinaus spielt die Antwortsicherheit eine entscheidende Rolle. Sie hängt vom domänenspezifischen Wissen ab und beeinflusst die Akzeptanz und Nutzung von Feedback (ebd.). Auf Seiten des Feedbacksenders existieren verschiedene Vorschläge, wie zugeschriebene Kompetenz operationalisiert werden kann. Strijbos, Narciss und Dünnebier (2010) entwickelten z. B. einen Fragebogen, welcher die wahrgenommene Adäquatheit (Fairness, Nützlichkeit, Akzeptanz) von den Lernenden einschätzen lässt.

¹⁵ Peer oder Peergroup bezeichnet eine Gruppe von Gleichen, mit besonderem Fokus auf das Merkmal der Altersgleichheit. Daher hat sich auch die Bezeichnung Gleichaltrigengruppe durchgesetzt. Peer-Groups entstehen im schulischen bzw. außerschulischen Kontext als informelle Gruppen und teilen gemeinsame Werte und Verhaltensweisen (Tenorth & Tippelt, 2012, S. 552).

Im Rahmen der *zwischenmenschlichen Beziehung* kann erneut zwischen vertikalen und horizontalen Konstellationen differenziert werden. Beim Abwärtsfeedback kann die von den Lernenden wahrgenommene zwischenmenschliche Nähe einen großen Einfluss haben. Auf der horizontalen Ebene kann eine freundschaftliche Beziehung zu Überbewertungen führen. Andererseits führt zwischenmenschliche Nähe möglicherweise zu einer geschützten Atmosphäre, in der kritische Äußerungen ohne größeres Risiko möglich sind. Das Rahmenmodell von Strijbos und Müller (2014) berücksichtigt als erstes Feedbackmodell interaktionale Aspekte und integriert damit alle am Feedbackprozess beteiligten Akteure mit ihren Wechselbeziehungen. Es dient daher als Basis für die weiteren Ausführungen.

2.1.4.4 Erweitertes Rahmenmodell zur Gestaltung und Verarbeitung von Feedback

Im folgenden Kapitel 2.1.4.4 wird das Rahmenmodell um die Inhalte des Kapitels 2.1.1 und die Gestaltungsprinzipien des ITFL-Modells von Narciss (2006) aus Abbildung 4 erweitert, um einen umfassenden Überblick über die Einflussfaktoren auf die Gestaltung und Verarbeitung von Feedback zu geben sowie wesentliche Gestaltungskriterien für Feedback abzuleiten.

Ausgangspunkt stellt das Rahmenmodell von Strijbos und Müller (2014) dar (vgl. Abbildung 5). Die Komponenten der Gestaltung des Feedbacks, die personalen Variablen sowie die Bedingungen der Situation bzw. des Kontextes werden dabei erweitert bzw. angepasst (vgl. Abbildung 6).

Die personalen Variablen lassen sich nach dem Feedbacksender und dem Feedbackempfänger differenzieren. Beim Feedbackempfänger beeinflusst der formale sowie der informelle Status die Feedbackgestaltung. Der informelle Status beinhaltet dabei zum einen die vorhandene und zum anderen die zugeschriebene Kompetenz, welche sich z. B. über die Adäquatheit des Feedbacks operationalisieren lässt. Hierbei wird auf die Ausführungen Strijbos u. a. (2010) zurückgegriffen. Kritisch zu hinterfragen ist jedoch, inwieweit der informelle Status tatsächlich über die Einschätzung der Adäquatheit des Feedbacks messbar gemacht werden kann, da neben der Kompetenz des Feedbacksenders noch andere Faktoren eine Rolle spielen können, welche an dieser Stelle nicht isolierbar sind. Auf Seiten des Feedbackempfängers werden die von Strijbos und Müller (2014) vorgeschlagenen Faktoren aufgegriffen. Neben dem informellen Status sind das die Attribution, die Selbstwirksamkeit sowie die Big 5. Die Lernmotivation lässt sich nach Schiefele

(2009, S. 164) in extrinsische und intrinsische Motivation, Leistungsmotivation, Interesse und Zielorientierung unterscheiden. Demnach kann die von Strijbos und Müller (2014) thematisierte Zielorientierung als personale Variable in Bezug auf den Feedbackempfänger dem von Narciss (2006, S. 75 f.) vorgeschlagenen Einflussfaktor der Lernmotivation untergeordnet werden. Der informelle Status beinhaltet das Vorwissen sowie die Antwortsicherheit (Strijbos und Müller, 2014). Einen weiteren Einflussfaktor stellen die Fertigkeiten dar. Diese werden bei Narciss (2006) nicht näher spezifiziert. Einige Studien, insbesondere im Bereich des Fremdsprachenlernens, untersuchten den Einfluss verschiedener Fähigkeiten, z. B. der Sprachfähigkeit, auf die Lernwirksamkeit des Feedbacks (z. B. Yilmaz & Granena, 2016; Li, 2013; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; DeKeyser, 2000). Neben den Fertigkeiten sollten daher die Fähigkeiten mit in das erweiterte Rahmenmodell einbezogen werden. Vor dem Hintergrund der Bedeutsamkeit einer positiven Fehlerkultur werden die Faktoren Fehlerangst sowie der Umgang mit Fehlern von Spychiger u. a. (2006) mit aufgenommen (vgl. Kapitel 2.1.1). Als interpersonaler Faktor wird die zwischenmenschliche Beziehung aufgegriffen (Strijbos und Müller, 2014).

Bei den situativen Faktoren wird die kulturelle Rahmung von Strijbos und Müller (2014) einbezogen. Die weiteren Einflussfaktoren sind eher der inhaltlichen Ausgestaltung des Feedbacks zuzuordnen. Von Narciss (2006) werden die Lernaufgaben (insbesondere die Aufgabenanforderungen) sowie die möglichen Quellen für Fehlkonzepte und Fehler übernommen. Hierbei spielt auch die Komplexität des Aufgabenformats eine große Rolle. Eine Lückentextaufgabe ist z. B. weniger geeignet, um komplexere Anforderungen zu überprüfen. Sobald höhere Anforderungen an eine Aufgabe gestellt werden, steigt die Aufgabenkomplexität und es ist folglich auch möglich, komplexere Ziele zu erreichen. Dabei muss dieses Ziel nicht notwendigerweise ein vorgegebenes Lernziel sein und somit die sachliche Bezugsnorm betreffen. Es kann auch ein Vergleich mit der Gruppe (soziale Bezugsnorm) oder dem eigenen Lernfortschritt (individuelle Bezugsnorm) vorgenommen werden (vgl. Kapitel 2.1.1). Der von Narciss (2006) vorgeschlagene Einflussfaktor Lernziele wird daher an dieser Stelle um die soziale und die individuelle Bezugsnorm erweitert. Des Weiteren wird Feedback davon beeinflusst, inwieweit die Leistungsbeurteilung einen formativen oder summativen Charakter hat (vgl. Kapitel 2.1.1). Unter Bezug auf Spychiger u. a. (2006) werden die zwei weiteren Dimensionen einer positiven Fehlerkul-

tur, die Fehlerfreundlichkeit und die Normtransparenz als situative Faktoren mit einbezogen (vgl. Kapitel 2.1.1). Aus Gründen der Vollständigkeit wird die Art der Lernsituation als ein weiterer Einflussfaktor aufgenommen. Hierbei kann es z. B. ausschlaggebend sein, ob in einer natürlichen oder computergestützten Umgebung gelernt wird.

Die Gestaltungskriterien Feedback-Inhalt, Feedback-Funktion sowie Feedback-Form werden von Narciss (2006) übernommen, da hierzu bei Strijbos und Müller (2014) keine genaueren Aussagen darüber getroffen werden. Bei der Ausgestaltung dieser drei Bereiche wird auf Kapitel 2.1.2 und 2.1.3 zurückgegriffen. Hinsichtlich der Komplexität des Feedback-Inhalts kann demnach zwischen einfachem Feedback (KP, KR, KCR) und elaborem Feedback unterschieden werden. Feedback kann kognitive (auf der Aufgaben- und Prozessebene), metakognitive (auf der Ebene der Selbstregulation) und motivationale Funktionen (auf der Selbstebene) haben. Bezüglich der Feedback-Form kann zwischen der Feedbackquelle, dem Feedbackempfänger, dem Zeitpunkt der Feedbackgabe, der Transparenz für den Feedbackempfänger, dem Präsentationsformat und der Anzahl erneuter Lösungsvorschläge differenziert werden.

Die in Abbildung 6 dargestellten (inter-)personalen und situativen Faktoren sowie die Gestaltungskriterien dienen als Strukturierungshilfe für die Darstellung der empirischen Ergebnisse zum Feedback in nachfolgendem Kapitel 2.2.

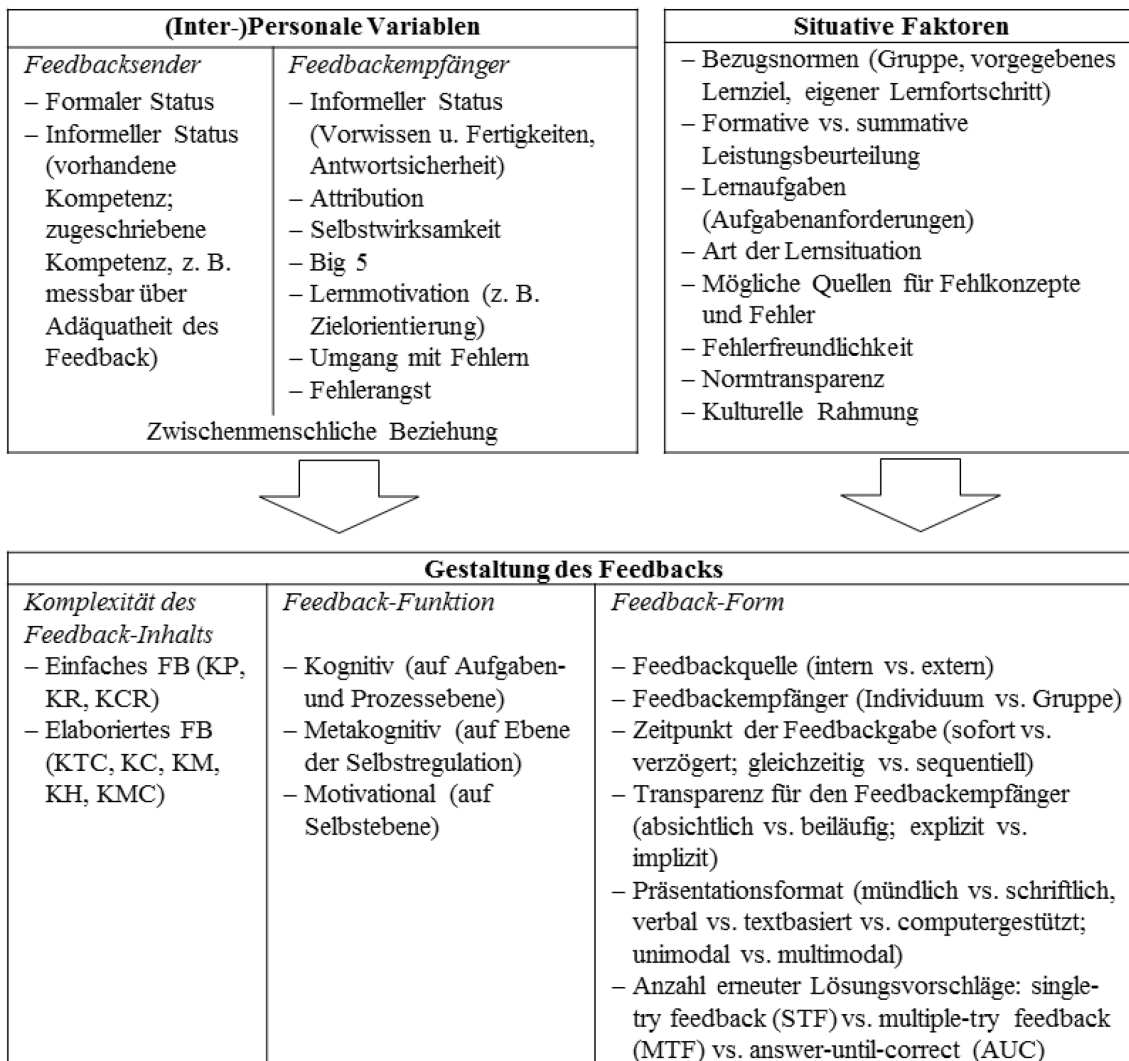


Abbildung 6: Erweitertes Rahmenmodell zur Gestaltung und Verarbeitung von Feedback

2.2 Empirische Befunde zur Lernwirksamkeit des Feedbacks

2.2.1 Überblick über Metaanalysen zur Lernwirksamkeit von Feedback

In einer sehr umfassenden Studie hat Hattie (2015) eine Synthese von mehr als 800 Metaanalysen vorgenommen und damit auf mehr als 50.000 Studien mit über 80 Millionen Lernenden zurückgegriffen. Er identifizierte 138 Faktoren, welche sich in sechs Bereiche (Lernende, Elternhaus, Schule, Curricula, Lehrperson, Unterrichten) gruppieren lassen und das Lernen beeinflussen.

Feedback ist ein Faktor im Bereich des Unterrichtens und gemäß Hattie (2015, S. 206 ff.) einer der größten Einflussfaktoren auf das Lernen. In insgesamt 23 Metaanalysen wurde der Effekt von Feedback auf das Lernen untersucht. Hierbei wurden 1.287 Studien herangezogen und 2.050 Effekte herausgefiltert. Feedback erzielte dabei eine Effektstärke von $d=0,73$.¹⁶ Die Effektstärken weisen jedoch eine beträchtliche Variabilität auf (Hattie, 2015, S. 207). Einen sehr kleinen Effekt erzielt dabei z. B. das Loben durch die Lehrperson ($d=0,12$). Ein mittlerer Effekt kann z. B. erreicht werden, sobald dem Lernenden die Lernziele bekannt sind ($d=0,63$). Eine sofortige Feedbackgabe hat einen sehr großen positiven Effekt auf das Lernen ($d=1,15$).

Zu den bedeutendsten Metaanalysen, welche verschiedene Studien im Hinblick auf die Lernwirksamkeit zusammenfassen, zählen die Metaanalysen von Bangert-Drowns u. a. (1991) sowie Kluger und DeNisi (1996), von Shute (2008, S. 156) als „two landmark meta-analyses“ bezeichnet. Diese werden nachfolgend vorgestellt.

Bangert-Drowns u. a. (1991) bezogen in ihre Metaanalyse Studien ein, welche die Effektivität von Feedback in verschiedenen Lehr-Lernsituationen untersuchten (vgl. auch Kapitel 2.1.4.1). Insgesamt berechneten die Autoren aus 40 Studien 58 Effektstärken, wobei diese stark variierten (von $-0,83$ bis $1,42$). Die durchschnittliche Effektstärke beträgt $d=0,26$. Circa ein Drittel der Studien wies negative Effekte auf. Um mögliche Erklärungen für diese großen Schwankungen der Effektstärken zu finden, führten Bangert-Drowns

¹⁶ Die Berechnung von Effektstärken empfiehlt sich als Ergänzung zur Durchführung von Signifikanztests, da die Stichprobengröße einen Einfluss auf die statistische Signifikanz hat (Döring & Bortz, 2016, S. 668). Bei großen Stichproben können sich auch sehr kleine Effekte als statistisch signifikant erweisen. Bei Cohen (1988, S. 24 ff.) gelten Werte von $d=0,2$ als klein, Werte von $d=0,5$ gelten als mittelgroß und Werte von $d=0,8$ gelten als groß. Hattie (2015, S. 11) geht im Rahmen seiner Synthese der Metaanalysen von einer anderen Einteilung aus: $d=0,20$ klein; $d=0,40$ mittel und $d=0,60$ groß.

u. a. (1991) eine Moderatoranalyse durch. Im Ergebnis ließen sich vier Moderatorvariablen¹⁷ festhalten (vgl. Tabelle 4):

- 1) Wurde die sog. presearch availability (vgl. Kapitel 2.1.4.1) kontrolliert, ergab sich im Mittel ein positiver Effekt ($d=0,46$), wurde sie nicht kontrolliert, ergab sich ein negativer Effekt ($d=-0,08$).
- 2) Die Wirksamkeit von Feedback variierte in Abhängigkeit von der Feedbackart. KR führte zu keinem Effekt ($d=-0,08$). Feedback, das dem Lernenden in irgendeiner Weise die richtige Antwort (KCR: $d=0,22$; AUC: $d=0,53$) sowie weitere Informationen in Form von elaboriertem Feedback ($d=0,53$) präsentierte, zeigte einen kleinen bzw. mittlere Effekte. Vergleichende Studien in Bezug auf die Komplexität des Feedbackinhalts konnten dabei keinen Vorteil des elaborierten Feedbacks gegenüber KCR nachweisen. Bangert-Drowns u. a. (1991, S. 234) argumentieren damit, dass elaboriertes Feedback eher beim Herausbilden von Begriffsstrukturen in komplexen Situationen zum Tragen kommt, die Inhaltsbereiche der berücksichtigten Studien jedoch eher zu einfach waren.
- 3) Die Durchführung eines Vortests reduzierte die Wirksamkeit von Feedback (nur Nachtest $d=0,40$; Vortest $d=-0,07$). Ein Vortest kann die Wirksamkeit von Feedback verringern, da dieser durch die Aktivierung des Vorwissens bereits ein mindful processing der Lernaufgaben (vgl. Kapitel 2.1.4.1) anregen kann (Bangert-Drowns u. a., 1991, S. 233). Infolgedessen werden die Lernmaterialien intensiver bearbeitet und die Fehlerrate so verringert. Das Feedback wird daher nicht so oft benötigt wie bei einer weniger intensiven Bearbeitung mit evtl. daraus resultierender hoher Fehlerrate.
- 4) Weiterhin beeinflusste die Art der Lernsituation die Wirksamkeit von Feedback. Im programmierten¹⁸ ($d=-0,04$) und computergestützten Unterricht¹⁹ ($d=0,22$) war Feedback weniger effektiv als beim verstehenden Lesen²⁰ ($d=0,48$) und herkömmlichen

¹⁷ Moderatorvariablen beeinflussen die Wirkung einer unabhängigen Variable auf eine abhängige Variable (Bortz & Döring, 2006, S. 3).

¹⁸ Programmierter Unterricht beinhaltet dabei eine Reihe von Abschnitten, welche zusammen einen Text bilden. Nach jedem Abschnitt werden Testfragen beantwortet und der Lernende erhält Feedback, bevor er zum nächsten Abschnitt gelangt (Bangert-Drowns u. a., 1991, S. 219).

¹⁹ Nicht alle in der Metaanalyse einbezogenen und am Computer durchgeführten Untersuchungen wurden als computergestützter Unterricht codiert, sondern ggf. den anderen drei Formen zugeordnet. Als computergestützt wurde lediglich solcher Unterricht bezeichnet, der Drill-And-Practice und tutorielle Programme enthielt und den anderen drei Formen nicht zugeordnet werden konnte (Bangert-Drowns u. a. 1991, S. 219).

²⁰ Die integrierten Studien zum verstehenden Lesen beinhalteten Textabschnitte, gefolgt von Fragen zum Textverständnis. Nach dem Beantworten der Fragen erhielten die Lernenden Feedback und setzten anschließend mit dem nächsten Textabschnitt fort (Bangert-Drowns u. a. 1991, S. 219).

Prüfungssituationen²¹ ($d=0,63$). Bangert-Drowns u. a. (1991) argumentieren, dass in programmiertem bzw. computergestütztem Unterricht nur sehr kleine Informationsteile präsentiert werden und anschließend mittels stark lenkender Fragen überprüft wird, was dabei gelernt wurde. Fragen beim verstehenden Lesen sowie in herkömmlichen Prüfungssituationen zielen darauf ab, größere und komplexere Inhaltsgebiete zu beleuchten.

Aus Tabelle 4 kann entnommen werden, wie groß die Effekte in Abhängigkeit von der Ausprägung der vier Variablen auf die Wirkung von Feedback sind. Dabei wurden lediglich die Faktoren berücksichtigt, dessen Einfluss gemäß Bangert-Drowns u. a. (1991) isoliert betrachtet werden konnte. Bei einigen Faktoren z. B. beim Zeitpunkt der Feedbackgabe konnte nicht genau identifiziert werden, inwieweit die Wirkung eher auf eine computergestützte Umgebung zurückzuführen ist.

Insgesamt zeigt sich, dass ein mindful processing besonders dann gegeben ist,

- wenn Feedback-Informationen nicht jederzeit zur Verfügung stehen,
- wenn das Feedback weitere Informationen in Form von elaboriertem Feedback umfasst,
- wenn kein Vortest eingesetzt wird, welcher das Vorwissen zusätzlich aktiviert,
- wenn das Feedback nach einem verstehenden Lesen oder herkömmlichen Prüfungssituationen gegeben wird.

All diese Faktoren zeigen mittlere Effekte.

²¹ Die Studien hierzu wurden im traditionellen Unterricht durchgeführt. In diesem Rahmen absolvierten die Lernenden einen Test und erhielten anschließend Feedback. Nach einer gewissen Zeitspanne wurde erneut ein ähnlicher Test durchgeführt. Das Ziel dieser Studien bestand darin, die Lernenden mittels Feedback beim Korrigieren ihrer Prüfungsleistungen zu unterstützen.

Tabelle 4: Mittlere Effektstärken verschiedener Moderatorvariablen auf die Feedbackwirkung

Moderatorvariable	mittlere Effektstärke
Kontrolle der presearch availability	
• unkontrolliert	-0,08
• kontrolliert	0,46
Feedbackart	
• KR	-0,08
• KCR	0,22
• AUC	0,53
• Elaboriertes Feedback	0,53
Methodik	
• nur Nachtest	0,40
• Vortest	-0,07
Art des Unterrichts	
• programmierter Unterricht	-0,04
• Computergestützter Unterricht	0,22
• Textverständnis	0,48
• Prüfung in traditionellem Unterricht	0,63

Gemäß Hattie (2015, S. 208) stellt die Metaanalyse von Kluger und DeNisi (1996) die systematischste Studie zu den Effekten verschiedener Feedbackarten dar. Aus insgesamt 131 Studien wurden 607 Effektstärken bestimmt. Die korrigierte mittlere Effektstärke beträgt $d=0,41$.²² Insgesamt zeigt sich jedoch eine beträchtliche Variabilität der Feedbackeffekte hinsichtlich der Vorzeichen und der Stärke. Die Autoren stellten fest, dass 33 Prozent der betrachteten Effekte negativ sind. Aus diesen Ergebnissen rechtfertigten Kluger und DeNisi (1996) eine Moderatorenanalyse. Ausgangspunkt hierfür bilden drei Zielebenen, die bei der Regulation der Aufgabenbewältigung beteiligt sind: die aufgabenbezogene Lernebene, die aufgabenbezogene Motivationsebene und die selbstbezogene Metaebene (vgl. Kapitel 2.1.1). Kluger und DeNisi (1996) gehen dabei u. a. davon aus, dass Feedbackeffekte auf die Leistung durch Feedbackinhalte, welche die Aufmerksamkeit auf die selbstbezogene Metaebene lenken (z. B. Schulnoten bzw. Lob oder Tadel), vermindert werden. Des Weiteren können Feedbackeffekte auf die Leistung erhöht werden, wenn der Feedbackinhalt die Aufmerksamkeit auf die aufgabenbezogene Motivations- und Lernebene lenkt (z. B. Präsentieren der richtigen Lösung (KCR), velocity).²³ Im Ergebnis der Metanalyse wurden einige Effektstärken ausgeschlossen, so dass letztendlich 470 Effektstärken berücksichtigt wurden, was eine korrigierte mittlere Effektstärke von

²² Die korrigierte Effektstärke nach Klauer (1993) berücksichtigt die Unterschiedlichkeit der Gruppengröße.

²³ Die weiteren drei Annahmen, welche im Rahmen der Feedback-Interventions-Theorie aufgestellt wurden, können bei Kluger und DeNisi (1996) nachgelesen werden. Die ersten beiden Annahmen konnten mittels der Metaanalyse bestätigt werden und werden daher im Rahmen der Arbeit erläutert.

$d=0,38$ ergab. Extern vermittelte Rückmeldungen führen also im Mittel zu moderaten Leistungsanstiegen. Nachfolgende Tabelle 5 zeigt die signifikanten Moderatorvariablen der Metaanalyse. Diese basieren auf Einstufungen der Variablenausprägungen von zuvor geschulten Beurteilern. Die Ausprägungen wurden mit der metaanalytisch bestimmten durchschnittlichen Effektstärke des Feedbacks korreliert. Signifikante positive bzw. negative Korrelationen deuten darauf hin, dass der durchschnittliche Feedbackeffekt durch die Moderatorvariablen verstärkt bzw. vermindert wird. Lediglich die Moderatorvariablen mit signifikanten Korrelationen wurden in Tabelle 5 berücksichtigt. Die dichotomen Variablen wurden von den Beurteilern in die Kategorien „Ja“ oder „Nein“ eingeordnet. Bei den kontinuierlichen Variablen werden die Effekte der oberen und unteren Quartile präsentiert.

Die Ergebnisse zeigen, dass sich die durchschnittlichen Feedbackeffekte auf die Leistung verstärken, wenn das Feedback die richtige Lösung präsentiert (KCR; $d=0,25$ vs. $d=0,43$), wenn computerbasiertes Feedback bereitgestellt wird ($d=0,23$ vs. $d=0,41$), wenn Informationen über die individuelle Leistungsentwicklung (velocity Feedback) gegeben werden ($d=0,28$ vs. $d=0,55$) und wenn Feedback mit Zielsetzungen bzw. -vereinbarungen kombiniert wird ($d=0,30$ vs. $d=0,51$). Dies bestätigt die Annahme, dass die Feedbackeffekte auf die Leistung durch Feedbackinhalte, welche die Aufmerksamkeit auf die aufgabenbezogene Lern- bzw. Motivationsebene lenken, erhöht werden. Hinsichtlich der Aufgabenkomplexität wird die Annahme bestätigt, dass Feedback zu komplexen Aufgaben ($d=0,03$) geringere Effekte als Feedback zu einfachen Aufgaben ($d=0,55$) erzielt. Kluger und DeNisi (1996) begründen dies mit der moderierenden Wirkung der Aufgabenkomplexität. Demnach kann Feedback zwar die Motivation erhöhen. Sobald eine Aufgabe jedoch eine hohe Komplexität aufweist und somit höhere kognitive Ressourcen erfordert, führt die durch das Feedback erlangte höhere Motivation zu keiner besseren Leistung. Hinsichtlich der Feedbackhäufigkeit sind ähnliche Effekte für Feedback, das selten ($d=0,39$) und häufig ($d=0,32$) bereitgestellt wird, erkennbar. Kluger und DeNisi (1996) trafen hierzu keine konkrete Annahme. Bei den Arten der Lernaufgaben zeigt sich ein mittlerer Effekt bei Gedächtnisaufgaben ($d=0,69$). Bei technischen Aufgaben führt Feedback sogar zu einer Leistungsverschlechterung ($d=-0,11$). Bei Aufgaben, die eine Regelbefolgung erfordern, ist ein geringer Effekt erkennbar ($d=0,19$).

Je mehr selbstwertbedrohende Informationen Feedback gibt (zwischen $d=0,47$ und $d=0,08$), wenn das Feedback die Feedbackempfänger entmutigt ($d=-0,14$ vs. $d=0,33$), Lob

zum Ausdruck bringt ($d=0,09$ vs. $d=0,34$) sowie in verbaler Form vermittelt wird ($d=0,23$ vs. $d=0,37$), schwächen sich die Feedbackeffekte auf die Leistung ab. Entmutigendes Feedback geht sogar mit einer Leistungsverschlechterung einher ($d=-0,14$). Die Annahme, dass die Feedbackeffekte auf die Leistung durch Feedbackinhalte, welche die Aufmerksamkeit auf die selbstbezogene Metaebene lenken, vermindert werden, kann somit ebenfalls bestätigt werden.²⁴

Tabelle 5: Mittlere Effektstärken signifikanter Moderatorvariablen auf die Feedbackwirkung

Moderatorvariable	Korrigierte mittlere Effektstärke	
Aufgabenbezogene Lernebene		
• KCR	<i>Ja</i>	0,43
	<i>Nein</i>	0,25
• Feedbackhäufigkeit		0,39 ^a – 0,32 ^b
• Aufgabenkomplexität		0,55 ^a – 0,03 ^b
• computerbasiertes Feedback	<i>Ja</i>	0,41
	<i>Nein</i>	0,23
Lernaufgabe		
• Gedächtnisaufgabe	<i>Ja</i>	0,69
	<i>Nein</i>	0,30
• technische Aufgabe	<i>Ja</i>	-0,11
	<i>Nein</i>	0,36
• Aufgaben, die Regelbefolgung erfordern	<i>Ja</i>	0,19
	<i>Nein</i>	0,36
Aufgabenbezogene Motivationsebene		
• Feedback über die individuelle Leistungsentwicklung (velocity Feedback)	<i>Ja</i>	0,55
	<i>Nein</i>	0,28
• zusätzliche Zielsetzungsmaßnahme	<i>Ja</i>	0,51
	<i>Nein</i>	0,30
Selbstbezogene Metaebene		
• Grad der Bedrohung des Selbstwertgefühls durch Feedback		0,47 ^a – 0,08 ^b
• entmutigendes Feedback	<i>Ja</i>	-0,14
	<i>Nein</i>	0,33
• Lob	<i>Ja</i>	0,09
	<i>Nein</i>	0,34
• verbales Feedback	<i>Ja</i>	0,23
	<i>Nein</i>	0,37

a ... unteres Quartil, b ... oberes Quartil

Feedback, das die Aufmerksamkeit auf die aufgabenbezogene Lern- und Motivationsebene lenkt, stellt somit die Aufgabe in den Fokus. Hierbei kann eine hohe Komplexität der Aufgabe jedoch die Leistung beeinträchtigen, da der Lernende mehr kognitive Ressourcen

²⁴ Insgesamt ist erkennbar, dass Kluger und DeNisi im Vergleich zu den erläuterten Funktionen in Kapitel 2.1.2 einige Faktoren anders zuordnen. So werden das entmutigende Feedback sowie das Lob der selbstbezogenen Metaebene zugeordnet, wohingegen Hattie (2014) dem Lob eher eine motivationale Funktion zuschreibt.

cen beanspruchen muss. Diese erhöhte Anstrengung kann die Motivation negativ beeinflussen und infolgedessen die Leistung verringern (z. B. Ackerman, 1987; Wood, Mento, & Locke, 1987).

Feedback, das die Aufmerksamkeit auf die selbstbezogene Metaebene bezieht, lenkt die Aufmerksamkeit weg von der Aufgabe hin zu anderen Zielen des Selbst und schwächt folglich die Leistung. Ein Beispiel dafür ist verbales Feedback. Im Gegensatz zu computergestütztem Feedback konzentriert sich die Aufmerksamkeit bei verbalem Feedback auf den Feedbackgeber, indem dessen Erwartungen sowie die Konsequenzen für die eigene Zielerreichung bewertet werden (Kluger und DeNisi, 1996, S. 267). Kluger und DeNisi (1996) verwenden dabei verbales und mündliches Feedback synonym.

Insgesamt kann geschlussfolgert werden, dass nicht davon ausgegangen werden kann, dass Feedback generell lernwirksam ist, unabhängig davon, wie es gestaltet wurde. Die Ergebnisse der Metaanalysen zeigen, dass verschiedene Variablen den Einfluss moderieren. Im nachfolgenden Kapitel 2.2.2 werden Studien zur Lernwirksamkeit verschiedener komplexer Feedbackarten analysiert.

2.2.2 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht

Der Inhalt des Feedbacks - insbesondere im Hinblick auf die Komplexität der Feedbacknachricht - ist der in der Feedbackforschung am meisten erforschte Bereich (Hattie & Wollenschläger, 2014, S. 138 f.; Strijbos u. a., 2010, S. 293; Shute, 2008, S. 292) und wird demzufolge in einem separaten Kapitel betrachtet. Die Befundlage hierzu ist eher uneinheitlich (vgl. Kapitel 2.2.1, weitere Befunde z. B. bei Shute, 2008; Mory, 2004).

Die in den nachfolgenden Studien verwendeten Feedbackarten werden anhand des Vorschlags von Narciss (2006) klassifiziert bzw. sind leicht auf diesen Klassifikationsvorschlag übertragbar. Daher werden diese Studien genauer vorgestellt. Da die Klassifikation nach Narciss (2006) im Vergleich zu den in Kapitel 2.1.3.2 vorgestellten Vorschlägen am umfassendsten und ausdifferenziertesten ist und als einzige strategisches und metakognitives Wissen berücksichtigt, werden auch die vorgestellten Studien anhand dieser Einteilung systematisiert. Darüber hinaus werden nachfolgende Studien in der Literatur häufig zitiert (z. B. bei Kopp & Mandl, 2014, S. 157). Dabei werden Studien, welche bei den Metaanalysen von Bangert-Drowns u. a. (1991) und Kluger und DeNisi (1996) sowie bei

der Synthese der Metaanalysen von Hattie (2015) einbezogen wurden, nicht berücksichtigt. Die Studien werden in Tabelle 6 zusammenfassend dargestellt.

McKendree (1990) untersuchte, welche Rückmeldungen am lernwirksamsten sind, nachdem ein Fehler begangen wurde. Er führte dazu einen Versuch mit 36 Schülern zum Thema geometrische Beweise einer High School durch. Ziel war es zu beweisen, dass zwei Dreiecke kongruent zueinander sind, d. h. den richtigen Kongruenzsatz für die gegebenen Voraussetzungen anzuwenden. Das verwendete PC-Programm strukturierte und überwachte die Schritte des Problemlösens der Studenten. Bei einem falschen Lösungsschritt unterbrach das System den Lernenden sofort, stellte KR zur Verfügung und ermöglichte einen weiteren Lösungsversuch. Bei zwei nacheinander auftretenden Fehlern gab das System Feedback, ohne einen weiteren Versuch zu fordern. Dabei gab es vier Feedbackgruppen: eine Kontrollgruppe [KG] und drei Experimentalgruppen [EG]. Die KG erhielt „minimal Feedback“ (KR). Der ersten EG wird die Diskrepanz zwischen den für den angenommenen Kongruenzsatz benötigten Voraussetzungen und den vom Schüler gewählten Voraussetzungen aufgezeigt. Dem Schüler wird also die Information gegeben, wann der gewählte Kongruenzsatz gelten würde. Diese Art von Feedback kann mit dem von Narciss (2006) vorgeschlagenen KM bezeichnet werden. Die zweite EG erhielt Informationen, was als nächstes zu tun sei, um der Aufgabenlösung näher zu kommen (vergleichbar mit KH) und die dritte EG erhielt KM und KH. Die Fehleranzahl für die Beweise der Treatmentphase sowie der Nachtestphase stellten die abhängigen Variablen dar (die Fehleranzahl des Vortests wurde als Kovariate einbezogen). Für die drei Experimentalgruppen zusammen konnte eine signifikant niedrigere Fehleranzahl sowohl für die Treatmentphase als auch für die Nachtestphase als bei der Gruppe KG nachgewiesen werden. Ein Vergleich des Anteils richtiger Schritte nach dem Feedback zeigte tendenzielle Unterschiede zwischen der einfachen Feedbackgruppe und den anderen drei Gruppen zusammen, die Effekte waren aber nicht signifikant. Eine Fehleranalyse zeigte, dass eine Information über die Fehlerart (bei allen drei EG gegeben) hilfreicher ist als das Wissen, dass man einen Fehler gemacht hat (KR). Insgesamt schienen KH sowie die Kombination aus KM und KH am vorteilhaftesten für das Lernen geometrischer Beweise.²⁵

Ziel der beiden von Huth (2004) durchgeführten Studien war es u. a., die Lernwirksamkeit von fehlerspezifischem informativem tutoriellem Feedback [ITF] zu untersuchen.

²⁵ Effektstärken wurden nicht angegeben. Diese lassen sich auch nicht berechnen, da keine Mittelwerte und Standardabweichungen bzw. Werte für das Eta-Quadrat angegeben wurden.

Beide Studien wurden mit Viertklässlern verschiedener Dresdner Grundschulen und Sechstklässlern verschiedener Dresdner Förderschulen zum Thema schriftliche Subtraktion mittels der computergestützten Lernumgebung „Subtratino“ durchgeführt. Dabei wurde zwei Wochen vor der Untersuchung ein Vortest durchgeführt und lediglich die Schüler mit gravierenden Problemen beim Lösen schriftlicher Subtraktionsaufgaben in die Studie einbezogen.

In der ersten Studie wurde ein Vortest-Nachtest-Vergleichsgruppen-Plan realisiert. Huth (2004) verglich dabei fehlerspezifisches ITF (Experimentalbedingung) mit KR-KCR (Vergleichsbedingung). Als abhängige Variablen wurden u. a. die Treatmentleistung (Anzahl erfolgreich bearbeiteter Aufgabentypen während der Treatmentphase), die Nachtestleistung (Anzahl korrekt bearbeiteter Aufgaben im Nachtest) und der Korrekturwahrscheinlichkeitsindex²⁶ herangezogen.²⁷ In der Vergleichsbedingung wurde den Schülern bei einem Fehler mitgeteilt, dass ihre Lösung falsch war (KR) und sie wurden aufgefordert, es noch einmal zu versuchen (Try Again Feedback). Nach einem zweiten fehlerhaften Lösungsversuch wurde den Schülern die richtige Lösung präsentiert (KCR). In der Experimentalbedingung erhielt der Schüler nach dem ersten fehlerhaften Lösungsversuch ebenfalls KR und wurde zu einem erneuten Lösungsversuch aufgefordert. Ab dem zweiten Versuch unterschied sich diese Bedingung von der Vergleichsbedingung. Der Schüler erhielt KR und KM, d. h. dem Schüler wurde mitgeteilt, dass seine Lösung noch Fehler enthält und der Fehlerort wurde farblich markiert. Darüber hinaus wurde bei systematischen Fehlern eine Lösungsstrategie vorgeschlagen, bei unsystematischen Fehlern wurden Lösungsbeispiele präsentiert. Bei einem dritten fehlerhaften Versuch wurde KR mit KH kombiniert, d. h. neben der Fehlermarkierung wurde die korrekte Lösungsprozedur visuell-auditiv demonstriert. Dabei wurde die fehlerhafte Schülerlösung nicht gelöscht, so dass der Schüler seine mit der korrekten Lösung vergleichen konnte. Das Präsentieren

²⁶ Das Konzept des Korrekturwahrscheinlichkeitsindex geht auf Phye und Bender (1989) zurück. Zur Bestimmung dieses Index wird die Anzahl der in einem Vortest falsch, in einem Nachtest jedoch richtig gelösten Aufgaben zur Gesamtzahl der in einem Vortest falsch gelösten Aufgaben ins Verhältnis gesetzt (Huth, 2004, S. 96).

²⁷ Als eine weitere abhängige Variable wurde der wahrgenommene Nutzen der angebotenen Feedbackinformationen bei der Experimentalgruppe bzw. der Feedbackwunsch bei der Vergleichsgruppe eingeschätzt. Des Weiteren wurde der Einfluss des ITF auf post-aktionale motivationale Parameter erfasst. Hierbei sollten die Schüler ihre eigene Leistung bewerten, die Zufriedenheit mit ihrer Leistung einschätzen, den Lerngewinn, die Aufgabenschwierigkeit sowie die Anstrengung bei der Aufgabenbearbeitung bewerten. Über die Erfassung der Freude beim Bearbeiten der Aufgaben konnte die intrinsische Motivation erfragt werden (Huth, 2004, S. 96 ff.)

der richtigen Prozedur wurde auch bei richtigen Lösungen zur Festigung vorgenommen. Die Schüler der Experimentalbedingung bearbeiteten signifikant mehr Aufgaben während der Treatmentphase erfolgreich als die Schüler der Gruppe KR-KCR und erzielten demnach eine höhere Treatmentleistung ($d=1,36^{28}$). Die Schüler der Experimentalbedingung bearbeiteten im Nachtest ebenfalls mehr Aufgaben richtig als die Schüler der KR-KCR-Bedingung (die Vortestleistung wurde als Kovariate einbezogen), wobei das Ergebnis nicht signifikant war ($d=0,55$). Der Vergleich des Korrekturwahrscheinlichkeitsindex zwischen beiden Feedbackgruppen zeigte, dass Schüler der Experimentalbedingung signifikant mehr Aufgaben im Nachtest richtig lösten, die sie im Vortest falsch hatten ($d=0,92$).

In einer zweiten Studie, einem Vortest-Nachtest-Vergleichsgruppen-Plan über zwei Messzeitpunkte, ging Huth (2004) der Frage nach, ob sich die positiven Wirkungen von ITF auch auf einem ökonomischeren Wege erreichen lassen, d. h. indem lediglich einmalig fehlerspezifische Informationen bereitgestellt werden. Dazu wurden in die Untersuchung zwei weitere Feedbackformen integriert, die sich zwischen der KR-KCR- und der ITF-Bedingung einordnen lassen. Die Feedbackformen wurden analog zu den von Narciss (2006) vorgeschlagenen elaborierten Feedbackformen als KR-KM bzw. KR-KH bezeichnet und gaben, anders als ITF, neben den korrekturrelevanten Informationen, die richtige Lösung an. Huth (2004) wollte analysieren, welche Feedbackform sich dabei als am lernwirksamsten erweist.

Beide Feedbackalgorithmen waren zweistufig, ähnlich dem KR-KCR Feedbackalgorithmus aus der ersten Studie. Nach dem ersten fehlerhaften Lösungsversuch wurden die Schüler informiert, dass ihre Lösung falsch war und aufgefordert, es erneut zu versuchen. Bei einem zweiten fehlerhaften Lösungsversuch wurde die Fehlerstelle markiert. Schüler in der KR-KM-Bedingung erhielten einen auditiven Hinweis auf den Fehler sowie die korrekte Lösungsstrategie. Unter der KR-KH-Bedingung wurde den Schülern die korrekte Lösungsprozedur visuell und auditiv vorgeführt. Gemeinsam mit den korrekturrelevanten Informationen wurde den Personen beider Bedingungen die richtige Lösung präsentiert. Die abhängigen Variablen waren dieselben wie in der ersten Untersuchung. Im Unterschied zu dieser wurde jedoch zu zwei Messzeitpunkten, im Abstand von einer Wo-

²⁸ Die Effektstärke d wurde von Omega-Quadrat (ω^2) umgerechnet (Cohen, 1988).

che, ein Nachtest durchgeführt. Für die Treatmentleistung konnte zu beiden Messzeitpunkten ein signifikanter Unterschied zwischen ITF und KR-KCR festgestellt werden ($d=1,18$; $d=0,86$). Die anderen Einzelvergleiche zeigten keine signifikanten Unterschiede. Es sind jedoch Tendenzen erkennbar, welche einen Vorteil der komplexeren Feedbackarten zeigen. Für die Ergebnisse des Nachtests zum 1. Messzeitpunkt zeigte sich ein signifikanter Unterschied des ITF gegenüber KR-KM ($d=0,74$) und KR-KCR ($d=0,71$). Bei den anderen Feedbackbedingungen waren auch hier lediglich Tendenzen zugunsten der komplexeren Feedbackarten erkennbar. Zum 2. Messzeitpunkt war die höhere Lernwirksamkeit des ITF nicht mehr statistisch nachweisbar, jedoch noch in der Tendenz erkennbar (d am höchsten gegenüber KR-KCR mit $0,40$). Huth (2004) begründet dies mit möglichen Deckeneffekten. Sie nimmt an, dass Schüler, welche zum 1. Messzeitpunkt bereits viele Aufgaben richtig gelöst hatten, kein Verbesserungspotential mehr hatten. Die Ergebnisse in Bezug auf den Korrekturwahrscheinlichkeitsindex sind im Wesentlichen mit den Resultaten hinsichtlich der Nachtestleistung vergleichbar. Zum 1. Messzeitpunkt unterschied sich ITF signifikant von KR-KCR ($d=1,04$) und KR-KM ($d=0,99$), zum zweiten Messzeitpunkt zeigten sich lediglich Tendenzen zugunsten der komplexeren Feedbackarten. Insgesamt ließ sich festhalten, dass die KR-KM- und KR-KH-Bedingungen nicht so lernwirksam waren wie ITF, aber lernwirksamer als KR-KCR. Somit kann die Annahme, dass „je informativer das Feedback und je stärker eine aktive Auseinandersetzung mit den präsentierten Informationen angeregt wird, desto stärker wird der Erwerb von Fertigkeiten der schriftlichen Subtraktion gefördert...“ (tendenziell) bestätigt werden (Huth, 2004, S. 120).

Moreno (2004) untersuchte in zwei Versuchen mit Collegestudenten eines einführenden Psychologiekurses u. a. die Lernwirksamkeit von zwei unterschiedlich komplexen Feedbackbedingungen im Bereich Botanik. Mit Hilfe eines multimedialen Lernspiels lernten die Teilnehmer, wie Pflanzen aufgebaut sein sollten, um bei verschiedenen Klimabedingungen überleben zu können. Dabei wollte Moreno überprüfen, ob komplexeres Feedback eine kognitive Überlastung²⁹ vermeiden kann. Diese kann entstehen, wenn Lernende keine Erfahrung im Umgang mit selbstgesteuerten multimedialen Lernspielen haben und lediglich korrekatives Feedback erhalten. Die Versuchsteilnehmer wurden aufge-

²⁹ Gemäß der Theorie des multimedialen Lernens (Mayer & Moreno, 2003) und der „Cognitive Load Theory“ (Sweller, 1999) kann eine kognitive Überlastung entstehen, wenn Lernende wenig Erfahrung im Umgang mit multimedialen Materialien haben und den Lernenden keine Anleitung bereitgestellt wird.

fordert, für die entsprechende Klimabedingung die entsprechenden Pflanzenteile zu kombinieren. Für ihre Auswahl erhielten sie korrekatives oder erklärendes Feedback. Bei dem erklärenden Feedback wurde den Lernenden in einem ersten Schritt erläutert, ob ihre Auswahl richtig war (KR). Zweitens wurde darlegt, warum eine ausgewählte Lösungsmöglichkeit richtig bzw. falsch war. Dieser Schritt lässt sich als KM-KC beschreiben, da neben der Erläuterung der Fehlerursachen auch Informationen zu den beschriebenen Klimabedingungen bzw. Pflanzenteilen bereitgestellt wurden, die aufgabenrelevantes konzeptuelles Wissen darstellten. Drittens wurde die richtige Lösung präsentiert (KCR). Unter der korrekativen Feedbackbedingung wurde den Lernenden mitgeteilt, ob ihre Auswahl richtig war (KR) und die richtige Lösung angegeben (KCR). In einem darauffolgenden Retention-Test sollten die Studenten alle Pflanzenarten aufschreiben, an die sie sich erinnern konnten. Im Transfer-Test wurden die Studenten aufgefordert, Pflanzen für verschiedene Klimabedingungen, die nicht behandelt wurden, zusammenzusetzen. Des Weiteren sollten sie aus angegebenen Pflanzenmerkmalen mögliche Klimabedingungen, in denen die charakterisierten Pflanzen überleben können, ableiten. Die Leistungen in diesen beiden Tests stellten u. a. die abhängigen Variablen³⁰ dar. Die erste Untersuchung zeigte sowohl im Retention-Test ($d=0,76$) als auch im Transfer-Test ($d=1,16$) einen signifikanten Vorteil für die KR-KM-KC-KCR-Bedingung, d. h. die Lernenden dieser Feedbackbedingung gaben in beiden Testteilen mehr richtige Antworten als die Lernenden der KR-KCR-Bedingung. In dem zweiten Versuch konnte Moreno nur für den Transfer-Test ($d=1,58$) einen signifikanten Vorteil für die KR-KC-KCR-Bedingung festhalten. Beide Gruppen erinnerten sich vergleichbar an Informationen aus der Lernphase. Die Ergebnisse beider Studien sprechen insgesamt für einen Vorteil hinsichtlich der Lernwirksamkeit von elaborem Feedback (KR-KM-KC-KCR) im Vergleich zu einfachem Feedback (KR-KCR) bei einer Reduzierung der kognitiven Überlastung für das erklärende Feedback.

³⁰ Weitere abhängige Variablen waren die Motivation, das Interesse, die Nützlichkeit sowie die Schwierigkeit. Die Schwierigkeit wurde über eine Selbsteinschätzung des Schwierigkeitsgrades des multimedialen Lernspiels erhoben. Über diese Selbsteinschätzung kann indirekt eine Aussage zur kognitiven Überlastung getroffen werden. Pass und van Merriënboer (1993) schlagen vor, die mentale Anstrengung mit Leistungskennzahlen ins Verhältnis zu setzen, um die Effektivität des Treatments zu bewerten. In beiden Versuchen zeigte sich eine hohe Effektivität für die KR-KM-KC-KCR Bedingung (eine niedrige kognitive Belastung bei hoher Leistung) und eine niedrige Effektivität (eine hohe kognitive Belastung bei niedriger Leistung) für die KR-KCR-Bedingung.

Narciss (2006) erforschte in zwei Studien in dem Bereich Konzeptlernaufgaben³¹ u. a. die Frage, welchen Einfluss unterschiedliche Feedbackarten sowie verschiedene kognitive und motivationale Eingangsvoraussetzungen der Lernenden auf die Lernwirksamkeit von Feedback haben. In der ersten Untersuchung nahmen Studenten verschiedener Fachrichtungen der Technischen Universität Dresden teil. Bei der Bearbeitung der Experimentalaufgaben war es den Probanden freigestellt, wie viele Aufgaben sie bearbeiteten, sie konnten auch Aufgaben abbrechen. Innerhalb der Bearbeitungsphase wurden drei Feedbackbedingungen realisiert. Unter der ersten Bedingung erhielten die Versuchsteilnehmer lediglich die Information, ob ihr Konzeptvorschlag richtig oder falsch war (KR) sowie ggf. die richtige Lösung (KCR). Unter der zweiten Bedingung wurden zusätzlich zu KR bei einer falschen Lösung Informationen über den Ort, die Art und die Ursache des Fehlers bereitgestellt (KM). Die dritte Feedbackbedingung entsprach ITF und gab den Versuchsteilnehmern zusätzliche fehlerspezifische strategische Informationen zur Korrektur des Fehlers (KH). Insgesamt beinhalteten beide Feedbackbedingungen eine detaillierte Anforderungs- und Fehleranalyse. Als abhängige Variable wurde u. a. ein Leistungsindex berechnet. Die individuellen Voraussetzungen der Lernenden wurden als Kontrollvariablen erfasst. Im Ergebnis konnte Narciss (2006) festhalten, dass die Wirkung von ITF insbesondere vom individuellen Faktor Kompetenzeinschätzung abhängt. Narciss (2006) konstatierte daher, dass bei Nichtberücksichtigen dieser Kontrollvariablen geschlussfolgert worden wäre, dass sich die drei Feedbackbedingungen bezüglich ihrer Wirksamkeit nicht wesentlich voneinander unterscheiden.

In der zweiten Studie waren die Versuchsteilnehmer Schüler eines technischen Berufsschulzentrums in Dresden. In dieser Studie wurde die Offenheit der Lernaufgabe eingeschränkt. Die Probanden sollten in der Zeit von 90 Minuten so viele Aufgaben wie möglich bearbeiten. Sie konnten dabei selbst entscheiden, wie viele Aufgaben sie bearbeiten, und sie konnten auch Aufgaben abbrechen. Darüber hinaus wurden nur zwei Feedbackbedingungen realisiert, KR und ITF (für die konkrete Ausgestaltung vgl. Studie 1). Die für diese Untersuchung relevanten abhängigen und unabhängigen Variablen sowie

³¹ Experimentelle Konzepte sind durch verschiedene Merkmalsdimensionen (z. B. Form, Farbe), die bestimmte Merkmalsausprägungen annehmen können sowie logische Verknüpfungsregeln zwischen diesen Merkmalsdimensionen definiert (Narciss, 2006, S. 134 f.). In den skizzierten Untersuchungen wurden den Versuchspersonen Stimuluskarten mit verschiedenen Merkmalsdimensionen in konkreten Ausprägungen und konjunktiver oder disjunktiver Verknüpfung präsentiert. Sie mussten herausfinden, welche Merkmalsdimensionen und –ausprägungen relevant sind und wie die Dimensionen miteinander verknüpft waren (Narciss, 2006, S. 135).

Kontrollvariablen entsprachen denen der ersten Studie. Als Ergebnis lässt sich festhalten, dass die Schüler unter der ITF-Bedingung signifikant bessere Leistungen als die Schüler der KR-Bedingung erbrachten.

Ziel der Untersuchung von Chase und Houmanfar (2009) war es, die Wirkungen von elaboriertem und einfachem Feedback auf die Lernleistung zu bestimmen. Die Untersuchung wurde mit Studenten der Psychologie während eines Einführungskurses zu Grundlagen der empirischen Sozialforschung über 10 Wochen hinweg durchgeführt. Die Kursinhalte wurden allen Teilnehmern durch sechs Referenten unter standardisierten Bedingungen präsentiert. Um den Kurs erfolgreich beenden zu können, mussten die Studenten, zusätzlich zu einer Prüfung in der Mitte und am Ende des Kurses, im Verlauf des Semesters zehn Tests für jedes Kapitel absolvieren, welche jeweils 15 Multiple-Choice Fragen umfassten. Zu den Testergebnissen erhielten die Studenten entweder einfaches oder elaboriertes Feedback. Das einfache Feedback enthielt die Information, zu wie viel Prozent eine Aufgabe richtig gelöst wurde (KP). Zusätzlich erfolgte nach einer falschen Antwort die Präsentation der richtigen Lösung (KCR). Elaboriertes Feedback wurde als schriftliches, verbales Feedback bestimmt, das dem Lernenden sofort nach der Eingabe seiner Lösung für jeden Test durch die computergestützte Testumgebung bereitgestellt wurde. Dabei wurde zwischen spezifischem und allgemeinem Feedback unterschieden. Spezifisches Feedback enthielt Informationen zu jeder Frage und gab an, welche Antwortmöglichkeiten richtig oder falsch waren (vergleichbar mit KM). Allgemeines Feedback enthielt konzeptuelle Informationen, Definitionen oder Hinweise, wo bestimmte Informationen im Text gefunden werden konnten (vergleichbar mit KC bzw. KTC). Als abhängige Variablen wurden die Korrektheit der Antwort, der Lernzuwachs, die Leistung bei schwierigen Fragen sowie die Notenverteilung der Prüfungen erfasst. Die Korrektheit der Antworten wurde für die 10 Tests, die Midterm- sowie die finale Prüfung erfasst. Um den Lernzuwachs zu ermitteln, erhielten die Studenten in den Prüfungen dieselben Fragen wie in den Tests. Im Ergebnis konnte festgehalten werden, dass sich beide Feedbackgruppen vom ersten Testzeitpunkt (Ergebnis des jeweiligen Zwischentests) zum zweiten Testzeitpunkt (Midterm-Prüfung bzw. finale Prüfung) verbesserten. Des Weiteren führte elaboriertes Feedback zu einer signifikanten Verbesserung der Leistung im Vergleich zu einfachem Feedback ($d=0,45$). Darüber hinaus wurde eine signifikante Wechselwirkung zwischen der Feedbackart und der Schwierigkeit der Lernaufgabe gefunden ($d=0,50$).³²

³² Die Effektstärken wurden von Eta-Quadrat (η^2) umgerechnet (Cohen, 1988).

Elaboriertes Feedback führte bei schwierigeren Fragen zu höheren Leistungen als einfaches Feedback ($d=1,02$).³³ Bei der Notenverteilung zeigte sich, dass einfaches Feedback bei 49 Prozent der Studenten zu guten bis sehr guten Noten (A und B) und bei 51 Prozent zu schlechteren bis sehr schlechten Noten (C, D und F) führte. Unter der elaborierten Feedbackbedingung hingegen erzielten 70 Prozent der Schüler gute bis sehr gute und 30 Prozent schlechtere bis sehr schlechte Noten. Elaboriertes Feedback zeichnet sich folglich durch eine höhere Lernwirksamkeit im Vergleich zu einfachem Feedback aus, insbesondere bei schwierigen Aufgaben.

Rakoczy, Klieme, Bürgermeister und Harks (2008) untersuchten u. a. wie die Bewertung von Lehrern die Lernprozesse der Schüler beeinflusst. Dabei analysierte die Studie das Feedback, welches den Schülern im Rahmen einer natürlichen Lehrer-Schüler-Interaktion gegeben wurde. Dabei wurde einfaches („evaluative“) und elaboriertes („informational“) Feedback unterschieden. Die Studie war Teil eines Schweizer-Deutschen Forschungsprojektes, welches die Qualität von Mathematikunterricht in mittleren Bildungsgängen der neunten Jahrgangsstufe untersuchte. Um zu analysieren, inwieweit elaboriertes oder einfaches Feedback lernwirksamer sind, wurde in mehreren Klassen die Einführung in den Satz des Pythagoras auf Video festgehalten und ausgewertet. Versuchsteilnehmer waren Schüler aus zwei Gymnasial- und acht Realschulklassen. Die Feedbackbedingungen wurden nicht vor dem Unterricht festgelegt, sondern aus der natürlichen Lehrer-Schüler-Interaktion analysiert. Einfaches Feedback signalisierte dabei den Schülern lediglich, ob eine Antwort richtig oder falsch war. Äußerungen wie „ja“ oder „genau“ deuteten beispielsweise auf eine richtige Antwort hin, Äußerungen wie „Das ist nicht richtig“ auf falsche. Elaboriertes Feedback beinhaltete zusätzliche Hinweise. Im Falle einer falschen Antwort wurde die Fehlerursache genannt und weitere inhaltsbezogene und/oder prozessbezogene Informationen angeboten, um die Antwort zu reflektieren und den Fehler zu korrigieren (vergleichbar mit KM, KC, KH). Bei richtigen Antworten beinhaltete das Feedback Informationen, wie man noch effektiver vorgehen könnte (vergleichbar mit KH). Die Leistung der Schüler wurde sowohl vor als auch nach dem Unterricht mit einem Wissenstest erfasst. Die Leistung des Nachtests stellte eine abhängige Variable dar, welche durch die Leistung im Vortest kontrolliert wurde. Der Vortest beinhaltete Voraussetzungen, die für ein tieferes Verständnis des Satzes von Pythagoras not-

³³ Auf den Einfluss der Aufgabenschwierigkeit wird im Kapitel 2.2.5 erneut Bezug genommen.

wendig waren. Der Nachtest fokussierte das konzeptuelle Verständnis des Satzes von Pythagoras und leichte Anwendungsfragen. Im Ergebnis konnten Rakoczy u. a. (2008) keinen signifikanten Einfluss der beiden Feedbackbedingungen auf die Leistungsentwicklung nachweisen.³⁴ Die Autoren führten dies darauf zurück, dass die mit dem Feedback bereitgestellten Informationen nicht ausreichend erschienen, um eine tiefere Auseinandersetzung mit dem Wissensinhalt anzuregen. Das elaborierte Feedback wurde in einer natürlichen Unterrichtsbedingung gegeben und beinhaltete somit eine große Vielfalt verschiedener Feedbackinterventionen. Dabei schlussfolgerten die Autoren, dass für zukünftige Untersuchungen die Situationen mehr standardisiert werden sollten.

Die Ergebnisse der Studien zeigen mit Ausnahme der Studie von Rakoczy u. a. (2008) einen Vorteil des elaborierten gegenüber dem einfachen Feedback. Werden die Ergebnisse der in Kapitel 2.2.1 erläuterten Metaanalyse von Bangert-Drowns u. a. (1991) mit hinzugezogen, so führt die Analyse der vergleichenden Studien in Bezug auf die Komplexität des Feedbackinhalts zu keinem Vorteil des elaborierten Feedbacks gegenüber KCR. Es stellt sich die Frage, wie diese uneinheitlichen Befunde zu erklären sind. Ein bedeutender Unterschied der Studie von Rakoczy u. a. (2008) zu den anderen in Kapitel 2.2.2 sowie der Metaanalyse von Bangert-Drowns u. a. (1991) ist die Art der Lernsituation. Die anderen Studien wurden computerbasiert, die Studie von Rakoczy u. a. (2008) in einer natürlichen Lernumgebung, durchgeführt. Des Weiteren deuten Bangert-Drowns u. a. (1991, S. 234) an, dass die Inhaltsbereiche der Studien der Metaanalyse zu einfach waren (vgl. auch Kapitel 2.2.1), weshalb auch von niedrigen Aufgabenanforderungen ausgegangen werden kann. Bei höheren Aufgabenanforderungen hätte der Vorteil des elaborierten Feedbacks ggf. nachgewiesen werden können.

Das deutet darauf hin, dass die Studien nicht pauschal nach der Komplexität des Feedbackinhalts unterschieden werden können, sondern dass es hinsichtlich der Gestaltung von Feedback weitere Einflussfaktoren geben kann. Im nachfolgenden Kapitel werden Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Feedback-Form dargestellt.

³⁴ Über Schätzungen von Regressionskoeffizienten konnte kein signifikanter Einfluss des einfachen Feedbacks (richtige Antwort: $p=0,92$; $\beta=0,01$; falsche Antwort: $p=0,86$; $\beta=0,01$) und des elaborierten Feedbacks ($p=0,97$; $\beta=0,00$) nachgewiesen werden. Als eine weitere abhängige Variable wurde der Einfluss der beiden Feedbackarten auf die Motivation untersucht. Hierbei konnte ein signifikanter Einfluss nachgewiesen werden.

Tabelle 6: Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht

Studie	Fachgebiet, Versuchsteilnehmer	Feedbackbedingung	ausgewählte, abhängige Variablen	ausgewählte Ergebnisse
McKendree (1990)	geometrische Beweise, Schüler einer High School	KG: KR EG 1: KM EG 2: KH EG 3: KM-KH	<ul style="list-style-type: none"> Fehleranzahl in der Treatmentphase und dem Nachtest Anteil richtiger Schritte nach dem Feedback 	<ul style="list-style-type: none"> EG 1 + EG 2 + EG 3 > KG^s EG 1 + EG 2 + EG 3 > KG^{ns}
Huth (2004) Studie 1 Studie 2	schriftliche Subtraktion, Grund- und Förderschüler	EG : ITF VG: KR-KCR EG: ITF VG 1: KR-KM VG 2: KR-KH VG 3: KR-KCR	<ul style="list-style-type: none"> Treatmentleistung Leistung im Nachtest Korrekturwahrscheinlichkeitsindex <p><u>1. Messzeitpunkt</u></p> <ul style="list-style-type: none"> Treatmentleistung Leistung im Nachtest Korrekturwahrscheinlichkeitsindex <p><u>2. Messzeitpunkt</u></p> <ul style="list-style-type: none"> Treatmentleistung Nachtestleistung Korrekturwahrscheinlichkeitsindex 	<ul style="list-style-type: none"> EG > VG^s; (d=1,36) EG > VG^{ns}; (d=0,55) EG > VG^s; (d=0,92) <p>1. Messzeitpunkt</p> <ul style="list-style-type: none"> EG > VG 2 > VG 1 > VG 3, wobei lediglich EG > VG 3^s (d=1,18); EG > VG 2 > VG 1 > VG 3, wobei EG > VG 1^s (d=0,74); EG > VG 3^s (d=0,71) EG > VG 2 > VG 1 > VG 3, wobei EG > VG 1^s (d=0,99); EG > VG 3^s (d=1,04) <p>2. Messzeitpunkt</p> <ul style="list-style-type: none"> EG > VG 2 > VG 1 > VG 3, wobei lediglich EG > VG 3^s (d=0,86); EG > VG 2 > VG 1 > VG 3^{ns} EG > VG 2 > VG 1 > VG 3^{ns}
Moreno (2004) Versuch 1 Versuch 2	Botanik, Collegestudenten	EG 1: KR-KM-KC-KCR EG 2: KR-KCR	<ul style="list-style-type: none"> Retention-Test-Leistung Transfer-Test-Leistung Retention-Test-Leistung Transfer-Test-Leistung 	<ul style="list-style-type: none"> EG 1 > EG 2^s (d=0,76) EG 1 > EG 2^s (d=1,16) EG 1 = EG 2 EG 1 > EG 2^s (d=1,58)
Narciss (2006) Studie 1 Studie 2	Konzeptlernaufgaben, Studenten Berufsschüler	EG 1: KR EG 2: KR-KM EG 3: KR-KM-KH (ITF) EG 1: KR EG 2: KR-KM-KH (ITF)	<ul style="list-style-type: none"> Leistungsindex 	<ul style="list-style-type: none"> kein sign. Haupteffekt für Feedback sign. Interaktionseffekt zwischen Feedback und Kontrollvariablen EG 1 > EG 2^s
Chase & Humanfar (2009)	Psychologie, Studenten	EG: KM-KTC-KC VG: KP-KCR	<ul style="list-style-type: none"> Antwortrichtigkeit Lernfortschritt Leistung bei schwierigen Fragen Notenverteilung 	<ul style="list-style-type: none"> EG > VG^d EG > VG^s (d=0,45) EG > VG^s (d=1,02) EG > VG^d
Rakoczy, Klieme, Bürgermeister & Harks (2008)	Satz des Pythagoras Sekundarstufe 1	EG: EF VG: KR	<ul style="list-style-type: none"> Leistungsentwicklung 	<ul style="list-style-type: none"> kein Einfluss beider Feedbackbedingungen auf die Leistungsentwicklung

^s signifikant, ^{ns} nicht signifikant, ^d deskriptive Auswertung

2.2.3 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Feedback-Form

Neben der Komplexität des Feedbackinhalts kann sich die Gestaltung des Feedbacks auch auf die Feedbackform beziehen. In der Metaanalyse von Kluger und DeNisi (1996) wurde schon sehr ausführlich auf die Feedbackwirkung in Abhängigkeit von den Feedbackfunktionen eingegangen (vgl. Kapitel 2.2.1). Dieser Aspekt wird daher nicht weiter thematisiert.

Hinsichtlich der *Feedbackquelle* kann zwischen *internem* und *externem* Feedback unterschieden werden. Wollenschläger, Möller und Harms (2012, 2011) fanden in naturwissenschaftlichen Studien mit deutschen Gymnasialschülern heraus, dass externes Feedback (Feedback vom Lehrer entsprechend der Kompetenzstufen) gegenüber internem Feedback (Selbsteinschätzungen der Leistung entsprechend der Kompetenzstufen) zu einer verbesserten Leistung beiträgt. Dabei wurde die wahrgenommene Feedbackeffektivität als Mediator identifiziert. Schüler, die externes Feedback erhalten, nehmen das Fremdfedback effektiver wahr und dies ist entscheidend für das bessere Abschneiden (Wollenschläger u. a., 2011).

Gabelica, Van den Bossche, Segers und Gijsselaers (2012) führten ein Literaturreview auf Basis von 59 empirischen Studien durch. Hierbei fassten sie u. a. die Ergebnisse von Studien zusammen, welche die Lernwirksamkeit von *Gruppen-* vs. *individuellem Feedback* untersuchten. Hierbei wurde untersucht, welche unterschiedlichen Effekte ein Feedback, das einer gesamten Gruppe, den einzelnen Gruppenmitgliedern bzw. sowohl der Gruppe als auch den Einzelpersonen gegeben wurde, hervorbrachte. Die Befundlage hierzu ist uneinheitlich. Die Untersuchungen lassen sich in vier Gruppen unterteilen. Die erste Gruppe konnte positive Effekte des Gruppenfeedbacks gegenüber dem individuellen Feedback nachweisen (Doerr, Mitchell, Klastorin, & Brown, 1996; Tindale, 1989). Weitere Untersuchungen stellten heraus, dass eine Kombination von Gruppen- und individuellem Feedback gegenüber dem alleinigen Geben von Gruppenfeedback oder individuellem Feedback lernwirksamer ist (z. B. Matsui, Kakuyama, & Onglatco, 1987). In der dritten Gruppe von Studien wurde die Überlegenheit von individuellem gegenüber Gruppenfeedback postuliert (z. B. Stone, 1971). In der letzten Gruppe ließ sich kein Effekt nachweisen (z. B. Sniezek, May, & Sawyer, 1990).

In Bezug auf den *Zeitpunkt der Feedbackgabe* ist die Befundlage nicht eindeutig. Einige Studien berichten von einer Überlegenheit des *verzögerten* gegenüber dem *sofortigen*

Feedback (Phye & Andre, 1989; Kulhavy & Anderson, 1972). Die Autoren begründen dies damit, dass die Verzögerung zum Vergessen der Anfangsfehler und einer Entschlüsselung der richtigen Antwort ohne Störung (Interferenz) führt. Darüber hinaus erhöht verzögertes Feedback die Aufmerksamkeit. Andere Studien weisen eine Überlegenheit des sofortigen Feedbacks nach (z. B. Corbett & Anderson, 2001). Weitere Studien betonen, dass der Zeitpunkt der Feedbackgabe auch von weiteren Faktoren wie dem Vorwissen oder der Aufgabenkomplexität abhängt (z. B. Clariana, Wagner, & Murphy, 2000). Im Kapitel 2.2.4 und 2.2.5 werden Ergebnisse zu Studien, die diese Einflussfaktoren beleuchten, genauer betrachtet. Insgesamt lassen sich die Studien hinsichtlich des Zeitpunktes der Feedbackgabe aufgrund nicht einheitlicher Definitionen von sofortigem bzw. verzögertem Feedback jedoch schwer vergleichen.

Im Zusammenhang mit elaboriertem Feedback zeigen Studien einen Vorteil, wenn die Informationen *sequentiell* gegeben werden (z. B. Heift, 2004; Narciss & Huth, 2004). Im Rahmen der Cognitive-Load-Forschung wird davon ausgegangen, dass eine sequentielle Präsentation von Informationen einer gleichzeitigen Präsentation überlegen sein müsste (Chandler & Sweller, 1992). Narciss und Huth (2006) kritisieren, dass elaboriertes Feedback häufig zusammen mit der richtigen Lösung und somit KCR präsentiert wird. Dies führt dazu, dass der Lernende die elaborierte Feedbackinformation möglicherweise als irrelevant erachtet und das Feedback lediglich oberflächlich verarbeitet. Demnach ist es von Vorteil, das elaborierte Feedback zunächst ohne die richtige Lösung und erst in einem nächsten Schritt KCR zu geben. Die in Kapitel 2.2.2 vorgestellten Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht verwenden alle mit Ausnahme der Studie von Rakoczy u. a. (2008) sequentielle Feedbackinformationen und weisen einen Vorteil des elaborierten gegenüber dem einfachen Feedback nach. Bei Rakoczy u. a. (2008) zeigte sich kein Einfluss der beiden Feedbackbedingungen auf die Leistungsentwicklung. Demnach könnte an dieser Stelle auch die Gestaltung der Feedbackinformationen eine Rolle gespielt haben.

Absichtliches und *beiläufiges* Feedback in einer empirischen Studie gegenüberzustellen um die Lernwirksamkeit zu analysieren, ist sehr schwierig, da beiläufiges Feedback aus einer zufälligen Interaktion mit der Umwelt resultiert und somit nicht als kontrollierbare Treatmentbedingung gestaltet werden kann. Bangert-Drowns u. a. (1991) beziehen daher auch nur Studien, welche absichtliches Feedback verwenden, in ihre Metaanalyse ein. In Bezug auf die Transparenz für den Feedbackempfänger können daher lediglich Studien

betrachtet werden, welche *implizites* und *explizites* Feedback gegenüberstellen. Denn implizit bedeutet nicht gleichzeitig, dass es unabsichtlich sein muss, es kann auch absichtlich gestaltet werden. Explizites Feedback wird absichtlich gegeben. In Bezug auf die Lernwirksamkeit des expliziten im Vergleich zum impliziten Feedback haben zahlreiche Studien keine eindeutigen Ergebnisse erzielt. Laborstudien haben gezeigt, dass implizites Feedback durch die Lernenden durchaus erfasst und genutzt werden kann (z. B. Egi, 2010; Mackey, Gass, & McDonough, 2000; Roberts, 1995). Studien, welche in natürlichen Unterrichtssituationen durchgeführt wurden, zeigten zum einen kleine oder keine Effekte für recasts (z. B. Sheen, 2010; Lyster, 2004), andere dagegen große Effekte (z. B. Doughty & Varela, 1998; Han, 2002). Insgesamt besteht weiterer Forschungsbedarf.

In Bezug auf das Präsentationsformat stellt Krause (2007, S. 54) die Vorteile des *mündlichen* gegenüber dem *schriftlichen* Feedback heraus (vgl. auch Kap. 2.1.3.1). Zusätzliche Kommunikationskanäle, d. h. para- und nonverbale Signale, erleichtern die Interpretation der Informationen. Ebenso stellen Kopp und Mandl (2014) fest, dass mündliches Feedback besser als schriftliches Feedback verarbeitet wird. In diesem Zusammenhang verweisen die Autoren auf eine Studie von Weaver (2006), welche zeigt, dass das schriftliche Feedback von den Probanden als zu vage und zu negativ formuliert erachtet wird und die Lernenden infolgedessen demotiviert. Die Hälfte der Befragten war sich darüber hinaus nicht über die Bedeutung der schriftlich verwendeten Formulierungen sicher. Studien zur Vergleichbarkeit beider Präsentationsformate in Bezug auf die Lernwirksamkeit sind hierbei nicht bekannt. Kluger und DeNisi (1996) konnten in ihrer Metaanalyse zeigen, dass Feedback, welches in *verbaler* Form vermittelt wird, die Leistung abschwächt, wohingegen *computerbasiertes* Feedback die Leistung erhöht. Sie argumentieren, dass sich bei verbalem Feedback im Gegensatz zu computerbasiertem Feedback die Aufmerksamkeit auf den Feedbackgeber konzentriert, indem dessen Erwartungen sowie die Konsequenzen für die eigene Zielerreichung bewertet werden. Hattie und Wollenschläger (2014, S. 138) führen lediglich Studien auf, welche Feedback entweder *textbasiert* (z. B. Dempsey, Driscoll, u. a., 1993), verbal (z. B. Krampen, 1987) oder computergestützt (z. B. van der Kleij, Eggen, Timmers, & Veldkamp, 2012) geben. Vergleichende Studien im Hinblick auf eine bessere Lernwirksamkeit werden dabei nicht aufgeführt.

Im Hinblick auf computergestütztes Feedback zeigen einige Studien den Vorteil der *bimodalen* gegenüber der *unimodalen* Präsentation von verbalen und piktorialen Infor-

mationen (z. B. Mayer & Moreno, 1998; Mayer & Anderson, 1992, 1991). In diesen Untersuchungen wurden die verbalen Informationen auditiv und zusätzlich die piktorialen Informationen visuell dargeboten. Es zeigte sich eine Überlegenheit der bimodalen Präsentation gegenüber der unimodalen Darbietung in visueller Form. Erklärungen hierfür finden sich in der kognitiven Theorie des multimedialen Lernens von Mayer und Moreno (2002) sowie der Cognitive-Load-Theorie von Sweller, van Merriënboer und Paas (1998). Die kognitive Belastung des Lernenden wird aufgrund der kombinierten Nutzung des auditiven und visuellen Informationsverarbeitungssystems reduziert und die Lernleistung verbessert. Bei einer unimodalen simultanen Präsentation verbaler und piktorialer Informationen in visueller Form wird nur der visuelle Kanal benutzt. Somit ist eine Aufteilung der verfügbaren Aufmerksamkeitsressourcen vonnöten (split-attention-effect), was den Lernenden überfordern kann.

Hinsichtlich der *Anzahl der Lösungsversuche* verwenden viele Studien lediglich einen Versuch (*STF*). Untersuchungen, welche mehrere Lösungsversuche anbieten, verwenden häufig *AUC* (für einen Review siehe Clariana, 1993). Clariana (1993) stellte dabei eine Überlegenheit aller Feedbacktypen gegenüber keinem Feedback fest, konnte jedoch keine generelle Überlegenheit von *AUC* bzw. *MTF* gegenüber *STF* nachweisen. Hierbei wird betont, dass die Lernwirksamkeit des Feedbacks mit einem oder mehreren Lösungsversuchen auch vom Vorwissen abhängt (vgl. Kapitel 2.2.4). Einen weiteren Einflussfaktor stellt die Aufgabenkomplexität dar (Clariana & Koul, 2006; vgl. Kapitel 2.2.5). Eine generelle Überlegenheit von *MTF* oder *AUC* gegenüber *STF* konnte dabei folglich nicht festgestellt werden.

In Bezug auf die Feedbackform ist die Befundlage somit ebenfalls uneinheitlich. Nachfolgend werden Studien in Bezug auf (inter-)personale und situative Einflussfaktoren analysiert, deren Bedeutung bereits im Kapitel 2.1.4 hervorgehoben wurde.

2.2.4 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von (inter-) personalen Faktoren

Studien in Abhängigkeit von Faktoren bezüglich des Feedbacksenders

Im Kapitel 2.2.4 werden zunächst Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von Einflussfaktoren, welche sich auf den Feedbacksender beziehen, analysiert. Hierbei werden zum einen der formale und zum anderen der informelle Status betrachtet. Strijbos und Müller (2014, S. 104 ff.) unterscheiden beim *formalen Status* zwischen einer vertikalen und einer horizontalen Sender-Empfänger-Konstellation (vgl. auch Kapitel 2.1.4.3). Da die Feedbackübermittlung in der vertikalen Sender-Empfänger-Konstellation typischerweise abwärtsgerichtet erfolgt, konzentriert sich auch die Forschung in diesem Bereich darauf, mit Hilfe welcher Formen der Lehrkrafrückmeldung eine effiziente und effektive Unterstützung der Lernprozesse der Schüler möglich ist (z. B. Hattie & Timperley, 2007; Narciss, 2006; Mory, 2004;). Bislang wurde nicht systematisch untersucht, inwieweit sich Lehrkräfte selbst als autoritäre Feedbackquellen einschätzen, ob sie den höheren Status beim Unterrichten berücksichtigen und welche Folgen dies auf die Schüler als Feedbackempfänger haben kann (Strijbos & Müller, 2014, S. 112). Zum aufwärtsgerichteten Feedback existiert bisher kaum Forschung, insbesondere in Bezug auf das Einfordern von Schülerfeedback durch die Lehrkräfte, um den eigenen Unterricht zu evaluieren und somit den Schülern eine stärkere Partizipation im Hinblick auf die Weiterentwicklung des Unterrichts zu ermöglichen (Gärtner, 2013). Bei der horizontalen Sender-Empfänger-Konstellation existiert beim Feedbackempfänger im Vergleich zu vertikalen Beziehungen eine höhere Unsicherheit in Bezug auf wahrgenommene Fairness und Nützlichkeit der Rückmeldung (Brown, Irving, Peterson, & Hirschfeld, 2009; Strijbos, Ochoa, Sluijsmans, Segers, & Tillema, 2009; Cheng & Warren, 1997). Einige Studien betonen in diesem Zusammenhang die Bedeutung von Begründungen beim Peer-Feedback (z. B. Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Chiu, 2008). Gielen u. a. (2010) stellten fest, dass das Begründen³⁵ von Feedback durch den Feedbackgeber zu einer Verbesserung der Leistung von Lernenden mit niedrigen Vortestleistungen führen kann. Zur

³⁵ Die Lernenden wurden dazu aufgefordert, einen schriftlichen Bericht für den Lehrer zu formulieren, in dem zusammengefasst wurde, welche Kommentare ausgewählt und wie diese übermittelt wurden und was durch das Geben von Peer-Feedback gelernt wurde. Des Weiteren sollten die Lernenden ihre eigene Leistung reflektieren.

Erforschung der konkreten Rolle von Begründungen besteht noch weiterer Forschungsbedarf.

Zu den Auswirkungen der Kompetenz einer Lehrkraft, d. h. des *informellen Status*, auf das Gestalten und Übermitteln von Feedback wurde bisher nur wenig geforscht. In einer Studie von Lee (2009) wurde analysiert, inwieweit sich die Überzeugungen zu Feedback in den tatsächlich gegebenen schriftlichen Rückmeldungen widerspiegeln. Es zeigte sich, dass sich Lehrkräfte zwar bewusst sind, wie konstruktives Feedback gestaltet werden sollte, ihr Wissen jedoch häufig nicht in einer angemessenen Feedbackübermittlung anwenden können. Es zeigt sich also, dass Lehrkräfte ihre tatsächliche Kompetenz in Bezug auf das Geben von Feedback oft überschätzen. Auswirkungen auf die Schülerleistungen wurden hierbei nicht untersucht. Aufgrund der zunehmenden Bedeutung der Peerfeedbackverfahren im Lehr-Lern-Kontext wurde auch die wahrgenommene bzw. zugeschriebene Kompetenz näher erforscht (vgl. auch Kapitel 2.1.4.3 und 2.1.4.4). Hierbei wurden verschiedene Fragebögen entwickelt (z. B. de Kleijn, Mainhard, Meijer, Brekelmans, & Pilot, 2013; Strijbos u. a., 2010; King, Schrod, & Weisel, 2009). Strijbos u. a. (2010) differenzieren bei der zugeschriebenen Kompetenz des Senders zwischen wahrgenommener Adäquatheit (Fairness, Nützlichkeit und Akzeptanz), Bereitschaft zur Verbesserung und affektiver Reaktion (z. B. Freude und Ärger). In einer begleitenden Studie wurden Studierenden verschiedene Szenarios vorgelegt, wobei die Feedbackinhalte (elaboriertes vs. allgemeines Feedback) sowie die tatsächliche Kompetenz des Senders³⁶ (hoch vs. niedrig) variiert wurden. Feedback, das eine kompetente Person gibt, wird dabei als adäquater gegenüber Feedback, das eine weniger kompetente Person gibt, eingeschätzt. Dabei bewerten Studierende elaboriertes Feedback von einem hochkompetenten Sender am adäquatesten. Es zeigten sich jedoch keine Auswirkungen auf die Leistung der Studierenden. Theoretisch kann angenommen werden, dass ein als nützlich wahrgenommenes Feedback dazu führt, dass der Lernende die Informationen auch tatsächlich nutzt, um sein Lernverhalten anzupassen und die Leistung infolgedessen verbessert wird. Darüber hinaus wird das Kompetenzerleben unterstützt und folglich Motivation und Interesse gefördert (Deci & Ryan, 1991). In den Studien von Harks, Rakoczy, Hattie, Besser und Klieme (2014) und Rakoczy, Harks, Klieme, Blum und Hochweber (2013) konnte ein

³⁶ Der Kompetenzlevel des Senders wurde während der Treatmentphase insgesamt drei Mal erwähnt, z. B. mit dem Satz: „The following feedback is provided by a peer who, thus far, has achieved high/low performances in writing“ (Strijbos u. a. 2010, S. 295).

positiver Zusammenhang zwischen der wahrgenommenen Nützlichkeit von Rückmeldungen und Leistungs- und Interessensänderung festgestellt werden.

Studien in Abhängigkeit von Faktoren bezüglich des Feedbackempfängers

Nachfolgend werden die Ergebnisse von Studien, welche den Einfluss des *informellen Status* (insbesondere des Vorwissens, der Sprachfähigkeit und der Antwortsicherheit), der *Attribution*, der *Selbstwirksamkeit*, der *Big 5* und der *Lernmotivation* (insbesondere der Zielorientierung) auf die Lernwirksamkeit des Feedbacks analysieren, zusammengefasst. Hinsichtlich des *Umgangs mit Fehlern* sowie der *Fehlerangst* sind keine Studien zum Einfluss auf die Lernwirksamkeit des Feedbacks bekannt. Daher werden diese Faktoren an dieser Stelle nicht näher betrachtet.

Dass die Kompetenz, d. h. der *informelle Status*, des Empfängers (z. B. das Vorwissen) einen großen Einfluss auf die Lernwirksamkeit von Feedback hat, wurde in zahlreichen Studien gezeigt (Narciss, 2014; Stark, Kopp, & Fischer, 2009; Hattie & Timperley, 2007; Krause, 2007; Butler & Winne, 1995).

In Bezug auf das Vorwissen liegt es nahe, dass Lernende mit hohem fachspezifischem Vorwissen weniger Fehler machen und daher Feedback auch seltener in Anspruch nehmen sollten. Sofern in Studien keine Lernwirksamkeit von Feedback nachweisbar ist, kann dies u. a. daran liegen, dass Lernende mit hohem Vorwissen das bereitgestellte Feedback nicht benötigen. Trotz der hohen Bedeutsamkeit des Vorwissens als Einflussfaktor, wurde es in vielen Untersuchungen nicht beachtet (Huth, 2004). Eine Erklärung hierfür gibt die Metaanalyse von Bangert-Drowns u. a. (1991). Demnach stellen Vortests eine Art Aktivierung von domänenspezifischem Wissen dar und können den möglichen Lernzuwachs durch Feedback verringern (vgl. auch Kapitel 2.2.1). Vor dem Hintergrund der möglicherweise geringeren Effektstärken bei Studien ohne Vortests, aufgrund von Probanden mit hohem Vorwissen, die kein Feedback benötigen, ist dieser Befund jedoch zu hinterfragen.

Nachfolgend werden einige Ergebnisse von Studien vorgestellt, welche das Vorwissen als Einflussfaktor in ihre Untersuchungen einbezogen. Krause (2007) sowie Stark, Kopp und Fischer (2009) stellten jeweils einen moderierenden Effekt des fachspezifischen Vorwissens fest, wobei eine Angleichung des Wissensniveaus durch elaboriertes Feedback lediglich bei Krause (2007) nachgewiesen werden konnte. Erwartungsgemäß wurde ein

geringeres Vorwissen durch das elaborierte Feedback kompensiert. Stark, Kopp und Fischer (2009) konnten dies nicht bestätigen. Diese Ergebnisse lassen auf einen umgekehrt propositionalen Zusammenhang zwischen dem Vorwissensniveau und der Komplexität des bereitgestellten Feedbacks schließen. Lernende mit einem geringen Vorwissen profitieren von der zusätzlichen externen instruktionalen Unterstützung, wohingegen Lernende mit hohem Vorwissen diese Informationen nicht benötigen. Clariana (1993) konnte eine Überlegenheit von MTF gegenüber KCR bei Lernenden mit hohem Vorwissen nachweisen. Bei vorwissensarmen Probanden erweist sich KCR gegenüber MTF als lernförderlicher. Gemäß Clariana (2000) lassen sich hierfür zwei Erklärungsansätze finden. Der affektive Erklärungsansatz betont, dass Lernende mit niedrigem Vorwissen viele Fehler machen und aufgrund der wiederholten fehlerhaften Lösungsversuche entmutigt werden. Lernende mit hohem Vorwissen präferieren das selbständige Finden der Lösung und lernen folglich auch mehr von MTF als von KCR. Der kognitive Erklärungsansatz bezieht sich auf die unterschiedliche Verarbeitung der Feedback-Informationen. Beim KCR wird die richtige Lösung vorgegeben. Das MTF erlaubt es dem Lernenden sich die richtige Lösung selbständig zu erschließen. Dabei erlaubt MTF insbesondere Lernenden mit hohem Vorwissen die richtige Lösung nach einer falschen Antwort eigenständig bei einem weiteren Versuch zu finden. Einige Studien konnten zeigen, dass KCR gegenüber MTF nicht lernwirksamer ist, insofern MTF mit elaboriertem Feedback kombiniert wird (z. B. Clariana, 1990; Nielson, 1990; zit. in Clariana, 1993).

Die Befundlage zum Einfluss der *Sprachfähigkeit* auf die Lernwirksamkeit des Feedbacks ist eher uneinheitlich. In diesen Studien wird das Feedback hinsichtlich seiner Form in explizites und implizites Feedback unterschieden. Yilmaz und Granena (2016) führten eine Studie mit türkischen Studenten durch, welche Englisch als Fremdsprache im Rahmen ihres Studiums vertieften, und fanden heraus, dass Probanden mit hoher Sprachfähigkeit eher vom expliziten als vom impliziten Feedback profitieren. Bei niedriger Sprachfähigkeit spielt es dagegen keine Rolle, ob die Probanden implizites oder explizites Feedback erhalten. Probanden mit hoher Sprachfähigkeit sind eher in der Lage mit explizitem und korrektivem Feedback umzugehen. Sie vertrauen auf ihre sprachlichen Fähigkeiten und sind somit eher in der Lage Regeln aus dem erhaltenen Feedback abzuleiten. Die Sprachfähigkeit war demnach in dieser Studie lediglich ein Prädiktor der Nachtestleistung für Probanden, welche explizites Feedback erhielten. Weiterhin zeigte sich, dass die Probanden, welche Feedback erhielten, gegenüber der Kontrollgruppe, unabhängig

von der Sprachfähigkeit, im Nachtest bessere Ergebnisse erzielen. Daraus kann geschlossen werden, dass der sprachliche Vorteil allein für eine Kompensation der Wirksamkeit des Feedbacks nicht ausreicht. Li (2013) führte eine Studie mit amerikanischen Studenten durch, welche Chinesisch im Rahmen ihres Studiums erlernten. Dabei konnte die Sprachfähigkeit als Prädiktor der Nachtestleistung für Probanden, welche implizites Feedback erhielten, identifiziert werden. Basierend auf den Ergebnissen seiner Studie und der insgesamt uneinheitlichen Befundlage stellte Li (2013) die Hypothese auf, dass eine hohe Sprachfähigkeit in Kombination mit implizitem Feedback beim Lernen einfacher, transparenter Strukturen innerhalb der eigenen Verarbeitungskapazität und in Kombination mit explizitem Feedback beim Lernen schwieriger, opaker Strukturen, bei dessen Verarbeitung hohe kognitive Anstrengungen vonnöten sind, lernwirksam ist. Insgesamt ist weitere Forschung nötig. Dabei stellt sich insbesondere die Frage, welchen Einfluss die Sprachfähigkeit auf das Lernen in anderen Domänen hat. Die Vergleichbarkeit der Studien gestaltet sich schwierig, da die Sprachfähigkeit mit unterschiedlichen Tests gemessen wird, was u. a. durch den Einsatz verschiedener Sprachen begründet ist.

Im Makromodell von Kulhavy und Stock (1989) wurde auf die *Antwortsicherheit* als Einflussfaktor eingegangen. Der Zusammenhang zwischen Diskrepanz und Feedbackbearbeitungszeit sowie zwischen Antwortsicherheit und Behaltensdauer konnte dabei empirisch bestätigt werden. Bei falschen Antworten entsteht eine hohe Diskrepanz zwischen der Antwort des Lernenden und der Feedbackinformation. Eine hohe Antwortsicherheit führt dabei aufgrund größerer Anstrengung zu einer höheren Feedbackbearbeitungszeit und infolgedessen auch zu einer längeren Behaltensdauer als eine niedrige Antwortsicherheit. Es konnte nicht empirisch bestätigt werden, dass sich die möglichen Wirkungen unterschiedlich komplexer Feedbackarten auf die Variable Antwortsicherheit zurückführen lassen (vgl. Kapitel 2.1.4.1).

Im Kapitel 2.1.4.3 wurde bereits die Bedeutung einer internalen, variablen und kontrollierbaren *Attribution* hervorgehoben. Es stellt sich dabei die Frage, welche Möglichkeiten es bei der Feedbackgestaltung gibt, damit der Feedbackempfänger dazu angeregt wird, „persönliche Verantwortung“ zu übernehmen und individuelle Anpassungsprozesse stattfinden. Es existieren wenige Studien, welche zwischen verschiedenen Attributionsformen, wie z. B. Wissen oder Zufall, unterscheiden (Strijbos & Müller, 2014, S. 95). In der Studie von Lyden, Chaney, Danehower und Houston (2002) wurde festgestellt, dass zwar die Gestaltung der Feedbacknachricht eine Rolle spielt, es aber dabei eher auf die Valenz

der Rückmeldung als auf weitere attributionale Hinweise ankommt. Die teilnehmenden Studierenden attribuieren Erfolg eher auf die eigenen Fähigkeiten (intern) und Misserfolg auf die Aufgabenschwierigkeit (extern), unabhängig davon, welche attributionalen Hinweise sie erhalten. Begründet wird dies u. a. mit der Reihenfolge der Hinweise. Die Rückmeldungen über Erfolg bzw. Misserfolg werden dabei zuerst gegeben und somit der Fokus hierauf gelenkt. Die Äußerungen über Erfolg und Misserfolg sollten demnach äußerst sensibel behandelt und attributionale Hinweise mehr in den Fokus gestellt werden, um positive Effekte auf die Leistung zu erzielen (Lyden u. a., 2002).

Der Einfluss der *Selbstwirksamkeitsüberzeugungen* (vgl. auch Kapitel 2.1.4.3) auf die Lernwirksamkeit von Feedback konnte in einigen Studien empirisch bestätigt werden (z. B. Lyden u. a., 2002; Stajkovic & Sommer, 2000). Kritische Rückmeldungen bei Lernenden mit einer niedrigen Selbstwirksamkeitserwartung führten bei Stajkovic und Sommer (2000) zu geringeren Leistungen. Lyden u. a. (2002) stellten fest, dass Feedback die Selbstwirksamkeitserwartungen und die nachfolgende Leistung lediglich bei einer kontrollierbaren Ursachenzuschreibung für die zugrundeliegende Leistung beeinflussen kann.

Der Einfluss der *Persönlichkeitsdimensionen (Big 5)* auf die Verarbeitung von Feedback wurde bislang wenig erforscht (Strijbos & Müller, 2014, S. 101). Die Befunde verweisen auf einen Zusammenhang der Big 5 mit der Feedbackverarbeitung, beschränken sich jedoch auf den Unternehmenskontext (z. B. Atwater & Brett, 2005; Bell & Arthur, 2008).

In Bezug auf die *Lernmotivation* existieren in der Motivationsforschung zwei Perspektiven, bei denen die Motivation zum einen als abhängige Variable und zum anderen als unabhängige Variable betrachtet wird (Vollmeyer & Rheinberg, 1998). Die Lernmotivation als abhängige Variable wurde vielfach erforscht, um zu analysieren, inwieweit Feedback die Lernmotivation beeinflusst (z. B. Rakoczy u. a., 2008; Vollmeyer & Rheinberg, 2005). Zur Frage, inwieweit sich die Lernmotivation auf Feedback auswirkt, wurde bisher wenig geforscht. DePasque und Tricomi (2015) untersuchten den Einfluss der intrinsischen Motivation auf die Feedbackverarbeitung während und nach dem Lernen von

Wortassoziationen. Die Lernphasen wurden hierzu mit KCR-Feedback kombiniert. Insbesondere sollte dabei mittels eines motivationsfördernden Interviews³⁷ (motivational interviewing) erforscht werden, inwieweit eine höhere Motivation auch zu einem besseren Lernergebnis führt. Hierzu wurde eine Experimentalgruppe mit einer Kontrollgruppe verglichen, mit welcher kein Interview durchgeführt wurde. Die Experimentalgruppe erzielte im Nachtest ein besseres Lernergebnis als die Kontrollgruppe. Dabei steigerte sich die Motivation jedoch nicht bei der gesamten Experimentalgruppe. Es konnte dennoch gezeigt werden, dass die Lernenden mit dem größten Motivationsanstieg auch den größten Leistungsanstieg erzielten. Diese Lernenden unternahmen größere Anstrengungen, um ihre Aufmerksamkeit auf einem hohen Niveau zu halten und während der gesamten Lernphase vom Feedback zu lernen, wohingegen weniger motivierte Probanden den Fokus am Ende der Treatmentphase verlieren. In der in Kapitel 2.2.2 vorgestellten ersten Studie von Narciss (2006) konnte gezeigt werden, dass die Wirkung von ITF insbesondere vom individuellen Faktor Kompetenzeinschätzung abhängt. Aktuelle Ansätze zur Lernmotivation betrachten die Einschätzung und das Erleben der eigenen Kompetenz als zentral für das Motivationsgeschehen (Narciss, 2006, S. 131). In der Studie von Narciss (2006) erzielten Versuchspersonen mit einer hohen Kompetenzeinschätzung unter der ITF-Bedingung bezüglich des Leistungsindexes überdurchschnittliche Werte, wohingegen Versuchspersonen mit niedriger Kompetenzeinschätzung, unabhängig von der Feedbackbedingung, unterdurchschnittliche Werte erzielten. Bei Versuchspersonen mit mittlerer Kompetenzeinschätzung ergab sich kein eindeutiges Bild. Daraus schlussfolgerte Narciss (2006), dass in offenen Lehr-Lernsituationen, bei denen es dem Lernenden frei gestellt ist, wie viel, wie lange und wie intensiv er sich mit Lernaufgaben beschäftigen, Lernende mit einer hohen Kompetenzeinschätzung von ITF profitieren. Für Lernende mit mittlerer bzw. geringer Kompetenzeinschätzung war ITF hinsichtlich der Leistung hingegen nicht wirksam.

Wie bereits in Kapitel 2.1.4.3 beschrieben, lassen sich zwei Zielorientierungsdimensionen unterscheiden: die Performanz- und die Lernorientierung. In einigen Studien konnte der positive Zusammenhang zwischen Lernorientierung und Leistung bestätigt werden

³⁷ Im Rahmen des motivationsfördernden Interviews sollten die Probanden ihre Motivation auf einer Skala (von 0 bis 10) mit der Frage, wie wichtig es ihnen ist, bei der Aufgabe ein gutes Ergebnis zu erzielen, einschätzen. Anschließend wurden sie vom Interviewer aufgefordert mindestens zwei Gründe für ihre Einschätzung zu nennen (DePasque & Tricomi, 2015).

(z. B. Heslin & Latham, 2004). Des Weiteren wurde erforscht, inwieweit die beiden Zielorientierungsdimensionen die Verarbeitung von positivem und negativem Feedback beeinflussen. Cianci, Schaubroeck und McGill (2010) stellten fest, dass eine Lernorientierung eher dazu beiträgt, mit negativem Feedback konstruktiv umzugehen und es nicht als selbstwertgefährdend wahrzunehmen. Performanzorientierung kann bei einer niedrigen Selbstwirksamkeitserwartung in Verbindung mit negativem Feedback eine bedrohliche Druck- bzw. Erwartungssituation hervorrufen (z. B. Ilgen & Davis, 2000; Dweck, 1986). Weitere Studien (Cianci u. a., 2010; Donovan & Hafsteinsson, 2006) zeigen, dass positives Feedback das Bedürfnis von performanzorientierten Menschen nach Kompetenzerleben befriedigt und eher zu Leistungssteigerungen führen kann als bei lernorientierten Personen, welche bei positiven Rückmeldungen wenig Lernpotential erkennen.

Studien zum Einfluss der zwischenmenschlichen Beziehung

Die zwischenmenschliche Beziehung in *vertikalen Konstellationen* wurde bisher eher im Unternehmenskontext erforscht. Erst in den letzten Jahren wurde auch der Beziehung zwischen Lehrenden und Lernenden im Feedbackkontext mehr Beachtung geschenkt. Dialoge finden dabei nach einer Studie von Blair und McGinty (2013) eher abwärtsgerichtet statt. Hierbei agiert der Lehrende als Experte und der Lernende erhält wenig Kontrolle über sein eigenes Lernen. Aus diesem Grund sollten Dialoge insgesamt kooperativ gestaltet werden, wodurch eine aufwärtsgerichtete Kommunikation sowie Dialoge zwischen den Lernenden ermöglicht werden. De Klejn, Meijer, Pilot und Brekelmans (2014) fanden heraus, dass die studentische Wahrnehmung der zwischenmenschlichen Nähe zum Betreuer der Masterarbeit einen großen Einfluss auf die Zufriedenheit mit der Betreuung hat. Dabei kann Feedback insbesondere bei Betreuungen mit weniger Nähe die Zufriedenheit erhöhen. Inwieweit dies einen Einfluss auf die Lernwirksamkeit hat, wurde dabei nicht untersucht.

In der Literatur zu Peer-Assessment, d. h. in *horizontalen Konstellationen* werden Verzerrungen, welche im Rahmen des Gebens von Feedbacks auftreten können, unter dem Begriff „reciprocity effects“ zusammengefasst (Strijbos u. a., 2009). Hierbei zeigen einige Studien, dass eine freundschaftliche Beziehung zu Überbewertungen, d. h. einer weniger strengen Einschätzung, führen kann (z. B. Cheng & Warren, 1997). Demgegenüber stehen Untersuchungen, welche postulieren, dass eine enge und vertraute Beziehung, eine geschützte Atmosphäre schafft, die kritische Äußerungen zur Leistung des Gegenübers begünstigt (z. B. van Gennip, Segers, & Tillema, 2010). Eine Untersuchung von Jehn &

Shah (1997), welche im Kontext des kooperativen Lernens durchgeführt wurde, zeigte ähnliche Ergebnisse. Hierbei erzielten Gruppen mit freundschaftlichen Beziehungen der Gruppenmitglieder auch bessere Ergebnisse bei kognitiven und motorischen Aufgaben im Vergleich zu Gruppen, die sich lediglich flüchtig kannten. Dabei spielt sicherlich der Aspekt eine große Rolle, ein gemeinsames Ziel zu erreichen. Inwieweit dem Gegenüber ein ehrliches und auch ggf. kritisches Feedback übermittelt wird, sofern die Leistung jedes einzelnen bewertet wird, wird hierbei vernachlässigt. Insgesamt ist die Befundlage also eher kontrovers.

2.2.5 Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit situativer Faktoren

Im Kapitel 2.2.5 werden Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von situativen Einflussfaktoren analysiert. Hierbei werden die *Bezugsnormen* (Gruppe, vorgegebenes Lernziel, eigener Lernfortschritt), die *formative vs. der summativen Leistungsbeurteilung*, die *Lernaufgaben* (insbesondere die Aufgabenanforderungen), die *Art der Lernsituation*, *mögliche Quellen für Fehlkonzepte und Fehler* sowie die *Fehlerfreundlichkeit* betrachtet. Hinsichtlich der *Normtransparenz* sowie der *kulturellen Rahmung*³⁸ wurden im Kontext der Arbeit keine Studien gefunden. Daher werden diese Faktoren nicht näher betrachtet.

Zum Einfluss der *formativen Leistungsbeurteilung* im Vergleich zur traditionellen *summativen Leistungsbeurteilung* wurde in den letzten Jahren viel geforscht. Zu nennen ist beispielsweise das Projekt Co²CA (Conditions and Consequences of Classroom Assessment), welches von der DFG gefördert wurde (Bürgermeister, Klieme, Rakoczy, Harks, & Blum, 2014). Im Rahmen dieses Projekts wurde u. a. eine Feldstudie im Mathematikunterricht durchgeführt, um die Wirksamkeit der formativen Leistungsbeurteilung mit einer traditionell unterrichteten Kontrollgruppe ohne spezifischer, vorgegebener Leistungsbeurteilung und Rückmeldung zu vergleichen. Hierbei zeigt sich eine eher adaptive und kompetenzunterstützende Wahrnehmung durch die Schüler der Experimentalgruppe im Vergleich zu den Schülern der Kontrollgruppe. Dies wirkte sich wiederum auf die Leis-

³⁸ Hierzu existieren lediglich Studien, welche kollektivistische (z. B. China) und individualistische Kulturen (z. B. USA) in Bezug auf unterschiedliche Präferenzen hinsichtlich der Gestaltung von Feedback vergleichen (Kung, 2008; Luque & Sommer, 2000). Inwieweit die kulturelle Rahmung dabei einen Einfluss auf die Lernwirksamkeit hat, wurde nicht untersucht.

tung und die Motivation förderlich aus (Bürgermeister u. a., 2014). Darüber hinaus existieren einige Metaanalysen, welche die Lernwirksamkeit der formativen Leistungsbeurteilung bestätigen (Kingston & Nash, 2011; Black & Wiliam, 1998).

Zur Untersuchung des Einflusses der *Bezugsnormen* auf die Lernwirksamkeit des Feedbacks zeigen mehrere Studien, dass individuelle und sachliche Bezugsnormen gegenüber der sozialen Bezugsnorm lernwirksamer sind (z. B. Rheinberg & Fries, 2010). Die Lernwirksamkeit wird dabei jedoch vom Leistungsstand der Lernenden moderiert. Krampen (1985) fand in einer Studie im Mathematikunterricht heraus, dass lediglich bei der sachlichen Bezugsnorm (z. B.: „Du kennst die Formel, aber hast falsch gerechnet.“) der Inhalt eines Feedbackkommentars allein für die Art der Wirkung ausschlaggebend ist. Bei der sozialen und individuellen Bezugsnorm hängt die Wirkung vom Leistungsstand der Schüler ab. Sozial orientiertes Lehrerfeedback, d. h. der Vergleich mit der Gruppe (z. B. „Du hast ein besseres Ergebnis als der Klassendurchschnitt erzielt.“), wirkt bei leistungsschwachen Schülern deutlich negativ, bei leistungsstärkeren zeigt sich dagegen eine neutrale oder positive Wirkung. Von individuell orientierten Kommentaren, d. h. in Bezug auf den eigenen Lernfortschritt (z. B. „Du hast deine Leistung im Vergleich zur vorherigen Klassenarbeit leicht verschlechtert.“), profitieren insbesondere die leistungsschwachen Schüler. Sie werden auf diese Weise nicht entmutigt, haben eine größere Hoffnung auf Erfolg und weniger Furcht vor Misserfolgen. Für leistungsstarke Schüler dagegen wirkt der Vergleich mit der Gruppe eher motivierend und lernwirksam. Die sachliche Bezugsnorm lenkt aufgrund ihrer klaren Zielbezogenheit die Aufmerksamkeit eher auf die Aufgabe und weniger auf das Selbst (vgl. auch Kapitel 2.1.2). Aus diesem Grund spielt der Leistungsstand bei dieser Bezugsnorm eine eher untergeordnete Rolle (Rheinberg & Fries, 2010).

Die Lernwirksamkeit des Feedbacks kann weiterhin von den *Aufgabenanforderungen der Lernaufgabe* abhängen. Wie bereits in Kapitel 2.1.4.4 erläutert, können komplexere Anforderungen auch zum Erreichen komplexerer Ziele führen, welche sich an der sachlichen, der sozialen oder der individuellen Bezugsnorm orientieren können. Empirische Untersuchungen zum Einfluss der Komplexität finden sich überwiegend in Bezug auf die sachliche Bezugsnorm. Bei der Metaanalyse von Kluger und DeNisi (1996) zeigten sich deutliche Effekte bei weniger komplexen Anforderungen ($d=0,55$). Der deutliche Feedbackeffekt bei weniger komplexen Aufgaben ist dabei nachvollziehbar, da Fehler im Anschluss an das Feedback leicht zu korrigieren sind. Darüber hinaus handelt es sich um

KCR und nicht um elaboriertes Feedback. Elaboriertes Feedback hätte wahrscheinlich auch bei komplexen Anforderungen einen höheren Effekt erzielt. In weiteren Studien zeigte sich KCR gegenüber elaboriertem Feedback als ausreichend, da aufgrund der niedrigen Aufgabenanforderungen eine vertiefte Verarbeitung der Feedbackinformationen nicht erforderlich war. Den Lernenden genügte die richtige Antwort zum Bearbeiten des Nachtests und sie sahen wahrscheinlich keine Notwendigkeit die zusätzlichen Informationen zu nutzen (z. B. Dempsey, Litchfield, & Driscoll, 1993; Kulhavy, White, Topp, Chan, & Adams, 1985; Phye, 1979). Schimmel (1983) fasste in einem Review die Ergebnisse von 11 Studien zusammen, die KR und KCR mit keinem Feedback beim Bearbeiten von verbal information tasks³⁹ vergleichen. Es zeigte sich ein Vorteil von KR gegenüber den anderen Bedingungen. In dieser Studie war folglich sogar die Information über die Richtigkeit oder Falschheit der Antwort ausreichend. Aufgrund der veränderten Anforderungen im Alltag und in der Arbeitswelt sollte es jedoch nicht im Fokus stehen, die Lernenden mit wenig komplexen Lerninhalten und -aufgaben zu konfrontieren, nur um von Anfang an weniger kognitive Ressourcen zu beanspruchen, sondern eher zu erforschen, inwieweit Feedback die Lernenden insbesondere bei komplexen Anforderungen (als vorherrschende Grundvoraussetzung) unterstützen kann.

Lee, Smith und Savenye (1991) fanden heraus, dass erklärendes und somit elaboriertes Feedback gegenüber KR und KCR in Bezug auf das Anwenden von Regeln in der Computerprogrammierung überlegen ist. Das Anwenden von Regeln⁴⁰ beinhaltet dabei nach

³⁹ Er stützt sich dabei auf Gagné (1980), welcher die Bedingungen menschlichen Lernens in einem Lehrbuch zusammenfasst. Dabei unterscheidet Gagné u. a. verschiedene Leistungsfähigkeiten und beantwortet damit die Frage: „Welche Faktoren bedingen wirklich einen Unterschied für Unterricht?“ (ebd., S. 9). Leistungsfähigkeiten beziehen sich dabei auf die Arten von Lernergebnissen und sind erlernte Dispositionen mit unterschiedlichen Eigenschaften und Strukturen. Dabei werden fünf Hauptarten unterschieden: intellektuelle Fähigkeiten, kognitive Strategien, verbale Informationen, motorische Fertigkeiten und Einstellungen. Das Lernen verbaler Informationen bezieht sich auf „eine große Vielfalt von Dingen, die wir Namen, Fakten oder Ideen nennen“ (ebd., S. 175). Verbale Informationen können dabei sprachlich formuliert werden. Gagné führt aus, dass diese Informationen in verschiedenen Komplexitätsgraden auftreten. Das Lernen von Namen und Bezeichnungen ist dabei noch wenig komplex. Gruppen von Fakten als zusammenhängender Diskurs (z. B. ein Text) zu erlernen, stellt die höchste Komplexitätsstufe dar (ebd., S. 176). Zum Lernen einzelner Wörter oder Wortpaare existieren gegenüber dem Erlernen komplexerer Informationen viele empirische Untersuchungen. Hieraus kann geschlussfolgert werden, dass die bei Schimmel (1983) einbezogenen Studien ebenso verbale Informationen auf einer geringen Komplexitätsstufe analysieren. Anzumerken ist an dieser Stelle, dass Gagné (1980) lediglich die Wissensdimensionen betrachtet. Er geht jedoch davon aus, dass z. B. das Wiedergeben des Sinns eines Textes gegenüber dem wortwörtlichen Erinnern von Namen eine höhere kognitive Anstrengung erfordert (ebd., S. 187). Demnach sind mit dem Lernen von Texten auch höhere Anforderungen an die Lernaufgabe verbunden.

⁴⁰ Gemäß Gagné (1980) steigen intellektuelle Fähigkeiten „im Komplexitätsgrad von den Diskriminationen zu den Begriffen, Regeln und Regeln höherer Ordnung an“ (S. 56). Regeln anzuwenden stellt demnach eine

Gagné (1980) hohe Aufgabenanforderungen. Ebenso stellten Moreno (2004) und Krause (2007, S. 176) in Bezug auf komplexere Lernziele fest, dass die Aneignung anwendbaren und transferfähigen Wissens durch elaboriertes Feedback begünstigt wird (Studien wurden bereits in Kapitel 2.2.2 vorgestellt).

KCR scheint folglich bei niedrigen Aufgabenanforderungen zu genügen, wohingegen bei komplexen Aufgaben elaboriertes Feedback lernwirksamer ist. Diese Studien analysierten jedoch Feedback entweder zu einfachen oder komplexen Aufgaben. Interessanter sind Untersuchungen, die Aufgaben verschiedener Komplexitätsstufen miteinander vergleichen. Hierzu wurde bisher wenig geforscht. Lindell (1976) untersuchte den Einfluss von „Cognitive Feedback“ im Vergleich zu „Outcome Feedback“⁴¹ bei einfachen und komplexen Aufgaben. Dabei zeigte sich eine Überlegenheit des „Cognitive Feedback“, d. h. des elaborierten Feedbacks, sogar bei einfachen Aufgaben. Bei komplexen Aufgaben war der größte Effekt des elaborierten gegenüber dem einfachen Feedback erkennbar. In weiteren Studien zeigte sich jedoch kein Vorteil des elaborierten gegenüber einfachem Feedback bei komplexen im Vergleich zu einfachen Aufgabenanforderungen (z. B. O. M. Lee, 1985; Wager, 1983).

Schimmel (1988) schlussfolgerte aus einem Review zur Feedbackforschung, dass für Aufgaben höherer Taxonomiestufen, welche nach Gagné (1980) das Anwenden von Konzepten und Regeln erfordern, nicht pauschal von einem Vorteil elaborierten Feedbacks ausgegangen werden kann. Schimmel (1988) stellt heraus, dass leistungsstarken Lernenden eine Wahlmöglichkeit zwischen verschiedenen Feedbackoptionen mit unterschiedlichen Mengen an Informationen gegeben werden sollte. Beim Vermitteln bestimmter Verfahrensweisen und somit prozeduralen Wissens ist bug-related Feedback (vgl. Kapitel 2.1.3.2) besonders effektiv, insofern Fehler zufällig identifiziert und durch Feedback korrigiert werden können. Die Schlussfolgerungen von Schimmel könnten eine Erklärung

höhere kognitive Anforderung an den Lernenden als lediglich Begriffe im Rahmen des Diskriminationslernens zu unterscheiden. Mory (2004, S. 777) differenziert in diesem Zusammenhang zwischen „lower-order learning“ (für das Diskriminationslernen) und „higher-order learning“ (für das Lernen von Begriffen, Regeln, Regeln höherer Ordnung bzw. Problemlösen).

⁴¹ „Outcome Feedback“ informiert den Lernenden lediglich über die richtige Lösung und ist somit mit KCR, also einfachem Feedback vergleichbar. „Cognitive Feedback“ erklärt dem Lernenden, wie und warum das gewünschte Ergebnis erreicht wurde bzw. wie das Ziel erreicht werden kann. Es stellt dem Lernenden mehr Informationen zur Verfügung. Dabei ist es möglich, dem Lernenden die richtige Lösung nicht von Anfang an zu präsentieren, um Denkprozesse anzuregen. Es kann somit als elaboriertes Feedback bezeichnet werden (Lindell, 1976).

für die divergierenden Ergebnisse der Vergleichsstudien in Bezug auf einfache im Vergleich zu komplexen Aufgaben sein. Die informelle Kompetenz der Lernenden scheint also die Feedbackwirkung, insbesondere der komplexen Aufgaben, zu moderieren. Das elaborierte Feedback könnte folglich in diesem Zusammenhang keine Wirkung gezeigt haben, da die Lernenden aufgrund ihrer Leistungsstärke eine so umfangreiche Rückmeldung gar nicht benötigt hätten.

Neben der Komplexität des Feedback-Inhalts können der Zeitpunkt der Feedbackgabe und die Anzahl der Lösungsversuche ebenso von den Aufgabenanforderungen der Lernaufgabe abhängen.

Clariana u. a. (2000) fanden heraus, dass verzögertes im Vergleich zu sofortigem Feedback die größten Effekte bei Aufgaben mit hoher Komplexität erzielt. Dabei ermöglicht verzögertes Feedback sich länger mit der Aufgabe auseinanderzusetzen, das Feedback stellt dabei eine erneute Lernphase dar. In Bezug auf die Anzahl der Lösungsversuche führten Clariana und Koul (2005) ein Review anhand von Feedback-Studien durch und stellten fest, dass MTF im Vergleich zu anderen Feedbackarten bei einfachen Aufgaben wenig effektiv ist. Für Aufgaben mit einer höheren Komplexität zeigte sich ein gegensätzliches Bild. Hier war MTF anderen Feedbackarten überlegen. Clariana und Koul (2006) integrierten in ihrer Studie sowohl einfache als auch komplexe Aufgaben und konnten den höchsten Effekt für MTF bei komplexen Aufgaben nachweisen.

Zum Einfluss der *Art der Lernsituation* fanden Bangert-Drowns u. a. (1991) in ihrer Metaanalyse heraus, dass Feedback in programmiertem ($d=-0,04$) und computergestütztem Unterricht ($d=0,22$) weniger effektiv ist als beim verstehenden Lesen ($d=0,48$) und herkömmlichen Prüfungssituationen ($d=0,63$). An dieser Stelle argumentieren die Autoren ebenfalls mit der Komplexität. Ihrer Meinung nach ist es beim verstehenden Lesen und in herkömmlichen Prüfungssituationen eher möglich, anhand von Fragen, größere und komplexere Inhaltsgebiete zu beleuchten. In computergestütztem bzw. programmiertem Unterricht können nur sehr kleine Informationsteile präsentiert werden, welche mittels stark lenkender Fragen im Anschluss überprüft werden (vgl. Kapitel 2.2.1). Baralt (2013) führte eine Studie mit Studierenden durch, welche Spanisch als Fremdsprache erlernten, wobei die Untersuchung (inkl. der Tests) zum einen computergestützt und zum anderen in einer natürlichen Umgebung (face-to-face) stattfand. Dabei sollten die Studierenden eine in Englisch gelesene Geschichte in Spanisch wiedergeben. Die Aufgabenkomplexität wurde dahingehend variiert, dass ein Teil der Studierenden als zusätzliche komplexere

Aufgabe für die Charaktere der Geschichte die Gründe für einige Handlungen angeben sollten. Diese Gründe waren aus dem Text nicht direkt ableitbar. Die Interaktion mit dem Versuchsleiter und somit auch das Feedback in Form von „recasts“ fanden dabei persönlich oder via Chat am Computer statt. Baralt (2013) stellte fest, dass Feedback beim Bearbeiten von komplexen Aufgaben lediglich in Face-To-Face-Bedingungen lernwirksam ist. Das Bearbeiten einfacher Aufgaben war in der computergestützten Umgebung am lernwirksamsten. Eine zu hohe Aufgabenkomplexität in der computergestützten Umgebung führt nach Meinung von Baralt (2013) zu einer hohen kognitiven Belastung. Die Ergebnisse sind dabei jedoch kritisch zu hinterfragen, da die gesamte Untersuchung computergestützt bzw. face-to-face stattfand. Es kann folglich nicht eindeutig geschlossen werden, inwieweit sich die Ergebnisse auf die Art der Lernsituation oder das Präsentationsformat des Feedbacks zurückführen lassen. Baralt (2013) argumentiert hier auch selbst, dass die Studierenden der computergestützten Umgebung wahrscheinlich nicht in der Lage waren, das Feedback tiefergehend zu verarbeiten. Des Weiteren wurde bereits in Kapitel 2.2.2 auf den möglichen Einfluss der Art der Lernsituation hingewiesen.

Einen weiteren Einflussfaktor stellen die *Quellen für Fehlkonzepte und Fehler* sowie die *Fehlerfreundlichkeit* dar. Hierzu führten Kline, Schumaker & Deshler (1991) eine Studie mit kanadischen Schülern zum schriftlichen Bilden von Sätzen durch. Hierbei wurde analysiert, inwieweit sich Feedback, welches konkrete Fehleranalysen implementierte, vom im Unterricht gegebenen Standardfeedback (Kontrollgruppe) unterschied. Die Studie umfasste zwei Experimentalgruppen mit speziellen Feedbackprozeduren („Feedback Routine“ und „Feedback-Plus-Acceptance-Routine“). Zur Gestaltung dieser Prozeduren stützten sich Kline u. a. (1991) auf zahlreiche Studien (z. B. Walberg, 1986; Kulhavy, 1977). Jede Feedbackprozedur bestand aus mehreren Stufen. Zunächst wurde hierzu in beiden Experimentalgruppen eine Fehleranalyse vorgenommen, auf deren Basis die Fehler kategorisiert, notiert und dem Schüler für eine Korrektur mitgeteilt wurden. Im nächsten Schritt wurden die Korrekturen der Schüler überprüft und ggf. verbessert. Anschließend wurde eine individuelle Feedback-Besprechung durchgeführt, in der jeder Schüler zunächst drei positive Statements zu seinem Ergebnis abgeben sollte. Falls dies dem Schüler nicht möglich war, unterstützte der Lehrer mit positiven Statements (z. B. „Du hast dich daran erinnert, das erste Wort in jedem Satz groß zu schreiben.“). Der Lehrer beschrieb anschließend eine Fehlerkategorie, nannte Beispiele und erläuterte Methoden,

um diesen speziellen Fehler in der Zukunft zu vermeiden. Der Schüler sollte diese Methode anschließend erfolgreich anwenden. Sobald alle Kategorien besprochen wurden, fasste der Lehrer die Fehler-Besprechung noch einmal zusammen. In die „Feedback-Plus-Acceptance-Routine“ wurde der Schüler in die Abschlussphase aktiv einbezogen, indem er dazu aufgefordert wurde, das Gelernte zusammenzufassen. Darüber hinaus sollte er sich während der gesamten Sitzung zur Feedbackakzeptanz durch verbales und nonverbales Verhalten äußern. Dies wurde vorab geübt. Insgesamt benötigten die Schüler der Kontrollgruppe signifikant mehr Versuche zum erfolgreichen Abschließen einer bestimmten Niveaustufe als die Experimentalgruppen. Darüber hinaus reduzierten die Schüler in beiden Experimentalgruppen signifikant mehr Fehler vom ersten zum zweiten Versuch. Aufgrund der detaillierten und individuellen Analyse sowie Besprechung der Fehlerkonzepte und Fehler mit jedem einzelnen Schüler konnte somit eine Überlegenheit gegenüber herkömmlichem Unterrichtsfeedback nachgewiesen werden. In weiteren Studien, in denen kognitive Anforderungs- und Fehleranalysen als Basis für elaboriertes Feedback dienten, konnten positive Wirkungen des elaborierten Feedbacks gegenüber einfachem Feedback nachgewiesen werden (z. B. Narciss, 2006; Narciss & Huth, 2006).

2.2.6 Zusammenfassung der empirischen Befunde zur Lernwirksamkeit des Feedbacks

Im nachfolgenden Kapitel 2.2.6 werden die in den Kapiteln 2.2.1 bis 2.2.5 dargestellten empirischen Befunde zur Lernwirksamkeit des Feedbacks zusammengefasst.

Feedback ist einer der größten Einflussfaktoren auf das Lernen (Hattie, 2015). Zur Erforschung der Lernwirksamkeit von Feedback existieren einige Metaanalysen, wobei die Metaanalysen von Bangert-Drowns u. a. (1991) sowie Kluger und DeNisi (1996) die bedeutendsten Metaanalysen darstellen. Beide Metaanalysen zeigen, dass Feedback nicht generell lernwirksam ist, sondern von verschiedenen Faktoren moderiert wird.

Feedback lässt sich anhand verschiedener Kriterien gestalten. Neben der Komplexität der Feedbacknachricht kann sich die Gestaltung dabei auch auf die Feedbackform (Feedbackquelle, Feedbackempfänger, Zeitpunkt der Feedbackgabe, Transparenz für den Feedbackempfänger, Präsentationsformat, Anzahl erneuter Lösungsversuche) beziehen. Die Lernwirksamkeit des Feedbacks in Abhängigkeit von der Komplexität der Feedbacknachricht wurde in der Feedbackforschung bisher am meisten erforscht (Hattie und Wollenschläger,

2014, S. 138 f.; Strijbos u. a. 2010, S. 293, Shute, 2008, S. 292). Es zeigte sich kein genereller Vorteil für das elaborierte gegenüber dem einfachen Feedback. Die uneinheitliche Befundlage lässt auf den Einfluss weiterer Faktoren schließen.

Hinsichtlich der *Feedbackquelle* sprechen die Studien für einen Vorteil des *externen* gegenüber dem *internen* Feedback, wobei die wahrgenommene Feedbackeffektivität als Mediator auf die verbesserte Leistung wirkt (z. B. Wollenschläger u. a., 2012). In Bezug auf den *Feedbackempfänger*, den *Zeitpunkt der Feedbackgabe* sowie die *Transparenz für den Feedbackempfänger* zeigten Studien insgesamt uneinheitliche Befunde. Hinsichtlich einer *sequentiellen* Präsentation gegenüber dem *gleichzeitigen* Präsentieren von Informationen wird insbesondere im Zusammenhang mit elaboriertem Feedback ein Vorteil nachgewiesen. An dieser Stelle wird empfohlen, das elaborierte Feedback zunächst ohne die richtige Lösung zu geben, um eine tiefere Verarbeitung des elaborierten Feedbacks zu ermöglichen (z. B. Narciss & Huth, 2006). Hinsichtlich des *Präsentationsformats* konnten einige Studien einen Vorteil von *mündlichem* gegenüber *schriftlichem* Feedback nachweisen, da zusätzliche Kommunikationskanäle die Verarbeitung der Informationen erleichtern (z. B. Krause, 2007). Andere Studien sprechen für einen Vorteil des *computergestützten* gegenüber *verbalem* Feedback, da verbales Feedback die Aufmerksamkeit auf den Feedbackgeber lenkt und somit zu sehr vom Feedbackinhalt ablenken kann (z. B. Kluger & DeNisi, 1996). Bei computergestütztem Feedback zeigen einige Studien eine Überlegenheit der *bimodalen* (auditiv und visuell) gegenüber der *unimodalen* Präsentation (visuell) von Informationen (z. B. Mayer & Moreno, 1998). Die kombinierte Nutzung des auditiven und visuellen Informationsverarbeitungssystems reduziert die kognitive Belastung des Lernenden, welche bei einer unimodalen simultanen Präsentation verbaler und piktorialer Informationen in visueller Form durch die Aufteilung der verfügbaren Aufmerksamkeitsressourcen sehr hoch ist (Mayer & Moreno, 2003). In Bezug auf die Anzahl der Lösungsversuche verwenden viele Studien lediglich einen Versuch (d. h. *STF*). Untersuchungen, welche mehrere Lösungsversuche anbieten, konnten keine generelle Überlegenheit von *MTF* oder *AUC* gegenüber *STF* nachweisen.

Hinsichtlich der Gestaltung von Feedback zeigt sich folglich eine uneinheitliche Befundlage. Insgesamt wurden daher verschiedene Studien zur Lernwirksamkeit des Feedbacks unter Berücksichtigung (inter-)personaler sowie situativer Faktoren, auf Grundlage des in Kapitel 2.1.4.4 dargestellten erweiterten Rahmenmodells, analysiert.

Studien zum *Feedbacksender* wurden in Bezug auf den *formalen* und den *informellen* Status analysiert. Feedback in einer vertikalen Sender-Empfänger-Konstellation wird typischerweise abwärtsgerichtet gegeben. Der Einfluss des *formalen Status* auf die Lernwirksamkeit des Feedbacks wurde bisher jedoch nicht systematisch untersucht. Inwieweit der formale Status beim aufwärtsgerichteten Feedback eine Rolle spielt, wurde ebenso nicht systematisch erforscht. Bei der horizontalen Sender-Empfänger-Konstellation zeigen die Feedbackempfänger eine höhere Unsicherheit in Bezug auf wahrgenommene Fairness und Nützlichkeit des Feedbacks (z. B. Brown u. a., 2009). In diesem Zusammenhang können Begründungen zum Peer-Feedback, in Form von schriftlichen Berichten für die Lehrkraft, insbesondere die Leistung von Lernenden mit geringem Vorwissen verbessern (z. B. Gielen u. a., 2010). An dieser Stelle besteht noch weiterer Forschungsbedarf.

In Bezug auf den *informellen Status* zeigt sich, dass Lehrkräfte ihre tatsächliche Kompetenz in Bezug auf das Geben von Feedback oft überschätzen. Einige Studien (z. B. Strijbos u. a. 2010) konnten zeigen, dass elaboriertes Feedback von einer als hochkompetent eingeschätzten Person am adäquatesten eingeschätzt wird. Die zugeschriebene Kompetenz kann folglich über die Einschätzung der Adäquatheit des Feedbacks (Fairness, Nützlichkeit, Akzeptanz) operationalisiert werden. Ein positiver Zusammenhang zwischen der wahrgenommenen Nützlichkeit des Feedbacks und einer Leistungsänderung konnte nicht generell nachgewiesen werden (z. B. bei Harks u. a. 2014). An dieser Stelle besteht weiterer Forschungsbedarf.

Studien zum Feedbackempfänger wurden in Bezug auf den *informellen Status* (Vorwissen, Sprachfähigkeit und Antwortsicherheit), die *Attribution*, die *Selbstwirksamkeit*, die *Big 5* und die *Lernmotivation* analysiert.

Dass der *informelle Status* des Empfängers einen großen Einfluss auf die Lernwirksamkeit von Feedback hat, wurde in zahlreichen Studien gezeigt (z. B. Hattie & Timperley, 2007; Butler & Winne 1995). Einige Studien stellten einen moderierenden Effekt des fachspezifischen *Vorwissens* fest. Zum Teil zeigte sich dabei ein umgekehrt propositionaler Zusammenhang zwischen dem Vorwissensniveau und der Komplexität des Feedbackinhalts (z. B. Krause, 2007). Elaboriertes Feedback ermöglicht in diesem Zusammenhang eine Angleichung des Wissensniveaus von vorwissenschwachen an das von vorwissensstarken Lernenden, wobei vorwissensstarke Probanden diese zusätzliche externe instruktionale Unterstützung nicht benötigen. Weiterhin konnte gezeigt werden,

dass Lernende mit hohem Vorwissen von MTF profitieren, wohingegen KCR bei Lernenden mit niedrigem Vorwissen lernwirksamer ist (Clariana, 1993). Dies könnte darin begründet liegen, dass Lernende mit geringem Vorwissen viele Fehler machen und schneller entmutigt sind, d. h. die richtige Lösung bevorzugen und auch kognitiv verarbeiten können. Vorwissensstarke Probanden dagegen sind motiviert und auch kognitiv in der Lage, sich die richtige Lösung selbst zu erschließen (Clariana, 2000).

Zum Einfluss der *Sprachfähigkeit* existieren lediglich Studien, welche explizites und implizites Feedback gegenüberstellten. Die Ergebnisse sind jedoch uneinheitlich.

Im Makromodell von Kulhavy und Stock (1989) konnte ein Zusammenhang zwischen der *Antwortsicherheit* und der Behaltensdauer nachgewiesen werden. Eine hohe Antwortsicherheit führt dabei aufgrund größerer Anstrengung zu einer längeren Feedbackbearbeitungszeit und infolgedessen auch zu einer längeren Behaltensdauer als eine niedrige Antwortsicherheit. Der Einfluss auf die Lernwirksamkeit wurde dabei nicht untersucht.

Hinsichtlich des Einflusses der *Attribution* gilt es beim Lernenden auf negatives Feedback insbesondere eine internale, variable und kontrollierbare Ursachenzuschreibung zu fördern. Hierzu kann die Lehrperson die Aufgabenschwierigkeit variieren (Strijbos & Müller, 2014, S. 94 f.). Inwieweit dies tatsächlich einen Einfluss auf die Lernwirksamkeit des Feedbacks hat, wurde bisher nicht untersucht. In einer Studie von Lyden u. a. (2002) wurde lediglich festgestellt, dass die Reihenfolge der Hinweise einen großen Einfluss auf die Attribution hat. Lehrpersonen sollten demnach mit den Äußerungen zum Erfolg der Leistung äußerst sensibel umgehen und attributionale Hinweise in den Vordergrund stellen (Lyden u. a., 2002).

Der Einfluss der *Selbstwirksamkeitsüberzeugungen* wurde in einigen Studien untersucht (z. B. Lyden u. a., 2002; Stajkovic & Sommer, 2000). Es konnte gezeigt werden, dass eine niedrige Selbstwirksamkeitserwartung bei negativem Feedback auch zu geringeren Leistungen führt. Des Weiteren zeigte sich, dass Feedback die Selbstwirksamkeitserwartungen und die nachfolgende Leistung lediglich bei einer kontrollierbaren Ursachenzuschreibung für die zugrundeliegende Leistung beeinflussen kann. Insbesondere bei negativem Feedback sollte daher eine positive Selbstwirksamkeitserwartung gefördert werden, um Lernende dabei zu unterstützen, Kritik zu akzeptieren, ohne die eigenen Ziele zu revidieren.

Die Befunde zum Einfluss der *Persönlichkeitsdimensionen* (Big 5) auf die Verarbeitung von Feedback beschränken sich bisher auf den Unternehmenskontext (Strijbos & Müller, 2014, S. 101).

Die *Lernmotivation* wurde bisher als abhängige Variable vielfach erforscht, um den Einfluss von Feedback auf die Lernmotivation zu untersuchen (z. B. Rakoczy u. a., 2008). Im Hinblick auf Lernmotivation als unabhängige Variable dagegen existieren lediglich wenige Studien. DePasque und Tricomi (2015) konnten zeigen, dass eine höhere intrinsische Motivation auch zu einem besseren Lernergebnis führt. Darüber hinaus scheinen Lernende mit einer hohen Kompetenzeinschätzung in offenen Lehr-Lernsituationen von elaboriertem Feedback zu profitieren, Lernende mit mittlerer bzw. geringer Kompetenzeinschätzung dagegen nicht (Narciss, 2006). Die Verarbeitung von positivem und negativem Feedback unterscheidet sich in Abhängigkeit von der Zielorientierung. Negatives Feedback ermöglicht einen konstruktiven Umgang bei einer Lernorientierung, ruft hingegen bei einer Performanzorientierung eine bedrohliche Druck- und Erwartungssituation hervor (Cianci u. a., 2010). Demgegenüber lässt positives Feedback bei lernorientierten Personen wenig Lernpotential zu, führt hingegen bei performanzorientierten Personen zu einer Bedürfnisbefriedigung und kann bei ihnen somit eher zur Leistungssteigerung beitragen.

Studien zum Einfluss der *zwischenmenschlichen Beziehung* existieren im Hinblick auf vertikale und horizontale Konstellationen. Studien zeigen, dass Dialoge bei vertikalen Konstellationen zwischen Lehrenden und Lernenden überwiegend abwärtsgerichtet stattfinden (z. B. Blair und McGinty, 2013). Ziel sollte es daher sein, Dialoge zukünftig kooperativer zu gestalten, um dem Lernenden mehr Kontrolle über ihr eigenes Lernen zu geben und auch eine aufwärtsgerichtete Kommunikation zu ermöglichen. Des Weiteren konnte im Hochschulkontext festgestellt werden, dass Feedback die Zufriedenheit mit der Betreuung von Masterarbeiten bei Betreuungen mit wenig Nähe erhöhen kann (De Klejn u. a., 2014). Studien zum Peer-Assessment, d. h. zu horizontalen Konstellationen zeigen gegensätzliche Ergebnisse. Zum einen zeigt sich, dass ein freundschaftliches Verhältnis Überbewertungen, d. h. eine weniger strenge Einschätzung, begünstigt (z. B. Strijbos u. a., 2009). Andere Studien gehen dagegen davon aus, dass eine enge und vertraute Beziehung gerade eine geschützte Atmosphäre schafft, welche gegenseitige kritische Äußerungen begünstigt (z. B. van Gennip u. a., 2010). Inwieweit diese Erkenntnisse die Lernwirksamkeit beeinflussen, wurde bisher noch nicht erforscht.

Des Weiteren wurden Studien zur Lernwirksamkeit des Feedbacks in Abhängigkeit von den situativen Einflussfaktoren *Bezugsnormen* (Gruppe, vorgegebenes Lernziel, eigener Lernfortschritt), *formative vs. der summative Leistungsbeurteilung*, *Lernaufgaben* (insbesondere der Aufgabenanforderungen), *Art der Lernsituation*, *mögliche Quellen für Fehlkonzepte und Fehler* sowie *Fehlerfreundlichkeit* betrachtet.

Zum Einfluss der *formativen* im Vergleich zur *summativen Leistungsbeurteilung* existiert eine sehr umfassende Befundlage, welche einen lernförderlichen Einfluss der formativen gegenüber der summativen Leistungsbeurteilung nachweist (z. B. Metaanalyse von Black & William, 1998).

Zur Untersuchung des Einflusses der *Bezugsnormen* auf die Lernwirksamkeit des Feedbacks zeigen mehrere Studien, dass die individuelle und die sachliche Bezugsnormen gegenüber der sozialen lernwirksamer sind (z. B. Rheinberg & Fries, 2010). Die Lernwirksamkeit wird dabei jedoch vom Leistungsstand der Lernenden moderiert (Krampen, 1985).

Die Lernwirksamkeit des Feedbacks kann weiterhin von den *Aufgabenanforderungen der Lernaufgabe* abhängen. Bei einfachen Anforderungen scheint dabei KCR (z. B. Kluger & DeNisi, 1996), teilweise sogar KR (Schimmel, 1983) zu genügen. Bei komplexen Anforderungen zeigen Studien keine generelle Überlegenheit des elaborierten gegenüber dem einfachen Feedback (z. B. Lee, 1985). Schimmel (1988) betont, dass Leistungsstarken eine Wahlmöglichkeit zwischen unterschiedlich komplexen Feedbackoptionen gegeben werden sollte. Die informelle Kompetenz der Lernenden scheint die Feedbackwirkung von komplexen Aufgaben zu moderieren. Nicht nur die Komplexität des Feedback-Inhalts, sondern auch der Zeitpunkt der Feedbackgabe sowie die Anzahl der Lösungsversuche beeinflussen die Aufgabenanforderungen der Lernaufgabe. Hierbei zeigt sich verzögertes gegenüber sofortigem Feedback bei komplexen Anforderungen überlegen, da verzögertes Feedback eine erneute Lernphase darstellt und somit ein längeres Auseinandersetzen mit der Aufgabe ermöglicht (Clariana u. a., 2000). Hinsichtlich der Anzahl der Lösungsversuche zeigt sich MTF insbesondere bei komplexen Aufgaben am effektivsten (Clariana & Koul, 2006).

In Bezug auf die *Art der Lernsituation* zeigt sich bei einer hohen Aufgabenkomplexität eine Überlegenheit von Untersuchungen, welche in einer natürlichen Umgebung (face-to-face) gegenüber Studien, welche computergestützt durchgeführt wurden (z. B. Baralt,

2013). Dies lässt sich auf eine hohe kognitive Belastung in der computergestützten Umgebung zurückführen.

Des Weiteren hängt die Lernwirksamkeit des Feedbacks von der Kenntnis über *Fehlkonzepte und Fehler* sowie von der *Fehlerfreundlichkeit* ab. Kline u. a. (1991) konnten nachweisen, dass Fehleranalysen sowie eine anschließende Fehlerbesprechung gegenüber herkömmlichem Unterrichtsfeedback überlegen sind. Des Weiteren zeigten sich in Untersuchungen, welche auf der Grundlage von Fehleranalysen elaboriertes Feedback entwickelten, positive Wirkungen von elaboriertem gegenüber einfachem Feedback (z. B. Narciss, 2006).

Insgesamt lässt sich schlussfolgern, dass vielschichtige Faktoren die Lernwirksamkeit des Feedbacks beeinflussen. Bei der Gestaltung des Feedbacks in Abhängigkeit von der Komplexität des Feedback-Inhalts, der Feedback-Funktion sowie der Feedback-Form sollten daher (inter-)personale und situative Faktoren umfassend berücksichtigt werden.

3 Concept Maps in Lehr-Lernprozessen

Concept Maps ermöglichen es, Zusammenhänge in Form eines Netzwerkes strukturiert darzustellen. Es existieren alternative Begriffe, z. B. Cognitive Maps oder Knowledge Maps, wobei sich der Begriff Concept Maps im Forschungskontext weitgehend durchgesetzt hat. In diesem Kapitel wird ein Überblick über theoretische Grundlagen zu Concept Maps in Lehr-Lernprozessen (Kapitel 3.1) sowie zu empirischen Befunden zur Lernwirksamkeit von Concept Maps (Kapitel 3.2) gegeben. Dazu wird im Kapitel 3.1.1 der Begriff Concept Map zunächst bestimmt und klassifiziert sowie im Kapitel 3.1.2 Anwendungsgebiete von Concept Maps vorgestellt. Kapitel 3.1.3 beleuchtet verschiedene Ansätze, um die Lernwirksamkeit von Concept Maps theoretisch zu begründen. Im Kapitel 3.2.1 werden zunächst die Ergebnisse der Studie von Hattie (2015) in Bezug auf den Einfluss des Concept Mapping auf das Lernen sowie die Ergebnisse der Metaanalysen von Nesbit und Adesope (2006) sowie Schroeder u. a. (2017) vorgestellt. Kapitel 3.2.2 betrachtet Studien zu Einflussfaktoren auf die Lernwirksamkeit von Concept Maps. Im Kapitel 3.2.3 werden abschließend Studien vorgestellt, welche Feedback als Ergänzung zum Concept Mapping einsetzen.

3.1 Theoretische Grundlagen zu Concept Maps in Lehr-Lernprozessen

3.1.1 Bestimmung des Begriffs und Klassifikation von Concept Maps

Der Begriff Concept Map wurde in den 1970ern von J. D. Novak geprägt. Concept Maps sind zweidimensionale Strukturdarstellungen, welche Wissen oder Informationen in Form von Netzwerken graphisch darstellen und organisieren können. Sie bestehen grundsätzlich aus *Konzepten*, welche durch gerichtete und bezeichnete *Relationen* miteinander verknüpft werden. Konzept-Relation-Konzept-Verbindungen werden als *Propositionen* bezeichnet. Sie sind Bedeutungseinheiten und lassen sich auf ihren Wahrheitsgehalt hin überprüfen (Novak & Cañas, 2008, S. 1; Ruiz-Primo & Shavelson, 1996, S. 570). Konzepte werden oft als Rechtecke oder Kreise gekennzeichnet und bezeichnen beispielsweise Objekte oder Zeitbegriffe. Relationen verbinden die Konzepte durch gerichtete Pfeile miteinander und können beispielsweise Verben, Adjektive oder Konjunktionen sein (Cañas u. a., 2003, S. 5). Verschiedene Wissensgebiete oder Abschnitte eines Concept Maps können anhand von Querverbindungen, den sogenannten *Cross-Links*, verbunden werden (Novak & Cañas, 2008, S. 2). Insofern Concept Maps lediglich einen Ab-

schnitt bzw. ein Wissensgebiet ansprechen, werden die Konzepte mit *Internal-Links* verbunden.⁴² Concept Maps sollten nach Novak und Cañas (2008, S. 11) auf eine *Kernfrage* hin ausgerichtet sein.

Der Begriff Concept Map bedeutet im Deutschen „*Begriffslandkarte*“. Hierbei sei darauf hingewiesen, dass die Konzepte und Relationen von Concept Maps im Hinblick auf ihre räumlichen Abstände jedoch beliebig angeordnet werden können. Es bedeutet folglich nicht, dass weiter voneinander entfernte Konzepte z. B. weniger zusammenhängen als näher zusammenliegende Konzepte. Infolgedessen wäre der Begriff Netzwerk eher passend (Fürstenau, 2011, S. 56). Obwohl der Begriff nicht entsprechend seiner Semantik verwendet wird, hat er sich in der internationalen Forschung durchgesetzt. Im Folgenden werden die Begriffe Concept Map, Map, Begriffsnetz und Netzwerk synonym verwendet. Nachfolgend wird der Begriff Concept Map von *Mind Maps*, *Knowledge Maps*, *Cognitive Maps* und *semantischen Netzwerken* abgegrenzt.

Eine weitere Möglichkeit, um Wissen in organisierter Form darzustellen, sind *Mind Maps*. Ausgehend von einem zentralen Schlüsselbegriff im Zentrum gehen mehrere Äste ab, die sich nach außen in zunehmend dünnere Äste verzweigen. Die Äste werden direkt mit weiteren Begriffen beschriftet. Beziehungen zwischen den verschiedenen Begriffen bleiben dabei im Gegensatz zum Concept Map unbestimmt. Zusammenhänge zwischen den Begriffen lassen sich beim Mind Map daher nicht verdeutlichen. Es dient eher der spontanen Ideenfindung, z. B. im Rahmen des Brainstormings. Bei Concept Maps kann mit einem zentralen Begriff begonnen werden. Dies hängt jedoch vom Thema ab und ist nicht immer vonnöten (Nückles, Gurlitt, Pabst, & Renkl, 2004, S. 4 f.).

Mit Hilfe von *Knowledge Maps* lässt sich Wissen ebenfalls organisiert darstellen. Insbesondere die Forschergruppe um Dansereau (z. B. Lambiotte, Skaggs, & Dansereau, 1993) verwendet diese Methode, welche auf den ersten Blick den Concept Maps sehr ähnlich erscheint. Knowledge Maps zeichnen sich jedoch durch strengere Konstruktionsregeln aus (McCagg & Dansereau, 1991, S. 319 f.). Sie stellen zweidimensionale Repräsentationen dar, welche sich ebenfalls wie Concept Maps aus Knoten und Verbindungen zusammensetzen. Hierbei werden drei Typen von Links vorgegeben: statische, dynamische und elaborierte Links (O'Donnell, Dansereau, & Hall, 2002, S. 72 ff.). Statische Links beschreiben strukturelle Beziehungen (z. B. „Finger ist Teil von Hand“), dynamische Links

⁴² Der Begriff wurde in Anlehnung an die Elaborationstheorie von Reigeluth (1983, S. 358) entwickelt.

beschreiben eine veränderte Beziehung (z. B. „Lernen führt zu guten Noten“), laborierte Links kennzeichnen vertiefende Informationen anhand von Beispielen, Kommentaren oder Analogien (z. B. „Einstein ist ein Beispiel für ein Genie“). Die verschiedenen Typen von Links werden meist durch Abkürzungen gekennzeichnet. Des Weiteren werden die Knowledge Maps hierarchisch, cluster- oder kettenförmig angeordnet (Lambiotte, Dansereau, Cross, & Reynolds, 1989, S. 338). Da für Concept Maps keine Vorgaben für Verbindungen existieren, können diese als Oberbegriff für Knowledge Maps angesehen werden.

Semantische Netzwerke sind Modelle zur Klassifikation von Wissen. Sie basieren auf der Annahme, dass das Gedächtnis netzwerkartig organisiert ist. Concept Maps entsprechen der angenommenen Repräsentation kognitiver Strukturen und können demnach als Externalisierung semantischer Netzwerke angesehen werden (Fürstenau, 2011, S. 46).

Der Begriff *Cognitive Map* ist auf Tolman (1948) zurückzuführen und bezeichnet die mentale Repräsentation geographisch räumlichen Wissens. Hierbei wird angenommen, dass sich Tiere oder auch Menschen mithilfe intern abgespeicherter Repräsentationen ihrer Umwelt im Raum orientieren. Das Lernen von Merkpunkten, Routen bzw. Wegstrecken (sog. Routenwissen) ermöglicht eine Orientierung im Raum anhand von landkartenähnlichen Bildern, den Cognitive Maps. Sofern Concept Maps entsprechend ihrer Semantik lediglich als „Begriffslandkarten“ zur Anwendung kämen, könnten diese als externalisierte Cognitive Maps angesehen werden (Wender, 1988, S. 101). Teilweise werden Cognitive Maps auch genutzt, um die vorhandenen individuellen kognitiven Strukturen eines Lernenden abzubilden (Novak & Gowin, 1984, S. 138 f.). In diesem Zusammenhang werden Concept Maps trotz vergleichbarer struktureller Elemente von Cognitive Maps abgegrenzt. Das in den Concept Maps abgebildete Wissen sollte dieser Ansicht nach so dargestellt werden, dass es Experten eines bestimmten Fachgebietes als gesichert erachten. Diese Definition widerspricht jedoch der mittlerweile üblichen Verwendung von Concept Maps. Für die Begriffsbestimmung ist es demnach nicht entscheidend, wer die Maps erstellt.

Für den weiteren Verlauf der Arbeit wird Literatur zu Knowledge Maps einbezogen, da der Begriff Concept Map hierbei als Oberbegriff verwendet werden kann. Den Mind Maps wird aufgrund ihrer strukturellen Unterschiede zu Concept Maps keine weitere Beachtung geschenkt. Cognitive Maps sowie semantische Netzwerke beziehen sich auf interne Repräsentationen kognitiver Strukturen. Da es im Rahmen der Arbeit nicht um die

graphische Darstellung geographischen Wissens geht, wird den Cognitive Maps keine weitere Beachtung geschenkt. Die semantischen Netzwerke werden noch einmal detaillierter im Kapitel 3.1.3 betrachtet, da sich dieser Ansatz für die Begründung der Lernwirksamkeit von Concept Maps eignet.

Als wesentliches Merkmal von Concept Maps postulieren Novak und Gowin (1984, S. 15) eine *hierarchische* Struktur von Concept Maps, von allgemeinen Konzepten an der Spitze der Maps hin zu spezifischeren Konzepten im unteren Bereich der Maps. Andere Forscher betonen hierbei, dass eine hierarchische Darstellung nicht in jedem Fall möglich und sinnvoll ist. Safayeni, Derbentseva und Cañas (2005) gehen in diesem Zusammenhang darauf ein, dass kausale bzw. dynamische Beziehungen als *zyklische* Concept Maps dargestellt werden können. Hierarchische Maps sind dagegen eher für statische Beziehungen sinnvoll. Die Konzepte in zyklischen Concept Maps sind voneinander abhängig, eine Veränderung eines Konzepts beeinflusst alle weiteren Konzepte dieses Concept Maps. Neben hierarchischen und zyklischen Maps existieren auch *systemische* Concept Maps (Fürstenau, 2011, S. 47). Systemische Maps sind multirelationale Netzwerke, d. h. jedes Konzept kann durch beliebig viele Relationen mit anderen Konzepten verbunden werden, wobei nicht notwendigerweise hierarchische Beziehungen dargestellt werden müssen (Fürstenau & Trojahnner, 2005, S. 196). Wiegmann, Dansereau, McCagg, Rewey und Pitre (1992) gehen des Weiteren auf die Möglichkeit ein, Maps als sogenannte „*Web Maps*“ spinnennetzartig darzustellen, mit einem zentralen Konzept in der Mitte umgeben von unterstützenden Konzepten.

Concept Maps lassen sich hinsichtlich des *Grades der Vorstrukturierung* in *selbsterstellte Maps*, *Lückenmaps* und *Expertenmaps* unterteilen (z. B. Hardy & Stadelhofer, 2006). In Anhängigkeit von den Anforderungen der Aufgabenstellung konstruieren die Lernenden ein Map selbst, vervollständigen ein Lückenmap oder arbeiten ein Expertenmap durch. Das Lückenmap ist dabei teilweise und das Expertenmap vollständig vorstrukturiert. Bei der Selbstkonstruktion ist es Ziel, individuelle Strukturen zu einem Themengebiet darzustellen. Daher sind die Lernenden weitgehend frei von einschränkenden Vorgaben (McCagg & Dansereau, 1991, S. 318). Die Concept Maps können entweder per Hand im *Paper & Pencil-Verfahren* oder *computergestützt* mit einer speziellen Software erstellt werden (Behrendt & Reiska, 2001, S. 11 f.; Tergan, Gräber, & Neumann, 2006, S. 329

ff.). Eine international häufig verwendete Software ist Cmap Tools.⁴³ Die Lernenden übernehmen während des Erstellens der Concept Maps eine aktive Rolle, was ein tiefergehendes Lernen fördern kann. Hierbei ist jedoch eine gewisse Vertrautheit mit der Methodik eine Grundvoraussetzung. Dem Erlernen der Technik sollte daher vorab eine entsprechende Bedeutung beigemessen werden (Fürstenau, 2018). Um die Freiheit der Lernenden nicht einzuschränken, sollte es ermöglicht werden, Konzepte und Relationen frei zu wählen. Die somit erhöhte Komplexität bei der Selbstkonstruktion kann jedoch zu einer kognitiven Überlastung der Lernenden führen. Aus diesem Grund kann es ggf. von Vorteil sein, Konzepte und/oder Relationen vorzugeben (Chang u. a., 2002, S. 6 f., 19 f.; Jüngst & Strittmatter, 1995, S. 202). Vorgaben lassen sich hierbei entweder an den Sprachgebrauch von Experten oder an den Alltagssprachgebrauch der Probanden anlehnen bzw. können Mischformen sein (Fürstenau, 2001, S. 254). Eine weitere Möglichkeit stellen „offene Pools“ da, d. h. neben vorgegebenen Konzepten und/oder Relationen, welche auch nicht notwendigerweise alle verwendet werden müssen, haben die Lernenden die Möglichkeit, weitere selbst gewählte Begriffe oder Verbindungen einzufügen (Weber & Schumann, 2000, S. 162).

Expertenmaps werden von Experten eines bestimmten Gebietes konstruiert, der Inhalt ist somit bereits vorselektiert. Die Aufmerksamkeit kann folglich auf die von den Experten als relevant eingeschätzten Zusammenhänge gelenkt werden. Beim Durcharbeiten der Expertenmaps erfolgt eher eine passive Informationsaufnahme. In diesem Zusammenhang besteht die Gefahr, dass die Lernenden die Inhalte auswendig lernen und folglich nur oberflächlich verarbeiten (Hardy & Stadelhofer, 2006, S. 177; Bernd & Jüngst, 1999; McCagg & Dansereau, 1991, S. 317 f.).

Lückenmaps (auch „Fill-in-the-Map“) nehmen eine Position zwischen selbsterstellten und Expertenmaps ein und bieten somit ein gewisses Maß an Vorstrukturierung. Hierbei kann zwischen „Fill-in-the-Nodes“ und „Fill-in-the-Lines“ unterschieden werden. Bei der Fill-in-the-Nodes-Technik fehlen die Bezeichnungen einiger Konzepte, wohingegen bei der Fill-in-the-Lines-Technik einige Relationen nicht beschriftet sind. Des Weiteren kann auch die Richtung einiger Relationen zu ergänzen sein. Die hinzuzufügenden Konzepte bzw. Relationen können entweder vom Lernenden selbst gefunden (Generate-and-Fill-in)

⁴³ Cmap Tools ist im Internet unter <http://cmap.ihmc.us> zum Download für Unterrichts- und nicht kommerzielle Zwecke frei verfügbar. Die Software verfügt über eine einfach bedienbare Benutzeroberfläche mit einer Vielzahl von Organisations- und Repräsentationsmöglichkeiten (Tergan, 2006, S. 315).

oder aus einer Liste ausgewählt werden (Select-and-Fill-in). Lückenmaps können als geschlossenes oder offenes Map gestaltet sein. Bei einer offenen Gestaltung ist es dem Lernenden freigestellt, weitere Konzepte und/oder Relationen zu ergänzen, wohingegen dies bei einem geschlossenen Lückenmap nicht möglich ist. Die Entscheidung, inwieweit Experten-, Lücken- oder selbsterstelltes Map von Vorteil sind, hängt von der Zielstellung sowie von weiteren Einflussfaktoren, z. B. den individuellen Voraussetzungen der Lernenden, ab (Hardy & Stadelhofer, 2006, S. 177 f.; Ruiz-Primo, Schultz, Li, & Shavelson, 2001, S. 262 ff.; Schau, Mattern, Zeilik, Teague, & Weber, 2001, S. 140).

Hinsichtlich der *Ausgestaltung* kann zunächst zwischen „*Whole Maps*“ und „*Stacked Maps*“ unterschieden werden. „*Whole Maps*“ stellen alle Inhalte in einem Map dar, wohingegen „*Stacked Maps*“ aus mehreren ineinander geschachtelten Maps bestehen (Wiegmann u. a., 1992, S. 138). Des Weiteren kann zwischen *statischen* bzw. *animierten Maps* sowie *Concept Maps mit Hyperlinks* unterschieden werden (Nesbit & Adesope, 2006). Bei statischen Maps werden alle Informationen in einem Map auf einmal dargestellt, bei animierten Maps werden Knoten und/oder Relationen nacheinander hinzugefügt bzw. modifiziert (z. B. farblich hervorgehoben). Dies erfolgt in der Regel computergestützt. Ein animiertes Concept Map kann mit einer geführten Tour durch das Map verglichen werden, was insbesondere einer kognitiven Überlastung entgegenwirken kann (Nesbit & Adesope, 2013). Concept Maps mit Hyperlinks ermöglichen einen interaktiven Zugriff auf Informationen in Datenspeichern, Datenbanken oder im Internet (Tergan, 2006, S. 311 f.).

Concept Mapping kann in Bezug auf die *Häufigkeit der Anwendung einmalig* bzw. *kontinuierlich* eingesetzt werden. Cañas (2003, S. 8) betont in diesem Zusammenhang, dass Concept Maps in den Lernprozess eingebettet werden und nicht nur als „Add-On“ am Beginn oder Ende des Prozesses eingesetzt werden sollten. Eine Studie von Pankratius (1990) bestätigt den Vorteil einer integrierten kontinuierlichen Anwendung gegenüber dem einmaligen Anwenden der Concept Mapping-Technik am Ende des Lernprozesses.

Concept Maps können *horizontal* oder *vertikal* ausgerichtet sein (O'Donnell, 1994).

Concept Maps können von *Einzelpersonen*, *Dyaden* oder *Gruppen* erstellt werden (Fürstenau, 2001, S. 254). Hierbei ist ggf. auch eine Kombination denkbar, bei der die Lernenden zunächst individuell Concept Maps erstellen und im Anschluss in der Gruppe über diese diskutieren (Nesbit & Adesope, 2006, S. 431).

Nachfolgende Abbildung 7 stellt die Grundlagen zur Begriffsbestimmung und Klassifikation von Concept Maps in einem Concept Map dar.

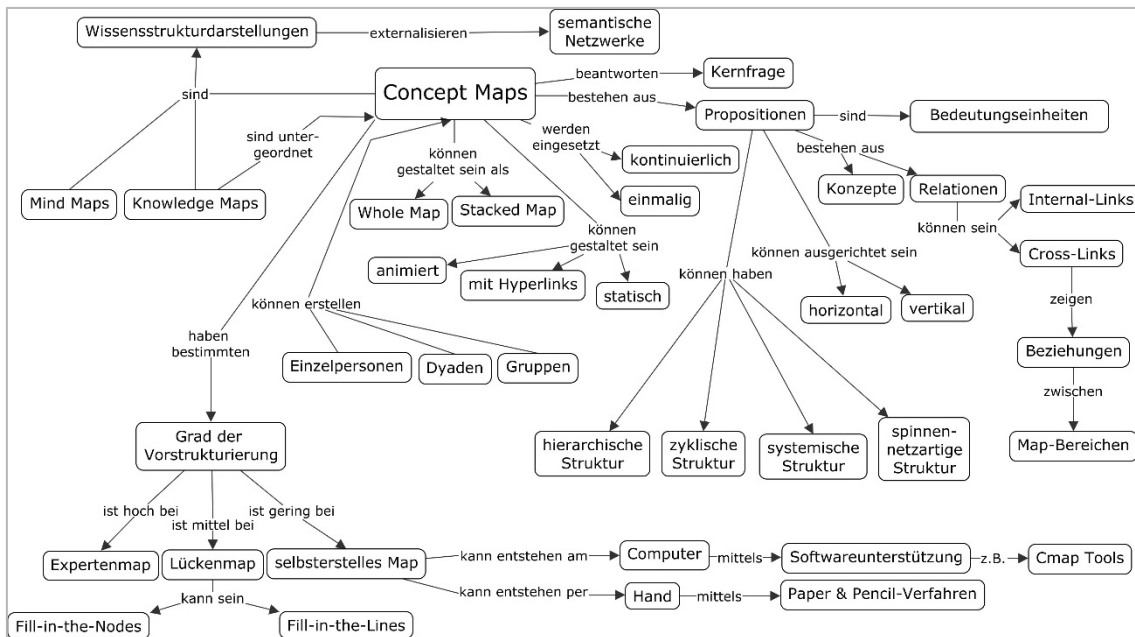


Abbildung 7: Concept Map über Concept Maps

3.1.2 Anwendungsgebiete von Concept Maps in Lehr-Lernprozessen

Concept Maps werden im Lehr-Lernkontext hauptsächlich zur Wissensdiagnose sowie zur instruktionalen Unterstützung verwendet (Fürstenau, 2011, S. 56; Mandl & Fischer, 2000, S. 6 ff.).

Anwendung zur Wissensdiagnostik

Bei der *Wissensdiagnostik* werden Concept Maps erhoben und bewertet, um Rückschlüsse auf das zu einem Zeitpunkt vorhandene Wissen schließen zu können bzw. um die Veränderung des Wissens zu messen. Die Ergebnisse können Hinweise geben, wie weitere Lehr-Lernprozesse zu gestalten sind (Fürstenau, 2011, S. 47).

Im Rahmen der Wissensdiagnostik werden *quantitative* und *qualitative Indikatoren* herangezogen. Quantitativ lassen sich Concept Maps zunächst anhand des Umfangs, d. h. der Anzahl der Konzepte, Relationen bzw. Propositionen, analysieren. Des Weiteren kön-

nen Strukturmaße, wie die Vernetztheit (z. B. Quotient aus Propositionen und Konzepten⁴⁴; durchschnittlicher Grad des Knotens⁴⁵; Verknüpfungsdichte⁴⁶) oder die Kohärenz⁴⁷ (auch innere Verbundenheit oder Zerklüftetheit) berücksichtigt werden (Fürstenau, 2011, S. 47; Weber, 1994, S. 67 ff.; Bonato, 1990). Die Aussagekraft quantitativer Indikatoren ist jedoch zu hinterfragen.

Für eine umfassende Bewertung von Concept Maps im Lehr-Lern-Kontext sollten daher quantitative und qualitative Indikatoren aufeinander bezogen werden. Als qualitative Indikatoren können die Anzahl richtiger Propositionen, die Anzahl schwacher Propositionen, Fehlvorstellungen, d. h. die Anzahl falscher Propositionen oder der Übereinstimmungsgrad mit dem Referenznetz analysiert werden. Schwache Propositionen spiegeln ein unvollständiges Verständnis für die Beziehung zwischen zwei Begriffen wider (Rebich & Gautier, 2005, S. 360; Novak & Gowin, 1984, S. 36 f.). Wadouh (2008) beurteilte die Richtigkeit der Concept Maps anhand zweier Indikatoren: fachliche Korrektheit (falsch, ungenau, richtig) und Fachgehalt (niedrig, mittel, hoch), um nicht nur zu bewerten, inwieweit die Zusammenhänge richtig dargestellt wurden, sondern auch den Fachgehalt zu analysieren. Auf diese Weise können „schwache“ Propositionen näher spezifiziert werden.

Ein weiterer qualitativer Indikator ist der Übereinstimmungsgrad mit dem Referenzmap, welchen bereits Novak und Gowin (1984, S. 36) berechneten (Criterion Map). Sie bezogen sich dabei jedoch auf die Struktur der Netzwerke und nicht auf die inhaltliche Richtigkeit. Ruiz-Primo (2000) berechnete das Verhältnis der richtigen Verknüpfungen im Lernermap zu allen Propositionen im Referenzmap, um den Grad der Übereinstimmung (Convergence Score) zu erhalten. Es stellte sich die Frage, wie ein Referenzmap entsteht

⁴⁴ Je höher der Quotient aus Propositionen und Konzepten ausfällt, desto höher ist auch die Vernetztheit. Sobald ein Lernender neue Zusammenhänge zwischen bestehenden Konzepten entdeckt, steigt die Anzahl der Propositionen, wohingegen die Anzahl der Konzepte dabei unverändert bleibt (Rebich & Gautier, 2005).

⁴⁵ Der Grad des Knotens spiegelt die Anzahl der benachbarten Knoten wider. Ein isolierter Knoten hat den Grad von 0. Ein Knoten kann maximal $(k-1)$ Nachbarknoten haben, wobei k für die Anzahl der Knoten steht (Bonato, 1990, S. 65).

⁴⁶ Allgemein wird die Verknüpfungsdichte als Verhältnis der Anzahl der verwendeten Propositionen zur Anzahl der maximal möglichen Propositionen definiert. Die Anzahl der maximal möglichen Propositionen errechnet sich anhand der Formel $n \times (n - 1)$, wobei n für die Anzahl der Propositionen steht (Weber, 1994, S. 74).

⁴⁷ Die Kohärenz untersucht, inwieweit ein Gesamtnetzwerk in Teilnetzwerke zerfällt. Operationalisiert wird die Kohärenz daher anhand der Anzahl der Zusammenhangskomponente bzw. Teilnetzwerke. Je kleiner dieses Maß ist, desto besser ist der Lernende in der Lage, differenzierte Themenbereiche eines übergeordneten Themas zu integrieren. Idealerweise sollte die Kohärenz daher den Wert 1 annehmen, d. h. es existieren keine isolierten Konzepte im Netzwerk (Bonato, 1990, S. 93 f.).

und auf welche Weise somit die Idealstruktur festgelegt wird. Referenzstrukturen sind auch unter dem Begriff Expertennetz bekannt. Teilweise werden beide Begriffe synonym verwendet (Stracke, 2004, S. 40). Beide Begrifflichkeiten können jedoch auch in Bezug auf den zu dienenden Zweck unterschieden werden. Expertenmaps werden z. B. im Rahmen der Expertiseforschung angewendet, um die Wissensstrukturen von Experten eines bestimmten Fachgebietes mit denen von Novizen zu vergleichen (Reinhold, Lind, & Friege, 1999). Referenznetze spiegeln das zu erreichende Lernziel und somit eine Idealstruktur wider. Zu diesem Zweck können sie anhand von Lehrtexten durch die Forscher (z. B. bei Bonato, 1990, S. 97) oder von unterrichtenden Lehrpersonen bzw. unabhängigen Fachleuten (Lehrpersonen oder Fachdidaktikern) erstellt werden. Ruiz Primo (2000) postuliert, dass unterrichtende Lehrpersonen, unabhängige Experten und Forscher an der Entwicklung des Referenznetzes beteiligt werden sollten. Hierbei sollten dann die gemeinsamen Propositionen in das Referenznetz aufgenommen werden.

Sofern das Wissen einer größeren Gruppe allgemein analysiert werden soll, ohne dass es um die Bewertung der inhaltlichen Richtigkeit geht, können Modalnetze erstellt bzw. prototypische Netze identifiziert werden. Modalnetze enthalten die am häufigsten gemeinsam genannten Propositionen. Die Anzahl der zu berücksichtigenden Propositionen bestimmt sich dabei als mittlere Häufigkeit aller genannten Propositionen in den Concept Maps der Versuchspersonen (Fürstenau & Trojahnner, 2005, S. 197; Becker, Oldenbürger, & Piehl, 1987, S. 464). Prototypische Netze sind allen anderen Probandennetzen inhaltlich und strukturell am ähnlichsten. Sie liegen jedoch im Gegensatz zu den Modalnetzen bereits genau in dieser Form vor (Fürstenau & Trojahnner, 2005, S. 198 ff.).

Anwendung im instruktionalen Kontext

Im *instruktionalen Kontext* sind Concept Maps zur Unterstützung der Wissensvermittlung bzw. -aneignung geeignet (Fürstenau, 2011, S. 47). Sie können hierbei sowohl beim Internalisieren neuen Wissens als auch beim Externalisieren erworbenen Wissens hilfreich sein. Die zu erwerbenden Lerninhalte können dabei durch den Lehrenden erläutert oder vom Lernenden selbst angeeignet werden. In diesem Zusammenhang können verschiedene Sozialformen zum Einsatz kommen. So ist nicht nur eine individuelle, sondern auch eine paar- oder gruppenweise Anwendung möglich (Mandl & Fischer, 2000, S. 7 f.). Hinsichtlich des Zeitpunkts im Lehr-Lernprozess können Concept Maps in allen Phasen des Lernprozesses angewendet werden. Zunächst können sie als Überblick, d. h. als vorstrukturierende Lernhilfe zum Einstieg in eine thematische Einheit (sog. Advance Organizer,

Kapitel 3.1.3) dienen. Des Weiteren ist ein wiederholtes Aufgreifen in Phasen der Wissenserarbeitung möglich (Jüngst & Strittmatter, 1995, S. 202; Jüngst, 1992, S. 61). Hierbei bietet sich z. B. das Durcharbeiten von Concept Maps an, d. h. sich auf verschiedenen Wegen durch das Netz zu bewegen (Aebli, 1983, S. 269), um das Wissen zu vertiefen. Abschließend können Concept Maps als Zusammenfassung, zum Üben und Wiederholen eingesetzt werden (Jüngst & Strittmatter, 1995, S. 202; Jüngst, 1992, S. 61).

Das Concept Mapping lässt sich als *Lernstrategie* interpretieren. Lernstrategien lassen sich nach Friedrich und Mandl (1992, S. 6) als Sequenzen von Handlungen, die darauf abzielen ein Lernziel zu erreichen, definieren. Es existieren verschiedene Klassifikationsvorschläge für Lernstrategien⁴⁸. Eine häufig verwendete Einteilung ist die von Friedrich und Mandl (2006, S. 1, 1992, S. 11 ff.) in kognitive, metakognitive (auch Kontrollstrategien) und affektiv-motivationale Strategien, welche auf Weinstein und Mayer (1986) zurückgeht. Die Klassifikation erfolgt nach ihrer Funktion für den Prozess der Informationsverarbeitung. Kognitive Lernstrategien lassen sich weiter unterteilen in *Wiederholungsstrategien*,⁴⁹ *Organisationsstrategien* und *Elaborationsstrategien*. *Wiederholungsstrategien* werden auch als Oberflächenstrategien bezeichnet, d. h. sie dienen dem Einprägen oberflächlicher Merkmale (Steiner, 2006, S. 101). Diese Strategien unterstützen ein aktives Memorieren, verhindern somit eine Verdrängung der neuen Informationen aus dem Arbeitsspeicher und erhöhen folglich die Chance einer Übernahme ins Langzeitgedächtnis. Beispiele hierfür sind das laute bzw. stille Wiederholen des Lernmaterials oder

⁴⁸ Lernstrategien lassen sich des Weiteren hinsichtlich der Analyseebene der Mikro-, Meso- oder Makroebene zuordnen (Friedrich & Mandl, 1992, S. 15), je nachdem ob die Lernstrategien kurz-, mittel- oder langfristige Prozesse der Informationsverarbeitung betreffen. In Bezug auf die Bandbreite der Einsatzmöglichkeiten können allgemeine und spezifische Lernstrategien unterschieden werden (Friedrich & Mandl, 1992, S. 10 f.). Allgemeine Strategien sind situations- und inhaltsübergreifend und somit nahezu immer einsetzbar (z. B. Zeitmanagement). Spezifische Strategien sind lediglich in abgegrenzten spezifischen Inhaltsgebieten anwendbar (z. B. die Addition zweier Zahlen, wobei die kleinere auf die größere Zahl aufzusummieren ist). Es existieren jedoch auch Strategien mit mittlerem Abstraktionsgrad, z. B. Verstehensstrategien beim Lesen von Texten (Friedrich & Mandl, 1992, S. 10 f.). Die Einteilung in Primär- und Stützstrategien ist auf Dansereau (1978) zurückzuführen und folgt der Funktion der Lernsteuerung. Primärstrategien beeinflussen die kognitiven Prozesse der Informationsverarbeitung direkt, wodurch kognitive Strukturen verändert und infolgedessen ein besseres Verständnis, Behalten und Abrufen der Informationen ermöglicht werden. Sie helfen unmittelbar dabei, eine bestimmte Aufgabe zu lösen (z. B. Zusammenfassungen von Texten in Worten oder anhand von Mapping-Techniken). Stützstrategien oder auch Ressourcenstrategien beeinflussen die Prozesse der Informationsverarbeitung indirekt, indem sie die motivationalen, kontrollierenden und organisierenden Prozesse für den Lernvorgang und die Lernumgebung steuern.

⁴⁹ In neueren Werken ordnen Mandl und Friedrich (2006) die Wiederholungsstrategien den Elaborationsstrategien unter. Beim Wiederholen findet eher eine oberflächliche Verarbeitung statt, wohingegen beim Elaborieren eine Tiefenverarbeitung gefördert wird. Aus diesem Grund ist eine Unterordnung eher ungünstig.

das Herausschreiben bzw. Unterstreichen wichtiger Textpassagen (Friedrich, 1995, S. 6 ff.; Friedrich & Mandl, 1992, S. 11). *Organisationsstrategien* werden verwendet, um neuen Lernstoff zu organisieren und zu strukturieren. Hierbei soll die Fülle des Stoffes auf ein Minimum reduziert werden (Friedrich & Mandl, 2006, S. 4). Mögliche Organisationsstrategien sind das Zusammenfassen von Textinformationen, Wissensschemata⁵⁰ sowie externe Visualisierungen (z. B. Mind Mapping oder Concept Mapping). *Elaborationsstrategien* fördern das Verstehen und Behalten neuen Wissens, indem neue Informationen mit bestehendem Wissen vernetzt werden. Ziel von Elaborationsstrategien ist somit das Integrieren neuer Informationen in bestehende kognitive Strukturen. Infolgedessen kann der spätere Abruf der neu gewonnenen Informationen erleichtert werden. Elaborationsstrategien sind Tiefenverarbeitungsstrategien, da sie im Gegensatz zu den Oberflächenstrategien den Fokus auf das Verstehen, d. h. den Sinn des zu lernenden Stoffes, legen. Mögliche Elaborationsstrategien sind Vorwissen aktivieren, Notizen machen oder Fragen stellen (Friedrich & Mandl, 2006, S. 2 ff.; Friedrich & Mandl, 1992, S. 12).

Das Concept Mapping wurde bereits als Beispiel für eine Organisationsstrategie aufgeführt. Diese Einordnung ist aufgrund der strukturellen Eigenschaften der Methode nahelegend. Des Weiteren kann das Concept Mapping jedoch auch eine Elaborationsstrategie sein. Hierbei ist eine Differenzierung zwischen selbsterstellten, Lücken- und Expertenmaps vonnöten. Bei der Selbstkonstruktion ist es unabdingbar, neue Informationen mit dem Vorwissen zu verknüpfen, um zu bestimmen, welche Konzepte relevant sind und wie diese verbunden werden. Dafür sollte der zu lernende Stoff auch verstanden worden sein. Lückenmaps sinnvoll zu vervollständigen, setzt ebenfalls ein tiefergehendes Auseinandersetzen mit dem zu lernenden Stoff sowie eine Verknüpfung mit dem Vorwissen voraus. Beim Durcharbeiten von Expertenmaps besteht die Gefahr einer oberflächlichen Verarbeitung, weshalb dies lediglich den Wiederholungsstrategien zuzuordnen wäre. Eine Elaboration kann dann entstehen, wenn die Lehrperson den Prozess des Durcharbeitens angemessen mit konkreten Anweisungen vor und Hinweisen während des Durcharbeitens für ein tiefgründiges Durchdenken unterstützt.

⁵⁰ Schemata sind Wissensstrukturen, in denen allgemeines Wissen, d. h. Wissen über typische Zusammenhänge eines Realitätsbereichs (z. B. wie normalerweise ein Auto auszusehen hat) organisiert ist. Diese kognitiven Strukturen beziehen sich auf unterschiedlichste Inhaltsbereiche sowie Abstraktionsebenen. Elementare Schemata, z. B. das Auto-Schema, können hierbei in Schemata höherer Hierarchieebenen, z. B. Verkehrs-Schema, eingebettet sein (Arbinger, 1997; Mandl, Friedrich, & Hron, 1988).

3.1.3 Theoretische Ansätze zur Lernwirksamkeit von Concept Maps

Die Lernwirksamkeit von Concept Maps wird sehr häufig über die Theorie der semantischen Netzwerke (Collins & Quillian, 1969), die Theorie des bedeutungsvollen rezeptiven Lernens (Ausubel, 1974) sowie die Theorie des Dual Coding (Paivio, 1986) begründet. Die Theorie der semantischen Netzwerke sowie die Theorie des bedeutungsvollen rezeptiven Lernens beschreiben die Art und Weise, wie Informationen im Gedächtnis abgespeichert werden. Die Theorie der semantischen Netzwerke geht von einem netzwerkartigen Aufbau des Gedächtnisses aus. Die Theorie des bedeutungsvollen rezeptiven Lernens erläutert, wie ausgehend von der Annahme einer hierarchisch organisierten kognitiven Struktur ein bedeutungsvolles rezeptives Lernen möglich ist. Diese Theorie bildet somit eine Brücke von der Strukturierung der Gedächtnisinhalte hin zum Lernen. Die Theorie des Dual Coding postuliert, dass unterschiedliche Informationen im Gedächtnis auch unterschiedlich kodiert abgespeichert werden. Des Weiteren kann die Theorie des multimedialen Lernens als Begründung für die Lernwirksamkeit von Concept Maps herangezogen werden, welche davon ausgeht, dass verschiedene Präsentationsformen (z. B. Text und Bild) das Lernen begünstigen (Fürstenau, 2011, S. 46). Die vier Theorien werden nachfolgend genauer vorgestellt.

Theorie der semantischen Netzwerke

Im Kapitel 3.1.1 wurde bereits darauf hingewiesen, dass semantische Netzwerke Modelle zur Klassifikation von Wissen sind und Concept Maps als Externalisierung semantischer Netzwerke angesehen werden können. Collins und Quillian (1969) haben das erste und zugleich einflussreichste Modell zum Aufbau semantischer Netzwerke entwickelt. Hierbei handelt es sich um gerichtete, bezeichnete Graphen, welche aus Knoten und Kanten bestehen. Knoten können als mentale Repräsentationen von Kategorien definiert werden. Kanten strukturieren die Gedächtnisinhalte anhand von Relationen. Knoten und Kanten werden als selbständige Behauptungen miteinander verbunden und bilden Propositionen, welche auf ihren Wahrheitsgehalt hin geprüft werden können (Wender, 1988, S. 60 ff.).

Die Wahrscheinlichkeit, dass ein Knoten aus dem Gedächtnis abgerufen wird, steigt mit dem Grad seiner Aktivierung. Die Aktivierung breitet sich entlang der Kanten aus. Je häufiger Verbindungen genutzt werden, desto stärker ist die Kante. Infolgedessen steigt auch die Aktivierung des assoziierten Knotens (Oberauer, Mayr, & Kluwe, 2006).

Bei der Theorie der semantischen Netzwerke wird also davon ausgegangen, dass das Wissen netzwerkartig im Gedächtnis strukturiert ist. Hierbei wird gemäß Dörner (1987) eine zweifach hierarchische Struktur, bestehend aus Komplexions- und Abstraktionshierarchie, angenommen. Mit Hilfe der Abstraktionshierarchie werden einzelne Begriffe verschiedener Hierarchieebenen anhand von „Ist-Ein-Relationen“ in Beziehung gesetzt. Der abstraktere Begriff „Gebäude“ könnte beispielsweise mit dem konkreteren Begriff „Haus“ durch die Relation „ist“ zu der Proposition „Haus ist Gebäude“ verbunden werden, welche durch das Notationssystem „sein (Haus, Gebäude)“ angegeben wird. Innerhalb der Komplexionshierarchie werden Elemente auf derselben Hierarchieebene mittels räumlicher, zeitlicher oder raumzeitlicher Relationen verbunden. Der Begriff „Dach“ lässt sich dabei z. B. mit dem Begriff „Haus“ durch die Relation „ist Teil von“ zu der Proposition „Dach ist Teil von Gebäude“ in Beziehung setzen, welche durch das Notationssystem „sein Teil von (Dach, Haus)“ angegeben wird. Somit wird ausgedrückt, dass das Dach ein Bestandteil des Hauses ist.

Informationen in semantischen Netzwerken können mithilfe der sog. Vererbungseigenschaft sehr ökonomisch gespeichert werden. Neues Wissen wird hierbei lediglich dem obersten zutreffenden Knoten zugeordnet und gleichzeitig auf alle Unterklassen übertragen (Wender, 1988, S. 61 f.). Beispielsweise lässt sich Frage, ob ein Bungalow ein Dach hat, bejahen. Die Eigenschaft „hat ein Dach“ ist beim Konzept „Gebäude“ gespeichert, „Haus“ ist eine Unterkategorie von „Gebäude“ und „Bungalow“ eine Unterklasse von „Haus“.

Concept Maps stehen im Einklang mit der inneren Repräsentation in Form von semantischen Netzwerken. Daher sind aufwändige Umkodierungsprozesse, z. B. bei der Überführung von linearen Texten in semantische Netzwerke, nicht notwendig (Fürstenau, 2011, S. 46). Infolgedessen scheint die Theorie der semantischen Netzwerke ein Erklärungsansatz für die Lernwirksamkeit von Concept Maps zu sein. Novak nutzte die Erkenntnisse zur Theorie der semantischen Netzwerke bei der Entwicklung der Concept Maps. Im Rahmen der Erforschung des Verständnisses von wissenschaftlichen Begriffen bei Kindern stellte er fest, dass es anhand von Interview-Transkripten schwierig ist, Veränderungen zu identifizieren. Die Annahme einer netzwerkartigen kognitiven Struktur wurde daher genutzt, um das Wissen der Kinder in Form von Concept Maps darzustellen (Novak & Cañas, 2008, S. 3).

Theorie des bedeutungsvollen rezeptiven Lernens

Novak und Gowin (1984, S. 15 ff.) aber auch weitere Autoren (Nesbit & Adesope, 2006, Jüngst, 1992) berufen sich bei der Begründung der Lernwirksamkeit von Concept Maps auf die Theorie des bedeutungsvollen rezeptiven Lernens von Ausubel (1974). Hierbei handelt es sich um eine kognitive Lerntheorie, bei der der Prozess der Aufnahme von Bedeutungen, die über sprachliches Material dargeboten werden, erklärt wird (Fürstenau, 2016, S. 27). Lernen kann rezeptiv oder entdeckend bzw. mechanisch oder bedeutungsvoll sein. Beim rezeptiven Lernen wird das zu lernende Material in seiner endgültigen Form präsentiert und von den Lernenden internalisiert. Bevor der Lernende beim Entdeckungslernen Informationen in seine kognitive Struktur integriert, muss er sie zunächst selbständig neu ordnen (Ausubel, 1974, S. 24 f.). Das rezeptive Lernen ist dabei typisch für schulische Lernprozesse, da eine rezeptive Übernahme des bereitgestellten Wissens im Vordergrund steht und nicht jeder Lernende in der Lage ist, entdeckend zu lernen (Ausubel, 1978, S. 61; Straka & Macke, 1979, S. 113). Mechanisches Lernen lässt sich mit dem „sturen“ Auswendiglernen von Sachverhalten vergleichen, welche anschließend nicht mit eigenen Worten wiedergegeben werden können. Der im Rahmen des bedeutungsvollen Lernens erworbene Inhalt ist dagegen reproduzierbar und auf neue Gegebenheiten übertragbar, da die Lernaufgabe zum einen zielgerichtet und zum anderen sinngemäß auf das Vorwissen bezogen werden kann (Fürstenau, 2016, S. 28). Somit ergeben sich vier Grundformen des Lernens nach Ausubel (1978).

Besonderen Fokus legt Ausubel (1978) dabei auf das bedeutungsvolle rezeptive Lernen, dessen Ziel es ist, die aktuelle Bedeutung einer Lernaufgabe für den Lernenden herauszuarbeiten. Somit wird der Lerninhalt flexibel und lässt sich auf neue Situationen übertragen. Hierzu muss ein Bezug zwischen Lernaufgabe und Vorwissen hergestellt werden. Eine aktuelle Bedeutung kann dabei nur individuell entstehen, weshalb es sich hierbei um einen idiosynkratischen Akt handelt. Im Fokus stehen also neue Bedeutungen, welche mit bestehenden in Beziehung gesetzt werden. Diesen Prozess beschreibt Ausubel als Subsumtionsprozess (derivate Subsumtion, korrelative Subsumtion, Überordnung, Kombination)⁵¹ im Rahmen seiner Assimilationstheorie (Fürstenau, 2016, S. 31 ff.). Ausubel

⁵¹ Neue Bedeutungen werden bei der derivaten Subsumtion als Spezialfälle unter bestehende Bedeutungen untergeordnet, d. h. neue Bedeutungen sind aus den vorhandenen direkt ableitbar. Bei der korrelativen Subsumtion werden bestehende Bedeutungen ausgeweitet, modifiziert, spezifiziert oder gar eingeschränkt. Insofern die neuen Bedeutungen allgemeiner als die bestehenden sind, findet Überordnungslernen statt. Beim

(1974, S. 99) geht dabei ebenfalls von einer hierarchisch organisierten kognitiven Struktur aus, von den allgemeinen Ideen an deren Spitze bis hin zu den spezifischen Ideen an der Basis. Insgesamt findet im Laufe der Zeit eine Anpassung an die stabilere Idee statt. Ausubel (1974, S. 99) nimmt an, dass die allgemeinere, inklusivere Idee auch die stabilere Idee ist. Es erfolgt somit eine Reduzierung (Assimilation) der speziellen auf die allgemeine Bedeutung. Für ein bedeutungsvolles rezeptives Lernen sollten drei Voraussetzungen erfüllt sein. Das Lernmaterial sollte erstens eine potentielle Bedeutung haben, zweitens sollte der Lernende eine sinnvolle Lerneinstellung mitbringen und drittens muss die vorhandene kognitive Struktur des Lernenden Anknüpfungspunkte, sog. Ankerideen, für den neuen Lerninhalt bieten (Ausubel, 1978, S. 61 ff.).

Um das bedeutungsvolle rezeptive Lernen zu unterstützen, entwickelte Ausubel (1974, S. 160 ff.) das expositorische Lehren, welches mehrere Prinzipien umfasst. Dabei sind die *Advance Organizer* (Organisationshilfen) das wichtigste und bekannteste Prinzip. Advance Organizer werden vor dem eigentlichen Lernmaterial präsentiert, um eine Verbindung zwischen neuen Informationen und vorhandenem Wissen herzustellen, insbesondere wenn in der kognitiven Struktur keine Ankerideen vorhanden sind bzw. nicht erkannt werden. Advance Organizer sind daher allgemeiner und inklusiver als das Lernmaterial. Sie geben einen Überblick über das zu Lernende, verdeutlichen die Relevanz des Materials sowie die Relevanz für die Lernenden und zeigen Unterschiede zwischen dem zu lernenden und dem vorhandenen Wissen auf. Ein weiteres Prinzip ist die *progressive Differenzierung*, bei der die allgemeinsten und umfassendsten Ideen zuerst gelehrt und im Anschluss weiter spezifiziert werden. Das *integrierende Verbinden* zielt darauf ab, Ähnlichkeiten, Unterschiede sowie Widersprüche zwischen neuen und bestehenden Bedeutungen aufzuzeigen und in mehreren Zusammenhängen bzw. an mehreren Stellen des Lernprozesses aufzugreifen, um einen hohen Verknüpfungsgrad der kognitiven Struktur zu erreichen. Das Prinzip der *Konsolidierung* beinhaltet das Festigen des Gelernten, insbesondere um eine Klarheit und Stabilität der Ankerideen für einen weiterhin erfolgreichen Lernprozess zu gewährleisten. Die Idee der *sequentiellen Organisation* von Lernsequenzen kann als Spezialfall der bisherigen Prinzipien angesehen werden. Hierbei geht es darum,

Kombinationslernen werden neue und bestehende Konzepte in Beziehung gesetzt, wobei die neue Idee weder spezifischer noch allgemeiner ist als die bestehende Bedeutung, wie z. B. die Beziehung von Angebot und Nachfrage (Ausubel, 1974, S. 56 f.; 109 ff.).

natürliche Abhängigkeiten einzelner Themenkomplexe zu nutzen, d. h. Themen, die andere voraussetzen, später zu behandeln (Ausubel, 1974, S. 166 ff.).

Die Theorie des bedeutungsvollen rezeptiven Lernens wird im Rahmen des Lernens mit Concept Maps zwar häufig erwähnt (z. B. Nesbit & Adesope, 2006; Jüngst, 1992), jedoch wird wenig aufgezeigt, welche Zusammenhänge genau bestehen. Concept Maps können den Assimilationsprozess unterstützen (Novak & Cañas, 2008, S. 3). Beispielsweise können am Anfang des Lernprozesses selbst erstellte Maps das Vorwissen aktivieren und offenlegen. Infolgedessen sind die Lernenden in der Lage, neues Material mit bestehenden Bedeutungen zielführender zu verbinden. In Bezug auf die Prinzipien des expositiven Lehrens können Concept Maps zunächst die *progressive Differenzierung* unterstützen. Hierzu können Referenznetze dienen, die anfänglich sehr allgemein gestaltet und immer detaillierter werden und somit weniger inklusiv sind. Ausubel (1974, S. 166) schlägt in diesem Zusammenhang den Einsatz einer Hierarchie von Advance Organizern vor. Hierbei können Referenzmaps also nicht nur am Anfang des Lernprozesses zur Anwendung kommen, sondern lassen sich auch im Rahmen der Vermittlung eines Themengebietes kontinuierlich einsetzen. Darüber hinaus können sie auch im Anschluss an bestimmte Lerneinheiten angewendet werden, um die Verknüpfung des gerade erworbenen Wissens mit neuen Wissens-elementen zu erleichtern. Einige Autoren sprechen in diesem Zusammenhang von sog. Post Organizers (z. B. Jüngst & Strittmatter, 1995, S. 202; McCagg & Dansereau, 1991, S. 317). Um eine Brücke zwischen Lerneinheiten zu bilden, eignet sich auch die Selbstkonstruktion. Beim Erstellen der Maps wird das Vorwissen bezüglich des soeben abgeschlossenen Themengebietes abgefragt und kann von der Lehrperson für die weitere Vermittlung aktiv mit neuem Wissen in Beziehung gesetzt werden. Sowohl die Selbstkonstruktion als auch der Einsatz von Referenzmaps kann des Weiteren dem *integrierenden Verbinden* dienen. In einem früheren Stadium erstellte bzw. besprochene Concept Maps können beispielsweise an späterer Stelle erneut eingesetzt werden, um Vergleiche anzustellen. Darüber hinaus lassen sich Concept Maps von verschiedenen Seiten bzw. Perspektiven durcharbeiten, um den Verknüpfungsgrad der kognitiven Struktur zu erhöhen. Letztlich können Concept Maps zur Konsolidierung des Gelernten eingesetzt werden. Hierbei eignet sich insbesondere die Selbstkonstruktion, um die Klarheit und Stabilität der Ankerideen zu überprüfen und ggf. zu korrigieren.

Dual Coding Theory

In seiner Dual Coding Theory stellte Paivio (1986, S. 53 ff.) fest, dass unterschiedliche Informationen im Gedächtnis auch unterschiedlich kodiert abgespeichert werden. Der Mensch speichert dabei Reizerfahrungen in zwei unabhängigen Gedächtnissystemen, welche jedoch miteinander verbunden sind, dem verbalen und dem nonverbalen System. Dabei werden verbale Informationen sequentiell verarbeitet und im verbalen System, nonverbale Informationen (bildliche, visuell-räumliche) im nonverbalen System gespeichert. Der Name der Theorie berücksichtigt die angenommenen Prozesse auf der referenziellen Ebene, auf der beide Systeme miteinander in Verbindung treten. Die Repräsentationen der Reize (Logogene, Imagene) können unter bestimmten Voraussetzungen dual kodiert werden, was gegebenenfalls die Erinnerungsleistung verbessert. Eine bildliche Darstellung eines Klassenraums ermöglicht z. B. eine gleichzeitige Speicherung vieler Details, da gesehenen Bildern automatisch verbale Informationen zugeordnet werden. Eine duale Speicherung ist auch bei verbalen Informationen möglich, sofern es sich um konkrete vorstellbare Inhalte handelt. Beispielsweise wird mit dem Hören oder Lesen des Begriffs „Buch“ zunächst das verbale System aktiviert, aber auch eine bildhafte Vorstellung induziert, wohingegen beim Begriff „Erfolg“ aufgrund dessen Abstraktheit eine bildhafte Vorstellung kaum möglich ist. Um die Behaltenswahrscheinlichkeit verbaler und nonverbaler Informationen zu erhöhen, sollte daher abstraktes Wissen konkretisiert oder durch bildliche Darstellung unterstützt werden (Clark & Paivio, 1991, S. 155 f.).

Concept Maps beinhalten oft eher abstrakte Begriffe (verbale Informationen). Eine duale Codierung dieser Begriffe allein wäre somit kaum möglich. Aufgrund der strukturellen Darstellung werden die abstrakten Begriffe bildlich zueinander in Beziehung gesetzt und somit zusätzlich nonverbale Informationen gegeben. Nesbit und Adesope (2013) gehen ebenfalls davon aus, dass Concept Maps sowohl verbale als auch visuell-räumliche Informationen beinhalten, welche separat gespeichert werden. Der Vorteil von Concept Maps zeigt sich darin, dass das verbale Arbeitsgedächtnis entlastet wird, indem der Lernende einen Teil der Weiterverarbeitung auf die kognitiven Ressourcen, die für die visuelle Verarbeitung zur Verfügung stehen, abgeben kann (Nesbit und Adesope, 2013, S. 309).

Kognitive Theorie Multimedialen Lernens

Mayer (2005, S. 2) unterscheidet drei Perspektiven zum Begriffsverständnis von Multimedia: „Delivery Media Perspektive“, „Presentation Modes Perspective“ sowie „Sensory Modalities Perspective“. Entsprechend der „Delivery Media Perspektive“ wird die Kombination von Ausgabemedien in den Mittelpunkt gestellt, z. B. Computerbildschirm und Stimme der Lehrperson. Die „Presentation Modes Perspective“ fokussiert das Format des Lernmaterials. Schnotz (2005, S. 52) unterscheidet hierbei deskriptionale Formate (z. B. Texte, Formeln) und depiktionale Formate (realistische bzw. logische Bilder).⁵² Mayer (2005, S. 2) differenziert zwischen verbalen Formaten (z. B. gedruckte oder gesprochene Texte) und piktorialen Formaten (z. B. Bilder, Fotos, Animationen, Videos). Die „Sensory Modalities Perspective“ spricht verschiedene Sinnesorgane an, üblicherweise das Auge (visuell) und das Ohr (auditiv). Die „Delivery Media Perspektive“ findet in der Forschung zum multimedialen Lernen wenig Beachtung, da hier der technische Aspekt fokussiert wird, es sich beim multimedialen Lernen jedoch um einen lernerzentrierten Ansatz handelt. Mittlerweile existiert ein Begriffsverständnis, welches Multimedia mit dem gleichzeitigen Darbieten von Worten (schriftlich oder mündlich) und Bildern verbindet und somit abhängig vom Forschungsinteresse die „Presentation Modes Perspective“ oder die „Sensory Modalities Perspective“ (Trojahnner & Fürstenau, 2016, S. 64) fokussiert.

Die „Kognitive Theorie Multimedialen Lernens“ beinhaltet den Zusatz „kognitiv“, da kognitive Prozesse zentral am Lernen mit Multimedia beteiligt sind (Selektion, Organisation und Integration von Informationen). Multimediales Lernen bezeichnet den Aufbau von kognitiven Schemata mittels gleichzeitiger Darbietung von Worten und Bildern (Mayer, 2005, S. 15). Um ein verstehendes Lernen zu fördern, erfolgt eine aktive Verknüpfung neuer Informationen mit bestehendem Wissen zu umfassenden Schemata (Trojahnner & Fürstenau, 2016, S. 74). Die Theorie orientiert sich folglich an Informationsverarbeitungsprozessen. Der „Kognitiven Theorie Multimedialen Lernens“ liegen drei Annahmen zugrunde: Erstens besteht das Gedächtnis aus sensorischem Speicher, Arbeits- und Langzeitgedächtnis, zweitens ist die Kapazität des Arbeitsgedächtnisses zeitlich und kapazitätsmäßig begrenzt, drittens werden die Informationen aktiv verarbeitet (Trojahnner &

⁵² Realistische Bilder, wie Zeichnungen, Fotos und Videos weisen eine große Ähnlichkeit zum Original auf, während logische Bilder, wie Diagramme die grundlegende Struktur des Originals in abstrakter Weise wiedergeben (Schnotz, 2011, S. 167).

Fürstenau, 2016, S. 74). Die Worte und Bilder einer multimedialen Nachricht werden separat wahrgenommen und im Arbeitsgedächtnis zu verbalen bzw. bildhaften Modellen verarbeitet (Mayer, 2009, S. 61 f.).

Gesprochene Worte werden mit den Ohren wahrgenommen, im auditiven sensorischen Speicher für wenige Sekunden gespeichert, im Arbeitsgedächtnis als Klangeindrücke (Sounds) zwischengespeichert und schließlich zu einem verbalen Modell zusammengesetzt. Bilder werden mit den Augen wahrgenommen, im visuellen sensorischen Speicher für wenige Sekunden gespeichert, im Arbeitsgedächtnis als Bildeindrücke (Images) zwischengespeichert und schließlich zu einem bildhaften Modell organisiert (Trojahnner & Fürstenau, 2016, S. 69 f.). Mayer (2009, S. 61 f.) differenziert dabei im Gegensatz zu Paivio (1986) zwischen gesprochenen und geschriebenen Worten und legt den Fokus also zunächst auf den Kanal, über den die Informationen aufgenommen werden (Sensory Modalities Perspective). Geschriebene Worte werden wie Bilder als Image zwischengespeichert und zu einem bildhaften Modell organisiert.

Hierbei kann es zu einer „Überkreuz-Verarbeitung von Informationen kommen. Einem gesprochenen Wort wird dabei das entsprechende Bild aus dem Langzeitgedächtnis zugeordnet bzw. der Anblick eines Hundes lässt umgekehrt innerlich das gesprochene Wort erklingen (vgl. Dual Coding Theory). Ein geschriebenes Wort wird zunächst als Folge von Buchstaben (Image) abgespeichert und erhält durch das innere Sprechen eine Bedeutung. Somit kann es in das verbale Modell einfließen. Um die Informationen ins Langzeitgedächtnis zu überführen, ist ein Abgleichen der Inhalte des auditiven und visuellen Arbeitsgedächtnisses sowie der Bezug zum Vorwissen vonnöten (Trojahnner & Fürstenau, 2016, S. 70 f.).

Multimediales Lernen kann die kognitive Belastung aufgrund der kombinierten Nutzung des auditiven und visuellen Informationsverarbeitungssystems reduzieren und infolgedessen die Lernleistung verbessern. Hierbei sind gemäß Mayer (2009, S. 267 f.) verschiedene Gestaltungsprinzipien zu berücksichtigen. Beispielsweise sagt das „Spatial Contiguity Principle“, dass geschriebene Texte und Bilder besser nah beieinander als weit voneinander entfernt präsentiert werden sollten.⁵³ Bei der Kombination aus geschriebenem

⁵³ Weitere Gestaltungsprinzipien sind das „Coherence Principle“, das „Signaling Principle“, das „Redundancy Principle“, das „Temporal Contiguity Principle“, das „Segmenting Principle“, das „Pre-training Principle“, das „Modality Principle“, das „Multimedia Principle“, das „Personalization Principle“, das „Voice Principle“ und das „Image Principle“ (Mayer, 2009, S. 267 f.).

Text und Bild gelangen erst alle Informationen durch den visuellen Kanal. Dabei kann lediglich eine Art von Informationen vom visuellen Kanal zur gleichen Zeit verarbeitet werden, da die Augen aufgrund des Split-Attention-Effects lediglich den textlichen oder bildlichen Informationen Aufmerksamkeit beimessen können. Werden Bilder und Worte räumlich nah beieinander präsentiert, kann die gleichzeitige Verfügbarkeit von bildlichen und verbalen Informationen im Arbeitsgedächtnis verbessert werden (Schnotz, 2005, S. 60 f.).

Bei Concept Maps werden aus der „Presentation Modes Perspective“ verbales und piktoriales Format gemäß Mayer (2005, S. 2) bzw. deskriptionales und depiktionales Format gemäß Schnotz (2005, S. 52) kombiniert. Hierbei wird eine bildhafte Vorstellung induziert und gleichzeitig das verbale System aktiviert. Entsprechend des „Spatial Contiguity Principle“ werden die Worte und Bilder räumlich nah beieinander präsentiert. Concept Maps können als logische Bilder interpretiert werden, da sie die grundlegende Struktur des Originals in abstrakter Weise wiedergeben (Schnotz, 2011, S. 167). Im Vordergrund stehen dabei die Organisations- und die Konkretisierungsfunktion. Die strukturelle Darstellung der Begriffe ermöglicht es einen Überblick über komplexe Sachverhalte zu geben und diese nachvollziehbar zu veranschaulichen. Bei Concept Maps findet eine besondere Form des multimedialen Lernens statt, da die verbalen Informationen in die bildhafte Darstellung integriert werden.

3.2 Empirische Befunde zur Lernwirksamkeit von Concept Maps

3.2.1 Überblick über Metanalysen zur Lernwirksamkeit von Concept Maps

Im Rahmen der Synthese von mehr als 800 Metanalysen identifizierte Hattie (2015, S. 200 f.) Concept Mapping als einen Faktor des Bereichs Unterrichten. In insgesamt sechs Metaanalysen wurde der Effekt auf das Lernen untersucht. Hierbei wurden 287 Studien herangezogen und 332 Effekte herausgefiltert. Das Concept Mapping erzielte dabei eine Effektstärke von $d=0,57$. Hattie (2015) fasst des Weiteren die wichtigsten Ergebnisse der Metaanalysen zusammen. Dabei berichtete er unter Bezug auf Moore und Readance (1984) von stärkeren Effekten für Concept Mapping, wenn die Lernenden mit dem Stoff vertraut sind. Bei der Selbstkonstruktion von Concept Maps verglichen mit traditioneller Instruktion zeigte sich ein größerer Vorteil als im Vergleich zu alternativer Instruktion (Nesbit & Adesope, 2006, S. 432). Hierbei sind nur geringe Unterschiede zwischen Concept Mapping und dem Notizen machen oder dem Schreiben von Zusammenfassungen

erkennbar ($d=0,19$). Größere Unterschiede zeigten sich hingegen für das Concept Mapping im Vergleich zu Vorlesungen oder Diskussionen ($d=0,74$). Das Durcharbeiten vorstrukturierter Maps zeigte gegensätzliche Effekte (Nesbit & Adesope, 2006, S. 432). Im Vergleich zu traditioneller Instruktion zeigte sich kein signifikanter Unterschied und gegenüber anderen Strukturierungsaufgaben - wie dem Durcharbeiten von Texten oder Zusammenfassungen - sind lediglich geringe Effekte erkennbar ($d=0,39$ bzw. $0,28$). In den Metaanalysen von Horton u. a. (1993) sowie Nesbit und Adesope (2006) wurde festgestellt, dass es keinen Einfluss zu haben scheint, ob die Maps von Einzelpersonen oder Gruppen erstellt werden. Im Hinblick auf den Grad der Vorstrukturierung zeigte sich ein Vorteil der Selbstkonstruktion gegenüber dem Durcharbeiten von vorstrukturierten bzw. Expertenmaps ($d=0,81$ gegenüber $d=0,36$ bei Nesbit und Adesope, 2006, S. 447). Hattie (2015) hebt darüber hinaus hervor, dass besonders Lernschwächere vom Concept Mapping zu profitieren scheinen.

In einer aktuellen Metaanalyse von Schroeder, Nesbit, Anguiano und Adesope (2017) wurden die Ergebnisse der Metaanalyse von Nesbit und Adesope (2006) mit aktuellen Studien (2005-2014) kombiniert. Insgesamt erzielte das Concept Mapping eine Effektstärke von $d=0,58$, wobei eine hohe Variabilität erkennbar war. Aus diesem Grund wurde eine Moderatorenanalyse durchgeführt. Im Vergleich zu traditionellen Instruktionsformaten, wie Vorlesungen oder Diskussionen, konnte das Lernen mit Concept Maps einen hohen Effekt von $d=1,05$ erzielen. Hierbei wurden Studien einbezogen, die sowohl das Erstellen als auch das Durcharbeiten von Concept Maps beinhalteten. Gegenüber alternativen Instruktionsformaten, wie dem Erstellen oder Durcharbeiten von Texten, Zusammenfassungen oder dem Notizen machen bzw. durcharbeiten, konnten lediglich Effekte zwischen $d=0,29$ und $d=0,57$ erzielt werden. Schroeder u. a. (2017) führte weiterhin getrennte Analysen für die Selbstkonstruktion und das Durcharbeiten von Maps im Vergleich zu traditionellen und alternativen Instruktionsformaten durch. Hierbei konnte nicht nur für die Selbstkonstruktion ($d=1,05$) ein großer Effekt für den Vergleich zu traditionellen Instruktionsformaten erzielt werden, wie es bei Nesbit und Adesope (2006) der Fall war, sondern auch für das Durcharbeiten der Maps ($d=1,09$). Verglichen mit dem Durcharbeiten bzw. Erstellen von Zusammenfassungen und dem Notizen machen bzw. durcharbeiten liegen die Effekte zwischen $d=0,29$ und $0,72$. Für die Selbstkonstruktion und das Durcharbeiten von Maps zeigte sich somit ein größerer Vorteil im Vergleich zu der traditionellen gegenüber der alternativen Instruktion.

Als weiterer Einflussfaktor wurde das Fachgebiet beleuchtet, in dem die Studien durchgeführt wurden. Hierbei wurden die Bereiche Naturwissenschaften, Technik, Ingenieurwesen und Mathematik zum Bereich „STEM“ zusammengefasst und den „Non-STEM“-Bereichen gegenübergestellt. Insgesamt waren keine signifikanten Unterschiede zwischen beiden Bereichen erkennbar.

Im Hinblick auf die Dauer der Intervention konnte gezeigt werden, dass das Lernen mit Concept Maps umso effektiver ist, je länger es eingesetzt wird (Dauer > 4 Wochen: $d=0,72$; Dauer 1-4 Wochen: $d=0,68$; Dauer <1 Woche: $d=0,36$). In der Metaanalyse von Nesbit und Adesope (2006) wurde hingegen der größte Effekt für eine Dauer von ein bis vier Wochen identifiziert. Dies wurde damit begründet, dass der anfängliche Neuigkeitseffekt verschwindet, sobald die Lernenden mit der Methode vertraut sind. Schroeder u. a. (2017) begründen ihre gegensätzlichen Ergebnisse damit, dass die kognitive Belastung, welche insbesondere mit dem Format der Selbstkonstruktion verbunden ist, mit der Zeit durch eine erhöhte Vertrautheit verringert wird und sich der Lernende besser auf die kognitive Verarbeitung der Lerninhalte fokussieren kann. An dieser Stelle sind Langzeitanalysen vonnöten, welche die Lernwirksamkeit von Concept Maps mindestens über ein Semester untersuchen (Schroeder u. a., 2017).

Des Weiteren wurden für die Selbstkonstruktion ($d=0,72$) signifikant höhere Effekte als für das Durcharbeiten vorstrukturierter Maps ($d=0,43$) erzielt. Schroeder u. a. (2017) merkten jedoch kritisch an, dass die meisten Studien, in welchen Concept Maps erstellt werden, mit passiven Lernaufgaben verglichen werden und das Durcharbeiten von Maps vermehrt aktiven Lernaufgaben gegenübergestellt wird. Da der Vorteil von aktivem gegenüber passivem Lernen bereits nachgewiesen wurde (Freeman u. a., 2014), könnten Verzerrungen auftreten.

Schroeder u. a. (2017) untersuchten darüber hinaus, inwieweit die Lernwirksamkeit der Selbstkonstruktion im Vergleich zum Durcharbeiten der Maps von bestimmten Faktoren beeinflusst wird. Hierbei zeigten sich keine Einflüsse des Fachgebietes, der Art der Maps (statisch, animiert oder interaktiv) sowie der Anzahl der erstellenden Personen (einzeln, Gruppe). Im Hinblick auf die Klassenstufe zeigte sich kein Einfluss bei der Selbstkonstruktion. Beim Durcharbeiten hingegen zeigten sich größere Effekte für Schüler der Primarstufe ($d=0,82$) und Sekundarstufe ($d=1,24$) gegenüber Schülern des tertiären Bereichs ($d=0,32$).

Die bisher durchgeführten Metaanalysen zeigen insgesamt den hohen Stellenwert des Lernens mit Concept Maps auf.

3.2.2 Studien zum Einfluss von Vorwissen, verbalen Voraussetzungen, kognitiven Fähigkeiten und dem Training der Methode auf die Lernwirksamkeit von Concept Maps

Einige Studien haben gezeigt, dass die Lernwirksamkeit von Concept Maps von verschiedenen Faktoren abhängig ist. Hattie (2015) betont die besonders hohe Lernwirksamkeit des Concept Mapping für Leistungsschwächere. Nesbit und Adesope (2006) heben den Einfluss von individuellen Voraussetzungen, wie den verbalen Fähigkeiten sowie dem Vorwissen, hervor. Des Weiteren spielt die Einstellung der Lernenden zur Methode des Concept Mapping eine bedeutende Rolle (Stracke, 2004). Darüber hinaus sollten die Lernenden mit der Methode vertraut sein, weshalb ein entsprechendes Training eine Grundvoraussetzung darstellt (Chang u. a., 2002; Hilbert & Renkl, 2008). Nachfolgend werden die Ergebnisse einiger Studien, welche den Einfluss des Vorwissens, der verbalen bzw. kognitiven Fähigkeiten, des Trainings sowie der Einstellung zur Methode des Concept Mapping analysieren, vorgestellt.

Vorwissen

In der Metaanalyse von Nesbit und Adesope (2006) wird im Hinblick auf den Einfluss des Vorwissens auf eine Studie von Lambiotte und Dansereau (1992) verwiesen, die eine Studie mit Studierenden im Bereich Biologie durchführten, um die Lernwirksamkeit von Knowledge Maps als Strukturierungshilfe im Vergleich zum Durcharbeiten von stichpunktartigen Zusammenfassungen oder Begriffslisten zu erforschen. Die Studierenden hörten Tonbandaufnahmen, während sie die Strukturierungshilfen studierten. Insgesamt zeigte sich ein Vorteil der vorstrukturierten Concept Maps bei Lernenden mit niedrigem Vorwissen gegenüber den anderen Strukturierungshilfen. Bei Lernenden mit hohem Vorwissen waren Begriffslisten effektiver als die anderen Strukturierungshilfen. Lambiotte und Dansereau (1992) verweisen hierbei auf die den Vorteil der Advance Organizer (Mayer, 1979; vgl. auch Kapitel 3.1.3). Die spezielle Makrostruktur, welche durch das Map präsentiert wird, unterstützt möglicherweise die Wissenskonstruktion von vorwissensärmeren Lernenden, steht jedoch im Konflikt mit den kognitiven Strukturen, welche bei vorwissensstärkeren Lernenden bereits aufgebaut wurden.

O'Donnell und Dansereau (2000) führten eine Studie mit amerikanischen Studierenden zum Thema vegetatives Nervensystem und Wahrscheinlichkeitstheorie durch. Ziel war es, die einzelnen bzw. kombinierten Effekte verschiedener Lernmaterialien (Knowledge Maps oder Texte) und Lehrmaterialien (Überblick-Maps und stichpunktartige Zusammenfassungen) im Kontext des kooperativen Lernens zu untersuchen. Die Probanden studierten zunächst die Knowledge Maps oder Texte als Lernmaterialien. Anschließend wurden die Lerninhalte in einer kooperativen Lernumgebung paarweise zusammengefasst, wobei jeweils ein Studierender die Lehrerrolle und der andere die Lernerrolle übernahm. Dabei konnten vorbereitete Maps oder stichpunktartige Zusammenfassungen zu Hilfe genommen werden. Für das zweite Thema wurde getauscht. Vor der Intervention erhielten die Probanden ein kurzes Training zur Nutzung von Knowledge Maps und Zusammenfassungen. Unter anderem wurden mit den Lernenden Beispielmats bzw. -zusammenfassungen über Inhalte ihrer Universität durchgearbeitet. Insgesamt zeigte sich ein Vorteil für Lernende mit hohem Vorwissen gegenüber Lernenden mit niedrigem Vorwissen im Nachtest, sowohl für Gruppen die mit Knowledge Maps lernten als auch für Gruppen, die die Überblick Maps als Zusammenfassung nutzten. Ebenso konnte ein Vorteil der Texte gegenüber den Knowledge Maps als Lernmaterial für Lernende mit niedrigem Vorwissen nachgewiesen werden. Hohes Vorwissen unterstützte die Probanden, welche Maps als Lehrmaterial nutzten besser als Probanden, welche stichpunktartige Zusammenfassungen als Lehrmaterial verwendeten. O'Donnell und Dansereau (2000) begründen ihre Ergebnisse mit der Neuheit und Schwierigkeit des Lehr- und Lernmaterials für Lernende mit niedrigem Vorwissen. An dieser Stelle können auch die Argumentationen der Metaanalyse von Schroeder u. a. (2017) zum Einflussfaktor Dauer herangezogen werden. Hierbei zeigten sich größere Effekte für Studien, die das Mapping über einen längeren Zeitraum verwendeten. Die Lernenden können sich mit der Zeit besser auf die kognitive Verarbeitung der Lerninhalte konzentrieren, die kognitive Belastung durch das Anwenden der Methode Concept Mapping verringert sich. Auch wenn dieser Effekt auf die Selbstkonstruktion bezogen war, sind die Begründungen dennoch übertragbar. Niedriges Vorwissen erhöht die kognitive Belastung beim Lernen, insbesondere aber beim Lehren anhand von vorstrukturierten Maps. Anderen Personen bestimmte Zusammenhänge anhand von Maps zu erläutern, setzt eine gewisse Vertrautheit mit der Methode voraus. Hohes Vorwissen könnte diese Defizite kompensieren, niedriges Vorwissen dagegen noch verstärken. Auf die Bedeutung des Trainings in diesem Kontext wird in diesem Kapitel noch eingegangen.

Ryssel und Fürstenau (2011) führten eine Studie zum Lernen betriebswirtschaftlicher Grundlagen im Rahmen eines Planspiels mit Schülern der Sekundarstufe I einer Mittelschule durch, um herauszufinden, ob das Erstellen von Concept Maps gegenüber dem Erstellen von Texten als ergänzende Lernstrategie lernwirksamer ist. Hierbei zeigte sich für Lernende mit niedrigem Vorwissen kein Vorteil des Concept Mapping gegenüber dem Erstellen von Texten im Hinblick auf den Wissenszuwachs. Unabhängig von der Versuchsgruppe (Erstellen einer Zusammenfassung, Selbstkonstruktion eines Concept Maps, Kontrollgruppe) erzielten Lernende mit niedrigem Vorwissen einen vergleichbar hohen Wissenszuwachs und umgekehrt. Die Durchführung des Planspiels führte zu einer Angleichung des Wissensniveaus. Es wurde jedoch festgestellt, dass vorwissenschwache Lernende Concept Maps mit niedriger Qualität erstellten. Hohes Vorwissen führte hingegen auch zu einer höheren Qualität der Maps. Bei den Probanden der Text-Gruppe hatte das Vorwissen einen nur geringen Einfluss auf die Qualität der Texte. Insgesamt war die Qualität der Maps bei vorwissenstarken Schülern deutlich höher als die Qualität der Texte. Um Concept Maps mit einer hohen Qualität zu erstellen, scheint demnach hohes Vorwissen eine Voraussetzung zu sein. Die Ergebnisse widersprechen den Resultaten von Lambiotte und Dansereau (1992) und stehen eher im Einklang mit den Befunden von O'Donnell und Dansereau (2000). Beide Studien verwenden im Gegensatz zu Ryssel und Fürstenau (2011) vorstrukturierte Maps. Dabei unterscheiden sich jedoch die Sozialformen, in denen die vorstrukturierten Maps zur Anwendung kommen. Bei Lambiotte und Dansereau (1992) waren die Lernenden lediglich aufgefordert, die Lerninhalte, welche eine Lehrperson über die Tonbandaufnahme vermittelte, anhand der Maps nachzuvollziehen. Bei O'Donnell und Dansereau (2000) hingegen mussten die Lernenden den Stoff selbstständig erarbeiten und anschließend sogar noch für einen Partner zusammenfassen. Infolgedessen spielt das Format des Lernmaterials als vorstrukturiertes Map wahrscheinlich gar keine Rolle, sondern die Sozialform könnte die Unterschiede eher erklären.

Verbale Voraussetzungen

Schmid und Telaro (1990) untersuchten die Lernwirksamkeit von Concept Maps als integrierenden Bestandteil im Biologieunterricht bei kanadischen Schülern einer High School und konnten zeigen, dass die Probanden mit den geringsten verbalen Fähigkeiten, welche nach jeder Lektion Concept Maps erstellten, gegenüber der Kontrollgruppe, welche keine Concept Maps anfertigte, bessere Ergebnisse im Nachtest erzielten. Stensvold und Wilson (1990) führten einen Laborversuch mit Schülern neunter Klassen zum Lernen

chemischer Zusammenhänge durch. Die Experimentalgruppe erstellte vor und nach der Durchführung der Laborexperimente Concept Maps, wohingegen die Kontrollgruppe lediglich die Experimente durchführte. Es zeigte sich, dass Schüler mit geringen verbalen Fähigkeiten vom Erstellen der Concept Maps im Vergleich zur Kontrollgruppe beim Nachtest profitierten. Bei Schülern mit hohen verbalen Fähigkeiten war die Experimentalgruppe der Kontrollgruppe jedoch unterlegen. Die Autoren begründeten das Ergebnis mit der „Aptitude Treatment Interaction“. Dabei hängt der Effekt bestimmter instruktionaler Maßnahmen von Merkmalen der Lernenden ab. Lernarrangements sollten demnach an die Fähigkeiten der Lernenden angepasst werden. Lernenden mit hohen verbalen Fähigkeiten sollten Freiheitsgrade bei der Steuerung ihres Lernens zur Verfügung gestellt werden, wohingegen bei niedrigen verbalen Fähigkeiten eher eine instruktionale Unterstützung lernförderlich ist (Snow, 1989).

Hilbert und Renkl (2008) konnten den Vorteil des Concept Mapping für Lernende mit geringen verbalen Fähigkeiten nicht bestätigen. Ziel ihrer Studie mit deutschen Studierenden zum Themenbereich Stammzellen war es, individuelle Unterschiede beim computergestützten Erstellen von Concept Maps zu identifizieren, um eine Strategie zum Vermitteln der Methode des Concept Mapping, welche an die Bedürfnisse der Lernenden angepasst ist, zu entwickeln. Mittels der Methode des lauten Denkens wurden hierzu unter anderem kognitive Prozesse analysiert. Im Hinblick auf die verbalen Fähigkeiten zeigte sich kein Zusammenhang zum Lernerfolg.⁵⁴ Bei Ryssel und Fürstenau (2011) konnte ebenfalls keine Korrelation zwischen der Sprachfähigkeit und dem Wissenszuwachs bzw. der Sprachfähigkeit und der Qualität der Maps festgestellt werden. An dieser Stelle besteht somit weiterer Forschungsbedarf.

Kognitive Fähigkeiten

Der Studie von Haugwitz, Nesbit und Sandmann (2010) zum Einfluss der kognitiven Fähigkeiten wurde in der Metaanalyse von Schroeder u. a. (2017) besondere Bedeutung beigemessen. Daher wird der Einfluss der kognitiven Fähigkeiten als weitere individuelle

⁵⁴ Zum Messen des Lernerfolgs wurde einerseits der Wissenszuwachs herangezogen. Hierzu beantworteten die Studierenden eine offene Frage vor dem Lesen von Zeitungsartikeln, dessen Antwort das erste Mal nach dem Lesen der Texte und ein zweites Mal nach dem Erstellen der Maps korrigiert werden konnte. Die Antworten wurden in eine propositionale Struktur überführt und anschließend die Differenz der richtigen und falschen Propositionen zwischen den drei Messzeitpunkten berechnet. Des Weiteren wurde ein Nachtest im Multiple-Choice-Format durchgeführt.

Voraussetzung nachfolgend näher beleuchtet. Haugwitz u. a. (2010) führten eine Untersuchung zum Lernen biologischer Zusammenhänge mit deutschen Schülern der Sekundarstufe I durch. Die Studie setzte sich aus einer kooperativen Lernaktivität und dem anschließenden gemeinsamen Zusammenfassen als Aufsatz oder dem Erstellen eines Concept Maps zusammen. Zur Erhebung der kognitiven Fähigkeiten wurde der Kognitive Fähigkeitstest (KFT) von Heller und Perleth (2000) herangezogen, der sich aus drei Subskalen zusammensetzt: sprachliches Denken, quantitative (numerische) Fähigkeiten und anschauungsgebundenes (figurales) Denken. Es zeigten sich keine signifikanten Unterschiede bei Lernenden mit einer hohen kognitiven Fähigkeit. Eine geringe kognitive Fähigkeit hingegen führte zu einer Überlegenheit des Erstellens von Concept Maps gegenüber dem Schreiben von Aufsätzen. Haugwitz u. a. (2010) begründen diese Ergebnisse damit, dass das Concept Mapping Lernenden mit einer geringen kognitiven Fähigkeit visuelle Hinweise gibt, inwieweit ein Concept Map vollständig ist oder nicht. So erkennen die Lernenden z. B. das Fehlen einer Relation zwischen zwei Konzepten. Das Bereitstellen von ergänzenden Hinweisen zum Planen und Kontrollieren des Mappingprozesses (sog. Prompts) kann dies zusätzlich unterstützen.

Training der Methode Concept Mapping

Bereits in Kapitel 3.1.1 wurde darauf hingewiesen, dass die Lernenden im Rahmen der Selbstkonstruktion eine aktive Rolle übernehmen und dem Erlernen der Technik eine entsprechende Bedeutung beigemessen werden sollte. Hilbert, Nückles, Renkl, Minarik, Reich und Ruhe (2008, S. 120) weisen darauf hin, dass bisher nur wenige Trainingsstudien existieren, um insbesondere Anfängern den Umgang mit der Methode zu erleichtern. Die Forschergruppe um Hilbert und Renkl unterscheiden zwei Ansätze, um mit Lernenden Concept Mapping zu trainieren. Dabei beziehen Sie sich auf Friedrich und Mandl (1997, S. 253 ff.), die im Rahmen der Förderung des selbstgesteuerten Lernens einen *direkten* und einen *indirekten Ansatz* vorschlagen.

Beim *direkten Ansatz* wird den Lernenden die Methode in Form eines Strategietrainings direkt vermittelt. An erster Stelle steht dabei das Planen des Mapping-Prozesses. Während der Mapperstellung sollten sich die Lernenden auf die Beziehungen zwischen den Konzepten konzentrieren. Der Fortschritt des Maps muss dabei ständig kontrolliert werden. Gegebenenfalls muss den Lernenden die Möglichkeit gegeben werden, ihr Concept Map zu überarbeiten. Wichtig ist es, das Üben der Methode anzubieten. Hierbei kann es hilfreich sein, anhand von Beispielen den Prozess des Mapperstellens zu demonstrieren.

An dieser Stelle sei auf das Lernen aus Beispielen verwiesen, dessen Lernwirksamkeit bereits vielfach nachgewiesen wurde (z. B. Renkl, Hilbert, & Schworm, 2007). Das direkte Training kann dabei zum einen anhand von Lösungsbeispielen selbstständig durchgeführt werden. Andererseits können die notwendigen Schritte auch von einer Lehrperson gezeigt und von den Lernenden nachvollzogen werden. Dabei sollte es nicht beim alleinigen Nachvollziehen bleiben, sondern die Lernenden sollten selbst tätig werden. Sie sollten insbesondere die Möglichkeit erhalten, im Anschluss das selbstständige Erstellen zu üben.

Leopold (2009, S. 240 ff.) führte eine Studie zum Lernstrategietraining durch, bei der den Lernenden schrittweise das Vorgehen beim Erstellen von Concept Maps aus Texten erläutert wurde. Hierbei wurde u. a. zwischen zwei Trainingsgruppen unterschieden: „Mappingtraining mit Selbstregulation“ vs. „Mappingtraining ohne Selbstregulation“. Allen Probanden wurde in einem ersten Schritt ein Fallbeispiel gezeigt, in welchem ein fiktiver Schüler auf die Idee kommt, sich die Zusammenhänge beim Lesen schwieriger Fachtexte als Concept Maps darzustellen. In einem zweiten Schritt wurde das Vorgehen beim Erstellen von Maps aus Texten erläutert (Textabschnitte lesen und markieren, Map erstellen und wiederholen, ggf. mit Expertenmap vergleichen) und anschließend geübt. Beim „Mappingtraining mit Selbstregulation“ wurde das Fallbeispiel (Schritt 1) um drei Phasen zur Selbstregulation ergänzt: Selbstbeobachtung zum Reflektieren der vorgenommenen Handlungen beim Mapping, Selbsteinschätzung zum Überprüfen der Vollständigkeit der Maps und Reaktion, um den Lernenden das unterschiedliche Vorgehen in Abhängigkeit vom Ergebnis der Selbsteinschätzung vor Augen zu führen. Die drei Phasen wurden beim „Mappingtraining mit Selbstregulation“ zusätzlich als dritter Schritt in das Training implementiert. Anschließend lasen beide Gruppen einen naturwissenschaftlichen Sachtext und erstellten zusätzlich Concept Maps. Beim lehrzielorientierten Verstehenstest war die Trainingsgruppe mit Selbstregulation der Trainingsgruppe ohne Selbstregulation signifikant überlegen. Darüber hinaus erstellten sie auch qualitativ bessere Maps. Es zeigte sich also, dass das Lernstrategietraining alleine nicht ausreichte, den Lernenden zu mehr Verständnis zu verhelfen, sondern lediglich in Kombination mit dem Selbstregulationstraining effektiv ist.

Beim *indirekten Ansatz* wird die Trainingsphase so gestaltet, „dass sie die Anwendung dieser Aktivitäten erfordert“ (Hilbert, Nückles, Renkl, u. a., 2008, S. 120). Dies bedeutet,

dass die Lernenden nach einer kurzen Einführung zur Methode sofort ein Map selbstständig erstellen sollen. Um den Prozess zu unterstützen, können Prompts integriert werden. Prompts sind Fragen oder Hinweise, die zum Planen und Kontrollieren des Mappingprozesses sowie zum bewussten Beschriften der Relationen auffordern, um produktive Lernaktivitäten anzuregen (Hilbert, Nückles, Renkl, u. a., 2008, S. 121).

In der bereits vorgestellten Studie von Hilbert und Renkl (2008) zeigte die Auswertung der Protokolle zum lauten Denken, die Analyse der Variablen zum Lernerfolg sowie einiger quantitativer und qualitativer Indikatoren (Anzahl der Konzepte, Relationen und Cross-Links; Anzahl korrekt bezeichneter Relationen) und eine anschließende Clusteranalyse, dass 50 Prozent der Lernenden zu den nicht erfolgreichen „Mappern“ (Non-Labeling Mappers und Non-Planning Mappers) zählen. Hierbei dachten die Non-Labeling-Mappers wenig über die Beziehungen zwischen den Konzepten nach und erstellten Maps mit vielen fehlenden Beschriftungen, jedoch planten und kontrollierten sie ihren Mapping-Prozess. Die Non-Planning-Mappers dachten zwar über die Beziehungen zwischen den Konzepten nach, planten und kontrollierten ihren Mapping-Prozess jedoch nicht. Infolgedessen schlussfolgerten Hilbert und Renkl (2008), dass nur eine Aktivität allein nicht zu genügen scheint. Als Ergebnis schlugen sie einen indirekten Trainingsansatz mit Prompts vor, der die Lernenden während des Erstellens der Maps dabei unterstützt, die Relationen sorgfältig zu beschriften und den Erstellungsprozess zu planen als auch zu kontrollieren.

Dieses Ergebnis griffen Hilbert, Nückles, Renkl u. a. (2008) auf und schlugen eine Kombination von *kognitiven* und *metakognitiven Prompts* zur Unterstützung des Erstellungsprozesses von Concept Maps vor. Nachfolgende Tabelle 7 zeigt die in der Studie von Hilbert, Nückles, Renkl u. a. (2008, S. 122) verwendeten kognitiven und metakognitiven Prompts.

Tabelle 7: Kognitive und Metakognitive Prompts

Kognitive Prompts	Metakognitive Prompts
<ul style="list-style-type: none"> – Verschaffen Sie sich einen Überblick über die zentralen Punkte und ihre Zusammenhänge («Links»)! – Formulieren Sie für die wichtigsten Punkte Konzeptknoten! – Ordnen Sie die Konzeptknoten räumlich an! – Wie können Sie die Konzeptknoten hierarchisch gliedern? – Finden Sie Zusammenhänge («Links») zwischen den einzelnen Konzeptknoten und beschriften Sie die gefundenen Links! – Erklären Sie sich die Bedeutung der Konzeptknoten und ihre Zusammenhänge! 	<ul style="list-style-type: none"> – Gibt es noch etwas, was an Ihrer Concept Map fehlt, zum Beispiel Links und/oder Konzepte? – Gibt Ihre Concept Map den Inhalt des Textes angemessen wieder? – Hat Ihre Concept Map noch Lücken? – Haben Sie die Begriffe (Konzepte) und Zusammenhänge (Links) in Ihrer Map auch gut verstanden? – Versuchen Sie, Schwierigkeiten im Verständnis von Konzepten und/oder Zusammenhängen (Links) Ihrer Map zu beseitigen! – Sind Sie mit der Gestalt (räumlichen Anordnung der Konzepte) Ihrer Concept Map zufrieden?

Hilbert, Nückles, Renkl u. a. (2008) führten eine Untersuchung mit deutschen Gymnasiasten der Oberstufe zum Thema Stammzellenforschung und Sterbehilfe durch. Ziel der Studie war es, den Nutzen von kognitiven und/oder metakognitiven *Prompts* in Kombination mit dem *indirekten Ansatz* auf den Lernerfolg beim Lernen mit Concept Mapping zu analysieren. Alle Probanden erhielten vorab eine kurze Einführung in das Concept Mapping, welche aus einer Erklärung der Begriffe Knoten und Verbindung sowie dem Zeigen eines Beispiel-Maps bestand. An zwei Terminen im Abstand von einer Woche erstellten die Probanden insgesamt zwei Concept Maps, nachdem sie Texte zum Thema Stammzellenforschung bzw. Sterbehilfe gelesen hatten. Dabei wurden die drei Experimentalgruppen nur während des 1. Mappings entweder mit kognitiven, metakognitiven oder einer Kombination aus kognitiven und metakognitiven Prompts unterstützt. Die Lernenden wurden mehrmals zum Beachten der Prompts aufgefordert. Die Kontrollgruppe erhielt keine Prompts. Das Mapping erfolgte computergestützt. Insgesamt zeigte sich ein signifikanter Vorteil im Nachtest für die Experimentalgruppen (Prompts) gegenüber der Kontrollgruppe (keine Prompts). Das Ergebnis konnte für die zweite Durchführung nicht bestätigt werden. Darüber hinaus zeigten sich auch keine Unterschiede zwischen den verschiedenen Arten der Unterstützung durch Prompts. Die Untersuchung von Hilbert, Nückles, Renkl u. a. (2008) bestätigte somit, dass Prompts als Hilfestellung bei der Anwendung der neuen Technik des Concept Mapping positive Effekte haben. Um das selbstständige Ausführen der Mapping-Technik zu verbessern, müssen Prompts jedoch mehr als einmal dargeboten werden. Darüber hinaus konnte der Vorteil der Kombination aus kognitiven und metakognitiven Prompts nicht bestätigt werden.

Hilbert und Renkl (2009) führten zwei Studien zum Erlernen des computerbasierten Concept Mapping anhand von heuristischen Lösungsbeispielen durch. Die erste Untersuchung *verglich den direkten mit dem indirekten Ansatz* und wurde an einer deutschen Polizeihochschule durchgeführt. Ziel der Studie war es herauszufinden, inwieweit das Training mit heuristischen Lösungsbeispielen (direkter Ansatz) lernwirksamer ist als das selbstständige Erstellen mittels Selbstkonstruktion von Concept Maps (indirekter Ansatz). Nach einer kurzen Einführung in das Concept Mapping analog zur Studie von Hilbert u. a. (2008) folgte die Trainingsphase. Eine Gruppe erhielt dabei papierbasierte heuristische Lösungsbeispiele, die einen fiktiven Studierenden zeigten, der anhand eines kurzen Textes zum Thema Kampfsport und posttraumatischer Belastungsstörung die notwendigen Schritte zum Erstellen idealtypischer Concept Maps erläuterte (Beispielgruppe). Die andere Gruppe wurde instruiert, auf Basis der gleichen Texte selbstständig Concept Maps im Paper & Pencil-Verfahren zu erstellen (Trainingsgruppe). Nach der Durchführung eines Vortests sowie einer kurzen Erläuterung der Concept Mapping-Software erstellten beide Gruppen in der Anwendungsphase auf der Basis von Zeitungsartikeln zum Thema Stammzellen selbstständig Concept Maps. Abschließend absolvierten die Probanden Posttests zum Thema Stammzellen sowie zum konzeptuellen Wissen über Concept Maps. Insgesamt zeigten sich keine signifikanten Gruppenunterschiede sowohl zum Thema Stammzellen als auch zum konzeptuellen Wissen über Concept Maps. Hilbert und Renkl (2009) nahmen an, dass die Lernenden der Beispielgruppe, das Material lediglich oberflächlich durcharbeiteten und somit die Kapazität ihres Arbeitsgedächtnisses nicht nutzten, um zu lernen, wie Concept Maps effektiv konstruiert werden können. Aus diesem Grund wurden in einer zweiten Studie Prompts eingesetzt. Wie in der ersten Studie war es zum einen Ziel, den Vorteil heuristischer Lösungsbeispiele (direkt) gegenüber dem selbstständigen Erstellen (indirekt) nachzuweisen. Darüber hinaus wurde angenommen, dass die Bereitstellung von Prompts zu den Beispielen gegenüber den Beispielen ohne Prompts lernwirksamer ist. Zusätzlich wurde die kognitive Belastung der drei Versuchsgruppen (Beispielgruppe mit Prompts, Beispielgruppe ohne Prompts, Trainingsgruppe) erhoben. Die Untersuchung wurde mit deutschen Gymnasiasten der Oberstufe durchgeführt. Das Training erfolgte analog zur ersten Studie jedoch zu anderen Themen (Kaufverträge und Abschreibung). Die Lernenden der Beispielgruppe mit Prompts wurden zusätzlich instruiert, sich selbst zu durchdenken, welcher Phase des Erstellungsprozesses die aktuellen Handlungen der fiktiven Studierenden jeweils zugeordnet werden könnte. Die Anwendungsphase wurde ebenfalls analog zur ersten Studie, jedoch zum

Thema Marketing, durchgeführt. Es zeigten sich keine signifikanten Unterschiede zwischen dem Training mit Beispielen und dem selbstständigen Erstellen im Hinblick auf das Wissen zum Thema Marketing. Beim konzeptuellen Wissen über Concept Mapping konnte jedoch ein signifikanter Vorteil des Trainings mit Beispielen gegenüber dem selbstständigen Erstellen nachgewiesen werden. Beim Vergleich der Beispielgruppen mit und ohne Prompts zeigte sich kein Unterschied beim konzeptuellen Wissen über Concept Mapping, jedoch ein signifikanter Vorteil der Beispielgruppe mit Prompts gegenüber der Beispielgruppe ohne Prompts beim domänenspezifischen Wissen. Zum Festigen der Methode ist folglich ein Training mit heuristischen Beispielen generell gegenüber dem selbstständigen Erstellen zu bevorzugen. Um jedoch den Erwerb domänenspezifischen Wissens zu fördern, sollten die Lernenden beim Durcharbeiten der heuristischen Lösungsbeispiele zusätzlich mit Prompts unterstützt werden. Die kognitive Belastung der Beispielgruppe ohne Prompts gegenüber der Trainingsgruppe konnte reduziert werden. Bei den Probanden der Beispielgruppe mit Prompts zeigte sich eine Erhöhung der kognitiven Belastung gegenüber der Beispielgruppe ohne Prompts. Dies zeigt, dass Lernende, die die Möglichkeit haben, heuristische Beispiele durcharbeiten eigentlich mehr Kapazitäten hätten, um aus ihnen zu lernen, als die Lernenden, welche die Methode selbstständig üben müssen. Dennoch können sie diesen Vorteil nicht nutzen. Um vollständig aus dem Lernen mit Beispielen profitieren zu können, müssen die Lernenden die Kapazität des Arbeitsgedächtnisses nutzen und die Beispiele aktiv durchdenken. Hierfür ist die zusätzliche Unterstützung mit Prompts nötig. Anderenfalls erfolgt das Durchdenken eher oberflächlich oder passiv. Ein Training auf der Basis von heuristischen Beispielen sollte daher immer mit Prompts unterstützt werden (Hilbert & Renkl, 2009). Um Aussagen für den indirekten Ansatz mit Prompts treffen zu können, hätte noch eine Trainingsgruppe mit Prompts einbezogen werden können. Der Vorteil der Beispielgruppen gegenüber der Trainingsgruppe hätte sich so möglicherweise relativiert.

Die Ergebnisse der vorgestellten Studien zum Einfluss des Trainings zeigen, dass ein direktes Training mit Möglichkeiten der Selbstregulation (Leopold, 2009) bzw. mit Prompts (Hilbert & Renkl, 2009) ergänzt werden sollte. Dabei können kognitive und metakognitive Prompts eingesetzt werden. Beim indirekten Training sollten Prompts mehr als einmal dargeboten werden, um einen längerfristigen Effekt zu erzielen (Hilbert, Nückles, Renkl, u. a., 2008).

Einschätzung des Umgangs mit Concept Mapping

Abschließend soll die Einschätzung des Umgangs mit Concept Mapping als Einflussfaktor betrachtet werden. Auf diese Weise können z. B. eine geringe Motivation, die fehlende Einsicht in die Nützlichkeit der Mapping-Technik oder Schwierigkeiten im Umgang als Ursachen für eine geringe Lernwirksamkeit identifiziert werden. Bisherige Studien zur Bewertung von Concept Maps zeigen ein uneinheitliches Bild. Teilweise stößt Concept Mapping als Lernhilfe bei den Lernenden auf eine positive Resonanz (Roth, 1994, S. 12), teilweise bewerten Lernende die Methode jedoch auch als eine schwierige Aufgabe, die wenig Spaß macht (Barenholz & Tamir, 1992, S. 48). Allerdings kann kein eindeutiger Zusammenhang zwischen der Einstellung gegenüber der Concept-Mapping-Technik und der späteren Lernleistung festgestellt werden. Die Studien von McClure, Sonak und Suen (1999) sowie Rice, Ryan und Samson (1998) fanden beispielsweise keinen Zusammenhang zwischen beiden Variablen. Hierbei muss jedoch die Aussagekraft der Studien hinterfragt werden. Rice, Ryan und Samson (1998, S. 1120 f.) verwendeten lediglich ein allgemeines Item zur Bewertung der Methode⁵⁵ und Mc Clure u. a. (1999, S. 481) operationalisierten die Familiarität mit der Methode über die Häufigkeit, mit der die Probanden zuvor mit Concept Maps gearbeitet hatten, was bei den meisten nie oder selten der Fall gewesen war.

Stracke (2004) führte eine Untersuchung mit deutschen Studierenden im Grundstudium durch. Ziel war es, eine entwickelte Concept-Mapping-Aufgabe zum Chemischen Gleichgewicht auf ihre Eignung zur Wissensdiagnose hin zu überprüfen. Ein Bestandteil der Untersuchung war die Beurteilung des computerbasierten Verfahrens durch die Studierenden und die Analyse des Einflusses der Beurteilung auf die Leistungen beim Vor- und Nachtest-Mapping (Stracke, 2004, S. 136 f.). Die Leistung beim Mapping wurde über die Anzahl der Relationen, die Zerklüftetheit sowie den einfachen und gewichteten Korrespondenzkoeffizienten⁵⁶ operationalisiert. Als Treatment erhielten die Probanden ein

⁵⁵ Die Probanden sollten am Ende der Studie anhand einer offenen Frage einschätzen, wie sie das Erstellen der Maps fanden. Die Antworten wurden in positiv und negativ eingeteilt. Anschließend wurden drei Gruppen gebildet: mehr positive Antworten, mehr negative Antworten und gleich viele positive wie negative Antworten. Es zeigten sich keine Unterschiede der drei Gruppen in Bezug auf die Leistung. Dabei wurden keine inhaltlichen Unterschiede hervorgehoben. Antworten waren bspw. hilfreich, einfach oder spaßig, was streng genommen verschiedene Kategorien anspricht.

⁵⁶ Eine Korrespondenzanalyse gibt allgemein Aufschluss über den Grad an Übereinstimmung zwischen Concept Maps. Der einfache Korrespondenzkoeffizient berechnet sich aus der Differenz zwischen Übereinstimmungen und Nicht-Übereinstimmungen im Verhältnis zur maximalen Anzahl an Verknüpfungen. Der gewichtete Korrespondenzkoeffizient beachtet die Chance eines verknüpfungsfreudigen Probanden

Lernangebot in Form einer Web-Vorlesung zum Chemischen Gleichgewicht. Der eingesetzte Fragebogen zur Bewertung des Concept Mapping wurde in Anlehnung an Eckert (1998) und Roth (1994) entwickelt. Er beinhaltete drei Bereiche, Interessantheit, Nützlichkeit und Handhabbarkeit der Methode, mit einer vierstufigen Skala von 1 = „trifft nicht zu“ bis 4 = „trifft voll zu“ (Stracke, 2004, S. 151 f.). Die Studierenden schätzten das Concept Mapping insgesamt als interessant ein ($M=2,85$; $SD=0,25$). Auch die Handhabbarkeit ($M=3,07$, $SD=0,42$) sowie die Nützlichkeit ($M=2,89$; $SD=0,56$) wurden positiv bewertet. Im Hinblick auf den Einfluss der Bewertung des Concept Mapping auf die Leistung im Vortest-Mapping zeigte sich ein signifikanter Zusammenhang zwischen der Handhabbarkeit und allen vier Concept-Mapping-Parametern. Die Nützlichkeit zeigte lediglich eine signifikante Korrelation mit der Anzahl der Relationen. Beim Nachtest-Mapping zeigte sich ein leicht verändertes Bild. Ein signifikanter Zusammenhang der Handhabbarkeit war lediglich mit der Anzahl der Relationen und zum gewichteten Korrespondenzkoeffizienten nachweisbar. Für eine weiterführende Analyse wurden die Teilnehmer durch einen Median-Split eingeteilt in Personen, welche die Handhabbarkeit als gut bzw. schlecht bewerteten. Eine Korrelation mit den Mapping-Parametern des Vortests zeigte lediglich signifikante Zusammenhänge für die Probanden, welche die Handhabbarkeit als gut einschätzten. Demnach fördert ein problemloser Umgang mit der Methode ein gutes Abschneiden im Vortest, eine niedrig eingeschätzte Handhabbarkeit hat jedoch keinen negativen Einfluss auf die Vortestleistung. Stracke (2004, S. 211) betont in diesem Zusammenhang die Bedeutung eines intensiveren Trainings der Methode vorab, insofern sich ein Zusammenhang der Einschätzung zur Handhabbarkeit mit der Leistung bei Probanden zeigen würde, welche die Handhabbarkeit als schlecht bewerteten. Die Einschätzungen zur Interessantheit und zur Nützlichkeit des Concept Mapping zeigten keinen oder nur einen geringen Zusammenhang mit der Leistung beim Mapping. Stracke (2004, S. 211) schlussfolgerte, dass sich eine geringe Motivation sowie eine skeptische Sicht in Bezug zum Nutzen des Concept Mapping somit nicht negativ auf die Mapping-Ergebnisse auswirken.

gegenüber einem vorsichtigen Probanden mehr Treffer zu erzielen und berücksichtigt daher die Verknüpfungsdichte von Probandennetzwerken im Verhältnis zum Expertennetz. Die genaue Berechnung kann bei Stracke (2004, S. 71) nachgelesen werden.

In der Studie von Ryssel und Fürstenau (2011) wurde bei den Probanden der Concept-Mapping-Gruppe im Nachtest ein Fragebogen zur Bewertung des Concept Mapping eingesetzt, der sich an den Fragebogen von Stracke (2004) anlehnte. Insgesamt zeigte sich kein signifikanter Zusammenhang der Interessantheit, der Nützlichkeit sowie der Handhabbarkeit mit der Qualität der Maps (Ryssel, Fürstenau & Thieme, 2018).

3.2.3 Studien zur Lernwirksamkeit von Feedback zu Concept Maps

Nachfolgend werden Ergebnisse zu Studien vorgestellt, die Concept Maps in Verbindung mit verschiedenen Feedback-Arten vergleichen. Bisher wurde wenig erforscht, welches Feedback als Ergänzung zum Concept Mapping am lernwirksamsten ist. Mehrheitlich verwenden die Untersuchungen computergestütztes Concept Mapping und verbinden dieses mit automatisiertem Feedback (z. B. Hwang u. a., 2011; Ifenthaler, 2010; Chang u. a., 2002). Einen Überblick über computergestützte Concept-Mapping-Tools, welche automatisiertes Feedback verwenden, geben Álvarez-Montero, Jacobo-García und Rocha-Ruiz (2015). An dieser Stelle wird auch hervorgehoben, dass papierhaft erstellte Concept Maps keine direkte Evaluation ermöglichen (Liu, Chen, & Chang, 2010; Yang, 2015, S. 274).

Zunächst werden Studien vorgestellt, die verschiedene Präsentationsformate für das Feedback (Wu u. a., 2012) bzw. für Concept Maps und Feedback (Yang, 2015) vergleichen.

In der Studie von Yang (2015) mit taiwanesischen Studierenden wurde die Lernwirksamkeit von *computergestütztem Concept Mapping* mit automatisiertem Feedback (Experimentalgruppe) im Vergleich zu *papierbasiertem Concept Mapping* mit Feedback durch die Lehrperson (Kontrollgruppe) analysiert. Eingesetzt wurde das Concept Mapping als Unterstützung beim Erstellen von Zusammenfassungen in englischer Sprache zur Verbesserung des Leseverständnisses. Nach einer acht wöchigen Trainingsphase (zwei Stunden pro Woche), in der die Lernenden angeleitet wurden, wie zentrale bzw. unterstützende Ideen aus einem Text herausgearbeitet werden, erfolgte die Erstellung eines ersten Entwurfs für eine Zusammenfassung. Anschließend wurde mit den Probanden der Experimentalgruppe weitere acht Wochen das computergestützte Identifizieren zentraler Konzepte aus einem Text, das Überführen in Concept Maps sowie dessen automatische Bewertung durch das System geübt. In der anschließenden Treatmentphase wurden die Ler-

nenden der Experimentalgruppe instruiert, die wichtigsten Konzepte aus ihrer ersten Zusammenfassung herauszuarbeiten, welche anschließend automatisch in ein Concept Map⁵⁷ überführt wurden. Das Concept Map wurde automatisch bewertet. Hierbei wurden die Maps der Lernenden mit einem Referenzmap⁵⁸ verglichen. Das Ergebnis wurde den Probanden als Feedback (KP) mitgeteilt. Des Weiteren konnten die Probanden auf das Referenzmap zugreifen (KCR), um es mit dem eigenen Map zu vergleichen. Auf dessen Basis wurde die finale Zusammenfassung erstellt. Diese wurde wiederum bewertet und das Ergebnis rückgemeldet, so dass die Lernenden ihre Leistung mit dem ersten Versuch vergleichen konnten. In der Kontrollgruppe erhielten die Lernenden im Anschluss an das Erstellen der ersten Zusammenfassung ein Training zum Erstellen von papierbasierten Concept Maps mit einem allgemeinen Lehrerfeedback. Im Anschluss an diese Übungsphase wurde ein Concept Map aus der ersten Zusammenfassung erstellt, vom Lehrer bewertet und das Ergebnis rückgemeldet. Dies diente der Erstellung der finalen Zusammenfassung. Auch hier konnten die Rückmeldungen mit dem Feedback zur ersten Zusammenfassung verglichen werden. Insgesamt zeigte die Experimentalgruppe im Leseverständnis-Test einen größeren Wissenszuwachs vom Vor- zum Nachtest. Des Weiteren verbesserte sich die Qualität der zweiten gegenüber der ersten Zusammenfassung bei der Experimentalgruppe stärker als bei der Kontrollgruppe (Yang, 2015). Die Ergebnisse sind jedoch kritisch zu hinterfragen. Die Probanden der Kontrollgruppe hatten nicht die Möglichkeit ihre Maps mit dem Referenzmap zu vergleichen und somit auf Basis der richtigen Lösung eine zweite Zusammenfassung zu erstellen. Des Weiteren erstellten die Probanden der Kontrollgruppe die Maps selbst, wohingegen das Mapping bei der Experimentalgruppe eher eine Art Lückenmap darstellte. Zum Lehrerfeedback werden in der Studie nur wenige Informationen gegeben. Es ist zu hinterfragen, inwieweit die Lernenden detaillierte Informationen über den Anteil richtiger Konzepte bzw. Verbindungen erhalten haben. Insgesamt ist daher anzuzweifeln, dass die besseren Ergebnisse der Experimentalgruppe auf das computergestützte Mapping sowie auf das automatisierte Feedback zurückzuführen sind.

⁵⁷ Hierbei entstanden dreischichtige Concept Maps bestehend aus dem zentralen Konzept in der Mitte, welches die zentrale Idee des Textes widerspiegelte. Die zweite Ebene stellten die Hauptideen dar, welche die Quintessenz der jeweiligen Absätze repräsentieren sollten. Die Anzahl der Konzepte für die Hauptaussagen richtete sich dabei nach der Anzahl der Absätze. Als dritte Ebene wurden den Hauptideen unterstützende Konzepte untergeordnet.

⁵⁸ Das Referenzmap wurde mithilfe des Tools Wikipedia Miner erstellt. Insgesamt konnten 100 Punkte erreicht werden. Die Konzepte der drei Ebenen erhielten hierbei eine unterschiedliche Gewichtung.

Wu, Hwang, Milrad, Ke und Huang (2012) führten eine vierwöchige Studie (3 Std. pro Woche) mit taiwanesischen Studierenden der Gesundheitspflege zum computergestützten Lernen anhand von Concept Maps mit Feedback durch. Hierbei wurde zwischen einer Experimentalgruppe (computergestütztes Concept Mapping und sofortiges computergestütztes Feedback) und einer Kontrollgruppe (computergestütztes Concept Mapping und verzögertes papierbasiertes Feedback) unterschieden. Nach einem lehrerzentrierten Unterricht zum Thema Herzversagen sowie einem computergestützten Concept Mapping-Training wurde ein Vortest durchgeführt. Anschließend lernten die Probanden in einer computergestützten Lernumgebung. Dabei erstellten die Lernenden computergestützte Maps anhand einer Liste von Konzepten. Zusätzlich konnte ergänzendes Lernmaterial abgerufen werden. Die Maps konnten überarbeitet sowie die verschiedenen Versionen miteinander verglichen werden. Sofern beim 2. Map keine Verbesserung erkennbar war, wurde der Status abgespeichert und Unterstützung durch die Lehrkraft angeboten. Die erstellten Maps der Experimentalgruppe wurden sofort bewertet und das Ergebnis vom System rückgemeldet. Das Feedback bestand aus der erreichten Punktzahl für das Map⁵⁹ (KP) sowie Hinweisen zu falschen oder fehlenden Konzepten und Relationen (KM und KH). Die Lernenden der Kontrollgruppe erhielten papierbasiertes Feedback erst vor der Durchführung des Nachtests. Zwischen dem Erstellen der Maps und dem Feedback lag somit eine Woche. Die Probanden der Experimentalgruppe erzielten signifikant bessere Ergebnisse im Nachtest als die Probanden der Kontrollgruppe. Darüber hinaus bewertete die Experimentalgruppe die Concept Mapping-Umgebung als geeigneter zum Lernen und auch als nützlicher. Das sofortige Feedback wirkt somit motivierender für die Lernenden, so dass die Lernumgebung eher akzeptiert wird und das Lernen besser unterstützt (Wu u. a., 2012). In den Beschreibungen der Studie wurden keine Angaben zum papierbasierten Feedback gemacht. Höchstwahrscheinlich wurde den Lernenden lediglich die erreichte Punktzahl und ggf. noch ein paar kurze Hinweise zu den Fehlern schriftlich mitgeteilt. Somit fand das Feedback bei der Experimentalgruppe viel elaborierter statt als bei der Kontrollgruppe. Die Lernenden der Experimentalgruppe erhielten während der Über-

⁵⁹ Jedes individuelle Map wurde mit einem Expertenmap, welches zuvor durch kompetente Lehrkräfte erstellt wurde, verglichen. Dem Expertenmap wurde eine maximal zu erreichende Punktzahl zugewiesen. Dabei erhielt jede Proposition eine bestimmte Punktzahl, wobei höhere Hierarchieebenen auch mehr gewichtet wurden. Sofern z. B. eine Relation falsch beschriftet wurde, konnte lediglich die Hälfte der Punktzahl für die Proposition erreicht werden. Die Gesamtpunktzahl für das individuelle Map wurde mit der Gesamtpunktzahl für das Expertenmap ins Verhältnis gesetzt (Wu, Hwang, Milrad, Ke, & Huang, 2012, S. 221).

arbeitung mehrmals Rückmeldung vom System und konnten die Maps besser durchdenken. Die Kontrollgruppe erhielt das Feedback lediglich einmal. Insgesamt ist es also fraglich, inwieweit die Gruppenunterschiede lediglich auf das unterschiedliche Präsentationsformat und den Zeitpunkt der Feedbackgabe zurückzuführen sind.

Nachfolgende Studien von Chang, Sung und Chen (2001; 2002) kombinierten *unterschiedlich stark vorstrukturierte Concept Maps* mit dem Geben von Feedback. Die Untersuchung von Chang u. a. (2002) wurde mit taiwanesischen Grundschulern durchgeführt und setzte Concept Maps als ergänzende Strategie zum Lernen aus Texten ein. Über einen Zeitraum von sieben Wochen sollten die Inhalte von sieben Texten durchgearbeitet werden. Hierbei wurde zwischen drei Experimentalgruppen (Selbstkonstruktion, Map-Korrektur⁶⁰ und Scaffold-Fading⁶¹) und einer Kontrollgruppe (nur Text-Verstehen) unterschieden. Den Experimentalgruppen wurde zusätzlich KP-Feedback entsprechend der Art der Concept Maps zur Verfügung gestellt: Prozentanteil der richtig gewählten Konzepte und Relationen (Selbstkonstruktion), Prozentanteil der im Vergleich zum Expertenmap richtigen Korrekturen (Map-Korrektur) sowie Prozentanteil richtig ausgefüllter Lücken bzw. Feedback analog zur Gruppe Selbstkonstruktion. Die Ergebnisse der Studie zeigten die signifikante Überlegenheit der Map-Korrektur gegenüber den anderen Versuchsgruppen sowie der Kontrollgruppe im Hinblick auf Textverständnis und Textzusammenfassungen. Chang u. a. (2002) betonen in diesem Zusammenhang die strukturellen Vorteile von Expertenmaps, welche bei der Map-Korrektur beibehalten werden, ohne die Lernenden jedoch kognitiv zu überlasten. Die Studie gibt keine Hinweise darauf, inwieweit das Feedback Einfluss auf die Ergebnisse hatte. Hilbert, Nückles und Matzel (2008) verglichen ihre Studie mit der von Chang u. a. (2002) und konnten die Überlegenheit der Map-Korrektur gegenüber der Selbstkonstruktion und dem Durcharbeiten von Expertenmaps nicht nachweisen. Sie geben das fehlende Feedback als einen möglichen Grund an.

⁶⁰ Die Map-Korrektur-Gruppe erhielt abgewandelte Expertenmaps, welche 30-40 Prozent fehlerhafte Konzepte und Relationen beinhalteten (Chang, Sung, & Chen, 2002, S. 10).

⁶¹ Bei der Scaffold-Fading-Gruppe wurde die Unterstützung im Laufe der Zeit zurückgenommen. In der ersten Lerneinheit wurde den Probanden ein Expertenmap zur Verfügung gestellt. In der zweiten Lektion wurde das vollständige Expertenmap mit Lücken präsentiert, welches die Probanden vervollständigen sollten. Die dritte Lerneinheit beinhaltet ein unvollständiges Expertenmap, welches vervollständigt werden sollte. In der vierten und fünften Lerneinheit wurden die Probanden instruiert, selbständig Concept Maps zu den Themen zu erstellen. Lerneinheit sechs und sieben enthielt keinerlei Unterstützung mehr. Die Lernenden arbeiteten lediglich die Texte durch (Chang u. a., 2002, S. 10 f.).

Chang, Sun, Chen (2001) führten eine Studie mit Schülern einer taiwanesischen Junior High School zum Lernen über das Thema Fortpflanzung mit Hilfe von computergestütztem Concept Mapping durch. Die Schüler wurden zufällig drei Versuchsgruppen zugeteilt: Selbstkonstruktion computergestützt, Lückenmap computergestützt oder Selbstkonstruktion im Paper & Pencil-Verfahren. Das Mapping-Verfahren wurde im Anschluss an den regulären Biologieunterricht eingesetzt. Beide computergestützten Gruppen erhielten Feedback. Zum einen konnten dabei Hinweise als unvollständige Propositionen (z. B. „Meiose führt zu ???“) abgerufen werden. Zum anderen konnten sich die Probanden die Bewertungen des eigenen Maps anzeigen lassen. Dazu wurde das eigene Map mit einem von erfahrenen Lehrern erstellten Expertenmap verglichen und die falschen Konzepte und Relationen angezeigt. Zusätzlich wurde eine Punktzahl in Anlehnung an die Scoringkriterien von Novak und Gowin (1984) angegeben, die die Übereinstimmung mit dem Expertenmap zeigte. Nach 30 Minuten Bearbeitungszeit konnten sich die Probanden des Weiteren das Expertenmap anzeigen lassen. Das Feedback für beide Gruppen war daher sehr umfassend. Die Probanden erhielten eine Rückmeldung über den Anteil der richtigen Lösungen (KP), aufgabenrelevante Hinweise (KC) sowie fehlerbezogene Informationen (KM). Schließlich war bei Bedarf auch die richtige Lösung abrufbar (KCR). Insgesamt zeigte sich ein Vorteil der Gruppe Lückenmap sowohl gegenüber der papierbasierten als auch gegenüber der computergestützten Selbstkonstruktion im Nachtest. Die beiden Selbstkonstruktionsgruppen unterschieden sich nicht signifikant. Chang u. a. (2001) schlussfolgern, dass das Vervollständigen von Lückenmaps die beste Strategie ist, da dies den Workload reduziert. Insgesamt bleibt zu hinterfragen, worauf genau die Unterschiede zurückzuführen sind. Der Unterschied zwischen beiden computergestützten Gruppen ist auf den Grad der Vorstrukturierung zurückzuführen, da beide Gruppen das gleiche Präsentationsformat verwenden und auch die gleiche Form des Feedbacks erhalten. Zwischen der Gruppe Lückenmap und der papierbasierten Selbstkonstruktion kann nicht eindeutig bestimmt werden, inwieweit die Unterschiede auf das Präsentationsformat des Concept Mapping (computergestützt vs. papierbasiert), auf das fehlende Feedback bei der papierbasierten Selbstkonstruktion oder auf den Grad der Vorstrukturierung der Maps zurückzuführen sind. Interessant ist, dass sich beide Selbstkonstruktionsgruppen im Nachtest nicht unterscheiden. Das zusätzliche Feedback scheint daher bei der Gruppe computergestützte Selbstkonstruktion nicht lernförderlich zu sein. Chang u. a. (2001) begründen dies mit der zusätzlichen kognitiven Belastung durch das Feedback. Eine weitere Ursache könnte aber auch das unterschiedliche Präsentationsformat sein. Die Probanden

waren möglicherweise mit dem Erstellen der Maps am Computer überfordert, da dies mehr Übung verlangt als das papierbasierte Erstellen der Maps. Für eine detailliertere Analyse des Präsentationsformats wäre es sinnvoll gewesen, eine Gruppe „papierbasierte Selbstkonstruktion mit Feedback“ einzubeziehen. Um herauszufinden, inwieweit das Feedback lernwirksam ist, wäre der Vergleich mit Probanden, welche lediglich die Maps am Computer erstellen bzw. vervollständigen und kein Feedback erhalten, sinnvoll gewesen. So lassen sich lediglich Vermutungen anstellen.

Nachfolgend werden Studien vorgestellt, welche Concept Mapping *mit und ohne Feedback* gegenüberstellen (Hwang u. a., 2011) bzw. verschiedene *Feedbackarten in Abhängigkeit von der Komplexität des Feedbackinhalts* (Ifenthaler, 2010) vergleichen.

Hwang, Wu und Ke (2011) führten eine Studie mit taiwanesischen Grundschulern zum Lernen der Ökologie von Schmetterlingen in einer natürlichen Lernumgebung kombiniert mit mobilem Lernen durch. Hierbei wurde zwischen einer Experimentalgruppe (interaktive Lernumgebung) und einer Kontrollgruppe (konventionelle Lernumgebung) unterschieden. Nach einer Instruktionsphase zum mobilen Lernen und dem Erstellen von Concept Maps fertigten die Probanden beider Gruppen zunächst computerbasierte Concept Maps basierend auf ihren Vorkenntnissen an. Anschließend hatten die Lernenden die Möglichkeit, Schmetterlinge in ihrer natürlichen Umgebung zu beobachten. Dabei konnten sie die zuvor erstellten Concept Maps über ein Tablet beliebig oft überarbeiten. Hierzu konnten sie im Internet recherchieren und ergänzendes Lernmaterial zu Hilfe nehmen. Die überarbeiteten Concept Maps der Experimentalgruppe wurden jeweils bewertet. Den Lernenden wurden sofort Hinweise und Kommentare zur Richtigkeit der einzelnen Propositionen rückgemeldet, ohne direkt das richtige Map zu präsentieren. Die Probanden der Kontrollgruppe erhielten keine Rückmeldung zu den erstellten und überarbeiteten Maps. Insgesamt zeigte sich eine Überlegenheit der Experimentalgruppe gegenüber der Kontrollgruppe im Nachtest. Darüber hinaus erstellten die Probanden der interaktiven Lernumgebung auch qualitativ bessere Concept Maps. Die Qualität der Maps korrelierte hoch mit den Ergebnissen des Nachtests. Darüber hinaus wurde die Nützlichkeit des Feedbacks als hoch eingeschätzt. Das sofortige Feedback im Rahmen der interaktiven Lernumgebung ermöglichte den Lernenden insgesamt gegenüber der Kontrollgruppe eine tiefgründige Auseinandersetzung mit den Lerninhalten (Hwang u. a., 2011). An dieser Stelle ist kritisch anzumerken, dass sich die Lernenden der konventionellen Lernumgebung selbst bewerten und sich dadurch internes Feedback geben. Darüber hinaus stellen

die ergänzenden Lernmaterialien sowie die Informationen, welche mittels Internetrecherche gewonnen werden können, nach Narciss (2006) Formen des elaborierten Feedbacks dar (KC). Insgesamt wurde somit eher internes und externes Feedback, jedoch nicht Feedback und kein Feedback, gegenübergestellt, wobei das Feedback zusätzlich mit KC-Feedback ergänzt wurde. Den Lernenden der Experimentalgruppe wurde zusätzlich KM- und KH-Feedback bereitgestellt.

Ifenthaler (2010) untersuchte in seiner Studie u. a. den Einfluss verschiedener Arten von Feedback nach dem Erstellen von Concept Maps und Texten auf das Lernen grundlegender Zusammenhänge zum Thema Klimaveränderungen bei deutschen Studierenden. Zunächst wurde der Vortest durchgeführt und eine Einführung zu Concept Maps sowie zur Software, welche zum Erstellen der Texte und Concept Maps genutzt wurde, gegeben. Anschließend erstellten die Lernenden zunächst ein Concept Map zum Thema Klimaveränderungen und verfassten direkt im Anschluss einen Text zu diesem Thema. Danach erhielten die Lernenden unterschiedliches vom System generiertes Feedback: „Cutaway Model-Based Feedback“, „Discrepancy Model-Based Feedback“ oder „Expert-Feedback“. Das „Cutaway Model-Based Feedback“ beinhaltet alle vom Probanden erstellten Propositionen, wobei richtige Konzepte markiert waren. Das „Discrepancy Model-Based Feedback“ beinhaltet lediglich die nicht mit der Expertenlösung übereinstimmenden Propositionen. Das „Expert-Feedback“ beinhaltet die Darstellung des Expertenmaps zum Thema. Zusätzlich zum Feedback wurde allen Probanden ein Text zum Thema zur Verfügung gestellt. Beides sollte durchgearbeitet werden. Abschließend absolvierten die Probanden den Nachtest. Es zeigten sich keine signifikanten Unterschiede im Wissenszuwachs zwischen den drei Feedbackarten (Ifenthaler, 2010). Dieses Ergebnis ist nicht sehr überraschend, wenn die drei Feedbackarten noch einmal genauer betrachtet werden. Sowohl das „Cutaway Model-Based Feedback“ als auch das „Discrepancy Model-Based Feedback“ stellen eine Form von KM-Feedback auf einem sehr niedrigen Niveau dar. Es werden zwar Informationen zu den Fehlern bzw. richtigen Propositionen gegeben, aber lediglich der Ort des Fehlers bzw. der richtigen Proposition angegeben und nichts über die Art oder Ursache der Fehler mitgeteilt. Daher ist der Unterschied zur dritten Gruppe, welcher die Expertenlösung als KCR zur Verfügung gestellt wird, möglicherweise zu gering, um Unterschiede in der Lernwirksamkeit hervorzurufen.

Abschließend lässt sich festhalten, dass bisher wenig Forschung im Rahmen des Lernens mit Concept Maps existiert, die die Wirkung von ergänzendem Feedback untersucht. In

den vorgestellten Studien wurde der Vorteil von computergestütztem im Vergleich zu papierbasiertem Feedback hervorgehoben. Des Weiteren wurden Concept Maps entsprechend des Grades ihrer Vorstrukturierung mit Feedback kombiniert und verglichen, wobei sich ein Vorteil für Lückenmaps gegenüber der Selbstkonstruktion bzw. ein Vorteil der Map-Korrektur gegenüber der Selbstkonstruktion und Expertenmaps zeigte. Es wurde lediglich eine Studie gefunden, die Feedback in Abhängigkeit von der Komplexität des Feedback-Inhalts vergleicht (Ifenthaler, 2010) und eine Studie, die Feedback und kein Feedback (Hwang u. a., 2011) gegenüberstellt. Ungenauigkeiten im Untersuchungsdesign führen jedoch dazu, dass die Versuchsgruppen bei Ifenthaler (2010) Feedback auf einem vergleichbaren Komplexitätsniveau erhalten und somit keine Unterschiede in der Lernwirksamkeit daraus entstehen konnten. Bei Hwang u. a. (2011) wurde vernachlässigt, dass sich die Probanden der Kontrollgruppe internes Feedback geben und das Bereitstellen ergänzender Lernmaterialien auch eine Form des Feedbacks darstellt. Überwiegend wird in den Studien dabei computergestütztes Concept Mapping und Feedback eingesetzt. Álvarez-Montero u. a. (2015) stellten in ihrem Literaturreview heraus, dass das Feedback im Rahmen von computergestützten Tools oft nur intuitiv gestaltet wird, ohne die Ergebnisse der Feedbackliteratur zu berücksichtigen. Darüber hinaus existieren zahlreiche Concept Mapping Tools, welche Feedback implementieren, jedoch die Wirksamkeit nicht evaluieren. Das papierbasierte Feedback als Ergänzung zum Concept Mapping wurde bisher lediglich im Rahmen von Studien im Vergleich zum computerbasierten Feedback untersucht. Methodische Defizite lassen hierbei jedoch keinen eindeutigen Schluss zu, der für einen Vorteil des computerbasierten Feedbacks spricht. Den Forschungsdefiziten soll mit der Konzeption der in den Kapiteln 4 und 5 vorgestellten Studie Rechnung getragen werden.

4 Empirische Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps – Konzeption, Erhebung und Auswertung

Im folgenden Kapitel wird die Konzeption, Erhebung und Auswertung der empirischen Untersuchung zur Lernwirksamkeit des Feedbacks zu Concept Maps vorgestellt. Hierzu werden in Kapitel 4.1 zunächst die Gestaltung der Untersuchung und die Fragestellungen, die Hypothesen, die Stichprobe sowie das Untersuchungsdesign erläutert. Kapitel 4.2 thematisiert die Gestaltung der Lernmaterialien und Interventionsmaßnahmen. Dabei wird in Kapitel 4.2.1 der Planspielunterricht mit Easy BusinessTM vorgestellt und die Gestaltung des im Rahmen der Studien eingesetzten Concept Mapping (Kapitel 4.2.2) und Feedbacks (Kapitel 4.2.3) erläutert. Die Erhebungsinstrumente werden im Kapitel 4.3 vorgestellt, wobei Kapitel 4.3.1 den Wissenstest als Paralleltest, Kapitel 4.3.2 den Sprachfähigkeitstest und Kapitel 4.3.3 den Fragebogen zur Einschätzung des Umgangs mit Concept Mapping, der Nützlichkeit des Feedbacks sowie des Umgangs mit Fehlern betrachten. Kapitel 4.4 geht auf die Auswertung des Wissenstests (Kapitel 4.4.1) und die Auswertung des Sprachfähigkeitstests (Kapitel 4.4.2) ein. Im Kapitel 4.5 wird die Auswertung der Concept Maps beschrieben. Hierzu wird zunächst eine kategoriale und strukturelle Inhaltsanalyse durchgeführt sowie eine Reliabilitätsanalyse des Codiervorgehens vorgenommen (Kapitel 4.5.1). Anschließend erfolgt eine qualitative Bewertung der Concept Maps (Kapitel 4.5.2). Im Kapitel 4.6 wird der Fragebogen zur Einschätzung des Umgangs mit Concept Mapping (Kapitel 4.6.1), des Umgangs mit Fehlern (Kapitel 4.6.2) sowie der Nützlichkeit des Feedbacks (Kapitel 4.6.3) ausgewertet. Die Ergebnisse der Auswertung der Tests, Fragebögen bzw. Concept Maps werden dabei als Überblick jeweils für alle Probanden insgesamt dargestellt. Die Gruppenvergleiche für eine Hypothesenprüfung werden erst im Kapitel 5 durchgeführt.

4.1 Gestaltung der Untersuchung

4.1.1 Fragestellungen und Hypothesen

Im Kapitel 2.2.2 wurde auf die uneinheitliche Befundlage von Studien hingewiesen, die die Lernwirksamkeit von Feedback in Abhängigkeit von der Komplexität der Feedbacknachricht untersuchen, was auf den Einfluss weiterer Faktoren schließen lässt. Dabei konnte der Vorteil des elaborierten gegenüber dem einfachen Feedback nicht eindeutig

nachgewiesen werden. Diesem Forschungsdefizit wird mit dieser Untersuchung begegnet. Im Rahmen der empirischen Untersuchung wird überprüft, inwieweit elaboriertes Feedback lernwirksamer ist als einfaches Feedback. Darüber hinaus wird der Einfluss verschiedener Faktoren untersucht. Dabei werden der Einfluss des Vorwissens und der Sprachfähigkeit als Komponenten des informellen Status des Feedbackempfängers aus dem erweiterten Rahmenmodell (vgl. Kapitel 2.1.4.4) analysiert. Ebenso wird untersucht, wie die Feedbackempfänger allgemein mit Fehlern umgehen und welchen Einfluss das auf die Lernwirksamkeit des Feedbacks hat. Des Weiteren wird die eingeschätzte Nützlichkeit des Feedbacks als Einflussfaktor erhoben, über die die zugeschriebene Kompetenz des Feedbacksenders operationalisiert werden kann (vgl. Kapitel 2.2.4). Ferner werden die Aufgabenanforderungen der Lernaufgabe als situativer Faktor analysiert.

Als Aufgabenformat für das Feedback dient die Methode des Concept Mapping. Im Kapitel 3.2.2 wurde gezeigt, dass die Lernwirksamkeit von Concept Maps ebenfalls von verschiedenen Faktoren wie Vorwissen, Sprachfähigkeit oder der Einstellung zur Methode des Concept Mapping abhängt. Die Einschätzung der Methode des Concept Mapping wird im Rahmen dieser Untersuchung ebenfalls erhoben. Zur Wirkung von ergänzendem Feedback wurde bisher wenig geforscht (vgl. Kapitel 3.2.3). Insbesondere die Notwendigkeit eines ergänzenden Feedbacks zu Concept Maps sowie der Einfluss des Vorwissens oder der verbalen Fähigkeiten auf die Lernwirksamkeit eines zusätzlichen Feedbacks wurden dabei bisher nicht ausreichend untersucht.

Somit ergeben sich folgende Forschungsfragen:

- F1) Inwieweit ist ergänzendes Feedback zu Concept Maps lernwirksamer als kein ergänzendes Feedback?
- F2) Ist elaboriertes Feedback lernwirksamer als einfaches Feedback?
- F3) Inwieweit beeinflussen das Vorwissen, die Sprachfähigkeit, die Einschätzung des Umgangs mit Concept Mapping, die Einschätzung der Nützlichkeit des Feedbacks, die Einschätzung des Umgangs mit Fehlern sowie die Aufgabenanforderungen der Lernaufgabe die Lernwirksamkeit des Feedbacks?

Zur Untersuchung der Forschungsfragen werden die Hypothesen H1-H39 formuliert. Diese werden nachfolgend beschrieben.⁶²

Hattie hat das Feedback als einen der größten Einflussfaktoren auf das Lernen identifiziert (vgl. Kapitel 2.2.1). Des Weiteren hat er auch dem Concept Mapping einen großen Einfluss zugeschrieben (vgl. Kapitel 3.2.1). Insgesamt stellt sich die Frage, inwieweit das Erstellen von Concept Maps allein genügt oder ob es eines zusätzlichen Feedbacks zu den in den Concept Maps dargestellten Inhalten bedarf. Dies soll in der Hypothese H1 überprüft werden. Um beurteilen zu können, inwieweit Feedback lernwirksamer als kein Feedback ist, werden in einem ersten Schritt die Gruppen *Einfaches Feedback [EiFB]* und *Elaboriertes Feedback [ElFB]* zur Gruppe *Feedback [FB]* zusammengefasst. Des Weiteren werden die Gruppen *EiFB* und *ElFB* jeweils *keinem FB* gegenübergestellt, um herauszufinden, auf welche Art des Feedbacks mögliche Unterschiede zurückzuführen sind.

H1: *FB* ist lernwirksamer als *kein FB*.

- a) *EiFB* ist lernwirksamer als *kein FB*.
- b) *ElFB* ist lernwirksamer als *kein FB*.

In der Feedbackforschung existieren zahlreiche Studien, welche Feedback unterschiedlicher Komplexität gegenüberstellen. Die Befundlage hierzu ist nicht eindeutig. Es zeigt sich kein genereller Vorteil des elaborierten gegenüber dem einfachen Feedback (vgl. Kapitel 2.2.2). Die Hypothese H2 analysiert mögliche Einflüsse der Komplexität des Feedbackinhalts.

H2: *ElFB* ist lernwirksamer als *EiFB*.

⁶² Im Kontext der Studie wird bewusst zunächst Feedback und kein Feedback als Ausgangssituation miteinander verglichen. Anschließend wird jeweils untersucht, welche Art des Feedbacks in Abhängigkeit von der Komplexität des Feedbackinhalts lernwirksamer ist. Bei der Untersuchung des Einflusses verschiedener Faktoren wird analog verfahren. Die Hypothesen werden in thematischen Blöcken gegliedert. Dabei weicht die Reihenfolge der Nummerierungen innerhalb der Blöcke teilweise von der Reihenfolge der ergänzenden Ausführungen ab. Dies liegt in dem Schwerpunkt der Arbeit, der auf die Analyse der Lernwirksamkeit des Feedbacks gerichtet ist, begründet. Die Feedbackforschung stellt die lernförderliche Wirkung des Feedbacks nicht in Frage und geht daher von einer Lernwirksamkeit des Feedbacks allgemein aus, lediglich die Gestaltung wird untersucht. Daher wird die Argumentation aus dem Kapitel 2 zu den theoretischen Grundlagen und empirischen Befunden zur Lernwirksamkeit des Feedbacks in Lehr-Lernprozessen bei der Beschreibung der Hypothesen für diese Studie oftmals vorangestellt. Anschließend werden diese um Erkenntnisse aus dem Kapitel 3 zu den theoretischen Grundlagen und empirischen Befunden zur Lernwirksamkeit von Concept Maps in Lehr-Lernprozessen ergänzt.

Die uneinheitliche Befundlage hinsichtlich der Komplexität des Feedbackinhalts lässt auf vielschichtige Einflussfaktoren auf die Lernwirksamkeit des Feedbacks schließen. Des Weiteren hat auch Hattie eine hohe Variabilität der Effektstärken identifiziert (vgl. Kapitel 2.2.2). Als personale Variablen des Feedbackempfängers werden im Rahmen der Studie der informelle Status (Vorwissen und Sprachfähigkeit) sowie der eingeschätzte Umgang mit Fehlern analysiert (vgl. Kapitel 2.1.4.4; Erweitertes Rahmenmodell, Abbildung 6).

Der informelle Status wurde als zentraler Einflussfaktor auf die Lernwirksamkeit des Feedbacks in zahlreichen Studien identifiziert (vgl. Kapitel 2.2.4). Zum Teil zeigt sich dabei ein umgekehrt propositionaler Zusammenhang zwischen dem Vorwissensniveau und der Komplexität des Feedbackinhalts (z. B. Krause, 2007). Elaboriertes Feedback führt zu einer Angleichung des Wissensniveaus von vorwissenschwachen an das von vorwissensstarken Lernenden, wobei Vorwissensstarke die zusätzlichen Informationen nicht benötigen. Die Befundlage ist hierbei jedoch nicht einheitlich, es besteht weiterer Forschungsbedarf. Im Rahmen der Hypothese H4 wird daher behauptet, dass Probanden mit einem *niedrigen Vorwissensniveau* vom elaborierten Feedback profitieren, wohingegen die Hypothese H6 postuliert, dass sich bei Probanden mit einem *hohen Vorwissensniveau* kein Vorteil des elaborierten gegenüber dem einfachen Feedback zeigt. In den im Kapitel 2.2.4 erläuterten Studien wird dabei nicht analysiert, inwieweit Probanden mit hohem Vorwissen überhaupt ein Feedback benötigen. Gerade beim Concept Mapping als Aufgabenformat stellt sich jedoch die Frage, ob bei den Lernenden mit hohem Vorwissen nicht schon allein das Auseinandersetzen mit den Lerninhalten während des Erstellens der Maps für den nötigen Wissenserwerb ausreicht. Diese Vermutungen bestätigen auch die Ergebnisse von Studien, welche den Einfluss des Vorwissens auf die Lernwirksamkeit von Concept Maps untersucht haben (vgl. Kapitel 3.2.2). Hierbei zeigte sich, dass insbesondere Lernende mit hohem Vorwissen vom Durcharbeiten der Maps profitieren (O'Donnell & Dansereau, 2000) bzw. Maps mit besserer Qualität (Ryssel & Fürstenau, 2011) erstellen. Möglicherweise ist die kognitive Belastung beim Erstellen der Maps für Lernende mit niedrigem Vorwissen aufgrund der Neuheit und Schwierigkeit der Methode zu hoch, weshalb ergänzendes Feedback das Lernen unterstützen kann. Das Erstellen der Maps stellt hohe kognitive Anforderungen an Lernende mit niedrigem Vorwissen. Daher sollte jegliche Art des Feedbacks, unabhängig von der Komplexität des Feedbackinhalts, lernwirksamer sein als kein Feedback, was Hypothese 3 postuliert. Lernende mit hohem

Vorwissen können dagegen diese Defizite kompensieren und benötigen kein zusätzliches Feedback. Daher wird in Hypothese 5 behauptet, dass sich *kein FB* und *FB* bei *hohem Vorwissen* nicht unterscheiden.

H3: Bei Probanden mit *niedrigem Vorwissen* ist *FB* lernwirksamer als *kein FB*.

- a) Bei Probanden mit *niedrigem Vorwissen* ist *EiFB* lernwirksamer als *kein FB*.
- b) Bei Probanden mit *niedrigem Vorwissen* ist *ElFB* lernwirksamer als *kein FB*.

H4: Bei Probanden mit *niedrigem Vorwissen* ist *ElFB* lernwirksamer als *EiFB*.

H5: Bei Probanden mit *hohem Vorwissen* unterscheiden sich *FB* und *kein FB* nicht.

- a) Bei Probanden mit *hohem Vorwissen* unterscheiden sich *EiFB* und *kein FB* nicht.
- b) Bei Probanden mit *hohem Vorwissen* unterscheiden sich *ElFB* und *kein FB* nicht.

H6: Bei Probanden mit *hohem Vorwissen* unterscheiden sich *ElFB* und *EiFB* nicht.

Sprache dient nicht nur als Medium für die Kommunikation fachlicher Inhalte und hat somit eine kommunikative Funktion, sondern stellt auch eine notwendige Voraussetzung für den Wissenserwerb dar. Sprache hat in diesem Zusammenhang eine kognitive Funktion. Sie dient als Werkzeug zur Erkenntnisgewinnung, ohne das eine Wissenserweiterung und –strukturierung nicht möglich ist (De Bleser, 2012, S. 424). Somit kann angenommen werden, dass eine ausreichende Sprachfähigkeit im Kontext der Studie notwendig ist, um die im Planspiel erlernten betriebswirtschaftlichen Inhalte im Rahmen des Erstellens der Concept Maps angemessen abzubilden. In diesem Zusammenhang sollte ein Feedback die kognitive Verarbeitung bei Lernenden mit niedriger Sprachfähigkeit unterstützen. In der Hypothese H7 wird daher postuliert, dass *FB* gegenüber *keinem FB* bei niedriger Sprachfähigkeit lernwirksamer ist. Dabei sollte sich ein Vorteil des elaborierten Feedbacks aufgrund der umfangreichen Informationen gegenüber dem Präsentieren der richtigen Lösung als einfaches Feedback zeigen. Daher wird in Hypothese H8 angenommen, dass bei *niedriger Sprachfähigkeit* *ElFB* lernwirksamer als *EiFB* ist. Im Gegensatz dazu sollte ein Feedback bei Lernenden mit hoher Sprachfähigkeit nicht notwendig sein, da sie ihren eigenen sprachlichen Fähigkeiten vertrauen und somit in der Lage sind, sich mit den erlernten Inhalten im Rahmen der Maperstellung intensiv auseinanderzusetzen, ohne ein Feedback zu erhalten. In der Hypothese H9 wird daher angenommen, dass sich *FB* und *kein FB* bei hoher Sprachfähigkeit nicht unterscheiden. Daher sollte es auch keine Unterschiede zwischen einfachem und elaboriertem Feedback geben. Hypothese H10 postuliert daher, dass sich die Gruppen *ElFB* und *EiFB* bei *hoher Sprach-*

fähigkeit nicht unterscheiden. An dieser Stelle kann nicht auf Erkenntnisse aus vorliegenden Untersuchungen zur Lernwirksamkeit von Feedback zurückgegriffen werden, da sich bisherige Studien zum Einfluss der Sprachfähigkeit lediglich auf den Unterschied zwischen explizitem und implizitem Feedback beziehen. Diese Untersuchung fokussiert explizites Feedback (vgl. Kapitel 2.2.4). Es lassen sich jedoch Bezüge zu Studien, die den Einfluss von verbalen Fähigkeiten auf das Lernen mit Concept Maps, untersuchen, herstellen (vgl. Kapitel 3.2.2). Einige Studien bestätigten dabei den Vorteil des Concept Mapping gegenüber traditioneller Instruktion insbesondere für Lernende mit geringen verbalen Fähigkeiten, wobei sich für Lernende mit hohen verbalen Fähigkeiten der Vorteil nicht bestätigte. Stensvold und Wilson (1990) begründen dies anhand der „Aptitude Treatment Interaction“. Lernenden mit niedrigen verbalen Fähigkeiten sollte eine zusätzliche Unterstützung, z. B. als Feedback angeboten werden, wohingegen Lernenden mit hohen verbalen Fähigkeiten Freiheitsgrade bei der Steuerung ihres Lernens zur Verfügung gestellt werden sollten. Zusätzliches Feedback sollte daher bei einer hohen Sprachfähigkeit keine Wirkung zeigen.

H7: Bei Probanden mit *niedriger Sprachfähigkeit* ist *FB* lernwirksamer als *kein FB*.

a) Bei Probanden mit *niedriger Sprachfähigkeit* ist *EiFB* lernwirksamer als *kein FB*.

b) Bei Probanden mit *niedriger Sprachfähigkeit* ist *ElFB* lernwirksamer als *kein FB*.

H8: Bei Probanden mit *niedriger Sprachfähigkeit* ist *ElFB* lernwirksamer als *EiFB*.

H9: Bei Probanden mit *hoher Sprachfähigkeit* unterscheiden sich *FB* und *kein FB* nicht.

a) Bei Probanden mit *hoher Sprachfähigkeit* unterscheiden sich *EiFB* und *kein FB* nicht.

b) Bei Probanden mit *hoher Sprachfähigkeit* unterscheiden sich *ElFB* und *kein FB* nicht.

H10: Bei Probanden mit *hoher Sprachfähigkeit* unterscheiden sich *ElFB* und *EiFB* nicht.

Studien zum Einfluss des eingeschätzten Umgangs mit Fehlern auf die Lernwirksamkeit des Feedbacks sind im Kontext der Studie nicht bekannt. Dennoch wurde im Kapitel 2.1.1 auf die Bedeutung einer positiven Fehlerkultur hingewiesen. Hierbei gilt es zwei Grundbedingungen zu beachten: eine Lernorientierung sowie ein positives Unterrichtsklima. Von den Lernenden sollte im Rahmen der Lernorientierung eine Fehlerbereitschaft ausgehen, d. h. sie sollten Fehler als bedeutsam einschätzen und konstruktiv mit ihnen umgehen. Der eingeschätzte Umgang mit Fehlern wurde daher als Einflussfaktor in das erweiterte Rahmenmodell (vgl. Kapitel 2.1.4.4) aufgenommen. Lernende, welche den Umgang mit Fehlern als nicht bzw. wenig konstruktiv einschätzen, sind auch nicht bereit

Feedback zu nutzen, um den Lernprozess zu optimieren. Die Hypothese H11 postuliert, dass bei Probanden mit einem *wenig konstruktiven Umgang mit Fehlern* sich *FB* und *kein FB* nicht unterscheiden. Die Art des Feedbacks ist dabei unbedeutend, Hypothese H12 behauptet daher, dass sich *ElFB* und *EiFB* nicht unterscheiden. Lernende mit einem konstruktiven Umgang mit Fehlern dagegen nutzen das Feedback aktiv, um ihren Lernprozess zu optimieren. Hypothese H13 postuliert daher, dass *FB* bei Probanden, welche den Umgang mit Fehlern als *konstruktiv* einschätzen, lernwirksamer ist als *kein FB*. Das elaborierte Feedback im Rahmen dieser Untersuchung beinhaltet u. a. die Besprechung eines Schülernetzes. An dieser Stelle werden die Fehler detailliert analysiert. Lernende mit einer hohen Fehlerbereitschaft sind eher bereit, diese Informationen für eine Anpassung des Lernprozesses zu nutzen, auch wenn es fremde Fehler sind. Hypothese H14 nimmt folglich an, dass *ElFB* lernwirksamer ist als *EiFB*.

H11: Bei Probanden, welche den *Umgang mit Fehlern* als *nicht bzw. wenig konstruktiv* einschätzen, unterscheiden sich *FB* und *kein FB* nicht.

- a) Bei Probanden, welche den *Umgang mit Fehlern* als *nicht bzw. wenig konstruktiv* einschätzen, unterscheiden sich *EiFB* und *kein FB* nicht.
- b) Bei Probanden, welche den *Umgang mit Fehlern* als *nicht bzw. wenig konstruktiv* einschätzen, unterscheiden sich *ElFB* und *kein FB* nicht.

H12: Bei Probanden, welche den *Umgang mit Fehlern* als *nicht bzw. wenig konstruktiv* einschätzen, unterscheiden sich *ElFB* und *EiFB* nicht.

H13: Bei Probanden, welche den *Umgang mit Fehlern* als *konstruktiv* einschätzen, ist *FB* lernwirksamer als *kein FB*.

- a) Bei Probanden, welche den *Umgang mit Fehlern* als *konstruktiv* einschätzen, ist *EiFB* lernwirksamer als *kein FB*.
- b) Bei Probanden, welche den *Umgang mit Fehlern* als *konstruktiv* einschätzen, ist *ElFB* lernwirksamer als *kein FB*.

H14: Bei Probanden, welche den *Umgang mit Fehlern* als *konstruktiv* einschätzen, ist *ElFB* lernwirksamer als *EiFB*.

Die Lernwirksamkeit des *Feedbacks* hängt möglicherweise auch von der Einschätzung zum *Umgang mit Concept Mapping* ab. Probanden, welche die *Nützlichkeit, Interessantheit und Handhabbarkeit* des *Concept Mapping* als gering einschätzen, haben höchst-

wahrscheinlich noch Schwierigkeiten im Umgang mit der Methode. Eine niedrige Einschätzung kann auch dann entstehen, wenn die Lernenden der Meinung sind, dass das ergänzende Concept Mapping für einen Wissenserwerb nicht notwendig ist. Infolgedessen ist die zusätzliche kognitive Belastung sehr hoch bzw. unnötig. Da das Feedback eng mit dem Aufgabenformat Concept Mapping verbunden ist – das einfache Feedback stellt z. B. das Durcharbeiten des Referenzmaps dar – kann zusätzliches Feedback das Lernen somit sogar behindern. Die Hypothesen H 15 und H 19 postulieren daher, dass bei Probanden, welche die *Interessantheit und Nützlichkeit* bzw. die *Handhabbarkeit* des Concept Mapping als *niedrig* einschätzen, *FB* weniger lernwirksam als *kein FB* ist. Je komplexer das Feedback dabei wird, desto weniger lernwirksam sollte es für Lernende mit einer niedrigen Einschätzung der Methode sein. Elaboriertes Feedback sollte weniger lernwirksam sein als einfaches Feedback (H16 und H20). Eine hohe Einschätzung der Interessantheit und Nützlichkeit sowie der Handhabbarkeit hat somit möglicherweise einen positiven Effekt auf die Lernwirksamkeit des Feedbacks. Die Lernenden sind dann eher in der Lage mit der zusätzlichen kognitiven Belastung durch das Feedback umzugehen. Je komplexer das Feedback ist, desto lernwirksamer sollte es somit für Lernende mit einer hohen Einschätzung sein. *Feedback* sollte folglich bei Probanden, welche die *Interessantheit und Nützlichkeit* sowie die *Handhabbarkeit* des Concept Mapping als *hoch* einschätzen lernwirksamer sein als *kein Feedback* (H17 und H21). Des Weiteren sollte *ElFB* lernwirksamer sein als *EiFB* (H18 und H22).

Bisherige Forschung konnte keinen eindeutigen Zusammenhang zwischen der Einschätzung der Concept-Mapping-Technik und der späteren Lernleistung allgemein nachweisen (vgl. Kapitel 3.2.2). Zum Einfluss der Einstellung gegenüber dem Concept Mapping auf die Lernwirksamkeit von Feedback sind keine Studien bekannt.

H15: Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *niedrig* einschätzen, ist *FB* weniger lernwirksam als *kein FB*.

- a) Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *niedrig* einschätzen, ist *EiFB* weniger lernwirksam als *kein FB*.
- b) Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *niedrig* einschätzen, ist *ElFB* weniger lernwirksam als *kein FB*.

H16: Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *niedrig* einschätzen, ist *ElFB* weniger lernwirksam als *EiFB*.

H17: Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *hoch* einschätzen, ist *FB* lernwirksamer als *kein FB*.

a) Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *hoch* einschätzen, ist *EiFB* lernwirksamer als *kein FB*.

b) Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *hoch* einschätzen, ist *ElFB* lernwirksamer als *kein FB*.

H18: Bei Probanden, welche die *Interessantheit und Nützlichkeit* des Concept Mapping als *hoch* einschätzen, ist *ElFB* lernwirksamer als *EiFB*.

H19: Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *niedrig* einschätzen, ist *FB* weniger lernwirksam als *kein FB*.

a) Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *niedrig* einschätzen, ist *EiFB* weniger lernwirksam als *kein FB*.

b) Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *niedrig* einschätzen, ist *ElFB* weniger lernwirksamer als *kein FB*.

H20: Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *niedrig* einschätzen, ist *ElFB* weniger lernwirksam als *EiFB*.

H21: Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *hoch* einschätzen, ist *FB* lernwirksamer als *kein FB*.

a) Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *hoch* einschätzen, ist *EiFB* lernwirksamer als *kein FB*.

b) Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *hoch* einschätzen, ist *ElFB* lernwirksamer als *kein FB*.

H22: Bei Probanden, welche die *Handhabbarkeit* des Concept Mapping als *hoch* einschätzen, ist *ElFB* lernwirksamer als *EiFB*.

Die Lernwirksamkeit des Feedbacks hängt des Weiteren von den Aufgabenanforderungen der Lernaufgabe als situativem Faktor ab (vgl. Kapitel 2.1.4.4). Da bei einfachen Anforderungen auch einfaches Feedback zu genügen scheint (vgl. Kapitel 2.2.5), sollte die Art des Feedbacks in Bezug auf die Komplexität der Feedbacknachricht keinen Einfluss auf das Erreichen von Lernzielen mit einem niedrigen Anspruchsniveau haben. Einige Studien stellten fest, dass komplexe Lernziele durch elaboriertes Feedback erreicht werden können (Krause, 2007; Moreno, 2004). Zum Vergleich von Aufgaben verschiedener Komplexitätsstufen wurde bisher wenig geforscht (vgl. Kapitel 2.2.5). Die ältere

Befundlage zeigt teilweise einen Vorteil des elaborierten gegenüber dem einfachen Feedback bei komplexen, aber auch bei einfachen Aufgaben (Lindell, 1976), andere Studien lassen keinen Vorteil erkennen (O. M. Lee, 1985; Wager, 1983). Möglicherweise wird die Lernwirksamkeit komplexer Aufgaben durch individuelle Eingangsvoraussetzungen moderiert. Insgesamt besteht hier weiterer Forschungsbedarf. Für diese Untersuchung wird angenommen, dass sich die Gruppen *EiFB* und *ElFB* hinsichtlich *einfacher Aufgaben* nicht unterscheiden (H24), wohingegen sich bei *komplexen Aufgaben* ein Vorteil des *ElFB* gegenüber dem *EiFB* zeigt (H26).

In der Feedbackforschung wird der Fokus darauf gelegt zu analysieren, welche Faktoren die Gestaltung des Feedbacks beeinflussen. Dabei wird von einer generellen Lernwirksamkeit des Feedbacks ausgegangen, was u. a. der von Hattie nachgewiesene hohe Effekt von Feedback belegt (vgl. Kapitel 2.2.1). Wie oben erläutert, wird hierbei lediglich analysiert, welches Feedback bei einfachen und komplexen Aufgaben sinnvoll ist. Dennoch stellt sich gerade vor dem Hintergrund des ebenfalls hohen Einflusses des Concept Mapping (vgl. Kapitel 3.2.1) die Frage, ob das Erstellen von Concept Maps wirklich zusätzlichen Feedbacks bedarf. Hierbei könnte die Komplexität der Lernaufgaben von Bedeutung sein. Bei Aufgaben mit niedrigen Anforderungen besteht die Möglichkeit, dass das Erstellen der Maps schon die Tiefenverarbeitung der erlernten Inhalte fördert und kein zusätzliches Feedback vonnöten ist. Bei hohen Anforderungen hingegen ist die kognitive Belastung der Lernenden sehr hoch, so dass zusätzliches Feedback diese Belastung reduzieren kann und somit das Verständnis der Lerninhalte fördert. Aus diesem Grund wird in Hypothese H23 angenommen, dass sich bei *einfachen Aufgaben* Lernende, welche Feedback erhalten, nicht von Probanden unterscheiden, denen kein Feedback gegeben wird. Hypothese H25 postuliert, dass *FB* bei *komplexen Aufgaben* lernwirksamer ist als *kein FB*.

H23: Bei Lernzielen mit einem *niedrigen Anspruchsniveau bzw. einfachen Aufgaben* unterscheiden sich *FB* und *kein FB* nicht.

- a) Bei Lernzielen mit einem *niedrigen Anspruchsniveau bzw. einfachen Aufgaben* unterscheiden sich *EiFB* und *kein FB* nicht.
- b) Bei Lernzielen mit einem *niedrigen Anspruchsniveau bzw. einfachen Aufgaben* unterscheiden sich *ElFB* und *kein FB* nicht.

H24: Bei Lernzielen mit einem *niedrigen Anspruchsniveau bzw. einfachen Aufgaben* unterscheiden sich *ElFB* und *EiFB* nicht.

H25: Bei Lernzielen mit einem *hohen Anspruchsniveau bzw. komplexen Aufgaben* ist *FB* lernwirksamer als *kein FB*.

a) Bei Lernzielen mit einem *hohen Anspruchsniveau bzw. komplexen Aufgaben* ist *EiFB* lernwirksamer als *kein FB*.

b) Bei Lernzielen mit einem *hohen Anspruchsniveau bzw. komplexen Aufgaben* ist *ElFB* lernwirksamer als *kein FB*.

H26: Bei Lernzielen mit einem *hohen Anspruchsniveau bzw. komplexen Aufgaben* ist *ElFB* lernwirksamer als *EiFB*.

In Kapitel 2.2.4 wurde aufgezeigt, dass die zugeschriebene Kompetenz des Feedbacksenders über die Einschätzung der Adäquatheit des Feedbacks (Fairness, Nützlichkeit, Akzeptanz) messbar gemacht werden kann. Im Rahmen der Studie soll es weniger darum gehen, die Kompetenz der Lehrpersonen zu analysieren, sondern die *wahrgenommene Nützlichkeit des Feedbacks* allgemein zu untersuchen. Infolgedessen kann möglicherweise gezeigt werden, inwieweit es einen Zusammenhang zwischen der wahrgenommenen Nützlichkeit und der Lernwirksamkeit des Feedbacks gibt. Huth (2004) analysiert in ihrer Studie nicht nur den wahrgenommenen Nutzen des Feedbacks, sondern auch den Feedbackwunsch. Im Rahmen dieser Studie soll ebenfalls der *Wunsch nach zusätzlichem Feedback* untersucht werden.

Das Erstellen der Concept Maps stellt hohe kognitive Anforderungen an die Lernenden. Des Weiteren steigt die kognitive Belastung aufgrund der Komplexität der Lerninhalte, welche in den Concept Maps abgebildet werden sollen. Daher ist anzunehmen, dass sich Probanden eine Rückmeldung zu den in den Concept Maps dargestellten Inhalten wünschen. Dabei erhalten Probanden der Gruppe *EiFB* bereits eine Rückmeldung in Form des Referenznetzes. Der Wunsch nach weiterem elaboriertem Feedback sollte infolgedessen nicht so hoch ausgeprägt sein, wie der Wunsch der Gruppe *kein FB* nach einer generellen Rückmeldung (Hypothese H27).

H27: Der *Wunsch nach (zusätzlichem) Feedback* ist bei der Gruppe *kein FB* höher als bei der Gruppe *EiFB*.

Das einfache Feedback sollte von den Probanden, welche im Anschluss elaboriertes Feedback erhalten, als nützlicher wahrgenommen werden, als von den Probanden, welche lediglich einfaches Feedback bekommen (Hypothese H28). Probanden der Gruppe *EIFB* sind wahrscheinlich interessierter an der richtigen Lösung in Form des Referenznetzes und auch eher motiviert, das eigene Map anhand des Referenznetzes zu korrigieren als die Probanden der Gruppe *EiFB*, da sie das einfache Feedback als Bestandteil einer umfassenden Rückmeldung betrachten.

H28: Probanden der Gruppe *EIFB* schätzen das *einfache Feedback als nützlicher* ein als die Gruppe *EiFB*.

Einfaches und elaboriertes Feedback sollten gemäß Hypothese H29 positiv zusammenhängen. Probanden, welche bereits das einfache Feedback nicht als nützlich erachten, werden sicherlich auch die Nützlichkeit des elaborierten Feedbacks als niedrig einschätzen und umgekehrt.

H29: Je höher (niedriger) die *Nützlichkeit des einfachen Feedbacks* eingeschätzt wird, desto höher (niedriger) wird auch die *Nützlichkeit des elaborierten Feedbacks* eingeschätzt.

In diesem Zusammenhang kann angenommen werden, dass lediglich die Probanden, welche bereits das einfache Feedback als nützlich einschätzen, auch zusätzliche Informationen als elaboriertes Feedback wünschen. Demnach sollte bei der Gruppe *EiFB* ein positiver Zusammenhang zwischen der eingeschätzten Nützlichkeit des einfachen Feedbacks und dem Wunsch nach zusätzlichem elaboriertem Feedback bestehen (Hypothese H30).

H30: Je größer der *Wunsch* der Probanden der Gruppe *EiFB* nach *zusätzlichem elaborierten Feedback* ist, desto höher wird auch die *Nützlichkeit des einfachen Feedbacks* eingeschätzt.

Des Weiteren kann vermutet werden, dass Probanden, welche sich eine (zusätzliche) Rückmeldung zu ihren erstellten Maps wünschen, mit dem alleinigen Erstellen der Maps (Gruppe *kein FB*) bzw. mit dem Erstellen der Maps und dem anschließenden einfachen Feedback nicht das notwendige Tiefenverständnis der erlernten Inhalte erlangt haben.

Demnach postuliert die Hypothese H31 einen negativen Zusammenhang zwischen dem Wunsch nach zusätzlichem Feedback und der Lernwirksamkeit.

H31: Es gibt einen negativen Zusammenhang zwischen dem *Wunsch nach zusätzlichem Feedback* und der *Lernwirksamkeit*.

- a) Je größer der *Wunsch* der Probanden der Gruppe *kein FB* nach *zusätzlichem Feedback* ist, desto geringer ist die Lernwirksamkeit.
- b) Je größer der *Wunsch* der Probanden der Gruppe *EiFB* nach *zusätzlichem elaboremtem Feedback* ist, desto geringer ist die Lernwirksamkeit.

Die wahrgenommene Nützlichkeit des Feedbacks sollte hingegen gemäß Hypothese H32 positiv mit der Lernwirksamkeit zusammenhängen.

H32: Es gibt einen positiven Zusammenhang zwischen der *eingeschätzten Nützlichkeit des Feedbacks* und der *Lernwirksamkeit*.

- a) Je *nützlicher* das *einfache Feedback* eingeschätzt wird, desto größer ist die Lernwirksamkeit.
- b) Je *nützlicher* das *elaborierte Feedback* eingeschätzt wird, desto größer ist die Lernwirksamkeit.

Die Lernwirksamkeit des Feedbacks durch den Wissenserwerb hängt möglicherweise von der Qualität der Maps ab. In Studien, welche Concept Mapping mit Feedback kombinieren, wurde der Zusammenhang bisher wenig erforscht. Hwang, Wu und Ke (2011) konnten eine positive Korrelation zwischen der Qualität der Maps und den Ergebnissen im Nachtest nachweisen. Dabei finden sich jedoch keine Hinweise auf Unterschiede zwischen den Gruppen mit und ohne Feedback (vgl. Kapitel 3.2.3). Für diese Studie wird angenommen, dass insbesondere Probanden, die qualitativ schlechtere Maps erstellen, vom Feedback profitieren. Das Feedback lenkt die Aufmerksamkeit auf die Fehler in den Maps. Die Probanden sind möglicherweise auch mehr motiviert, ihre Fehler zu korrigieren und nehmen das Feedback folglich besser auf. Reflexionsprozesse werden somit angeregt, was letztlich den Wissenserwerb positiv beeinflussen kann. Hypothese H33 postuliert daher, dass *FB* lernwirksamer als *kein FB* für Probanden mit einer *niedrigen Mapqualität* ist. *EiFB* sollte dabei *EiFB* überlegen sein (H34). Eine hohe Qualität der Maps führt möglicherweise dazu, dass Feedback nicht benötigt wird. Die Probanden lenken ihre Aufmerksamkeit somit nicht auf das Feedback und sind auch weniger motiviert.

Infolgedessen profitieren sie nicht von zusätzlichem Feedback. In Hypothese H35 wird daher angenommen, dass sich *FB* und *kein FB* bei *hoher Mapqualität* nicht unterscheiden. Dabei sollte es auch keinen Unterschied zwischen *EiFB* und *ElFB* geben (H36). Innerhalb der Feedbackgruppen *EiFB* und *ElFB* sollten die Probanden mit einer *niedrigen Mapqualität* gegenüber denen mit einer *hohen Mapqualität* profitieren (H37 und H38). In der Gruppe *kein FB* sollte jedoch *hohe Mapqualität* lernwirksamer sein als *niedrige Mapqualität* (H39). Die kognitive Belastung wird bei den Probanden dieser Gruppe, welche bessere Maps erstellen, reduziert, da sie eher in der Lage sind die Lerninhalte zu verarbeiten und auch zu behalten.

H33: Bei Probanden, welche *qualitativ schlechtere Concept Maps* erstellen, ist *FB* lernwirksamer als *kein FB*.

- a) Bei Probanden, welche *qualitativ schlechtere Concept Maps* erstellen, ist *EiFB* lernwirksamer als *kein FB*.
- b) Bei Probanden, welche *qualitativ schlechtere Concept Maps* erstellen, ist *ElFB* lernwirksamer als *kein FB*.

H34: Bei Probanden, welche *qualitativ schlechtere Concept Maps* erstellen, ist *ElFB* lernwirksamer als *EiFB*.

H35: Bei Probanden, welche *qualitativ bessere Concept Maps* erstellen, unterscheiden sich *FB* und *kein FB* nicht.

- a) Bei Probanden, welche *qualitativ bessere Concept Maps* erstellen, unterscheiden sich *EiFB* und *kein FB* nicht.
- b) Bei Probanden, welche *qualitativ bessere Concept Maps* erstellen, unterscheiden sich *ElFB* und *kein FB* nicht.

H36: Bei Probanden, welche *qualitativ bessere Concept Maps* erstellen, unterscheiden sich *ElFB* und *EiFB* nicht.

H37: Probanden der Gruppe *ElFB*, welche *qualitativ schlechtere Concept Maps* erstellt haben, sind gegenüber denen, die *qualitativ bessere Concept Maps* erstellt haben, überlegen.

H38: Probanden der Gruppe *EiFB*, welche *qualitativ schlechtere Concept Maps* erstellt haben, sind gegenüber denen, die *qualitativ bessere Concept Maps* erstellt haben, überlegen.

H39: Probanden der Gruppe *kein FB*, welche *qualitativ bessere Concept Maps* erstellt haben, sind gegenüber denen, die *qualitativ schlechtere Concept Maps* erstellt haben, überlegen.

Die Lernwirksamkeit als abhängige Variable wird operationalisiert durch den Wissenserwerb sowie die Qualität der erstellten Maps. Hinsichtlich des Wissenserwerbs werden die Ergebnisse des 1. bzw. 2. Nachtests sowie der Wissenszuwachs (Vortest-1. Nachtest, Vortest-1. Nachtest-2. Nachtest) herangezogen. Die Qualität der Maps wird für das 1., 2. und 3. Mapping analysiert. Zusätzlich wird die Entwicklung der Mapqualität über die drei Mapping-Aktivitäten hinweg betrachtet.

Es handelt sich somit um eine hypothesenprüfende Untersuchung, die Quer- und Längsschnittstudie kombiniert.

Nachfolgende Tabelle 8 stellt die Zuordnung der formulierten Hypothesen zu den Forschungsfragen dar.

Tabelle 8: Zuordnung der Hypothesen zu den Forschungsfragen

Forschungsfrage	Hypothese
F1: FB>kein FB	H1
F2: E1FB>EiFB	H2
F3: Einflussfaktoren	H3-H6 Vorwissen; H7-H10 Sprachfähigkeit; H11-H14 Einschätzung des Umgangs mit Fehlern; H15-H22 Einschätzung des Umgangs mit Concept Mapping; H23-H26 Aufgabenanforderungen der Lernaufgabe (einfache Aufgaben/komplexe Aufgaben); H27-H32 Einschätzung der Nützlichkeit des Feedbacks

Die Hypothesen H33-H39 untersuchen den Einfluss der Mapqualität auf den Wissenserwerb. Die Hypothesen lassen sich dabei keiner der drei Forschungsfragen zuordnen. Sowohl die Qualität der Maps als auch der Wissenserwerb stellen Operationalisierungen der Lernwirksamkeit dar. Die Mapqualität wird hierbei als unabhängige Variable und der Wissenserwerb als abhängige Variable betrachtet.

4.1.2 Stichprobe und Untersuchungsdesign

Stichprobe

An der Feldstudie nahmen 83 Schüler achter und neunter Klassen zweier sächsischer Mittelschulen teil. Das Alter der Schüler lag zwischen 13 und 16 Jahren, das Durchschnittsalter betrug 13,9 Jahre. Die Stichprobe setzte sich aus 40 männlichen und 43 weiblichen Probanden zusammen. Im Rahmen des Lehrplans für das Fach Wirtschaft-Technik-Haushalt/Soziales werden in der Klassenstufe 7 im Lernbereich 2 die Fertigung materieller Güter, in der Klassenstufe 8 im Lernbereich 2 die Produktion von Gütern im Unternehmen behandelt, so dass von grundlegenden betriebswirtschaftlichen Vorkenntnissen ausgegangen werden kann (Sächsisches Staatsministerium für Kultus, 2009). Im Umgang mit Planspielen hatten die Probanden noch keine Erfahrungen.

Untersuchungsdesign – Aufbau

Die Untersuchung wurde im Rahmen eines Seminars mit Studierenden im Bachelor und Master Wirtschaftspädagogik durchgeführt. Die Studierenden waren im Rahmen der Studie als Lehrpersonen tätig.

Vor der Intervention wurde der erste Teil eines Paralleltests als Wissenstest eingesetzt, um das Vorwissen zu erfassen. Zusätzlich wurde die Sprachfähigkeit (Analogien und gleiche Wortbedeutungen des Subtests des Wilde-Intelligenz-Tests 2 von Kersting, Althoff, & Jäger, 2008) überprüft.

Vor der Durchführung des Planspiels wurden die Probanden einer Kontrollgruppe (n=27), welche lediglich das Planspiel durchführte, oder einer von drei Experimentalgruppen - *kein FB* (n=18), *EiFB* (n=18) oder *EIFB* (n=20) - zugeordnet. Das Erstellen der Concept Maps wurde in den Planspielunterricht integriert. Dabei wurden insgesamt drei Mapping-Aktivitäten durchgeführt.

Das 1. Concept Map sollte dabei die Abläufe in einem Produktionsunternehmen abbilden. Im 2. Concept Map stellten die Schüler die Zusammenhänge zwischen linearer Abschreibung, Gewinn- und Verlustrechnung [GuV] und Bilanz her. Beim 3. Mapping sollten die Zusammenhänge zwischen dem ersten und dem 2. Concept Map dargestellt werden. Den Probanden der Experimentalgruppen wurden Konzepte und Relationen jeweils in Form einer Liste zur Verfügung gestellt, mit deren Hilfe sie papierbasierte Concept Maps zu den wichtigsten Zusammenhängen des Spiels konstruieren sollten. Die Gruppe *kein FB*

erstellte lediglich die drei Concept Maps. Die Gruppe *EiFB* erhielt zusätzlich einfaches Feedback in Form eines Vergleichs mit dem jeweiligen Referenzmap als richtige Lösung. Nach jeder der drei Mapping-Aktivitäten wurde ein Referenzmap zur Verfügung gestellt. Die Schüler sollten sich intensiv mit dem Referenzmap auseinandersetzen und konnten in einer Kopie ihrer zuvor angefertigten Concept Maps Änderungen vornehmen. Die Gruppe *ElFB* erhielt ebenfalls einfaches Feedback in der beschriebenen Form. Darüber hinaus wurden Fehler exemplarisch anhand eines Schülernetzes sowie die drei Referenznetze in einem Lehrer-Schüler-Gespräch besprochen.

Im Nachtest wurde der zweite Teil des Paralleltests als Wissenstest eingesetzt. Zusätzlich zum Nachtest erhielten die Probanden einen Fragebogen, um die Methode des Concept Mapping und die Nützlichkeit des Feedbacks zu bewerten sowie den allgemeinen Umgang mit Fehlern einzuschätzen. Nach 9 Wochen wurde ein 2. Nachtest durchgeführt.

Untersuchungsdesign - Ablauf

In Vorbereitung auf die Untersuchung wurde die Methode des Concept Mapping mit der entsprechenden Gruppe in zwei Unterrichtseinheiten á 45 Minuten geübt. Als Hilfestellung erhielten die Probanden eine schriftliche Anleitung, welche die Vorgehensweise beim Erstellen eines Concept Maps anhand eines Beispiels demonstrierte (Ryssel, Fürstenau, Förster-Kuschel, Thieme & Lützner, 2018).

Die Untersuchung erstreckte sich über zwei Tage. Das Planspiel wurde in Kleingruppen á vier bis fünf Schülern durchgeführt. Die Zuteilung erfolgte per Los. Eingesetzt wurde das Planspiel Easy Business™, welches von der Business Training International GmbH entwickelt wurde (vgl. Kapitel 4.2.1).

Am ersten Tag in der ersten Planspielphase wurden die Interessengruppen eines Unternehmens erarbeitet sowie das erste Jahr des fiktiven Unternehmens Easy Business™ gespielt, um die Abläufe in einem Produktionsunternehmen kennen zu lernen. Im Anschluss an diese Phase erstellten die Probanden der drei Experimentalgruppen Concept Maps zu den eben kennen gelernten Abläufen in einem Produktionsunternehmen. Die beiden Feedbackgruppen erhielten anschließend Feedback. In der nächsten Planspielphase wurde der Jahresabschluss erstellt, das erste Geschäftsjahr zusammengefasst und reflektiert.

Am zweiten Tag stellten die drei Experimentalgruppen das am Tag zuvor erlernte Wissen zu der linearen Abschreibung, der GuV, der Bilanz sowie den Zusammenhängen zwischen diesen drei Themengebieten in einem Concept Map dar. Dies stellte die 2. Concept Mapping-Aktivität dar. Den beiden Feedbackgruppen wurde erneut Feedback gegeben. Die anschließende Planspielphase bestand aus einer kurzen Wiederholung, der Besprechung der Rationalisierungsmaßnahmen sowie dem Spielen des zweiten Geschäftsjahres. Danach erfolgte eine 3. Mapping-Aktivität, in der die Schüler zusammenhängende Begriffe der ersten beiden Concept Maps verbinden sollten, um somit die Gesamtzusammenhänge des Planspiels darzustellen. Auch dazu erhielten die Schüler der Feedbackgruppen eine Rückmeldung. In der letzten Planspielphase wurde der Jahresabschluss des zweiten Geschäftsjahres erstellt, die Ergebnisse des ersten und zweiten Geschäftsjahres verglichen sowie ein Abgleich des Planspiels mit der Realität vorgenommen.

Der genaue Ablauf des Concept Mapping-Trainings sowie die Instruktionsanweisungen zu den Concept Mapping-Aktivitäten werden bei Ryssel, Fürstenau, Förster-Kuschel u. a. (2018, Abschnitt 4.2) vorgestellt. Abbildung 8 fasst den Aufbau und den Ablauf der Untersuchung noch einmal zusammen. Zusätzlich sind die Ablaufpläne, welche den Lehrpersonen während der Durchführung als Orientierung dienten, in den Anhängen 4.1.2.-1 bis 4.1.2.-4 dargestellt.

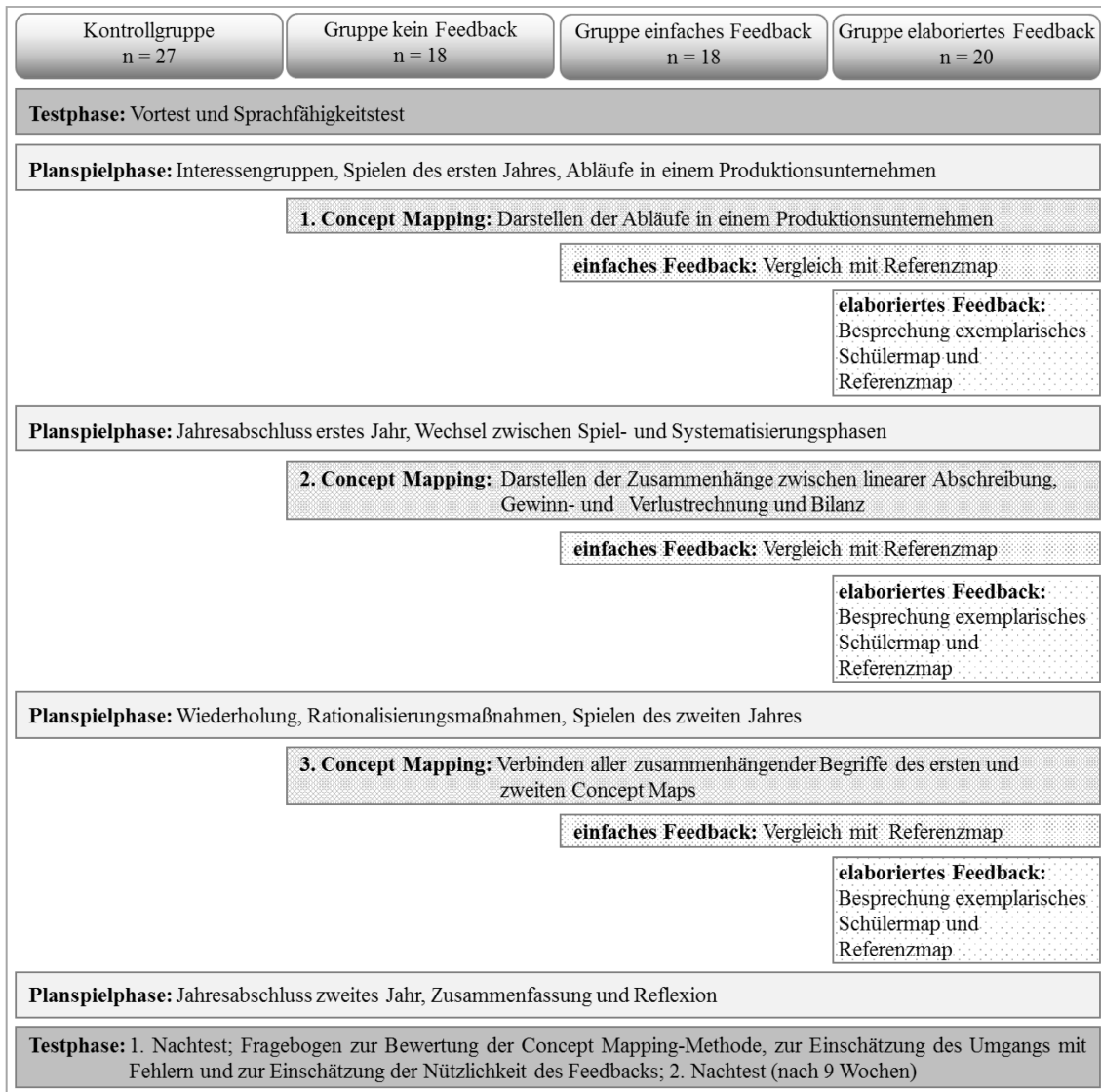


Abbildung 8: Aufbau und Ablauf der empirischen Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps

Der Planspielunterricht sowie die Gestaltung des Concept Mapping und der Feedbackmaßnahmen werden in Kapitel 4.2 unter Bezugnahme auf die in Kapitel 2 und 3 erläuterten theoretischen Grundlagen vorgestellt.

4.2 Gestaltung der Lernmaterialien und Interventionsmaßnahmen

4.2.1 Planspielunterricht mit Easy Business™

Bevor die Gestaltung des Planspielunterrichts dargelegt wird, müssen zunächst einige theoretische Vorüberlegungen zum handlungsorientierten Lernen mit Planspielen getätigt werden.

Handlungsorientiertes Lernen mit Planspielen

Neue technische, ökonomische und gesellschaftliche Entwicklungstendenzen haben ein grundlegendes Überdenken der beruflichen Erstausbildung und Weiterbildung ausgelöst. Anforderungen an junge Berufsanfänger sind durch zunehmende Komplexität gekennzeichnet. Die Berufsausbildung muss zukunftsfähig gestaltet werden, d. h. den vielfältigen Anforderungen an die Auszubildenden gerecht werden (Achtenhagen, 1995). Im Rahmen didaktischer Reformprogramme erfolgten Neuordnungen vieler Ausbildungsberufe sowie die Neugestaltung der Rahmenlehrpläne. Ziel dieser Neuordnungen war es, berufliche Handlungskompetenz zu fördern. Handlungskompetenz beinhaltet die Fähigkeit, (situationsangepasst) immer wieder „neuartige Handlungen zu generieren, d. h. sie zu planen, auszuführen und zu kontrollieren“ (Tramm & Rebmann, 1999, S. 238). Die Lernenden sollten daher bereits während ihrer Ausbildung vollständige Handlungen durchlaufen. Nach Tramm und Rebmann (1999, S. 238) existieren dabei sechs Phasen einer vollständigen Handlung: Situationswahrnehmung und –bewertung, Zielbildung, Suche nach Handlungsalternativen, Bewertung der Alternativen und Entscheidung, Handlungsentschluss, Regulation der Handlungsausführung.

Hierzu sollte handlungsorientiertes Lernen ermöglicht werden. Grundlage des Lernprozesses stellt nach Preiß (1995) vorhandenes Vorwissen dar, zusätzlich sollten auch Erfahrungen aus dem privaten und beruflichen Umfeld einbezogen werden. Am Anfang des Lernprozesses steht eine überschaubare, problemhaltige, authentische und komplexe Lernsituation. Diese stellt eine leichte Überforderung für die Lernenden dar, da sie nicht über alle Kenntnisse und Fertigkeiten zur Lösung verfügen. Um die Lernenden zu motivieren, muss die Lehrperson entsprechende Methoden wählen. An dieser Stelle können Komplexe Lehr-Lernarrangements, z. B. Planspiele eingesetzt werden. Die Lernenden agieren beispielsweise in einem Modellunternehmen, sie übernehmen konkrete Rollen aus Sicht des Unternehmens und durchlaufen vollständige Handlungen. Anschließend folgt die Phase der Abstraktion und Systematisierung. Im Rahmen der Abstraktion wird

fachliches und überfachliches Wissen erarbeitet. Mittels der Systematisierung wird neues Wissen in bisherige Wissensbestände eingeordnet. Die Kenntnisse und Fertigkeiten sind jedoch noch stark an die Lernsituation gebunden und noch nicht gefestigt. Aus diesem Grund sollte in einer ähnlichen oder neuartigen Anwendungssituation, z. B. durch Analogien, das Wissen geübt, angewendet und wiederholt werden, um eine Rekontextualisierung zu erreichen. Hierbei sollte das Leistungsvermögen der Schüler beachtet werden (Preiß, 1995). Das handlungsorientierte Lernen nach Preiß (1995) ist in nachfolgender Abbildung 9 dargestellt.

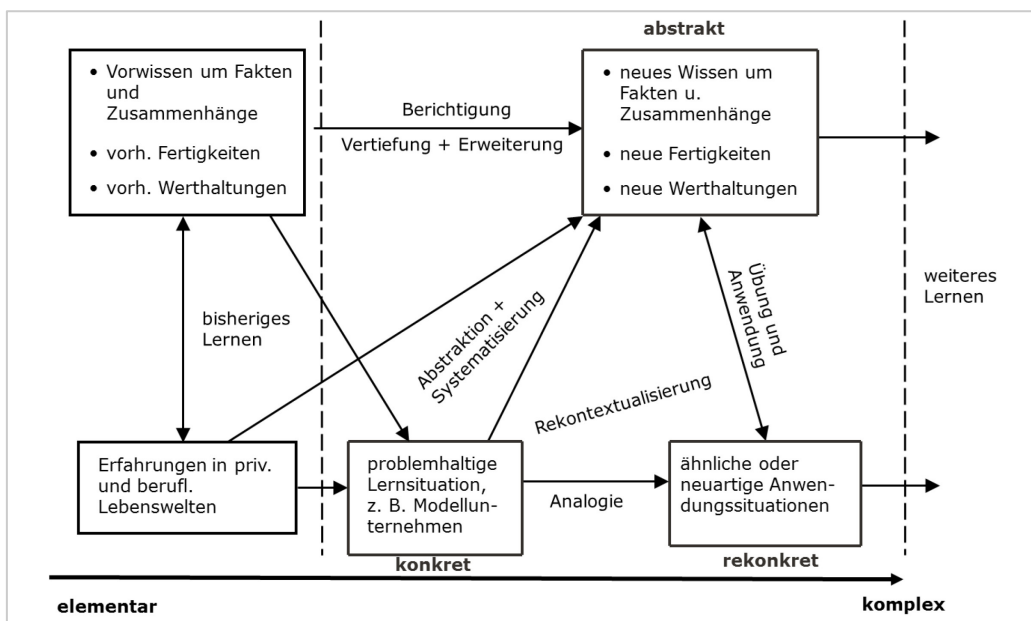


Abbildung 9: Handlungsorientiertes Lernen (Preiß, 1995, modifiziert)

Es stellt sich nun die Frage, wie Lernprozesse beim handlungsorientierten Lernen sinnvoll strukturiert bzw. sequenziert werden sollten. Hierbei wird zwischen Makro- und Mikrosequenzierung unterschieden (z. B. Sievers, 1984)⁶³. Makrosequenzierung bezieht sich auf die zeitliche Anordnung einzelner Curriculumelemente zu ganzen Unterrichtsreihen oder Kursen vom Elementaren zum Komplexen. Mikrosequenzierung bezeichnet die Strukturierung einzelner Curriculumelemente, z. B. Begriffe, mit dem Sequenzierungsansatz konkret-abstrakt-rekonkret (Sievers, 1984; Achtenhagen u. a., 1992, S. 113; Tramm & Rebmann, 1997, S. 26). Zusätzlich ist der Lernprozess ein von elementaren zu

⁶³ Die Ausgestaltung der Makro- und Mikrosequenzierung geht auf Dörner (1987, S. 32) zurück. Er geht davon aus, dass das Gedächtnis netzwerkartig organisiert ist und aus Abstraktions- und Komplexionshierarchie besteht. Aus dem Aufbau der Abstraktionshierarchie ergibt sich die Anordnung der Mikrosequenzierung, aus dem Aufbau der Komplexionshierarchie die Makrosequenzierung.

komplexen Strukturen voranschreitender Konstruktionsprozess. Dies bedeutet, dass zunächst wenig differenzierte Ganzheiten zunehmend ausdifferenziert werden (Achtenhagen u. a., 1992, S. 107). Lernende lernen beispielsweise am Anfang des Lernprozesses ein Modellunternehmen als Ganzheit kennen, jedoch werden die Bereiche, Aufgaben usw. des Unternehmens lediglich in Grundzügen vermittelt. Im Laufe des Lernprozesses werden mehr und mehr Details und Zusammenhänge verdeutlicht. Die Komplexität wird somit angereichert. Eine Makrosequenz setzt sich aus mehreren Mikrosequenzen zusammen. Die Sequenzierungsansätze für die Mikro- und Makrosequenzierung sind in Abbildung 9 mit dargestellt. In der Abbildung zeigt sich, dass bereits innerhalb der Mikrosequenz eine Komplexitätssteigerung stattfinden kann, beispielsweise indem das Vorwissen als elementare Vorstruktur immer mehr angereichert und somit komplexer wird.

Das Konzept des handlungsorientierten Lernens kann durch den Einsatz von Komplexen Lehr-Lern Arrangements, z. B. Planspielen, Fallstudien oder Lernbüros realisiert werden. „Ein Planspiel ist eine besondere Form eines dynamischen Modells, in dem sich die Spielteilnehmer und Spielteilnehmerinnen mit konfliktären bzw. problemhaltigen Situationen innerhalb eines durch Regeln festgelegten Kontextes unter einer bestimmten Zielsetzung kollektiv und individuell handelnd auseinander setzen“ (Rebmann, 2001, S. 10). Planspiele lassen sich durch zwei zentrale Komponenten charakterisieren: den Plan (das Modell) und das Spiel. Ein Modell ist ein Abbild der Realität. Im Rahmen eines Planspiels wird ein Problem aus der Realität bewältigt. Das Problem wird dabei vereinfacht unter Beachtung der Zielgruppe dargestellt (Fürstenau, 2009, S. 240). Während des Planspiels findet ein „Lernen im Modell“ durch die Übernahme von Rollen, das konkrete Ausführen von Tätigkeiten sowie Reflexionsprozessen, statt. Am Ende des Spiels sollte dieses um ein „Lernen am Modell“ ergänzt werden, um Unterschiede zur Realität zu thematisieren (Achtenhagen u. a., 1992, S. 128 f.). Ein Planspiel ermöglicht es, im Rahmen des Modells Spielentscheidungen zu treffen. Dabei wird der Spielcharakter durch Merkmale wie Wettbewerb oder Kooperation, Regeln, Spieler, Rollenübernahme, Aktivität reglementierten Ablauf und Freude verdeutlicht (Rebmann, 2001, S. 14 ff.).

Das Planspiel Easy Business™

Als Lernmaterial für die empirische Untersuchung wurde das von der Business Training International GmbH entwickelte Planspiel Easy Business™ eingesetzt. Das Planspiel wurde insbesondere für Schüler in Abschlussklassen allgemeinbildender Schulen (Haupt-

und Realschulen sowie Gymnasien u. ä.) oder für Schüler in Eingangsklassen beruflicher Schulen entwickelt, um betriebswirtschaftliche Grundlagen zu vermitteln. Die Schüler lernen spielerisch den Wertschöpfungsprozess eines Produktionsunternehmens kennen und erleben anschließend die Folgen ihres Handelns in der jährlichen Rechnungslegung. Hierbei lernen sie, eine GuV sowie eine Bilanz zu erstellen. Darüber hinaus werden Interessengruppen eines Produktionsunternehmens thematisiert sowie Auswirkungen von Rationalisierungsmaßnahmen veranschaulicht. Zusätzlich werden die Erwartungen der Anteilseigner des Unternehmens diskutiert und mit Rentabilitätskennziffern verdeutlicht. Insgesamt werden zwei Geschäftsjahre gespielt, wobei entsprechende Rationalisierungsmaßnahmen zu einer Steigerung des Gewinns im zweiten Jahr führen. Der Themenbereich Rentabilitätskennziffern wird nicht in die Untersuchung einbezogen. Das Planspiel wird mit Oberschülern der achten und neunten Klasse durchgeführt. Um diese mit der Komplexität nicht zu überfordern, ist diese Reduzierung sinnvoll.

Das Planspiel ist als Brettspiel konzipiert. Auf einem großen Spielbrett sind alle wichtigen Bereiche eines Produktionsunternehmens abgebildet. Die jeweiligen Güter- und Werteströme werden durch das Verrücken von Werteträgern optisch und haptisch verdeutlicht. Die entsprechenden Spielzüge sind einem Begleitskript (vgl. Anhang 4.2.1-1⁶⁴) zu entnehmen. Neben den Arbeitsanweisungen sind darin auch theoretische Erläuterungen enthalten. Bei der Planspieldurchführung verfügte jede Kleingruppe über ein Spielbrett. Das Begleitskript wurde den Schülern während der Durchführung sukzessive ausgegeben.

Gestaltung des handlungsorientierten Planspielunterrichts mit Easy Business™

Die gesamte Planspieldurchführung stellt die Makrosequenz der empirischen Untersuchung dar (vgl. Anhänge 4.2.1-2 bis 4.2.1-5). Den Schülern wurde als Einstieg erläutert, was ein Planspiel ist. Sie wurden anschließend mit der Struktur eines Produktionsunternehmens in vereinfachter Form konfrontiert. In diesem Zusammenhang wurden sie darauf hingewiesen, dass das Planspiel aus der Perspektive eines Produktionsunternehmens durchgeführt werden soll.

In diesem Zusammenhang wurden die Interessengruppen sowie die Erwartungen, die ein Unternehmen an diese hat bzw. die Erwartungen, welche die Interessengruppen an ein

⁶⁴ Das von der BTI GmbH entwickelte Skript wurde dabei an den Planspielunterricht mit der Zielgruppe dieser Untersuchung angepasst.

Unternehmen haben, erarbeitet. Danach wurde die Spielsituation vorgestellt sowie der Spielbau erläutert. Die Schüler spielten anschließend das erste Jahr des fiktiven Unternehmens Easy BusinessTM und lernten dabei die Abläufe in einem Produktionsunternehmen kennen. Anhand von zu bearbeitenden Kundenaufträgen wurden den Schülern die Phasen des Wertschöpfungsprozesses von der Zulieferung über die Fertigung bis hin zum Verkauf bewusst gemacht. Anschließend wurden lineare Abschreibungen ermittelt, eine GuV sowie die Bilanz erstellt. Eine Komplexitätssteigerung ist somit erkennbar. Die Spielphasen wurden um Abstraktions- bzw. Systematisierungsphasen zur linearen Abschreibung, GuV und Bilanz im Lehrer-Schülergespräch ergänzt, um neue Begrifflichkeiten vertiefend zu erarbeiten. Somit fand ein Wechsel von konkreten und abstrakten Phasen statt. Dabei wurden die Schülergruppen bewusst instruiert, die Arbeitsaufträge selbstständig zu bearbeiten, bevor die jeweilige Theorie vermittelt wurde. Zum Abschluss des ersten Tages wurde das erste Jahr von Easy BusinessTM zusammengefasst. Das Ergebnis der GuV des gespielten Geschäftsjahres wurde mit dem Vorjahresergebnis verglichen. Insgesamt konnte im Vergleich zum Vorjahr ein geringerer Gewinn erzielt werden. Die Schüler erhielten als Hausaufgabe den Auftrag, sich Maßnahmen zur Verbesserung der Geschäftssituation von Easy BusinessTM zu überlegen. Am zweiten Tag wurden zunächst die Ergebnisse des ersten Geschäftsjahres von Easy BusinessTM wiederholt und die Hausaufgabe sowie die Rationalisierungsmaßnahmen, die das Spiel vorgibt, besprochen. Anschließend sollten die Schüler die Kundenaufträge des zweiten Geschäftsjahres selbstständig bearbeiten sowie den Jahresabschluss erstellen. Dies stellte das komplexeste Moment der Planspieldurchführung dar. Gleichzeitig ist dies eine ähnliche Anwendungssituation, um das Wissen zu wiederholen, zu üben und anzuwenden und somit eine Rekontextualisierung zu erreichen. Danach wurden die Ergebnisse des ersten und zweiten Geschäftsjahres verglichen. Die Rationalisierungsmaßnahmen konnten zu einer Gewinnsteigerung beitragen. Abschließend wurde ein Lernen am Modell durchgeführt, indem die Lehrpersonen in einem Lehrer-Schüler-Gespräch Unterschiede des Planspiels zur Realität erarbeiteten. Beispielsweise sollten die Schüler dabei erkennen, dass im Planspiel weniger Produkte als in der Realität hergestellt werden und es keine Konkurrenz gab.

Für die Themenbereiche lineare Abschreibung, GuV und Bilanz wurden detaillierte Ablaufpläne erarbeitet, um die Vermittlung der Begrifflichkeiten zu standardisieren und den für die Zielgruppe sehr komplexen Begrifflichkeiten auf diese Weise besondere Bedeu-

tung beizumessen. Bei der Entwicklung der Mikrosequenzen wurde dabei die Sequenzierungsansatz konkret-abstrakt-rekonkret umgesetzt. Die Ablaufpläne der drei Mikrosequenzen sind in Anhang 4.2.1-6 dargestellt. Zusätzlich wurden fachliche Lernziele formuliert, welche dem Anhang 4.2.1-7 entnommen werden können.

4.2.2 Gestaltung des Concept Mapping

Nachfolgend wird die Gestaltung des Concept Mapping mit Bezug auf die theoretischen Grundlagen und empirischen Befunde zu Concept Maps im Lehr-Lern-Kontext aus Kapitel 3 erläutert.

Bei den von den Probanden zu erstellenden Concept Maps handelte es sich um *systemische* Concept Maps, d. h. jedes Konzept konnte durch beliebig viele Relationen mit anderen Konzepten verbunden werden. Hierarchische Beziehungen mussten dabei nicht notwendigerweise hergestellt werden. Demzufolge konnten die Maps sowohl *vertikal* als auch *horizontal* ausgerichtet sein. Die Probanden wurden instruiert, Concept Maps selbstständig per Hand mittels des *Paper & Pencil-Verfahrens* zu erstellen. Somit wurde die *Selbstkonstruktion* angewendet, welche den geringsten *Grad der Vorstrukturierung* darstellt. Aufgrund der papierbasierten Darstellung wurden die Probanden angehalten, alle Inhalte in einem Map darzustellen (*Whole Map*) und es handelt sich daher auch um *statische Maps*. Um die kognitive Belastung der Lernenden zu reduzieren, wurden den Lernenden Konzepte und Relationen jeweils in Form einer Liste zur Verfügung gestellt, mit deren Hilfe sie die Concept Maps konstruieren sollten. Hierbei handelte es sich um einen offenen Pool, da die Probanden selbst Begriffe und/oder Relationen einfügen konnten und auch nicht notwendigerweise alle vorgegebenen Konzepte und Relationen verwenden mussten. Die vorgegebenen Konzepte und Relationen wurden den Referenzmaps (vgl. Anhänge 4.2.2-1 bis 4.2.2.-3) entnommen. Die drei Referenzmaps wurden auf Basis der Lerninhalte des Planspiels entwickelt und spiegeln die zu erreichenden fachlichen Lernziele wider (vgl. Anhang 4.2.1-7). Dabei wurden die Referenzmaps auf Grundlage der Erfahrungen vorangegangener Untersuchungen optimiert (Ryssel & Fürstenau, 2011; Fürstenau, Ryssel, & Kunath, 2010; Sommer u. a., 2009). Das für die Voruntersuchungen entwickelte Referenzmap wurde erweitert bzw. angepasst, um die im Planspiel vermittelten Fachinhalte umfassender abprüfen zu können. Anstelle von einem in den Voruntersuchungen eingesetzten Mapping wurden jedoch drei Mapping-Aktivitäten durchgeführt,

um eine Überforderung der Probanden zu verhindern und die Selbstkonstruktion zusätzlich, wie von Cañas u. a. (2003) empfohlen, kontinuierlich in den Lernprozess einzubetten.

Das erste Referenzmap spiegelt die Abläufe in einem Produktionsunternehmen wider. Im 2. Referenzmap sind die Zusammenhänge zwischen linearer Abschreibung, GuV und Bilanz dargestellt. In der 3. Mapping-Aktivität sollten die Probanden Verbindungen zwischen dem ersten und 2. Map finden. Beide Referenzmaps wurden vollständig vorgegeben, es sollten nur Relationen eingezeichnet werden. Somit wurden für die 3. Mapping-Aktivität lediglich Relationen vorgegeben (Ryssel, Fürstenau, Förster-Kuschel u. a., 2018, S. 71). Das im Anhang 4.2.2.-3 dargestellte 3. Referenzmap besteht somit aus dem 1. und 2. Referenzmap sowie den durch fettgedruckte Linien hervorgehobenen Relationen, die beide Referenzmaps miteinander verbinden.

Schroeder u. a. (2017) weisen darauf hin, dass das Lernen mit Maps umso effektiver ist, je länger es eingesetzt wird (vgl. Kapitel 3.2.1). Da die Mapperstellung in die Planspiel-durchführung integriert werden sollte und das Planspiel für eine Dauer von zwei Tagen vorgesehen wurde, ist ein längerer Einsatz an dieser Stelle nicht zielführend. Die Mapping-Aktivitäten verteilten sich somit ebenfalls auf zwei Tage.

Die Concept Maps wurden von jeder *Einzelperson* selbstständig erstellt.

Im Kapitel 3.1 wurde bereits darauf eingegangen, dass die Vertrautheit mit der Methode Concept Mapping eine Grundvoraussetzung darstellt. Daher sollten die Probanden vorab ausreichend geschult werden. Mit den Probanden dieser Studie wurde hierzu ein Training (2 Unterrichtseinheiten á 45 Minuten) durchgeführt (Ryssel, Fürstenau, Förster-Kuschel u. a., 2018, S. 63 ff.). Es handelt sich dabei um ein *direktes Training*, welches durch *kognitive und metakognitive Prompts* unterstützt wurde (vgl. Kapitel 3.2.2). Im Kapitel 3.2.2 wurde bereits auf die Bedeutung von Prompts hingewiesen. Lediglich eine Kombination des direkten Trainingsansatzes mit Prompts zeigte sich lernwirksam (Hilbert & Renkl, 2009). Des Weiteren wurde darauf hingewiesen, dass beim indirekten Training Prompts nicht nur bei der Erstellung des ersten Maps angeboten werden sollten, sondern den Mappingprozess kontinuierlich begleiten sollten (Hilbert, Nückles, Renkl, u. a., 2008). Aus diesem Grund wurden die Instruktionsanweisungen aller drei Mappings mit Prompts ergänzt (Ryssel, Fürstenau, Förster-Kuschel u. a., 2018, S. 69 f.).

Die Mapping-Aktivitäten stellten dabei eine *Lernstrategie* dar, um den Planspieleinsatz zu unterstützen. Das Concept Mapping ist zum einen eine *Organisationsstrategie*, um die während des Planspielunterrichts erlernten Inhalte zu organisieren und zu strukturieren. Zum anderen kann es auch als *Elaborationsstrategie* angesehen werden. Beim Erstellen der Maps werden die neu erlernten Inhalte mit dem Vorwissen sowie mit dem bereits im Planspiel erworbenen Fachwissen verknüpft und somit neue Informationen in bestehende kognitive Strukturen integriert. Weiterhin dienten die Concept Maps der *Wissensdiagnose*, d. h. sie erfassten, was die Schüler jeweils gelernt hatten. In Kapitel 3.2.3 wurde auf Studien verwiesen, die auf den Vorteil von computergestütztem Concept Mapping mit Feedback im Vergleich zu papierbasiertem Mapping hinweisen, da papierbasiertes Mapping kein sofortiges Feedback ermöglicht. Dieser Kritik wurde im Rahmen der Studie Rechnung getragen, indem sofortiges Feedback bereitgestellt wird. Hierbei war es nicht möglich, jedes individuelle Map sofort zu bewerten und das Ergebnis zurückzumelden. Daher wurde den Feedback-Gruppen einfaches Feedback in Form des Referenznetzes oder elaboriertes Feedback bereitgestellt. Das elaborierte Feedback beinhaltete neben dem Bereitstellen der richtigen Lösung u. a. die beispielhafte Besprechung eines Schülernetzes für jedes der drei Mappings. Die konkrete Ausgestaltung des Feedbacks wird in Kapitel 4.2.3 dargestellt. Zusätzlich wurden die erstellten und korrigierten Maps (Feedback-Netze) im Rahmen der Auswertung einer detaillierten qualitativen Analyse unterzogen (vgl. Kapitel 4.5.2).

4.2.3 Gestaltung des Feedbacks

Nachfolgend wird die Gestaltung des Feedbacks mit Bezug auf die Elemente und empirischen Befunde zum erweiterten Rahmenmodell aus Kapitel 2 erläutert.

Im Rahmen der Studie wurde das Feedback den Schülern nach der Aufgabenbearbeitung, d. h. dem Erstellen der Concept Maps, von den jeweiligen Lehrpersonen als externe Informationsquelle angeboten (vgl. Kapitel 2.1.1).

Hinsichtlich der Komplexität des Feedbackinhalts kann dabei zwischen *einfachem* und *elaboriertem Feedback* differenziert werden (vgl. Kapitel 2.1.3.2). Die den Schülern im Anschluss an das Erstellen der Maps zum Vergleichen bereitgestellten Referenznetze repräsentierten *KCR*, da den Schülern auf diese Weise die richtige Lösung präsentiert wurde.

Das elaborierte Feedback lässt sich den Formen KM, KC und KH entsprechend der Klassifikation von Narciss (2006) zuordnen (vgl. Kapitel 2.1.3.2).

Zunächst wurden Fehler anhand von Schülernetzen in einem Lehrer-Schüler-Gespräch exemplarisch besprochen. Für jedes der drei Mappings wurde hierzu vom Versuchsleiter im Anschluss an das Erstellen der Maps ein Schülernetz ausgewählt und auf Folie kopiert. Die Rahmenbedingungen ließen es jedoch nicht zu, alle Schülermaps einer detaillierten Analyse zu unterziehen und somit eine systematische Auswahl vorzunehmen. Zu diesem Zweck hätte die Untersuchung nach jedem Mapping für eine längere Zeit unterbrochen werden müssen. Der Versuchsleiter wählte eher spontan anhand von Kriterien wie Lesbarkeit und Nachvollziehbarkeit der dargestellten Propositionen aus. Trotzdem soll durch dieses Vorgehen der Kritik Rechnung getragen werden, dass papierbasiertes Mapping kein sofortiges Feedback zulässt (vgl. Kapitel 3.2.3). Eines der verwendeten Schülermaps ist in Anhang 4.2.3-1 dargestellt. Die Lehrpersonen erarbeiteten gemeinsam mit den Schülern in einem Lehrer-Schüler-Gespräch am Overheadprojektor [OHP] Ort, Art und eventuelle Ursachen der Fehler, markierten diese und korrigierten sie gemeinsam mit den Schülern. Diese Art des Feedbacks lässt sich daher *KM* sowie *KH* zuordnen. Neben fehlerbezogenen Informationen erhielten die Lernenden fehlerspezifische Korrekturhinweise und Lösungsstrategien. In diesem Zusammenhang war es zwar nicht möglich, vorab eine systematische Fehleranalyse vorzunehmen, jedoch half das interaktive Vorgehen beim Geben des Feedbacks dabei, die Gedankengänge der Schüler beim Erstellen der dargestellten Propositionen zu erfragen und somit besser nachzuvollziehen. Durch dieses Vorgehen sollte ebenfalls der Kritik, einiger in Kapitel 3.2.3 vorgestellter Studien, Rechnung getragen werden, dass papierbasiertes Mapping kein sofortiges Feedback zulässt.

Anschließend wurden die Referenznetze in einem Lehrer-Schüler-Gespräch besprochen. Zu diesem Zweck wurden die Referenznetze auf OHP-Folien kopiert und die OHP-Folien in Themenblöcke zerschnitten. Auf diese Weise konnten die Lehrpersonen die Inhalte Schritt für Schritt mit Hilfe des OHP aufbauen, wobei die Schüler in einem Lehrer-Schüler-Gespräch aktiv einbezogen wurden. Auf Verständnisprobleme der Schüler konnte so jeweils eingegangen werden. Diese Art des Feedbacks lässt sich daher *KC* sowie *KH* zuordnen. Die Schüler erhielten Informationen zu den Fachbegriffen, Erläuterungen, Beispiele zu den Begriffen sowie aufgabenspezifische Lösungshinweise und letztlich das Lösungsbeispiel in Form des Referenznetzes.

Narciss (2006) betont, dass elaboriertes Feedback unabhängig von der richtigen Lösung gegeben werden sollte. Beim Concept Mapping als Aufgabenformat ist dieses Vorgehen jedoch nicht generell lernwirksamer. Die Komplexität der Concept Maps erschwert dem Lernenden das eigenständige Erschließen der richtigen Lösung, z. B. nur mit Hilfe zusätzlicher Informationen zu einzelnen Fachbegriffen und deren Zusammenhänge. Selbst die richtige Lösung, das Referenzmap, muss vom Lernenden umfassend durchdacht werden, da es mehrere Teilantworten bzw. Zusammenhänge, welche durch die Propositionen repräsentiert werden, beinhaltet. Das Präsentieren der richtigen Lösung trägt an dieser Stelle eher zur Unterstützung des Lernprozesses bei, als diesen zu hemmen.

Das im Rahmen der Studie gegebene Feedback hat primär *kognitive Funktionen* (vgl. Kapitel 2.1.2). Insgesamt muss dabei zwischen einfachem und elaboriertem Feedback differenziert werden. Im Rahmen des einfachen Feedbacks werden die Lernenden auf der Aufgabenebene über die Richtigkeit der erstellten Concept Maps informiert und ggf. richtige Propositionen bestätigt. Das elaborierte Feedback bezieht sich sowohl auf die *Aufgaben-* als auch auf die *Prozessebene*. Der Lehrende informiert über die Richtigkeit der Concept Maps, bestätigt ggf. richtige Propositionen und bietet zusätzliche Informationen zur Aufgabenbearbeitung an. Das Feedback zielt weiterhin auf Strategien und Prozesse ab. In diesem Zusammenhang werden dem Lernenden die richtigen Verknüpfungen zwischen den Konzepten detailliert aufgezeigt, Fehler analysiert und korrigiert. Das elaborierte Feedback bezieht sich darüber hinaus auf die *Selbstebene* und hat damit ebenso eine *motivationale Funktion*. Die Lehrpersonen stehen dabei vor der Herausforderung insbesondere die Leistungen der Schüler, deren Maps exemplarisch besprochen werden, angemessen wertzuschätzen, d. h. eine fehleroffene Lernumgebung zu schaffen. Letztlich ermöglicht das dreimalige Geben des elaborierten Feedbacks, dass die erste und zweite Feedback-Aktivität für die Gruppe *EIFB* die jeweils nachfolgende Phase des Planspiels und folglich auch das Erstellen und Korrigieren der nachfolgenden Maps im Hinblick auf das Anwenden von *metakognitiven* Strategien auf der *Ebene der Selbstregulation*⁶⁵ unterstützt. Die Feedback-Hinweise zum vorangegangenen Mapping ermöglichen dem Ler-

⁶⁵ Nach Flavell (1979) lassen sich zwei Dimensionen für Metakognition unterscheiden: metakognitives Wissen und metakognitive Steuerung bzw. Kontrolle (Selbstregulation). Das metakognitive Wissen stellt dabei eine Voraussetzung für eine metakognitive Steuerung und Kontrolle dar. Wissen über das eigene Denken, die Beschaffenheit und die Anforderungen einer Aufgabe sowie die Kenntnis von Lösungsstrategien müssen demnach für eine spätere Steuerung und Kontrolle des Lernprozesses vorhanden sein. Im Rahmen der Feedbackliteratur wird lediglich die Ebene der Selbstregulation fokussiert.

nenden bei den nachfolgenden Mapping-Aktivitäten das Lernen zu bewerten, zu kontrollieren und selbst zu regulieren. Auf diese Weise sind die Lernenden bereits vor dem jeweils extern dargebotenen elaborierten Feedback in der Lage, internes Feedback zur Korrektur der erstellten Maps mithilfe des jeweiligen Referenznetzes (einfaches Feedback) zu nutzen. Grundvoraussetzung hierfür ist es, dass der Lehrende auch metakognitive Strategien vermittelt. Im Rahmen der Studie wurden hierzu die einzelnen Propositionen des jeweiligen Schülermaps sowie des Referenznetzes Schritt für Schritt analysiert. Den Schülern wurde dabei das Zustandekommen der einzelnen Verbindungen mit Hilfe von Hinweisen auf verwendete Medien, wie Tafelbilder oder Abbildungen im Planspielskript, verdeutlicht. Auf diese Weise konnten die Schüler ihre erstellten Maps bei nachfolgenden Mapping-Aktivitäten anhand des Referenznetzes ebenfalls systematisch analysieren und ggf. korrigieren. Beim elaborierten Feedback wird folglich *externes* um *internes Feedback* ergänzt. Das einfache Feedback wird den Schülern mittels der Referenznetze von einer externen Informationsquelle angeboten.

Hinsichtlich des *Feedbackempfängers* kann ebenfalls zwischen einfachem und elaboriertem Feedback unterschieden werden. Das einfache Feedback richtet sich an jeden einzelnen Schüler (*individuelles Feedback*), wohingegen das elaborierte Feedback als *Gruppenfeedback* gegeben wird. Empirische Befunde zum Einfluss der Art des Feedbackempfängers auf die Lernwirksamkeit von Feedback sind uneinheitlich (vgl. Kapitel 2.2.3). Bei der Studie handelt es sich um eine Feldstudie, bei der die Rahmenbedingungen vor Ort beachtet werden mussten. Daher wurde das elaborierte Feedback als Gruppenfeedback gegeben.

In Bezug auf den *Zeitpunkt der Feedbackgabe* wurde sowohl das einfache als auch das elaborierte Feedback *sofort* nach dem Erstellen der Maps gegeben. Eine Verzögerung wäre aufgrund der Komplexität des Aufgabenformats nicht sinnvoll gewesen. Die Informationen im Rahmen des elaborierten Feedbacks werden, wie von Narciss und Huth (2006) empfohlen, gemäß der Cognitive-Load-Theorie *sequentiell* gegeben. Die jeweiligen Propositionen der Referenznetze werden systematisch nacheinander am OHP aufgebaut. Aufgrund der Komplexität des Aufgabenformats wurde vorab die richtige Lösung, d. h. das einfache Feedback, präsentiert.

Beide Feedbackarten werden *absichtlich* und *explizit* gegeben.

Bezüglich des Präsentationsformats wurden im Kapitel 2.1.3 mündliches und schriftliches Feedback sowie textbasiertes, verbales und computergestütztes Feedback unterschieden. Das im Rahmen der Studie gegebene einfache Feedback erfolgt *schriftlich* in Form des Referenznetzes. Das elaborierte Feedback wird zum einen *mündlich* in Form eines Lehrer-Schüler-Gesprächs gegeben. Zum anderen ist es auch *schriftlich*, da die Schülermaps sowie die drei Referenznetze mittels OHP-Folie besprochen werden.

Die in der Feedbackliteratur vorherrschende Unterscheidung zwischen textbasiertem, verbalem und computergestütztem Feedback ist nicht sinnvoll, insbesondere wenn die theoretischen Hintergründe zum multimedialen Lernen von Mayer berücksichtigt werden (vgl. Kapitel 3.1.3). Das einfache Feedback dieser Studie wird mittels eines Arbeitsblattes als Ausgabemedium als depiktionales bzw. piktoriales Format präsentiert und spricht somit das Auge als Sinnesorgan an. Das elaborierte Feedback wird mittels der auf OHP-Folien kopierten Schülermaps und der Referenznetze sowie der Stimme der Lehrpersonen als Ausgabemedien sowohl deskriptional als auch depiktional bzw. verbal und piktorial übermittelt und spricht somit das Auge und das Ohr als Sinnesorgane an.

Die Vorteile der bimodalen Präsentation im Rahmen des computergestützten Feedbacks lassen sich somit auf diese Studie übertragen. Beim elaborierten Feedback werden gleichzeitig piktoriale Informationen visuell und verbale Informationen auditiv wahrgenommen und somit die kognitive Belastung, gegenüber einer ausschließlichen Präsentation in visueller Form, möglicherweise reduziert.

Hinsichtlich der *Anzahl der Lösungsvorschläge* wird lediglich ein Versuch angeboten (*STF*). Im Anschluss an das Erstellen der Maps wird den Schülern die richtige Lösung präsentiert.

Im Kapitel 2.1.1 sowie 2.2.5 wurde bereits auf die Bedeutung des Feedbacks im Rahmen einer *formativen Leistungsbeurteilung* hingewiesen. Im Anschluss an das einfache Feedback ist es Aufgabe der Schüler, ihre erstellten Maps selbst in Bezug auf Richtigkeit zu beurteilen und zu korrigieren. Die Leistung wird somit mittels Selbstbeurteilung diagnostiziert und der Lernprozess an den aktuellen Könnensstand (*individuelle Bezugsnorm*) sowie das Referenzmap (*sachliche Bezugsnorm*) angepasst. Insgesamt wäre das einfache Feedback als Referenzmap ohne dem Korrigieren daher nicht lernförderlich.

Während des Lehrer-Schüler-Gesprächs im Rahmen des elaborierten Feedbacks können die Schüler ihr korrigiertes Map mit den Ausführungen bzw. dem am OHP sukzessive

aufgebauten Referenzmap vergleichen, was die *sachliche Bezugsnorm* widerspiegelt. Infolgedessen wird der Lernprozess reguliert. Fehler werden zugelassen. Die Lehrpersonen waren insgesamt angehalten, auf *Fehlerfreundlichkeit* zu achten.

Das einfache Feedback inklusive der selbständigen Mapkorrektur sowie das elaborierte Feedback können möglicherweise den weiteren Lehr-Lernprozess unterstützen. Dies kann sich zum einen in den Spielzügen des Planspiels widerspiegeln. Das Feedback zum 1. Mapping im Anschluss an das erste Geschäftsjahr kann möglicherweise das Verständnis der Inhalte zum Berechnen der Abschreibung sowie zum Erstellen des Jahresabschlusses verbessern. Infolgedessen werden ggf. qualitativ bessere Maps beim 2. Mapping erstellt. Dies wiederum erleichtert möglicherweise die Bearbeitung des zweiten Geschäftsjahres und erhöht die Qualität der Maps bei der 3. Mapping-Aktivität. Das aktive Auseinandersetzen mit den Lerninhalten im Rahmen des Feedbacks optimiert also insgesamt den Lernprozess.

Erster und zweiter Nachtest dienen der abschließenden *summativen Leistungsbeurteilung*. Wie bereits im Kapitel 2.1.1 erläutert, sollte dabei das Abprüfen von konzeptuellem Wissen fokussiert werden, um das Verständnis grundlegender Begriffszusammenhänge zu kontrollieren. Infolgedessen lässt sich eine summative Leistungsbeurteilung auch formativ nutzen. Bei der Konstruktion der Tests wurden diese Aspekte berücksichtigt (vgl. Kapitel 4.4.1).

Während des Lehrer-Schüler-Gesprächs im Rahmen des elaborierten Feedbacks konnten jederzeit die Ursachen für begangene Fehler erforscht sowie Fehlkonzepte identifiziert werden (*Quellen für Fehlkonzepte und Fehler*).

Das Erstellen der Concept Maps diene als Grundlage für das Geben von Feedback und somit als *Lernaufgabe* dieser Studie. Die Selbstkonstruktion von Concept Maps stellt hohe kognitive Anforderungen an die Lernenden und kann somit als ein *komplexes Aufgabenformat* angesehen werden. Da bei einfachen Anforderungen auch einfaches Feedback zu genügen scheint (vgl. Kapitel 2.2.5), stellt ein komplexes Aufgabenformat eine Grundvoraussetzung dar, um die Lernwirksamkeit des elaborierten im Vergleich zum einfachen Feedback analysieren zu können.

Im Kapitel 2.2.5 wurde bereits darauf hingewiesen, dass eine natürliche Lernumgebung bei einer hohen Aufgabenkomplexität gegenüber computergestützten Lernumgebungen lernwirksamer zu sein scheint. Die *Art der Lernsituation* im Rahmen dieser Studie stellt

ebenso eine *natürliche Umgebung* dar. Das Planspiel „Easy BusinessTM“ selbst ist ein haptisches Planspiel. Die integrierten Instruktionsphasen wurden als Lehrer-Schüler-Gespräche gestaltet und mit Medien wie OHP, Flipchart, Tafel und Metaplanwand unterstützt. Die Lernumgebung wurde fehleroffen gestaltet. Den Lernenden wurde es somit ermöglicht, in einer *fehlerfreundlichen* Lernumgebung zu agieren. Aus diesem Grund kann auch davon ausgegangen werden, dass die *Fehlerangst* minimiert wurde. Die Variable wurde jedoch nicht explizit erhoben. Der *Umgang mit Fehlern* wird auf Basis des von Spychiger u. a. (2006) entwickelten Fragebogens erhoben (vgl. Kapitel 4.3.3).

In Bezug auf die *Normtransparenz* sollten die Lehrpersonen soziale Normen und Regeln transparent vermitteln. Im Rahmen der Studie wurden die Schüler vorab durch die Schule belehrt, dass sie sich wie im regulären Unterricht verhalten sollten. Darüber hinaus trug die Durchführung in den gewohnten Klassenräumen dazu bei, dass die Schüler kein abweichendes Verhalten zeigten. Insgesamt wurden jedoch einige Regeln gelockert. Da die Studie viele Gruppenarbeitsphasen beinhaltete, herrschte eine lockere Atmosphäre, die Schüler waren angehalten zu interagieren. Während der Instruktionsphasen konnten die Schüler z. B. auch ohne Meldung Antworten geben, ohne Ermahnungen fürchten zu müssen.

Die Schüler hatten keinen Migrationshintergrund. Aus diesem Grund mussten sich die Lehrpersonen nicht auf andere Kulturen, beispielsweise im Umgang mit Fehlern, einstellen. Die *kulturelle Rahmung* hatte daher für die vorliegende Studie keine Bedeutung.

Der *formale Status des Feedbacksenders* wird im Rahmen der Studie nicht untersucht. Die Lehrpersonen waren Studierende der Wirtschaftspädagogik. Der formale Status kann in allen Versuchsgruppen als vergleichbar angesehen werden.

Der *informelle Status des Feedbacksenders* wird ebenfalls nicht direkt erhoben. Im Kapitel 2.2.4 wurde beschrieben, dass die zugeschriebene Kompetenz z. B. über die eingeschätzte Adäquatheit des Feedbacks (Fairness, Nützlichkeit und Akzeptanz) gemessen werden kann. Im Rahmen dieser Studie wird daher die eingeschätzte Nützlichkeit des Feedbacks als Einflussvariable mit erhoben (vgl. Kapitel 4.3.3).

In Bezug auf *den informellen Status des Feedbackempfängers* wird das *Vorwissen* über die Ergebnisse des Vortests (vgl. Kapitel 4.3.1) und die *Sprachfähigkeit* über die Ergebnisse des Wilde-Intelligenztest-2 (Kersting u. a., 2008; vgl. Kapitel 4.3.2) operationalisiert.

Die Faktoren Antwortsicherheit, Attribution, Selbstwirksamkeit, Big 5 sowie Lernmotivation wurden im Rahmen der Studie nicht berücksichtigt.

Hinsichtlich der *zwischenmenschlichen Nähe* im Rahmen vertikaler Konstellationen wurde in Kapitel 2.2.4 darauf hingewiesen, dass auch aufwärtsgerichtete Kommunikation ermöglicht werden sollte. Die Lehrer-Schüler-Gespräche im Rahmen des elaborierten Feedbacks wurden so gestaltet, dass der Lernende ohne Hemmungen auch Fragen an die Lehrpersonen richten konnte.

4.3 Erhebungsinstrumente

4.3.1 Wissenstest als Paralleltest

Paralleltest und Paralleltestreliabilität

Die verwendeten Wissenstests waren als Paralleltests konzipiert. Paralleltests werden auch als Alternativ- oder Äquivalenztests bezeichnet. Sie bestehen aus zwei gleichwertigen Testversionen ein und desselben Tests (Parallelform A und B) und stellen somit Operationalisierungen desselben Konstrukts dar. Die jeweiligen Aufgaben der Formen A und B ähneln sich inhaltlich und formal, sie stellen Itemzwillinge dar. Mit dem Einsatz eines Paralleltests können Übungs- und Erinnerungseffekte aus dem Vortest verhindert werden (Lienert & Raatz, 1998, S. 297; Döring & Bortz, 2016, S. 466). Aufgrund der Dauer der Studie von lediglich zwei Tagen können durch den Einsatz eines Paralleltests somit die Übungs- und Erinnerungseffekte reduziert werden.

Inwieweit beide Formen tatsächlich vergleichbar sind, kann anhand der Paralleltestreliabilität bestimmt werden. Zur Überprüfung der Testreliabilität schlagen Lienert und Raatz (1998, S. 180 ff.) folgendes Vorgehen vor:

- Eine Stichprobe erhält zwei Parallelformen eines Tests sofort hintereinander.
- Eine Hälfte der Stichprobe erhält dabei zunächst Form A, die andere Form B.
- Nach Art eines Überkreuzungsplans wird bei der Wiederholung entsprechend gewechselt.
- Die Rohwertpaare werden korreliert.

Bei der Berechnung des Reliabilitätskoeffizienten (r_{tt}) schlagen Lienert und Raatz (1998, S. 181) folgende Formel vor:

$$r_{tt} = \frac{N \sum_i X_{1i} X_{2i} - \sum_i X_{1i} \sum_i X_{2i}}{\sqrt{[N \sum_i X_{1i}^2 - (\sum_i X_{1i})^2][N \sum_i X_{2i}^2 - (\sum_i X_{2i})^2]}}$$

Dabei stellen X_1 die Rohwerte der Testform A, X_2 die Rohwerte der Testform B und N die Anzahl der Probanden dar.

Zur Überprüfung der Paralleltestreliabilität wurde ein auf Basis der Erfahrungen vorangegangener Untersuchungen optimierter Wissenstest eingesetzt (Ryssel & Fürstenau, 2011; Fürstenau u. a., 2010; Sommer u. a., 2009). Beispielsweise wurde eine Aufgabe zur GuV mit aufgenommen, um die im Planspiel behandelten Fachinhalte umfassender zu berücksichtigen.

Die Überprüfung der Paralleltestreliabilität wurde entsprechend des vorgeschlagenen Vorgehens von Lienert und Raatz mit 40 Auszubildenden im ersten Ausbildungsjahr vorgenommen (Bürokaufmann, Kaufmann für Bürokommunikation und Kaufmann für Einzelhandel). Insgesamt konnte ein Reliabilitätskoeffizient von $r_{tt}=0,79$ erzielt werden. Döring und Bortz (2016, S. 83) geben an, dass in der Forschung ein Wert von 0,8 oft als Mindestmaß für Reliabilitätskoeffizienten angesehen wird. Sie betonen, dass die Beurteilung jedoch unter Berücksichtigung der Besonderheiten der jeweiligen Studie erfolgen muss. Lienert und Raatz (1998, S. 14) stellen heraus, dass ein Reliabilitätskoeffizient zwischen 0,5 und 0,7 bereits genügt, insofern der Test als Forschungsinstrument zum Vergleich von Gruppen dienen soll. Infolgedessen kann von vergleichbaren Testformen ausgegangen werden.

Um für die Hauptuntersuchung ein zuverlässiges Testinstrument zu erhalten, wurden zusätzlich die erreichten Punkte jeder Aufgabe über alle Probanden zwischen den Formen A und B verglichen. Für einige Aufgaben waren deutliche Unterschiede erkennbar. Infolgedessen wurden u. a. die Bewertungen einiger Aufgaben und somit das Codebuch angepasst. Beispielsweise wurde die Frage nach den Erwartungen des Lieferanten an das Unternehmen (Form B) gegenüber der Form A (Erwartungen des Unternehmens an die Kunden) zu streng bewertet (Frage 2). Des Weiteren zeigte sich bei einer Frage zum Wertschöpfungsprozess, dass die Probanden eher in der Lage waren, Folgen dafür abzuleiten, dass ein Lieferant das Material nicht liefert, als dass der Kunde die Rechnung nicht bezahlt. Für die Hauptuntersuchung wurde daher für beide Formen die fehlende Materiallieferung verwendet, nur die Bezeichnung des Materials (Sägeblätter und Bohrerensätze) variiert.

Die endgültig entwickelte Testform A wurde als Vortest, die Testform B als 1. Nachtest eingesetzt. Die Tests sind in den Anhängen 4.3.1-1 und 4.3.1-2 dargestellt. Für den

2. Nachtest wurden Fragen aus dem Vortest (Fragen 1, 2, 9) bzw. aus dem 1. Nachtest (Fragen 3, 8, 11) verwendet. Für die Rechenaufgaben wurden im Vergleich zu Form A und B andere Zahlenwerte eingesetzt (Aufgaben 5, 6 und 7). Der erste Teil der Aufgabe 4 bzw. 10 wurde für den 2. Nachtest aus dem 1. Nachtest bzw. Vortest übernommen, in der zweiten Teilaufgabe sollten die Probanden den jeweils dargestellten Prozess noch einmal verbal beschreiben (vgl. Anhang 4.3.1-3).

Aufbau der Wissenstests in Bezug auf Taxonomie von Anderson u. a. (2001)

Vortest, 1. und 2. Nachtest bestanden aus insgesamt elf Aufgaben und enthielten Aufgabenformate mit freien Antwortformulierungen⁶⁶. Die Aufgaben prüften dabei die mit dem Planspiel verbundenen fachlichen Lernziele (vgl. Anhang 4.2.1-7) ab. Nach Klauer (1987, S. 3 f.) handelt es sich daher um kriteriumsorientierte Tests, wobei eine Lernzielnorm vorliegt. Bei Leistungstests werden die leichten Fragen aus motivationalen Gründen häufig an den Anfang gestellt. Die Schwierigkeit nimmt dann im Laufe des Tests zu (Pospeschill, 2010, S. 68). Bei der für diese Untersuchung vorliegenden Zielgruppe muss jedoch von einer sinkenden Konzentration während des Tests ausgegangen werden. Aus diesem Grund wurden die schwierigsten Aufgaben ungefähr in der Mitte des Tests platziert. Die Schwierigkeit nimmt zum Ende hin wieder ab.

Für die Entwicklung der Aufgaben wurde die Lernzieltaxonomie nach Anderson und Krathwohl (2001) genutzt.⁶⁷ Die verwendete Lernzieltaxonomie besteht aus zwei Dimensionen, der Wissensdimension, welche sich auf den Inhaltsaspekt von Lernzielen bezieht, und der Dimension der kognitiven Prozesse, welche den Verhaltensaspekt repräsentiert (Anderson & Krathwohl, 2001, S. 28 ff.). Die Wissensdimension besteht aus vier Kategorien: „Faktenwissen“, „konzeptuelles Wissen“, „prozedurales Wissen“ und „metakognitives Wissen“. Die Dimension der kognitiven Prozesse beinhaltet sechs Kategorien: „Erinnern“, „Verstehen“, „Anwenden“, „Analysieren“, „Bewerten“ und „Erschaffen“. Insgesamt zeichnet sich die Dimension der kognitiven Prozesse durch eine steigende

⁶⁶ Bei Aufgaben mit freier Antwortformulierung können die Probanden die Lösung mit eigenen Worten formulieren (Lienert & Raatz, 1998, S. 20). Dabei werden Ergänzungsaufgaben, Kurzantwortformen und Kurzaufsatzformen unterschieden (Ingenkamp & Lissmann, 2008).

⁶⁷ Die Taxonomie stellt eine Weiterentwicklung der Taxonomie von Bloom (1956) aufgrund empirischer Untersuchungen und Praxiserfahrungen dar. Die Taxonomie stieß auf internationale Resonanz und dient bis heute als anwendungsorientiertes Hilfsmittel für die Lernzielplanung, die Durchführung von Unterricht sowie die Evaluation der Lernziele (Moseley u. a., 2005). Infolgedessen stellt sie auch ein geeignetes Instrument zur Aufgabenkonstruktion im Rahmen der Testentwicklung dar.

Komplexität aus. Die Hauptkategorien lassen sich noch in weitere Subkategorien unterteilen. Beide Dimensionen werden in einer Matrixdarstellung verbunden (4x6). Für die vorliegende Untersuchung wurden jeweils die ersten drei Kategorien berücksichtigt. Dies kann mit der Zielgruppe sowie der Dauer der Untersuchung begründet werden. Die Probanden lernen während des zweitägigen Planspielunterrichts grundlegende betriebswirtschaftliche Zusammenhänge kennen. In diesem kurzen Zeitraum können die Themenbereiche lediglich überblicksartig behandelt werden. Infolgedessen kann davon ausgegangen werden, dass die Probanden im Anschluss an den Planspielunterricht Wissen erinnern, verstehen und anwenden können, jedoch keine höheren Stufen erreicht werden. Vor dem Hintergrund der zu prüfenden Hypothesen H23 bis H26 stellen das „Faktenwissen erinnern“ sowie das „Konzeptuelle Wissen erinnern“ einfache Aufgaben dar, womit Lernziele eines niedrigen Anspruchsniveaus erreicht werden können. Das „Konzeptuelle Wissen verstehen“ sowie das „Prozedurale Wissen anwenden“ repräsentieren komplexe Aufgaben, womit Lernziele eines hohen Anspruchsniveaus erreichbar sind.

Die Zuordnung der Aufgaben der Wissenstests zur Taxonomie nach Anderson und Krathwohl (2001) ist in nachfolgender Tabelle 9 dargestellt.

Tabelle 9: Einordnung der Wissenstestaufgaben in die Lernzieltaxonomie nach Anderson und Krathwohl (2001) für den Vortest und 1. Nachtest

		Dimension der kognitiven Prozesse					
		Erinnern	Verstehen	Anwenden	Analysieren	Bewerten	Erschaffen
Wissensdimension	Faktenwissen	Aufgabe 1 Aufgabe 7a					
	konzeptuelles Wissen	Aufgabe 2 Aufgabe 5a Aufgabe 6a Aufgabe 10	Aufgabe 3 Aufgabe 4 Aufgabe 8 Aufgabe 9 Aufgabe 11				
	prozedurales Wissen			Aufgabe 5b Aufgabe 6b Aufgabe 7b			
	metakognitives Wissen						

In Frage 1 galt es, sich an die Bezeichnung „Material“ (Vortest) bzw. „Produkt“ (1. Nachtest) zu erinnern. In Frage 7a sollte die Formel zur Berechnung der linearen Abschreibung wiedergegeben werden. Insgesamt sollte folglich terminologisches Wissen aus dem Gedächtnis abgerufen werden, so dass die beiden Fragen zu den Kategorien „Faktenwissen erinnern“ zugeordnet werden können. In Frage 2 sollte erläutert werden, was ein Unternehmen von seinen Kunden erwartet (Vortest) bzw. was ein Unternehmen von seinen Lieferanten erwartet (1. Nachtest). In Aufgabe 5a sollten vorgegebene Bestandteile der

GuV entweder dem Aufwand oder dem Ertrag zugeordnet werden. Aufgabe 6b beinhaltete das Zuordnen von Bestandteilen der Bilanz zur Aktiv- bzw. Passivseite. Bei Aufgabe 10 galt es, fehlende Elemente des Wertschöpfungsprozesses in einem lückenhaften Flussdiagramm zu ergänzen.⁶⁸ Insgesamt stand somit das Erinnern an Kategorien und Klassifikationen sowie ihren Zusammenhängen innerhalb einer größeren Struktur im Vordergrund (Anderson & Krathwohl, 2001, S. 48). Daher können die Aufgaben 2, 5a, 6a und 10 den Kategorien „Konzeptuelles Wissen erinnern“ zugeordnet werden. Die Frage 3 beinhaltete das Zuordnen eines Geräts zur Herstellung von Schlagbohrmaschinen (Vortest) bzw. eines Förderbands (1. Nachtest) zur Bilanz, wobei die Bilanzseite sowie die Bilanzposition genannt werden sollten. Hierbei war es notwendig, dass die Probanden den Aufbau der Bilanz verstanden hatten. Nur dann war es möglich, korrekte Antworten zu geben. Aufgabe 4 beinhaltete das Vervollständigen eines Lückenmaps zu den Geldströmen zwischen dem Unternehmen und seinen Lieferanten.⁶⁹ Dies setzte voraus, dass die Probanden die Hintergründe der getätigten Spielzüge verstanden hatten. Frage 8 stellte die Situation dar, dass ein Lieferant bestimmte Materialien nicht liefert. Die Probanden sollten mögliche Folgen ableiten. In Frage 9 sollten die Probanden Fehler in einem Text zu den Abläufen im Wertschöpfungsprozess erkennen und korrigieren. Beide Fragen setzten das Verständnis des Wertschöpfungsprozesses voraus. In Frage 11 galt es, für einen gegebenen Geschäftsvorfall zu erkennen, dass durch eine Vorortbezahlung des Lieferanten (Vortest) bzw. des Kunden (1. Nachtest) Kasse und Bankguthaben sinken bzw. steigen. Die Fragen 3, 4, 8, 9 und 11 lassen sich daher den Kategorien „Konzeptuelles Wissen verstehen“ zuordnen. In Frage 5b wurde ausgehend von den in 5a durchgeführten Zuordnungen der Verlust (Vortest) bzw. Gewinn (1. Nachtest) berechnet. In Frage 6b sollten ausgehend von den vorgenommenen Zuordnungen in Aufgabe 6a die Bilanzsumme sowie das Fremdkapital (Vortest) bzw. Eigenkapital (1. Nachtest) errechnet werden. Frage 7b beinhaltete das Berechnen des Abschreibungsbetrages. Voraussetzung für das korrekte Beantworten dieser Fragen war es, fachspezifische Fähigkeiten und Algorithmen anwenden zu können. Die Probanden mussten in der Lage sein, die notwendigen Schritte zum Lösen

⁶⁸ Im 2. Nachtest wurde die Aufgabe 10 in 10a und 10b unterteilt. Die Probanden sollten den Wertschöpfungsprozess zusätzlich zum Ergänzen des Flussdiagramms verbal beschreiben. Auf diese Weise sollte der langfristige Behaltenseffekt analysiert werden. Die Aufgabe 10b des 2. Nachtests lässt sich somit eher den Kategorien „Konzeptuelles Wissen verstehen“ zuordnen.

⁶⁹ Im 2. Nachtest wurde die Aufgabe 4 in 4a und 4b unterteilt. Die Probanden sollten den im Lückenmap vervollständigten Prozess noch einmal beschreiben. Auf diese Weise sollte die langfristige Behaltensleistung überprüft werden. Die Aufgabe 10b des 2. Nachtests lässt sich ebenfalls den Kategorien „Konzeptuelles Wissen verstehen“ zuordnen.

der Rechenaufgaben auszuführen (Anderson & Krathwohl, 2001, S. 52 f.; 77. f.). Daher können die Aufgaben 5b, 6b und 7b den Kategorien „Prozedurales Wissen anwenden“ zugeordnet werden.

4.3.2 Sprachfähigkeitstest

Zur Erhebung der Sprachfähigkeit wurde vor der Untersuchung ein Modul des Wilde-Intelligenz-Tests 2 [WIT-2] von Kersting, Althoff und Jäger (2008) eingesetzt. Wilde begann 1954 mit den Entwürfen und der Arbeit an der Entwicklung des Wilde-Intelligenz-Tests, welcher nach seinem Tod von einem Psychologenteam der Deutschen Gesellschaft für Personalwesen e. V. weitergeführt wurde. Nach circa 40 Jahren wurde nach einer grundlegenden Analyse des Tests eine zweite, vollständig überarbeitete Version erarbeitet. Diese zweite Version (WIT-2) erfasst differenziert klar unterscheidbare und theoretisch fundierte kognitive Fähigkeiten von Jugendlichen und Erwachsenen. Der Test kann z. B. im Bereich der Intelligenzdiagnostik oder Eignungsdiagnostik (Personalauswahl und –entwicklung) eingesetzt werden. Weiterhin hat er als Kontrollvariable für Bildungsstandmessungen eine hohe Bedeutung für die empirische Pädagogik (Kersting u. a., 2008).

Der WIT-2 umfasst 8 Module mit insgesamt 11 Subtests zur Messung unterschiedlicher intellektueller Fähigkeiten (vgl. Tabelle 10).

Tabelle 10: Aufbau des WIT-2

Modul	Subtest
1: Sprachliches Denken	AL - Analogien
	GW - Gleiche Wortbedeutung
2: Rechnerisches Denken	GR - Grundrechnen
	ER - Eingekleidete Rechenaufgaben
3: Räumliches Denken	AW - Abwicklungen
	SP - Spiegelbilder
4: Schlussfolgerndes Denken	AL - Analogien
	ZN - Zahlenreihen
	AW - Abwicklungen
5: Merkfähigkeit	MF - Merkfähigkeit
6: Arbeitseffizienz	EM - E-Mails bearbeiten
7: Wissen Wirtschaft	WW - Wissen Wirtschaft
8: Wissen Informationstechnologie	WI - Wissen Informationstechnologie

Dabei kann jedes Modul einzeln oder in Kombination mit anderen erfasst werden. Die beiden Subtests Analogien und Abwicklungen haben eine Doppelfunktion und kommen in jeweils zwei Modulen vor. Die Analogien werden zum Erfassen des sprachlichen sowie des schlussfolgernden Denkens, die Abwicklungen zum Erfassen des räumlichen sowie des schlussfolgernden Denkens eingesetzt (Kersting u. a., 2008, S. 30).

Im Rahmen der Untersuchung wurde das Modul 1 des WIT-2 eingesetzt (Dauer: 12 Minuten). Dieses Modul konzentriert sich auf das sprachliche Denken und umfasst die Subtests Analogien (AL) und Gleiche Wortbedeutung (GW). Dabei soll die Fähigkeit erfasst werden, mit sprachlichen Konzepten umzugehen, wobei Sprachverständnis, sprachlogisches Denken und der Wortschatz Beachtung finden (Kersting u. a., 2008, S. 29). Beide Subtests umfassen jeweils 20 Multiple-Choice-Aufgaben.

Bei den Analogieaufgaben sind auf der linken Seite eines Gleichheitszeichens zwei Wörter vorgegeben, welche eine bestimmte Beziehung zueinander widerspiegeln (z. B. Stadt: Grünanlage). Auf der rechten Seite soll eines von fünf Wörtern ausgewählt werden, das eine analoge Beziehung herstellt (z. B. Wasser, Welle, Schiff, Insel oder Ufer; richtig: Insel). Beim Subtest „Gleiche Wortbedeutung“ gilt es, zu einem vorgegebenen Wort (z. B. verbinden) aus fünf Vorschlägen das sinnähnlichste Wort herauszusuchen (z. B. heiraten, ausgleichen, kombinieren, vermitteln oder analysieren; richtig: kombinieren).

4.3.3 Fragebogen zur Einschätzung des Umgangs mit Concept Mapping, der Nützlichkeit des Feedbacks sowie des Umgangs mit Fehlern

Am Ende des zweiten Tages der Untersuchung wurde allen vier Gruppen ein Fragebogen vorgelegt, welcher in Abhängigkeit vom jeweiligen Treatment variierte.

Die Probanden gaben ihre Antworten mit Hilfe einer vierstufigen Ratingskala (trifft nicht zu, trifft eher nicht zu, trifft etwas zu, trifft zu). Um zu vermeiden, dass die Probanden eine neutrale Position einnehmen, wurde eine gerade Anzahl von Antwortkategorien gewählt und auf die Kategorie weder/noch verzichtet. Somit lässt sich auch die Validität steigern (Raab, Poost, & Eichhorn, 2008, S. 55).

Nachfolgend werden die jeweiligen Teil-Fragebögen erläutert.

Teil-Fragebogen zur Einschätzung des Umgangs mit Concept Mapping

Den Gruppen *kein FB*, *EiFB* sowie *ElFB* wurde ein Fragebogen zur Bewertung der Concept Mapping-Methode vorgelegt. Die Probanden sollten hierbei Aussagen zur Akzeptanz der Methode des Concept Mapping treffen. Der Teil-Fragebogen zum Umgang mit Concept Mapping orientiert sich an einem von Ryssel, Fürstenau und Thieme (2018) in Anlehnung an Stracke (2004, S. 151) und Eckert (1998) entwickelten Fragebogen, der sich in die Bereiche Interessantheit, Nützlichkeit und Handhabbarkeit unterteilt. Zusätzlich wurde noch ein Item eingefügt (Frage 6), welches überprüfen sollte, inwieweit die Probanden das zusätzliche Erstellen der Concept Maps nützlich fanden, d. h. ob ihnen die Inhalte schon durch das Planspiel deutlich geworden sind oder nicht. Die Fragerichtung der Items wurde variiert. Die Fragen 4, 5, 6 sowie 9 und 10 (bei der Gruppe *kein FB* Frage 10 und 11⁷⁰) sind negativ gepolt, um bestimmte Antworttendenzen durch Ermüdungsercheinungen zu vermeiden (Mummendy & Grau, 2008, S. 70 ff.). Der Teil-Fragebogen ist in nachfolgender Tabelle 11 dargestellt. Die Fragen 1, 4 und 7 lassen sich dem Bereich Interessantheit zuordnen, die Fragen 2, 3 und 6 beziehen sich auf den Bereich Nützlichkeit und die Fragen 5, 8, 9 und 10 repräsentieren den Bereich Handhabbarkeit.

Tabelle 11: Teil-Fragebogen zur Einschätzung des Umgangs mit Concept Mapping

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
1) Das Erstellen der Concept Maps hat mir Spaß gemacht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2) Durch das Erstellen der Concept Maps sind mir meine Wissenslücken bewusst geworden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3) Durch das Erstellen der Concept Maps konnte ich feststellen, ob ich die Zusammenhänge zwischen den einzelnen Begriffen wirklich verstanden habe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4) Das Erstellen der Concept Maps fand ich langweilig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5) Beim Erstellen der Concept Maps musste ich mich anstrengen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6) Das Erstellen der Concept Maps empfand ich als überflüssig, da mir die Inhalte schon durch das Planspiel deutlich geworden sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7) Ich habe mir beim Erstellen der Concept Maps große Mühe gegeben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8) Das Erstellen der Concept Maps ist mir insgesamt leicht gefallen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9) Beim Erstellen der Concept Maps fand ich es schwierig, die Begriffe auf dem Papier in eine sinnvolle Ordnung zu bringen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10) Beim Erstellen der Concept Maps fand ich es schwierig, mich zu entscheiden, welche Begriffe ich miteinander verbinde.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Interessantheit: Items 1, 4 und 7; Nützlichkeit: Items 2, 3 und 6; Handhabbarkeit: Items 5, 8, 9 und 10

⁷⁰ Bei der Gruppe *kein FB* wurde gefragt, inwieweit sich die Probanden zusätzliches Feedback gewünscht hätten. Da diese Gruppe kein Feedback erhielt, wurde dieses Item als Frage 7 in den Fragebogen zum eingeschätzten Umgang mit Concept Mapping integriert.

Teil-Fragebogen zur Einschätzung der Nützlichkeit des Feedbacks

Um herauszufinden, ob die Probanden das Feedback nützlich fanden bzw. inwieweit sie sich zusätzliches Feedback gewünscht hätten, wurden bei den drei Experimentalgruppen weitere Items erhoben.

Zunächst sollten die Gruppen *EiFB* und *EIFB* einschätzen, wie nützlich sie das Vergleichen ihrer Concept Maps mit den Referenzmaps fanden und somit die Nützlichkeit des einfachen Feedbacks beurteilen. Die ersten beiden Fragen lehnen sich bei der Formulierung an den Bereich Nützlichkeit des Fragebogens zum Umgang mit Concept Mapping an. Des Weiteren wurde ein Item eingefügt, welches Gründe erforschen sollte, warum die Probanden das einfache Feedback ggf. als überflüssig einschätzen (vgl. Tabelle 12).

Tabelle 12: Fragen zur Einschätzung der Nützlichkeit des einfachen Feedbacks

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
11) Durch das Vergleichen meiner Concept Maps mit den Expertenmaps sowie dem anschließenden Korrigieren sind mir meine Wissenslücken bewusst geworden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12) Das Vergleichen meiner Concept Maps mit den Expertenmaps sowie das anschließende Korrigieren halfen mir, die Zusammenhänge zwischen den Begriffen besser zu verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13) Das Vergleichen mit den Expertenmaps empfand ich als überflüssig. <i>Falls Sie gerade "trifft eher zu" oder "trifft voll zu" angekreuzt haben, entscheiden Sie zwischen a, b oder c. Ansonsten geht es mit Frage 14 weiter.</i> Gründe:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
a. Mir waren die Zusammenhänge auch schon nach dem Erstellen der Concept Maps klar. ODER	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Trotz des Vergleichens habe ich die Beziehungen zwischen den Begriffen nicht verstanden. ODER	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. anderer Grund:.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Die Gruppe *EiFB* sollte darüber hinaus einschätzen, inwieweit sie sich eine zusätzliche Fehlerbesprechung und somit elaboriertes Feedback gewünscht hätten (vgl. Tabelle 13).

Tabelle 13: Frage zum gewünschten elaborierten Feedback (Gruppe einfaches Feedback)

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
14) Ich hätte mir eine zusätzliche Besprechung der Fehler gewünscht, da mir dies gezeigt hätte, was die Gründe für meine Fehler waren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Die Gruppe *EIFB* sollte Aussagen bezüglich der Nützlichkeit der zusätzlichen Fehlerbesprechung treffen. Hierbei galt es einzuschätzen, inwieweit das elaborierte Feedback das

Aufdecken von Fehlerursachen sowie das Verständnis der Zusammenhänge unterstützte. Des Weiteren wurde auch hier ein Item hinzugefügt, mit dem Ziel, Gründe zu erforschen, die erklären, warum die Probanden das elaborierte Feedback ggf. als überflüssig einschätzen (vgl. Tabelle 14).

Tabelle 14: Fragen zur Einschätzung des elaborierten Feedbacks

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
14) Durch das zusätzliche Besprechen der Fehler sowie die Erklärungen konnte ich die Ursachen meiner Fehler finden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15) Die Fehlerbesprechung sowie die ergänzenden Erklärungen waren überflüssig. <i>Falls Sie gerade "trifft eher zu" oder "trifft voll zu" angekreuzt haben, entscheiden Sie zwischen a, b oder c. Ansonsten geht es mit Frage 16 weiter.</i> <u>Gründe:</u> a. Mir waren die Zusammenhänge auch schon nach dem Vergleichen mit den Expertenmaps klar. b. Trotz der Fehlerbesprechung habe ich die Beziehungen zwischen den Begriffen nicht verstanden. c. anderer Grund.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16) Die zusätzliche Fehlerbesprechung sowie die Erklärungen halfen mir dabei, die Beziehungen zwischen den Begriffen zu verstehen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Da die Gruppe *kein FB* kein zusätzliches Feedback erhielt, wurde lediglich gefragt, inwieweit sie sich eine Rückmeldung darüber gewünscht hätten, was sie richtig und was sie falsch gemacht haben (vgl. Tabelle 15). Hierbei sollten Rückschlüsse auf ein eventuell fehlendes Feedback gezogen werden. Diese Frage wurde als Frage 7 in den Teil-Fragebogen zum Umgang mit Concept Mapping integriert.

Tabelle 15: Frage zum gewünschten Feedback (Gruppe kein Feedback)

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
7) Nach dem Erstellen der Concept Maps hätte ich mir noch eine Rückmeldung darüber gewünscht, was ich richtig und was ich falsch gemacht habe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Teil-Fragebogen zum Umgang mit Fehlern

Der Teil-Fragebogen wurde in Anlehnung an Spychiger, Oser, Mahler und Hascher (1998) sowie Spychiger, Kuster und Oser (2006) für alle vier Versuchsgruppen eingesetzt, um einzuschätzen, wie die Schüler und Lehrpersonen im Allgemeinen mit Fehlern umgehen. Dazu wurde die Skala „Lernorientierung“ von Spychiger u. a. (1998, 2006) übernommen. Zum einen stellte diese Dimension bei Spychiger u. a. (1998, 2006) die gewichtigste Dimension dar und zum anderen kann mit dieser Skala ein Zusammenhang zwischen dem Umgang der Probanden mit Fehlern und der Lernwirksamkeit des Feedbacks hergestellt werden. Die drei anderen Dimensionen Fehlerangst, Fehlerfreundlichkeit und Normtransparenz wurden bei der Gestaltung des Feedbacks beachtet (vgl. Kapitel 2.1.1 bzw. 4.2.3).

Spychiger u. a. (1998) führten eine Befragung mit Schülern vierter bis neunter Klassen aus der Schweiz zur Fehlerkultur durch. Der ursprüngliche Fragebogen mit zehn theoretischen Dimensionen wurde mittels einer Faktorenanalyse auf drei Faktoren reduziert: „Lehrerverhalten“, „Selbstfaktor kognitiv“ und „Selbstfaktor emotional“. Die Skala „Selbstfaktor kognitiv“ enthält neun Items zur Selbsteinschätzung im Umgang mit Fehlern im kognitiven Bereich und stellt den für die hier vorliegende Studie relevanten Faktor dar. In einer weiteren Untersuchung im Jahr 2003 mit 12 bis 15-jährigen Jugendlichen aus der Schweiz wurde mittels einer Faktorenanalyse ein Item dieses Faktors eliminiert, so dass die Skala, welche als „Lernorientierung“ bezeichnet wurde, letztlich acht Variablen enthielt (Spychiger u. a., 2006).⁷¹ Für die hier vorliegende Untersuchung wurden die ursprünglichen neun Items verwendet, um möglichst umfassend zu erheben, wie die Schüler mit Fehlern umgehen. Einige Items wurden lediglich sprachlich an die Zielgruppe angepasst. Der Teil-Fragebogen ist in nachfolgender Tabelle 16 dargestellt.⁷²

⁷¹ Diese Variablen konnten zu drei Subskalen verdichtet werden, welche jeweils unterschiedliche Aspekte der Lernorientierung repräsentieren: einstellungsbezogene, leistungsmotivationale und reflexive Aspekte (Spychiger, Kuster, & Oser, 2006).

⁷² Bei den anderen Versuchsgruppen wurde die Nummerierung entsprechend angepasst: *KG*:1-8; *kein FB*: 12-20; *EiFB*: 15-23.

Tabelle 16: Teil-Fragebogen zur Einschätzung des Umgangs mit Fehlern

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft voll zu
17) Fehler, die ich während des Unterrichts gemacht habe, schaue ich mir zu Hause genau an.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18) Fehler in Klassenarbeiten werden von mir immer freiwillig verbessert, auch wenn es der Lehrer nicht extra sagt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19) Einen Fehler, den ich im Unterricht gemacht habe, behalte ich mir auch, um ihn nicht wieder zu machen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20) Falsche Lösungen in Aufgaben überdenke ich mehrmals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21) Ich setze mich im Unterricht mit Fehlern, die ich gemacht habe, auseinander und will diese begreifen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22) Es macht mir Spaß, bei einer Aufgabe verschiedene Lösungswege auszuprobieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23) Wenn ich im Unterricht etwas ungeschickt mache, nehme ich dies als Gelegenheit wahr, daraus zu lernen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24) Es macht mir Freude, mir durch Fehler neues Wissen anzueignen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25) Fehler im Unterricht helfen mir, es hinterher besser zu machen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4.4 Auswertung der Wissenstests und des Sprachfähigkeitstests

4.4.1 Auswertung der Wissenstests

Die Auswertung der Wissenstests erfolgte mittels der evaluativen qualitativen Inhaltsanalyse⁷³ (Kuckartz, 2014, S. 98 ff.). Nach Mayring (2015, S. 11 ff.) bietet die Inhaltsanalyse ein Verfahren zur systematischen, regelgeleiteten und somit intersubjektiv überprüf-
baren Auswertung von „fixierter Kommunikation“. Bei der evaluativen qualitativen Inhaltsanalyse stehen die „Einschätzung, Klassifizierung und Bewertung von Inhalten durch die Forschenden“ im Vordergrund (Kuckartz, 2014, S. 98 ff.). Insgesamt stellt sich für die Untersuchung die Frage, mit welcher Punktzahl die Schülerantworten bewertet werden können. Hierzu konnte auf ein in Voruntersuchungen entwickeltes und optimiertes Codierhandbuch⁷⁴ zurückgegriffen werden (Ryssel & Fürstenau, 2011; Fürstenau u. a., 2010; Sommer u. a., 2009). Das Codierhandbuch wurde dabei an den im Anschluss

⁷³ Diese Form der Inhaltsanalyse kann auch als skalierende Strukturierung bezeichnet werden (Mayring, 2015, S. 106 ff.).

⁷⁴ Im Rahmen der Auswertung der Wissenstests vorangegangener Untersuchungen wurden zunächst deduktiv Bewertungskategorien festgelegt, welche durch die Fachinhalte des Planspiels vorgegeben waren. Das Sichten der Schülerantworten führte zusätzlich zur induktiven Bildung einiger Kategorien.

an die Überprüfung der Paralleltestreliabilität optimierten Wissenstest angepasst (vgl. Kapitel 4.3.1).

Das Codierhandbuch beinhaltet die jeweiligen Kategorien zu jeder Frage sowie zwei Ausprägungen der Kategorien. Die erste Ausprägung entspricht den vollständig richtigen Antworten, wofür jeweils die maximale Punktzahl vergeben wurde. Dabei wurden in Abhängigkeit von der Aufgabe 1 bis 5,5 Punkte vergeben. Im Codierhandbuch wurden neben den zu erreichenden Punktzahlen die Anforderungen an eine richtige Antwort formuliert sowie Definitionen für die Kategorien angegeben. Ankerbeispiele sollten zusätzliche Sicherheit im Umgang mit den Kategorien schaffen. Als zweite Ausprägung wurden Antworten, welche das Verstehen der inhaltlichen Zusammenhänge widerspiegeln, die jedoch keine Fachtermini beinhalten, berücksichtigt. Diese Ausprägung wurde mit aufgenommen, da die Schüler oft nicht in der Lage waren, Fachtermini zu verwenden, jedoch die Zusammenhänge verstanden hatten. Dies war der Tatsache geschuldet, dass die Schüler in einem relativ kurzen Zeitraum mit vielen komplexen Begrifflichkeiten konfrontiert wurden. Die „halb“ richtigen Antworten wurden mit der Hälfte der maximalen Punktzahl bewertet. Im Codierhandbuch wurden neben der Angabe der Punktzahlen, die Anforderungen an eine solche Antwort, Definitionen für die „halb“ richtigen Kategorien sowie Ankerbeispiele angegeben. Für falsche Antworten wurden teilweise Ankerbeispiele mit aufgeführt. Diese wurden mit 0 Punkten bewertet. Beispielsweise konnte die Antwort „Arbeitsmaterial“ mit der vollen Punktzahl von 1 Punkt bewertet werden. Gefordert war die Kategorie „Material“. Für die Antwort „Sägeblätter“ konnten nur die Hälfte der Maximalpunktzahl, d. h. 0,5 Punkte gegeben werden. Die Antwort „Plastik“ wurde mit 0 Punkten bewertet, da sie zu allgemein war. Die Codierhandbücher zur Bewertung des Vortests, 1. und 2. Nachtests sind in den Anhängen 4.4.1-1 bis 4.4.1-3 dargestellt.

Ergänzend zum Codierhandbuch wurden verschiedene Codierregeln formuliert. Jede Kategorie konnte pro Proband und Frage nur einmal codiert werden. Bei einer Nennung mehrerer Kategorien in unterschiedlicher Qualität, wurde die Kategorie, die zu der höheren Punktzahl führte, codiert. Für den Fall, dass mehrere Antworten gefordert waren, sich diese aber nur einer Kategorie zuordnen ließen, konnte auch nur eine Antwort mit der entsprechenden Kategorie bewertet werden. Beispielsweise beantwortete der Proband die Frage 8 (welche Folgen es für das Unternehmen haben kann, wenn der Lieferant nicht liefert) mit: „die Produktion kann nicht weitergehen“ und „Produktion verzögert sich“. Beide Antworten konnten der Kategorie 2 (Fertigungsengpass) zugeordnet werden. Es

sollten aber 2 verschiedene Kategorien angesprochen werden. Daher konnte hier lediglich 1 Punkt (von 2 Punkten) vergeben werden.

Den Tests wurde entsprechend der Versuchsgruppe ein Code zugeordnet und die Antworten in Excel erfasst. Anschließend erfolgte die Bewertung auf Basis des Codierhandbuches und der Codierregeln. Im Anschluss an die Erstcodierung erfolgte eine Zweitcodierung durch eine weitere fachkundige Person. Trotz der Codierung des Materials nach festgelegten Regeln kam es zu geringen Abweichungen bei der Punktevergabe zwischen Erst- und Zweitcodierer, was auf eine mit dem Datenmaterial verbundene Interpretation auf semantischer Ebene durch den jeweiligen Codierer zurückgeführt werden kann. Um die Übereinstimmung der Bewertungen beider Codierer und somit das verwendete Kategoriensystem zu überprüfen, wurde die sog. Intercoderreliabilität berechnet. Mit Hilfe der Korrelation nach Spearman wurde für die Intercoderreliabilität ein Wert von $r_s=0,968$ ($p=0,000$) ermittelt. Dies bedeutet, dass zwischen beiden Codierern bei unabhängiger Codierung der Tests eine Übereinstimmung von 96,8 Prozent erzielt wurde. Abschließend fand bei strittigen Antworten eine Konsensbildung zwischen Erst- und Zweitcodierer statt, sodass im Ergebnis eine von beiden Codierern akzeptierte Punkteverteilung für die weitere Auswertung vorlag.

Insgesamt konnten im Vortest sowie im 1. Nachtest 30 Punkte erreicht werden. Im 2. Nachtest wurde für die Aufgabe 4 und 10 jeweils eine zusätzliche Teilaufgabe mit je 2 Punkten eingeführt, so dass die maximale Punktzahl bei 34 lag. Nachfolgend dargestellte Boxplots zeigen das Minimum, Maximum, das 25 und 75 Prozent-Quartil sowie den Median der erreichten Punktzahlen des Vortests, 1. Nachtests und 2. Nachtests⁷⁵ (vgl. Abbildung 10). Im 1. Nachtest sowie 2. Nachtest ist ein Ausreißer erkennbar (Proband E107), der insgesamt 26,6 Punkte bzw. 25,25 Punkte erreichte.

⁷⁵ Der 2. Nachtest wurde zunächst ohne die Aufgabe 4b und 10b einbezogen, um eine bessere Vergleichbarkeit zwischen Vortest 1. Nachtest und 2. Nachtest zu erreichen. Des Weiteren wurde ein Boxplot über alle Aufgaben des 2. Nachtests erstellt.

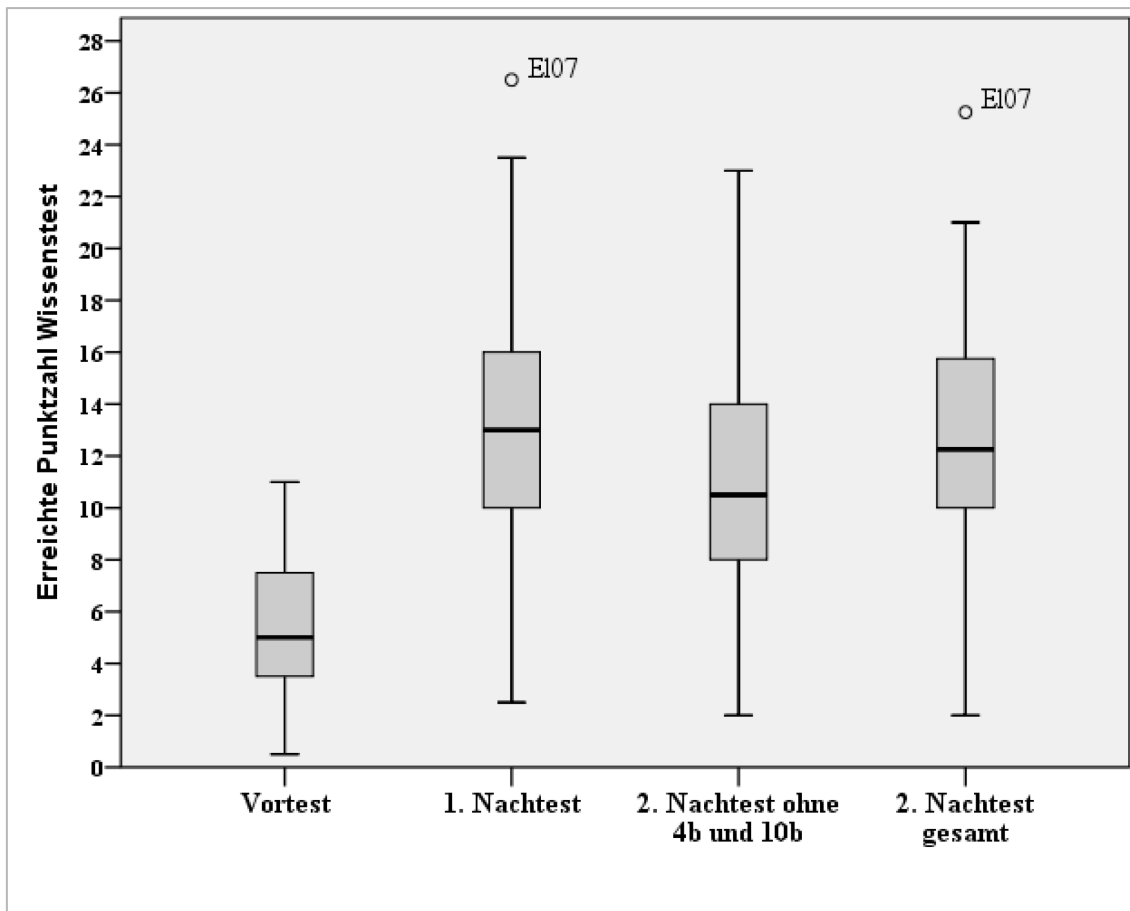


Abbildung 10: Boxplots für den Vortest, 1. Nachtest und 2. Nachtest (ohne Aufgaben 4b und 10b)

Der Median für den Vortest lag bei 5 Punkten, für den 1. Nachtest bei 13 Punkten, für den 2. Nachtest ohne Aufgaben 4b und 10b bei 10,5 Punkten und für den 2. Nachtest gesamt bei 12,25 Punkten. Nachfolgende Tabelle 17 zeigt die Mittelwerte und Standardabweichungen für die Tests.

Tabelle 17: Mittelwerte und Standardabweichungen für die Ergebnisse des Vortests, 1. Nachtests und 2. Nachtests

	Vortest	1. Nachtest	2. Nachtest
Mittelwert	5,46	13,15	12,52
Standardabweichung	2,54	4,71	4,52

4.4.2 Auswertung des Sprachfähigkeitstests

Die Auswertung der Sprachfähigkeitstests erfolgte nach dem von Kersting u. a. (2008) vorgeschlagenen standardisierten Vorgehen. Die Testergebnisse werden dabei anhand eines Vergleichsmaßstabes, z. B. einer festgelegten Normgruppe, interpretiert. Hierzu muss zunächst festgelegt werden, welche Personen eine sinnvolle Normgruppe darstellen. Für die Auswertung und anschließende Interpretation der Testwerte stellt der WIT-2 drei

bildungsspezifische und sechs altersspezifische Normen zur Verfügung. Die drei bildungsspezifischen Normen sind:

- 1a) Gesamtgruppe (bildungsrepräsentativ, 40% Abitur; 60% kein Abitur; die entsprechenden Daten wurden durch Stratifizierung erzeugt),
- 1b) Personen mit (fachgebundenem) Abitur und
- 1c) Personen ohne (fachgebundenes) Abitur.

Die sechs altersspezifischen Normen sind:

- 2a) Gesamt (ohne Altersdifferenzierung),
- 2b) 14-17 Jahre,
- 2c) 18 Jahre,
- 2d) 19-22 Jahre,
- 2e) 23-27 Jahre,
- 2f) 28 Jahre und älter.

Insgesamt stehen folglich 18 (3x6) Normgruppen pro Subtest zur Verfügung. Die Auswahl einer bestimmten Normgruppe hängt nach Kersting u. a. (2008, S. 44 f.) von der Forschungsfrage der jeweiligen Untersuchung ab. Eine Anwendung der alters- und/oder bildungsspezifischen Normen empfiehlt sich dann, wenn Alter und/oder Bildungsgrad der Probanden eine große Spannweite haben und Unterschiede analysiert werden sollen. Die Probanden der Untersuchung sind alle zwischen 13 und 16 Jahren alt und streben den Haupt-/ bzw. Realschulabschluss an. Kersting u. a. (2008, S. 45) empfehlen bei der Auswertung der Testdaten nicht zu differenzieren (d. h. die Kategorien 1a bzw. 1b heranzuziehen), sofern keine begründeten und rechtfertigenden Annahmen für eine Verwendung bestehen. Im Rahmen der Untersuchung wird der Argumentation gefolgt.

Für jeden Probanden wurde zunächst die Anzahl der richtigen Antworten in den Subtests „Analogien“ und „Gleiche Wortbedeutung“ ermittelt. Infolgedessen ergaben sich Rohwerte, welchen anhand der entsprechenden Normtabelle ein Standardwert zugeordnet wurde. Hierbei wurde die Normtabelle für die Kombination der Kategorien 1a und 2a herangezogen (Kersting u. a., 2008, S. 130). Erzielte ein Proband beispielsweise 6 richtige Antworten im Subtest Analogien, so konnte ein Standardwert von 98 zugeordnet werden. Dies bedeutete, dass 2 Prozent der Vergleichsgruppe gleich gute oder bessere Ergebnisse erzielten. Da sich die Dimension „Sprachliches Denken“ aus den Subtests „Analogien“ und „Gleiche Wortbedeutung“ zusammensetzt, wurden die Standardwerte der beiden Subtests für jeden Schüler addiert und durch 2 geteilt, um einen Mittelwert für

die Dimension „Sprachliches Denken“ zu erhalten. Insgesamt wurden gerundete Mittelwerte für die Dimension „Sprachliches Denken“ zwischen 84 und 106 ermittelt. Der Mittelwert über alle 83 Schüler betrug 94 (Standardabweichung: 5,38). Nachfolgendes Histogramm zeigt die Verteilung der Mittelwerte zum „Sprachlichen Denken“ entsprechend ihrer Häufigkeiten (vgl. Abbildung 11).

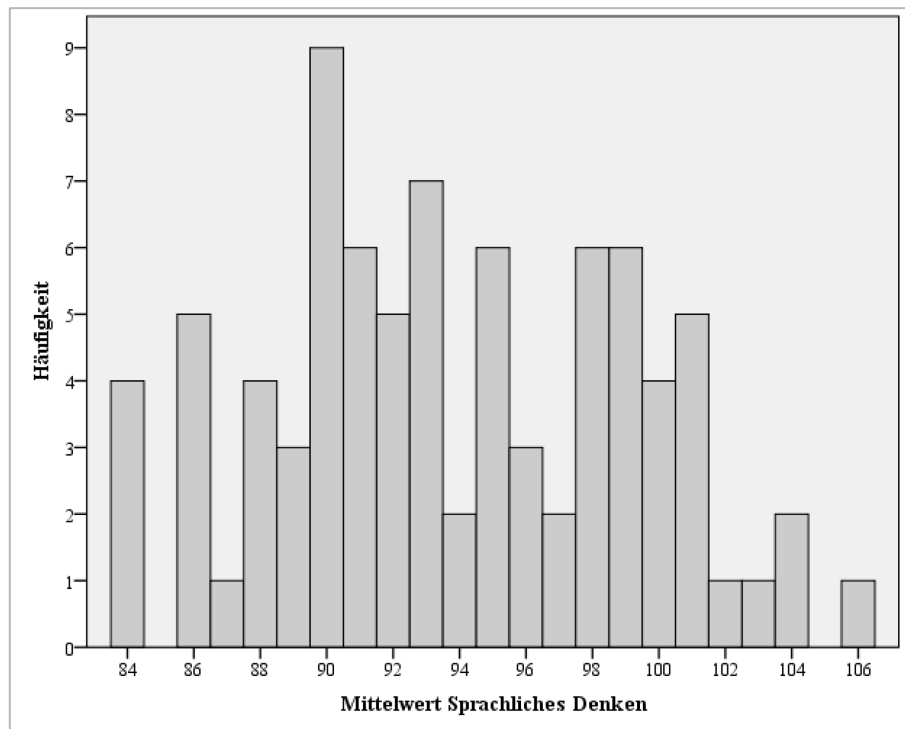


Abbildung 11: Histogramm zur Verteilung der Mittelwerte zum „Sprachlichen Denken“ entsprechend ihrer Häufigkeiten

4.5 Auswertung der Concept Maps

4.5.1 Kategoriale und strukturelle Inhaltsanalyse

Um die erstellten Concept Maps analysieren zu können, müssen sie hinsichtlich Inhalt und Struktur miteinander vergleichbar gemacht werden. Zu diesem Zweck werden eine *kategoriale* und eine *strukturelle Inhaltsanalyse* durchgeführt.

Kategoriale Inhaltsanalyse

Bei der *kategorialen Inhaltsanalyse* werden die Konzepte und Relationen getrennt voneinander betrachtet. Ziel ist es, eine inhaltliche Vergleichbarkeit der Concept Maps zu erreichen. Zu diesem Zweck werden die Konzepte und Relationen hinsichtlich ihrer se-

semantischen Ähnlichkeit überprüft und im Anschluss zu überschneidungsfreien Konzepten und Relationen zusammengefasst. Hierbei gilt es zu beachten, dass eine starke Aggregation einerseits die Vergleichbarkeit erhöht, aber andererseits die Aussagekraft der sich anschließenden strukturellen Inhaltsanalyse verringert. Da den Probanden eine Liste mit Konzepten und Relationen zur Verfügung gestellt wurde, waren diese nur durch zusätzlich von den Probanden genannte Konzepte bzw. Relationen zu ergänzen. Bestand hierbei jedoch semantische Ähnlichkeit mit bereits vorgegebenen Konzepten oder Relationen, wurden diese in vorhandene Kategorien integriert. Anderenfalls wurde eine neue Kategorie aufgenommen. Insgesamt wurde bei der kategorialen Inhaltsanalyse auf eine Codierung möglichst nah am Referenzmap geachtet, um eine qualitative Bewertung der Maps zur Überprüfung der Hypothesen zu ermöglichen (Fürstenau & Trojahnner, 2005, S. 196 f.).

Ergebnis der kategorialen Inhaltsanalyse waren die Basislexika der Konzepte und Relationen, welche in den Anhängen 4.5.1-1 bis 4.5.1-3 dargestellt sind. Für die 1. und 2. Mapping-Aktivität wurde je ein Basislexikon für die Konzepte und ein Basislexikon für die Relationen erstellt. Bei der 3. Mapping-Aktivität galt es, lediglich Verbindungen zwischen dem 1. und 2. Concept Map zu finden, so dass auch nur ein Basislexikon für die Relationen erstellt wurde. Die Tabellen enthalten den jeweiligen Kategoriennamen sowie eine dreistellige Codierung. Die Anzahl der Konzepte und Relationen der Basislexika ist in nachfolgender Tabelle 18 dargestellt. Es ist erkennbar, dass bei der 1. Mapping-Aktivität die meisten Konzepte und Relationen hinzugefügt wurden. Dies kann mit dem Themenbereich begründet werden. Die Darstellung des Wertschöpfungsprozesses liegt am ehesten im Erfahrungsbereich der Schüler, so dass an dieser Stelle am häufigsten eigene Ideen eingebracht wurden.

Tabelle 18: Anzahl der Konzepte und Relationen der Basislexika

	Anzahl Konzepte			Anzahl Relationen		
	gesamt	vorgegeben	hinzugefügt	gesamt	vorgegeben	hinzugefügt
Mapping 1	34	19	15	28	11	17
Mapping 2	24	20	4	17	9	8
Mapping 3	-	-	-	16	3	13

Strukturelle Inhaltsanalyse

Mithilfe der strukturellen Inhaltsanalyse können Concept Maps in Bezug auf ihre Struktur vergleichbar gemacht werden. Zu diesem Zweck wird die vollständige Proposition als Analyseeinheit betrachtet (Fürstenau & Trojahnner, 2005, S. 197). In Anlehnung an das Vorgehen von Röhle (2007, S. 50 f.) wurden die nachfolgend erläuterten Veränderungen der Probandennetze vorgenommen, um die strukturelle Vergleichbarkeit aber auch eine generelle Auswertbarkeit der Concept Maps zu gewährleisten. Hierbei wurde erneut auf eine Codierung möglichst nah am Referenzmap geachtet.

1) Umwandlung von passiv formulierten in aktiv formulierte Aussagen

Von Probanden passiv formulierte Aussagen wurden in aktiv formulierte umgewandelt und gleichzeitig die Richtung der Relation gedreht. Beispielsweise wurde die Aussage „Produkte werden gekauft von Kunden“ in „Kunden kaufen Produkte“ umformuliert (Proband E114, 1. Mapping).

2) Aufteilung von Doppelpfeilen

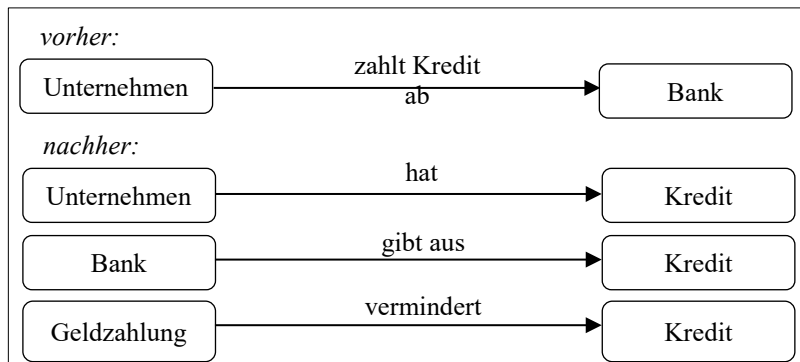
Insofern die Schüler wechselseitige Abhängigkeiten der Konzepte mittels Doppelpfeilen darstellten, wurden diese in zwei einfache gegeneinander gerichtete Pfeile geteilt. Beispielsweise stellte ein Proband eine wechselseitige Verbindung zwischen den Konzepten „Bilanz“ und „Passivseite“ mit der Relation „ist Teil von“ her. Um nah am Referenznetz zu bleiben, wurde eine Proposition in „Bilanz hat Aktivseite“ verändert. Die andere wurde als „Bilanz ist Teil von Aktivseite“ codiert (Ei04, 2. Mapping).

3) Ergänzung fehlender Beschriftungen bzw. Richtungen von Relationen

Nicht beschriftete Pfeile wurden mit Bezeichnungen aus den Basislexika ergänzt, falls die entsprechende Beschriftung aus dem Kontext erschlossen werden konnte. Andernfalls wurde die Beschriftung „beeinflusst“ ergänzt. Beispielsweise zeichnete ein Proband einen Pfeil zwischen „Kasse und Bankguthaben“ und „gute Konditionen“ ein, ohne diesen zu beschriften (Ei04; 1. Mapping). Die Verbindung wurde mit der Relation „beeinflusst“ ergänzt. Die Ergänzung der Pfeilspitzen erfolgte ebenfalls nur dann, wenn die Richtung der Relation aus dem Zusammenhang erkennbar war. Beispielsweise verknüpfte ein Proband die Konzepte „Geldzahlung“ und „Gewinn- und Verlustrechnung“, ohne eine Beschriftung vorzunehmen bzw. eine Richtung anzugeben (Ei15). Die Verbindung wurde durch Ergänzen der Richtung und der Relation „beeinflusst“ in die Proposition „Geldzahlung beeinflusst Gewinn- und Verlustrechnung“ überführt.

4) Trennung von Wortgruppen

Für den Fall, dass ein Proband ganze Wortgruppen als Relationen verwendete, wurden diese so getrennt, dass die ursprüngliche Bedeutung möglichst erhalten blieb, was folgendes Beispiel (Proband E114, 1. Mapping) verdeutlicht.



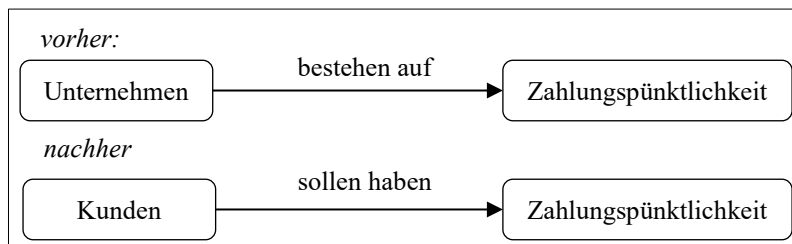
In dem Beispiel wurde die Wortgruppe „zahlt Kredit“ als Relation verwendet. Im Rahmen der strukturellen Inhaltsanalyse wurde „Kredit“ als Konzept und „gibt aus“ als Relation eingeführt und die Aussage anhand von drei Propositionen dargestellt.

5) Weitere Codierregeln

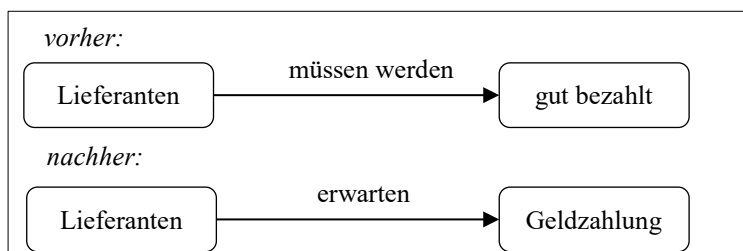
Nachfolgend werden weitere Codierregeln vorgestellt, welche sich aus der möglichst nah am Referenzmap vorgenommenen Codierung ergaben. Dabei wurden zum Teil auch Konzepte oder Relationen verändert bzw. neu eingeführt und somit eine kategoriale Inhaltsanalyse vorgenommen. Für die drei Mapping-Aktivitäten wurden dabei separate Regeln aufgestellt:

Concept Mapping 1:

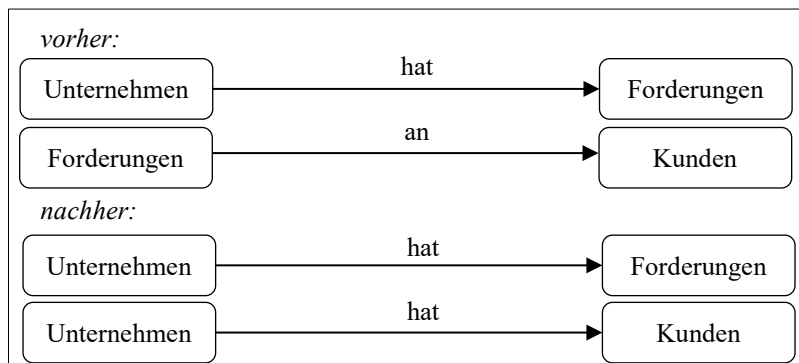
- In das überarbeitete Referenznetz wurden nur Propositionen aufgenommen, die sich mit den Zielen des Planspiels und den darin behandelten Inhalten verbinden ließen. Demzufolge wurden von den Schülern genannte Propositionen zu den Unternehmensbereichen nicht in das Referenznetz aufgenommen. Verknüpfungen zum Thema Interessengruppen wurden vollständig aufgenommen.
- Erwartungen des Unternehmens an die Lieferanten, Kunden bzw. Mitarbeiter wurden immer folgendermaßen codiert: „Lieferanten/ Kunden/Mitarbeiter sollen haben ...“. Beispielsweise wurde die Proposition „Unternehmen bestehen auf Zahlungspünktlichkeit“ daher an entsprechende Struktur angepasst und in „Kunden sollen haben Zahlungspünktlichkeit verändert (CM06).“



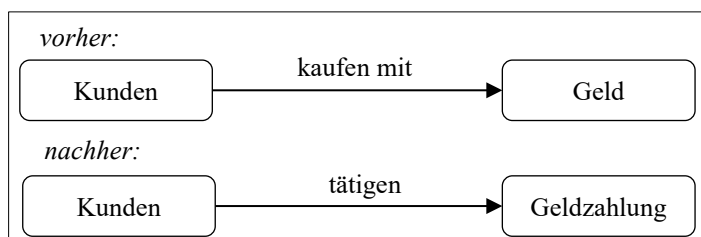
- Erwartungen des Kunden an das Unternehmen sowie des Lieferanten an das Unternehmen wurden immer folgendermaßen codiert: „Kunden/ Lieferanten erwarten...“. Beispielsweise wurde die Proposition „Lieferanten müssen werden gut bezahlt“ in „Lieferanten erwarten Geldzahlung“ verändert (CM16).



- Teilweise wurde das Vorhandensein von Forderungen bzw. Verbindlichkeiten in den Schülernetzen durch die Präposition an/zu in Verbindung mit der entsprechenden Interessengruppe dargestellt. Nachfolgendes Codierbeispiel veranschaulicht das Vorgehen in diesem Fall (z. B. bei CM05):

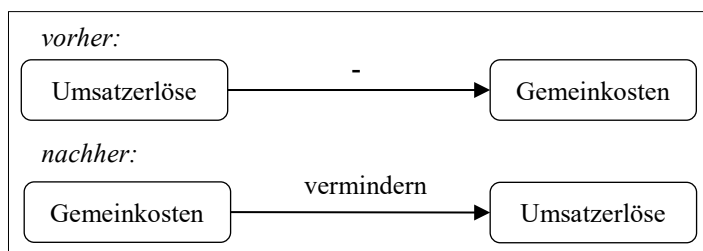


- Das Konzept „Geld“ wurde der Kategorie „Geldzahlungen“ zugeordnet. Dementsprechend ergaben sich auch Veränderungen der ursprünglichen Propositionen in den Schülernetzen (z. B. bei CM05):



Concept Mapping 2:

- Die Relation „führt zu“ wurde in „verursacht“ umcodiert. Dieses Vorgehen ergab sich aus dem Ziel einer referenznahen Codierung. Im Referenznetz kam die Relation „führt zu“ nicht vor.
- Das Subtraktionszeichen wurde, wenn aus dem Kontext möglich, als „vermindern“ codiert. Dabei musste ein Richtungswechsel vorgenommen werden (z. B. bei CM18).



Concept Mapping 3

- Bei Verbindungen, die im Zusammenhang mit den Anschaffungskosten, Gemeinkosten oder Herstellungskosten stehen, wurde die Relation „führt zu“ in „verursacht“ umkodiert, um nah am Referenznetz zu bleiben.

Erweiterung der Referenznetze zum 1. und 3. Mapping

Während der Analyse der Probandennetze fiel auf, dass die Schüler Inhalte darstellten, die im Rahmen des Planspiels thematisiert wurden, jedoch nicht in den vor der Untersuchung entwickelten Referenznetzen abgebildet waren. Um die Inhalte des Planspiels sowie die Lernziele bestmöglich durch die Referenzmaps widerzuspiegeln und somit eine zuverlässige Grundlage für die qualitative Bewertung der Probandennetze zu schaffen, wurden auf induktivem Wege einige zusätzliche Propositionen in die Referenznetze aufgenommen. Änderungen wurden hierbei in den Referenznetzen zum 1. und 3. Mapping vorgenommen. Im Rahmen des 2. Mappings spiegelte das ursprüngliche Referenzmap die Planspielinhalte angemessen wider. Die erweiterten Referenznetze zum ersten und 3. Mapping sind in den Anhängen 4.5.1.-4 und 4.5.1-5 dargestellt. Dabei wurden hinzugefügte Konzepte und Relationen mit grau hinterlegter Schrift und grau schraffierten Linien gekennzeichnet. Im ersten erweiterten Referenzmap wurden die Abkürzungen für die Relationen verwendet, um eine übersichtliche Darstellung zu ermöglichen. Im ersten Referenzmap wurden die Konzepte „Unternehmen“, „Mitarbeiter“ und „gute Qualität“ hinzugefügt. Viele Probanden wählten den Begriff „Unternehmen“ als zentrales Konzept

aus, um die Verbindung zu den Interessengruppen „Lieferanten“ und „Kunden“ herzustellen. In diesem Zusammenhang wurde auch die Verknüpfung zum Begriff „Mitarbeiter“ aufgeführt. Im ursprünglichen Referenznetz waren lediglich die Erwartungen des Unternehmens an die Kunden bzw. Lieferanten dargestellt. Die Probanden stellten jedoch auch häufig die Erwartungen der Interessengruppen an das Unternehmen dar. Diese Zusammenhänge wurden in das erweiterte Referenznetz mit aufgenommen. Daher wurde auch das Konzept „gute Qualität“ hinzugefügt. Im 3. Referenzmap wurden Verknüpfungen mit dem Konzept „Gemeinkosten“ mit aufgenommen. Zusätzliche Propositionen waren: „Fertigwarenlager verursacht Gemeinkosten“, „Fertigung verursacht Gemeinkosten“ und „Materiallager verursacht Gemeinkosten“.

Entwicklung der Modalnetze

Auf Basis der kategorialen und strukturellen Inhaltsanalyse lassen sich Modalnetze erstellen. Diese enthalten die von den Probanden am häufigsten übereinstimmend genannten Propositionen. Die Anzahl der zu berücksichtigenden Propositionen im Modalnetz bestimmt sich nach der mittleren Häufigkeit aller in den Probandennetzen genannten Propositionen (Fürstenau & Trojahnner, 2005, S. 197; Becker u. a., 1987, S. 464). Im Rahmen der Untersuchung wurden die Probanden instruiert, drei Concept Maps zu erstellen. Daher ergeben sich drei zu erstellende Modalnetze. Nachfolgende Tabelle 19 stellt die Anzahl der genannten Propositionen sowie den durchschnittlichen Umfang der Netze dar.

Tabelle 19: Anzahl der genannten Propositionen und durchschnittlicher Netzzumfang

	Anzahl der genannten Propositionen	Durchschnittlicher Netzzumfang (56 Probanden)
Mapping 1	1099	19,6
Mapping 2	1106	19,8
Mapping 3	357	6,4

Die von den Probanden genannten Propositionen werden entsprechend der Häufigkeit ihrer Nennung absteigend sortiert. In das Modalnetz sind nun so lange Propositionen aufzunehmen, bis der zuvor berechnete Durchschnittsnetzzumfang erreicht ist. Die Grenze liegt dabei selten am Ende einer Häufigkeitsklasse. Zuerst aufgelistete Propositionen dürfen nicht vorgezogen werden. Aus diesem Grund sind entweder alle oder keine der Propositionen mit dieser Häufigkeit einzubeziehen (Fürstenau & Trojahnner, 2005, S. 198).

Nachfolgende Tabelle 20 zeigt die Häufigkeitsklassen der Probandennetze sowie die kumulierte Anzahl der Propositionen. Beispielsweise wurden die Propositionen *Bilanz hat*

Passivseite und *Bilanz hat Aktivseite* 38 bzw. 37 Mal genannt und stellen somit die häufigsten Propositionen des 2. Mappings dar. Im 1. Mapping sind die Propositionen *Lieferanten sollen haben Liefertreue* und *Kunden sollen haben Treue* die häufigsten Propositionen mit einer 36-fach-Nennung.

Das Modalnetz zum 1. Mapping könnte entweder 18 oder 21 Propositionen enthalten, da der durchschnittliche Netzzumfang von 19,6 zwischen den 14-fach und 13-fach-Nennungen liegt. Hierbei liegt die Anzahl von 21 Propositionen näher am durchschnittlichen Netzzumfang, so dass die 13fach genannten Propositionen mit in das Modalnetz aufgenommen werden. Für das 2. Mapping ergibt sich auf diese Weise ein Modalnetz mit 22 Propositionen. Beim Modalnetz des 3. Mapping werden 7 Propositionen aufgenommen. Die 6-fachen Nennungen liegen zwar näher am durchschnittlichen Netzzumfang, jedoch sind generell sehr wenige Propositionen im Modalnetz enthalten, so dass eine zusätzliche Aufnahme der 7fach genannten Proposition die Aussagekraft erhöht.

Tabelle 20: Kumulierte Anzahl der Propositionen zur Ermittlung der Modalnetze

	Mapping 1	Mapping 2	Mapping 3
Häufigkeitsklasse	Kumulierte Anzahl der Propositionen		
38-fach	-	1	-
37-fach	-	2	-
36-fach	2	-	-
32-fach	3	-	-
31-fach	4	-	-
30-fach	5	-	-
29-fach	7	-	-
28-fach	8	-	-
25-fach	9	3	-
23-fach	10	-	-
21-fach	11	4	-
18-fach	12	-	-
17-fach	13	5	-
16-fach	14	-	-
15-fach	17	-	1
14-fach	18	6	2
13-fach	21	8	-
11-fach		9	3
10-fach		12	-
9-fach		14	4
8-fach		-	6
7-fach		17	7
6-fach		22	-
Abbildungsleistung	42,68 %	26,04 %	20,17 %

Des Weiteren lässt sich die Abbildungsleistung als Quotient aus den berücksichtigten Propositionen im Modalnetz und den insgesamt genannten Propositionen in den individuellen Netzen (vgl. Tabelle 19) berechnen. Die Abbildungsleistung beträgt für das 1.

Mapping 42,68 Prozent, für das 2. Mapping 26,04 Prozent und für das 3. Mapping 20,17 Prozent. Das Modalnetz des 1. Mapping repräsentiert folglich die individuellen Netze am besten. Von den 21 Propositionen des Modalnetzes des 1. Mappings sind 18 Propositionen des Referenznetzes enthalten. Beim Modalnetz des 2. Mappings sind es lediglich 16 von 22 und beim Modalnetz des 3. Mapping 1 von 8 Propositionen. Für das 1. Mapping verwenden die Probanden folglich zu 85,71 Prozent Propositionen des Referenzmaps. Die Propositionslisten für die drei Modalnetze sind im Anhang 4.5.1-6 dargestellt.

Überprüfung der Inter- und Intracoderreliabilität

Um das Codiervorgehen zu überprüfen, wurde für 11 Prozent⁷⁶ der Probandennetze die Inter- und Intracoderreliabilität berechnet. Als Grundlage dienten 18 der insgesamt 168 erstellten Concept Maps (3 Mappings x 56 Probanden), welche zufällig ausgewählt wurden. Die Netze wurden für die Berechnung der Intercoderreliabilität von einem zweiten Codierer sowie für die Berechnung der Intracoderreliabilität vom Erstcodierer zu einem späteren Zeitpunkt (nach ca. 5 Monaten) erneut inhaltsanalytisch ausgewertet. Die codierten Propositionen (insgesamt 267) beider Codierer bzw. beider Zeitpunkte wurden anschließend auf Übereinstimmung geprüft. Hierzu wurde der Überschneidungskoeffizient nach Holsti (1969) herangezogen. Dieser setzt die Anzahl der übereinstimmenden Entscheidungen verschiedener Codierer in Relation zu der Summe aller getroffenen Entscheidungen und wird folgendermaßen berechnet (Mayring, 2015, S. 127):

$$C = \frac{(\text{Zahl der Kodierer}) \times (\text{Zahl der übereinstimmenden Urteile})}{(\text{Zahl aller Codierurteile})}$$

Für die Intercoderreliabilität konnte ein Koeffizient von $C=0,83$, für die Intracoderreliabilität ein Koeffizient von $C=0,81$ ermittelt werden. Döring und Bortz (2016, S. 83) geben an, dass in der Forschung ein Wert von 0,8 oft als Mindestmaß für Reliabilitätskoeffizienten angesehen wird. Die inhaltsanalytische Auswertung der Concept Maps erforderte ein aufwendiges Codiervorgehen, weshalb ein Wert größer als 0,8 als zufriedenstellend angesehen werden kann. Das Codiervorgehen sowie dessen Anwendung kann daher als zuverlässig eingeschätzt werden.

⁷⁶ Laut Döring und Bortz (2016, S. 566) werden konventionell 10-20 Prozent des Materials für eine Reliabilitätsprüfung herangezogen.

Auswertung der Feedback-Netze

Im Kapitel 4.1.2 wurde bereits erläutert, dass die Probanden der beiden Feedbackgruppen einfaches Feedback in Form des jeweiligen Referenzmaps erhielten, welches sie mit dem selbsterstellten Map vergleichen sollten. Die Schüler konnten in einer Kopie ihrer zuvor angefertigten Concept Maps Änderungen vornehmen. Die so überarbeiteten Netze werden nachfolgend als Feedback-Netze bezeichnet. Die Auswertung der Feedback-Netze erfolgte analog zur Auswertung der erstellten Maps mittels der kategorialen und strukturellen Inhaltsanalyse und auf Basis der für das Auswerten der erstellten Maps formulierten Codierregeln. Einige Unterschiede wurden dennoch bei der Codierung vorgenommen:

- 1) Bei nicht beschrifteten Relationen wurde stets *beeinflusst* ergänzt.
- 2) Eine Ergänzung bzw. eine Änderung der Richtung erfolgte nur, falls dies aus dem Kontext heraus geschlossen werden konnte.
- 3) Sofern die Propositionen der selbstkonstruierten Netze im Rahmen der strukturellen Inhaltsanalyse so umcodiert wurden, dass sie denen des Referenznetzes entsprachen, fanden die von den Probanden entsprechend hinzugefügten oder korrigierten Propositionen der Feedback-Netze keine Berücksichtigung mehr.

Bei der Erfassung der Feedback-Netze wurde zudem in Excel farblich kenntlich gemacht, was die Schüler beim Vergleichen ihres jeweiligen Concept Maps mit dem Referenznetz verändert hatten. Dabei wurde zwischen dem Streichen bzw. als falsch Kennzeichnen von Propositionen, dem Hinzufügen von Propositionen oder dem Korrigieren falscher Propositionen unterschieden. Auf diese Weise konnten Strategien, die die Probanden zur Verarbeitung des einfachen Feedbacks nutzen, identifiziert werden. Bei der Strategie des Hinzufügens von Propositionen zeigte sich, dass die Schüler die Propositionen des Referenznetzes nicht immer richtig übernahmen. Häufig wurden Propositionen falsch oder unvollständig in die eigenen Netze übertragen. Bei der Auswertung wurde daher zwischen dem Hinzufügen richtiger, falscher und „halb“ richtiger Propositionen unterschieden. Im Folgenden werden diese drei Kategorien zusammengefasst. Nachfolgende Abbildung 12 zeigt die relativen Häufigkeiten der vier Verarbeitungsstrategien insgesamt. Beispielsweise wurde die Strategie des Löschens falscher Propositionen bei den drei Mapping-Aktivitäten von den 39 Probanden der Gruppen EiFB und ElFB insgesamt 40 Mal angewendet. Insgesamt wendeten diese Probanden 230 Mal Verarbeitungsstrategien an. Somit ergibt sich ein Anteil von 17 Prozent. Bei der Berechnung spielt es keine Rolle,

wie viele Propositionen gelöscht wurden. Vielmehr ist interessant, welche der Strategien wie häufig angewendet wurde.

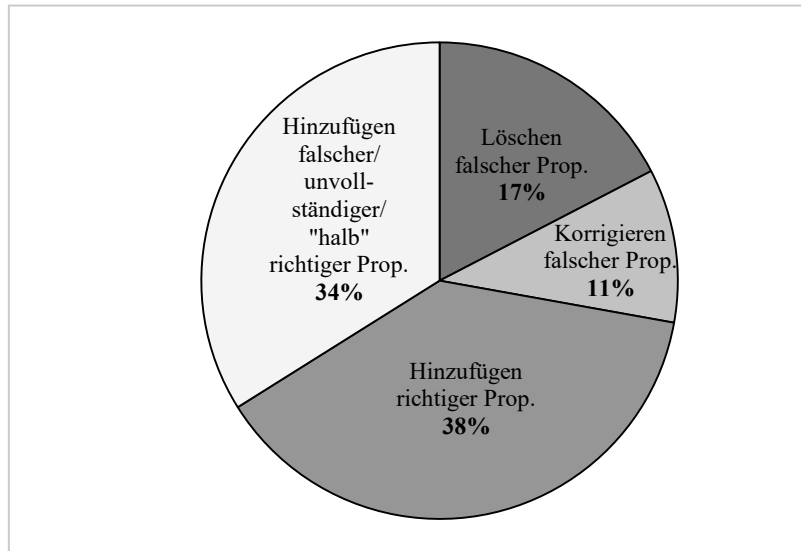


Abbildung 12: Relative Häufigkeiten der Verarbeitungsstrategien des einfachen Feedbacks

Es ist ersichtlich, dass das Hinzufügen von Propositionen aus dem Referenznetz mit 72 Prozent die am häufigsten angewandte Strategie darstellt, wobei jedoch nur 38 Prozent auf das Hinzufügen richtiger Propositionen entfällt. Von den insgesamt angewandten Strategien wurde das Löschen falscher Propositionen zu 17 Prozent und das Korrigieren falscher Propositionen zu 11 Prozent genutzt. Dies zeigt, dass die Schüler häufiger nur Propositionen des Referenzmaps übernehmen und sich weniger intensiv mit ihren erstellten Maps auseinandersetzen, indem sie falsche Propositionen löschen oder korrigieren.

4.5.2 Qualitative Bewertung der Concept Maps

Die Lernwirksamkeit als abhängige Variable wird neben dem Wissenserwerb durch die Qualität der erstellten Maps operationalisiert. Im Kapitel 3.1.2 wurde bereits auf die Bewertung der Concept Maps zur Wissensdiagnostik anhand von qualitativen Indikatoren eingegangen. Dabei stellt sich die Frage, welcher Indikator die Qualität der erstellten Maps am zuverlässigsten misst. Unter Berücksichtigung der in Kapitel 3.1.2 vorgestellten Indikatoren erscheint die Berechnung des Übereinstimmungsgrades mit dem Referenzmap am sinnvollsten. An dieser Stelle ist jedoch kritisch anzumerken, dass das hierzu vorgestellte Verfahren von Ruiz Primo (2000) zwar die richtigen Verknüpfungen im Lernermap zu allen Propositionen im Referenzmap ins Verhältnis setzt, um den Grad der Übereinstimmung zu erhalten. Es berücksichtigt jedoch nicht, dass ein Proband, welcher

ein Map mit sehr vielen Propositionen und einer hohen Vernetztheit konstruiert, auch eine höhere Chance hat, richtige Propositionen darzustellen, da die Trefferquote einfach höher ist.

Aus diesem Grund wird das Vorgehen von Fürstenau und Trojahnner (2005) und Oldenbürger (1986) zur Identifizierung prototypischer Netze auf die Berechnung des Übereinstimmungsgrades übertragen. Hierbei werden auf der Basis paarweiser Vergleiche der Probandennetze Distanzmatrizen erstellt. Anschließend wird ein Netz allen anderen gegenübergestellt, die Mittelwerte einer jeden Datenreihe berechnet und somit die Distanz des jeweiligen Netzes zu allen anderen bestimmt. Das prototypische Netz ist schließlich das Netz mit dem niedrigsten Wert (0 = keine Distanz, 1 = vollkommene Distanz). Zur Identifizierung prototypischer Netze verwenden Fürstenau und Trojahnner (2005) das von Oldenbürger (1992) entwickelte Software-Tools „Netzwerkzeug 1“. Für die Berechnung der Distanzen nutzt das Programm die Galanter-Metrik als relatives Distanzmaß, welches die Restle-Metrik ins Verhältnis zur Vereinigungsmenge setzt.

Berechnung des Übereinstimmungsgrades mit dem Referenznetz anhand der Galanter-Metrik

Um den Grad der Übereinstimmung zum Referenznetz zu berechnen, wird das Referenznetz wie ein weiteres Probandennetz mit aufgenommen, allen Probandennetzen gegenübergestellt und die Distanz zu jedem Map berechnet. Die vom Programm „Netzwerkzeug 1“ hierzu vorgenommenen Berechnungen werden nachfolgend vorgestellt.

Zunächst werden allgemein zwei Mengen A und B betrachtet. Die Werte $m(A)$ und $m(B)$ stellen die Mächtigkeit bzw. die Kardinalzahl der beiden Mengen dar. Für beliebige Mengen A und B werden folgende Mengenoperationen definiert (Lau, 2011, S. 10 f.):

Durchschnitt von A und B: $A \cap B := \{x|x \in A \wedge x \in B\}$

Vereinigung von A und B: $A \cup B := \{x|x \in A \vee x \in B\}$

Differenz von A und B: $A \setminus B := \{x|x \in A \vee x \notin B\}$

Symmetrische Differenz von A und B: $A \Delta B := (A \setminus B) \cup (B \setminus A)$

Die Mächtigkeit der symmetrischen Mengendifferenz erfüllt die Eigenschaften einer Metrik und wird als Restle-Metrik bezeichnet (Restle, 1959, S. 207 ff.):

$$d_R(A, B) = m(A \Delta B) = m(A) + m(B) - 2m(A \cap B)$$

Die Restle-Metrik stellt das absolute Distanzmaß dar (Oldenbürger, 1986, S. 32 f.).

Im Rahmen der Untersuchung erstellten die Probanden der Gruppen *kein FB*, *EiFB* und *EIFB* jeweils drei Concept Maps. Jedes erstellte Map besteht aus einer Propositionsmenge, deren Mächtigkeit die Anzahl der Propositionen für jede Versuchsperson ist. Die Restle-Metrik für zwei Propositionsmengen gibt also die Anzahl der insgesamt in beiden Netzen vertretenen Propositionen an, die jedoch nicht den beiden Propositionsmengen gleichzeitig angehören. Beispielsweise enthält das Referenznetzwerk zum 1. Mapping 47 Propositionen und das Concept Map vom Probanden Ei04 21 codierte Propositionen. Davon sind 13 Propositionen im Referenznetz enthalten. Für den Wert der Restle-Metrik ergibt sich: $47 + 21 - 2 \cdot 13 = 42$. Das bedeutet, dass sich das Concept Map des Probanden Ei04 und das Referenznetzwerk hinsichtlich der 42 Propositionen voneinander unterscheiden und in 13 Propositionen übereinstimmen. Die nachfolgende Abbildung 13 veranschaulicht die Bestimmung von der Restle-Metrik für dieses Beispiel:

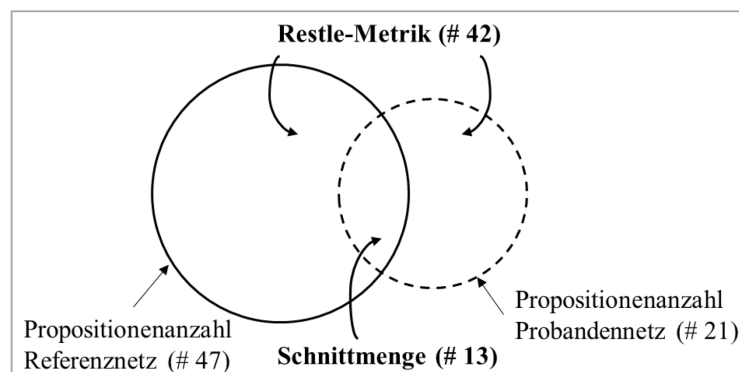


Abbildung 13: Ein Beispiel zur Bestimmung des absoluten Distanzmaßes

Da die Restle-Metrik sehr anfällig auf Veränderungen der absoluten Propositionen-Anzahl reagiert, wird zur Dämpfung des Effekts für weitere Berechnungen die Galanter-Metrik, ein relatives Distanzmaß, verwendet. Hierbei wird die Restle-Metrik als Maß der symmetrischen Mengendifferenz ins Verhältnis zum Maß der Vereinigungsmenge gesetzt (Oldenbürger, 1986, S. 33):

$$d_G(A, B) = \frac{m(A \Delta B)}{m(A \cup B)} = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B) - m(A \cap B)}$$

Die Werte der Galanter-Metrik liegen dabei zwischen 0 und 1, wobei der Wert 0 keine Distanz und der Wert 1 vollkommene Distanz bedeutet. Die Ergebnisse der Distanzmaße können als Distanzmatrix dargestellt werden. Die so ermittelten Werte spiegeln wider, wie weit die jeweiligen Individualnetze vom Referenznetz entfernt sind.

Für das Probandennetz Ei04 und das Referenznetz zum 1. Mapping resultiert danach das relative Distanzmaß in Höhe von: $\frac{42}{47+21-13} = 0,7636$.

Dabei empfiehlt es sich die Distanz-Werte mittels der Formel $\ddot{U}G = 1 - d_G$ umzurechnen, so dass der Wert 0 keine Übereinstimmung und der Wert 1 eine vollkommene Übereinstimmung widerspiegelt. Auf diese Weise wird der Grad der Übereinstimmung [$\ddot{U}G$] ermittelt. Der $\ddot{U}G$ beträgt in diesem Beispiel 0,2364.

Auf diese Weise wurde der $\ddot{U}G$ aller erstellten Concept Maps der drei Mapping-Aktivitäten auf Basis des jeweiligen Referenzmaps (Referenzmap 1: 47 Propositionen, Referenzmap 2: 24 Propositionen, Referenzmap 3: 12 Propositionen) berechnet.

Entwicklung eines differenzierten Scoringverfahrens auf Basis der Galanter-Metrik

Die Schüler wurden im Rahmen der Studie innerhalb kürzester Zeit mit komplexen Inhalten konfrontiert, die normalerweise in der dargebotenen Tiefe nicht im Lehrplan enthalten sind. Daher ist anzunehmen, dass viele Zusammenhänge lediglich oberflächlich erfasst werden können. Aus diesem Grund ist es auch nicht ungewöhnlich, wenn sogenannte naive Alltagsvorstellungen resistent bleiben und sich den wissenschaftlichen Vorstellungen, repräsentiert durch das Referenznetz, erst langsam über Zwischenvorstellungen annähern. Um den wissenschaftlichen Konzepten komplett zu entsprechen, ist der kurze Zeitraum von zwei Tagen nicht ausreichend.

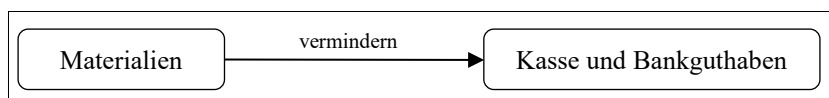
Bei der qualitativen Bewertung anhand der Galanter-Metrik wird für eine Proposition im Probandennetz nur dann eine Übereinstimmung gewertet, wenn diese mit einer Proposition des Referenznetzes vollkommen übereinstimmt. Im Rahmen der Auswertung der erstellten Concept Maps fiel auf, dass die Probanden häufiger mehrere Konzepte „überspringen“ und in der Folge Propositionen aufstellten, die es so nicht im Referenznetz gibt, welche aber weder vollständig richtig noch falsch sind. Diese „halbrichtigen“ Verbindungen werden bei nachfolgendem Scoringverfahren berücksichtigt. Des Weiteren werden oftmals Zusammenhänge zwischen zwei Konzepten erkannt, jedoch die falsche Relation bzw. die falsche Relation und Richtung verwendet. Auch diese Propositionen werden in das Scoringverfahren aufgenommen.

Im Folgenden werden daher vier Varianten für das Bewerten der Concept Maps unterschieden:

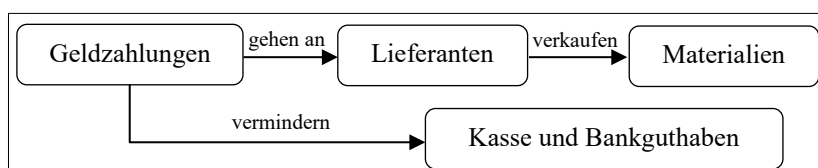
- 1) Die Proposition ist im Referenznetz enthalten und wird daher mit *1 Punkt* bewertet. Sie stellt zusammenhängende Konzepte mit der richtigen Relation und in der richtigen Richtung dar.
- 2) Die Proposition ist nicht im Referenznetz vorhanden, jedoch wurde der bestehende Zusammenhang zwischen beiden Konzepten erkannt. Des Weiteren sollte die Sinnhaftigkeit erkennbar sein. Hierbei werden zwei Fälle unterschieden:
 - a. $\frac{1}{2}$ *Punkt* für zusammenhängende Konzepte mit falscher Relation aber richtiger Richtung⁷⁷,
 - b. $\frac{1}{3}$ *Punkt* für das Erkennen zusammenhängender Konzepte, aber Richtung und Relation ist falsch.
- 3) Die Proposition ist nicht im Referenznetz vorhanden, der Zusammenhang kann jedoch über eine bestimmte Anzahl an Propositionen im Referenznetz nachvollzogen werden. Der Proband „überspringt“ somit gedanklich einige Konzepte. Die dargestellte Proposition im Individualnetz steht demnach für eine bestimmte Anzahl an Propositionen im Referenznetz. Für eine richtige Proposition könnte maximal 1 Punkt vergeben werden. Somit ergibt sich folgende Bewertung:

$$\frac{1}{\text{Anzahl von Propositionen im Referenznetz für dargestellten Zusammenhang}}$$

Dabei sollte die Sinnhaftigkeit dieser Proposition erkennbar sein. Beispielsweise stellte ein Proband im 1. Mapping die Proposition „Materialien vermindern Kasse und Bankguthaben“ dar. Bei der ursprünglichen Bewertung hätte er für diese Proposition 0 Punkte erhalten.



Im Referenznetz wird der Zusammenhang anhand von 3 Propositionen dargestellt. Somit ergibt sich eine Bewertung von $\frac{1}{3}$ Punkt.



⁷⁷ Propositionen, bei denen die richtige Relation aber die falsche Richtung verwendet wurde, mussten als falsch und somit mit 0 Punkten bewertet werden, da der richtige Zusammenhang nicht erkannt wurde.

4) Sofern die Regeln 1-3 nicht angewendet werden können, wird die Proposition mit 0 Punkten bewertet.

Der Scoringwert, als Grad der Übereinstimmung [$\ddot{U}G_{diff}$], wird auf Basis der Galantermetrik unter Berücksichtigung der differenzierten Bewertungsregeln ermittelt. Auf Grundlage der differenzierten Bewertungsregeln ergibt sich beispielsweise für das erste erstellte Concept Map des Probanden Ei04 eine Schnittmenge $m(A \cap B)$ von 15. Nachfolgende Tabelle 21 zeigt die Bewertung der Propositionen des ersten Maps für den Probanden Ei04 sowie deren Begründung. Das erstellte Map des Probanden beinhaltet 13 Propositionen, welche im Referenznetz vorhanden sind und somit mit 1 Punkt bewertet werden können. Des Weiteren können 2 Propositionen mit $\frac{1}{2}$ Punkt und 2 Propositionen mit $\frac{1}{3}$ Punkt bewertet werden, da die vom Probanden dargestellten Zusammenhänge über 2 bzw. 3 Propositionen im Referenzmap darstellbar sind. Eine Proposition wird mit $\frac{1}{3}$ Punkt bewertet, da der dargestellte Zusammenhang im Referenzmap enthalten ist, jedoch die Relation und die Richtung falsch sind.

Das relative Distanzmaß beträgt für den Probanden Ei04: $\frac{47+21-2 \cdot 15}{47+21-15} = 0,7170$. Der Übereinstimmungsgrad $\ddot{U}G_{diff} = 0,2830$. Auf diese Weise wurden alle Probandennetze mit Hilfe von Excel bewertet. Beim Vergleichen von $\ddot{U}G$ mit $\ddot{U}G_{diff}$ ist erkennbar, dass mit dem differenzierten Verfahren beim Probanden Ei04 eine 4,66 Prozent höhere Übereinstimmung ermittelt werden konnte.

Tabelle 21: Bewertung des ersten Maps des Probanden Ei04 anhand des entwickelten Scoringverfahrens

Proposition	Bewertung	Begründung
Fertigung erzeugt Produkte	1	Proposition aus Referenzmap 1
Kunden erwarten guten Service	1	Proposition aus Referenzmap 1
Kunden sollen haben große Abnahmemengen	1	Proposition aus Referenzmap 1
Kunden sollen haben Treue	1	Proposition aus Referenzmap 1
Kunden sollen haben Zahlungspünktlichkeit	1	Proposition aus Referenzmap 1
Kunden tätigen Geldzahlungen	1	Proposition aus Referenzmap 1
Lieferanten sollen haben Liefertreue	1	Proposition aus Referenzmap 1
Lieferanten verkaufen Materialien	1	Proposition aus Referenzmap 1
Materiallager gibt heraus Materialien	1	Proposition aus Referenzmap 1
Materialien gehen an Fertigung	1	Proposition aus Referenzmap 1
Materialien gehen an Materiallager	1	Proposition aus Referenzmap 1
Maschinen und Anlagen sind Voraussetzung für Fertigung	1	Proposition aus Referenzmap 1
Produkte gehen an Fertigwarenlager	1	Proposition aus Referenzmap 1
Fertigwarenlager beliefert Kunden	1/2	korrekter Weg im Referenzmap mit 2 Propositionen (Fel ghe Pro, Pro gan Kun)
Kasse und Bankguthaben begleicht Fertigung	1/2	korrekter Weg im Referenzmap mit 2 Propositionen (Gez erh KuB, Gez gan Fer)
gute Konditionen sind Voraussetzung für Fertigung	1/3	korrekter Weg im Referenzmap mit 3 Propositionen (Lie sha gKo, Lie vek Mat, Mat iVf Fer)
Kasse und Bankguthaben begleicht Materialien	1/3	korrekter Weg im Referenzmap mit 3 Propositionen (Gez gan KuB, Gez gan Lie, Lie vek Mat)
Rechnungen sind Voraussetzung für Geldzahlungen	1/3	falsche Relation und falsche Richtung (Proposition im Referenzmap: Gez beg Rec)
Forderungen sind Liefertreue	0	falsche Verwendung des Begriffs Forderungen (Problem: robuste Alltagsvorstellungen)
Kunden haben Forderungen	0	falsche Verwendung des Begriffs Forderungen (Problem: robuste Alltagsvorstellungen)
Zahlungspünktlichkeit führt zu Vertrauen	0	Konzept nicht Teil des Referenzmaps, Zusammenhang über Lie erw Zpt bzw. Kun sha Zpt abgedeckt

Überprüfung der Intercoderreliabilität für das differenzierte Scoringverfahren

Um das differenzierte Bewertungsverfahren zu überprüfen, wurde bei 16 Prozent der Probandennetze die Intercoderreliabilität berechnet. Als Grundlage dienten 27 der insgesamt 168 erstellten Concept Maps, welche zufällig ausgewählt wurden. Die Netze wurden von einem zweiten Codierer anhand der vorliegenden Bewertungsregeln erneut bewertet. Mit Hilfe der Korrelation nach Spearman wurde für die Intercoderreliabilität ein Wert von $r_s=0,926$ ($p=0,000$) ermittelt, was bedeutet, dass zwischen beiden Codierern bei unabhängiger Bewertung der Maps eine Übereinstimmung von 92,6 Prozent erzielt wurde. Abschließend fand bei strittigen Bewertungen eine Konsensbildung zwischen Erst- und

Zweitcodierer statt, sodass im Ergebnis eine von beiden Codierern akzeptierte Bewertung der Maps für die weitere Auswertung vorlag.

Evaluation des differenzierten Scoringverfahrens

Für eine weitere Analyse der Concept Maps stellt sich nun die Frage, welches Bewertungsverfahren angewendet werden sollte. Dabei empfiehlt es sich nur eines der beiden Maße heranzuziehen, sofern diese hoch korrelieren. Alle Werte für ÜG wurden dazu den Werten für ÜG_{diff} aus dem differenzierten Bewertungsverfahren gegenübergestellt. Mit Hilfe der Korrelation nach Spearman wurde ein Wert von $r_s=0,979$ ($p=0,000$) für Spearman's Rho ermittelt. Somit liegt eine sehr hohe Korrelation vor. Für weitere Analysen werden die Werte für ÜG_{diff} herangezogen, welche im Folgenden als Scoringwerte bezeichnet werden. Diese Werte berücksichtigen den Aspekt, dass die Schüler nach so kurzer Zeit noch nicht in der Lage sind, alle wissenschaftlichen Zusammenhänge in der geforderten Tiefe zu verstehen und somit eher oberflächliche Strukturen erkannt haben. Somit spiegelt das differenzierte Scoringverfahren das tatsächliche Wissen der Probanden besser wider.

Nachfolgend dargestellte Boxplots zeigen das Minimum, Maximum, das 25 und 75 Prozent-Quartil sowie den Median der Scoringwerte des 1., 2. und 3. Mappings (vgl. Abbildung 14). Im 1. Mapping ist ein Ausreißer erkennbar (Proband E107), für dessen Map ein Scoringwert von 0 ermittelt wurde, da lediglich isolierte Konzepte dargestellt wurden. Im 2. Mapping können 2 Ausreißer nach oben identifiziert werden. Für den Probanden CM05 wurde ein Scoringwert von 0,40 ermittelt, für den Probanden E107 ein Wert von 0,43. Im 3. Mapping hob sich erneut Proband E107 mit einem Scoringwert von 0,42 von den anderen Probanden ab. An dieser Stelle können die Ergebnisse mit denen der Wissenstests verglichen werden (vgl. Kapitel 4.4.1). Im 1. Nachtest und 2. Nachtest wurde ebenfalls der Proband E107 als Ausreißer identifiziert. Die Scoringwerte des 1. Mapping sind insgesamt am höchsten, die Werte des 3. Mapping am geringsten. Dies kann auf die Themenbereiche der Mappingaktivitäten zurückgeführt werden. Das 1. Mapping zu den Abläufen in einem Produktionsunternehmen war für die Probanden am einfachsten zu erstellen, da sie hier ihre Erfahrungen und das Vorwissen einbeziehen konnten (Median bei 0,22). Das 2. Mapping zu den Zusammenhängen zwischen linearer Abschreibung, GuV und Bilanz stellte ein vollkommen neues Inhaltsgebiet für die Lernenden dar, so dass eine

Erstellung deutlich schwieriger war (Median bei 0,15). Das 3. Mapping stellte die Probanden vor die größten Herausforderungen, da sie die Zusammenhänge zwischen dem ersten und dem 2. Map herstellen sollten (Median bei 0,07).

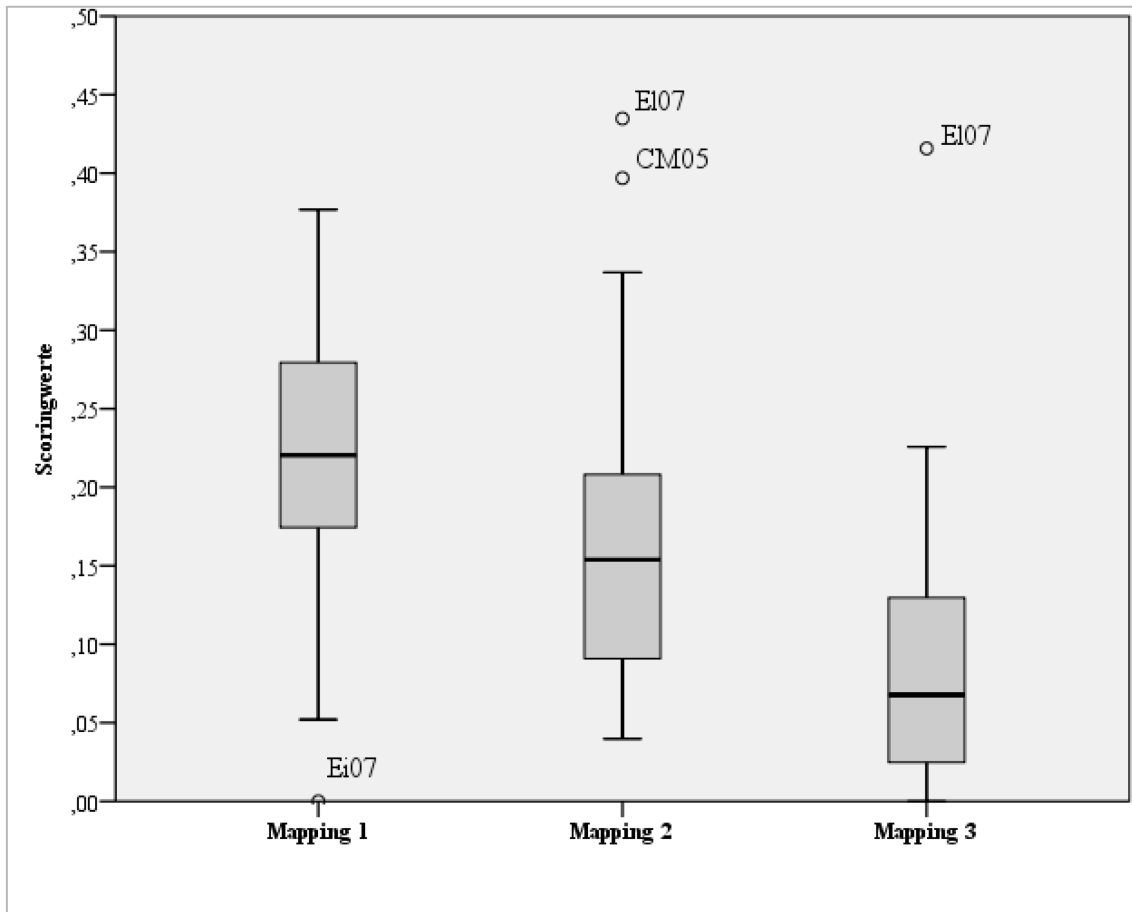


Abbildung 14: Boxplots der Scoringwerte für das Mapping1, Mapping 2 und Mapping 3

Nachfolgende Tabelle 22 stellt die Mittelwerte und Standardabweichungen für die Scoringwerte dar.

Tabelle 22: Mittelwerte und Standardabweichungen für die Scoringwerte der drei Mapping-Aktivitäten

	Mapping 1	Mapping 2	Mapping 3
Mittelwert	0,2195	0,1630	0,0815
Standardabweichung	0,0784	0,0916	0,0771

Die Qualität der Maps ist beim 1. Mapping am höchsten und beim 3. Mapping am geringsten, wobei jeweils eine starke Streuung erkennbar ist.

4.6 Auswertung des Fragebogens zur Einschätzung des Umgangs mit Concept Mapping, des Umgangs mit Fehlern sowie der Nützlichkeit des Feedbacks

4.6.1 Auswertung des Teil-Fragebogens zur Einschätzung des Concept Mapping

Zunächst wurden die Fragebögen in Excel erfasst. Hierzu war es vonnöten die Items der jeweiligen Kategorie, Interessantheit [I], Handhabbarkeit [H] und Nützlichkeit [N], zuzuordnen. Darüber hinaus wurden die ausgewählten Antworten der Probanden mit Zahlen kodiert (trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4). Anschließend erfolgte eine Übertragung der Daten in SPSS. Für eine einheitliche Richtung bei der Auswertung wurden im nächsten Schritt die negativ gepolten Antworten umkodiert (vgl. Kapitel 4.3.3).

In einem weiteren Schritt wurde Cronbachs Alpha⁷⁸ für die drei Bereiche berechnet: Für den Bereich Interessantheit liegt dieses bei 0,776, für die Handhabbarkeit bei 0,693 und für die Nützlichkeit bei 0,287. Diese Werte zeigen, dass die interne Konsistenz für den Bereich Interessantheit und Handhabbarkeit akzeptabel ist, wohingegen die interne Konsistenz für den Bereich Nützlichkeit daraufhin deutet, dass die Items unzureichend zwischen denen differenzieren, die das Concept Mapping nützlich oder nicht nützlich fanden. Cronbachs Alpha konnte auch durch den Ausschluss von Variablen nicht erhöht werden.

Aus diesem Grund wurde anschließend eine Faktorenanalyse durchgeführt. Diese sollte Aufschluss darüber geben, inwieweit die Variablen wirklich auf die drei Dimensionen Interessantheit, Handhabbarkeit und Nützlichkeit reduziert werden können (Backhaus, Erichson, Plinke, & Weiber, 2015, S. 385 ff.). Die Faktorenanalyse ist eine Interdependenzanalyse, welche ein metrisches (Intervall- oder Verhältnisskala) Skalenniveau voraussetzt. In der üblichen Forschungspraxis wird davon ausgegangen, dass die Stufen der Ratingskalen intervallskaliert sind (Bortz & Schuster, 2016, S. 23). Die Daten des Fragebogens zum Umgang mit Concept Mapping erfüllen somit diese Voraussetzung. Eine Datenverdichtung ist sinnvoll, wenn die Variablen eine hohe statistische und inhaltliche Nähe aufweisen. Hierzu gilt es einige Kriterien zu prüfen. Anhand des Bartlett-Tests wird die Nullhypothese geprüft, ob die Variablen in der Grundgesamtheit unkorreliert sind. Der Test setzt jedoch eine Normalverteilung der Daten voraus (Backhaus u. a., 2015, S.

⁷⁸ Cronbachs Alpha ist ein Maß für die Reliabilität (Zuverlässigkeit) der Messvariablen, welche einen gemeinsamen Faktor messen. Ab welchem Richtwert die interne Konsistenz als akzeptabel gilt, ist in der Literatur umstritten. Dabei wird jedoch häufig der von Nunnally (1978, S. 245) vorgeschlagene Mindestwert von 0,7 angegeben.

397). In der vorliegenden Studie zeigt der Kolmogorov-Smirnov-Anpassungstest für alle Items zur Einschätzung des Concept Mapping eine signifikante Abweichung von der Normalverteilung (Signifikanzniveau 5%⁷⁹). Neben der Prüfung mit statistischen Tests sollen jedoch auch deskriptive Analysen einbezogen werden (Janssen & Laatz, 2017, S. 248 ff.). Die Histogramme aller Variablen zeigen augenscheinlich eine Normalverteilung. Des Weiteren lassen sich die Schiefe (Maß für die Asymmetrie einer Verteilung) und Kurtosis (Steilheit der Wölbung einer Verteilung) als Kennzahlen für die Form einer Werteverteilung interpretieren. Werte von Null entsprechen dabei einer Normalverteilung (Janssen & Laatz, 2017, S. 213 f.). Die Werte für die Schiefe reichen von -0,541 bis 0,377, die Werte für die Kurtosis von -1,066 bis 0,132 und sind somit nahe Null, so dass insgesamt annähernd von einer Normalverteilung ausgegangen werden kann.

Der Bartlett-Test erbrachte einen Signifikanzwert von 0,000 und die Nullhypothese, dass die Variablen in der Grundgesamtheit unkorreliert sind, muss somit abgelehnt werden. Das erste Kriterium ist folglich erfüllt.

Zur weiteren Beurteilung lässt sich die Anti-Image-Kovarianzmatrix heranziehen.⁸⁰ Dabei dürfen maximal 25 Prozent der Nicht-Diagonal-Elemente größer 0,09 sein (Backhaus u. a., 2015, S. 398). Dieses Kriterium wird hier nicht erfüllt. Die Analyse kann trotzdem weitergeführt werden, da es sich nicht um ein Ausschlusskriterium handelt. Beim Kaiser-Meyer-Olkin-Test wird ermittelt, wie hoch die Variablen mit den Faktoren korrelieren. Dabei wird der Test für die gesamte Korrelationsmatrix [KMO] und für einzelne Variablen (Measure of sampling adequacy, [MSA]) durchgeführt. Der kritische Wert von 0,5 sollte nicht unterschritten werden (Backhaus u. a., 2015, S. 399). Das KMO-Kriterium ist mit einem Wert von 0,618 erfüllt. Die MSA-Werte zweier Variablen unterschreiten mit 0,471 bzw. 0,498 knapp den kritischen Wert, wobei alle anderen das Kriterium erfüllen. Aufgrund dessen, dass die Grenze nur knapp unterschritten wird, werden die Variablen nicht eliminiert.

⁷⁹ Für alle statistischen Analysen wurde ein Signifikanzniveau von 5% gewählt.

⁸⁰ Das Anti-Image der Variablen ist der Teil der Varianz, der sich nicht durch andere Variablen erklären lässt und sollte nahe Null sein (Backhaus, Erichson, Plinke, & Weiber, 2015, S. 398).

Das vorliegende Datenmaterial kann folglich als geeignet eingeschätzt und somit die Faktorenextraktion vorgenommen werden. Hierzu erweist sich die Hauptkomponentenanalyse⁸¹ mit einer rechtwinkligen Rotation nach dem Varimax-Verfahren⁸² am gebräuchlichsten (Bühl, 2012, S. 589). Anhand des Kaiser-Kriteriums wird die Anzahl der zu extrahierenden Faktoren bestimmt. Dabei dürfen lediglich Faktoren mit einem Eigenwert⁸³ größer als eins extrahiert werden (Überla, 2013, S. 125). Insgesamt werden vier Faktoren mit einem Varianzerklärungsanteil von 72,812 Prozent extrahiert. Alle Faktoren sollten zusammen mindestens 50 Prozent der Varianz erklären, was folglich gegeben ist (Homburg & Giering, 1996, S. 12).

Des Weiteren ist die Kommunalität als der durch alle Faktoren erklärte Varianzanteil einer Variablen zu prüfen. Die Kommunalitäten der Variablen sollten größer als 0,5 sein (Litfin, Teichmann, & Clement, 2000, S. 285), was an dieser Stelle gegeben ist.

Als weiteres Gütekriterium wird die Faktorladungsstruktur betrachtet. Jede Variable sollte auf lediglich einem Faktor mit einer Faktorladung größer als 0,4 laden (Homburg & Giering, 1996, S. 8). Die rotierte Komponentenmatrix zeigt Doppelladungen mit ähnlichen Werten (vgl. Anhang 4.6.1-1). Daher schlägt Bühner (2011, S. 350) als eine Möglichkeit vor, weniger Faktoren zu extrahieren.

Infolgedessen wurde eine weitere Faktorenanalyse durchgeführt, bei der drei Faktoren als feste Anzahl vorgegeben wurden. Das Ergebnis dieser Analyse zeigt immer noch Doppelladungen mit ähnlichen Werten (vgl. Anhang A-4.6.1-2). Aus diesem Grund wurden in einem nächsten Schritt zwei Faktoren vorgegeben.

Die Kriterien zur Prüfung des Datenmaterials brachten folgende Ergebnisse. Der *Bartlett-Test* zeigt mit 0,000 einen signifikanten Wert, das Kriterium ist folglich erfüllt. In der *Anti-Image-Kovarianz-Matrix* sind mehr als 25 Prozent der nichtdiagonalen Elemente größer als 0,09. Das Kriterium ist somit nicht erfüllt. Mit 0,618 wird der kritische Wert

⁸¹ Bei der Hauptkomponentenanalyse wird davon ausgegangen, dass sich die Varianz aller Variablen vollständig durch die Extraktion von Faktoren erklären lässt, es folglich keine Einzelrestvarianz in den Variablen gibt. Ziel ist es also, die Daten möglichst umfassend durch wenige Faktoren zu reproduzieren (Backhaus u. a., 2015, S. 412 f.).

⁸² Um eine Einfachstruktur der Variablen, d. h. dass alle Variablen nur auf einem Faktor laden, zu erreichen, wird eine Rotation durchgeführt. Das Varimax-Verfahren geht davon aus, dass die Faktoren untereinander nicht korrelieren und somit in einem rechten Winkel zueinander stehen (Backhaus u. a., 2015, S. 435).

⁸³ Der Eigenwert gibt an, wie viel Varianz aller Variablen durch einen Faktor erklärt wird (Backhaus u. a., 2015, S. 359).

des *KMO-Kriteriums* nicht unterschritten. Die *MSA-Werte* für zwei Variablen unterschreiten mit 0,471 bzw. 0,498 nur knapp die Grenze von 0,5 und werden aus diesem Grund nicht aus der Analyse ausgeschlossen. Alle anderen *MSA-Werte* liegen über dem kritischen Wert. Das Datenmaterial ist somit für eine Faktorextraktion geeignet. Die *Hauptkomponentenanalyse* mit einer Rotation nach dem *Varimax-Verfahren* resultierte in den zwei vorgegebenen Faktoren mit einem *Varianzerklärungsanteil* von 50,885 Prozent. Die *Kommunalitäten* der Variablen sollten größer als 0,5 sein, was an dieser Stelle nicht gegeben ist. Lediglich fünf der zehn Variablen erfüllen diese Voraussetzung, eine Variable unterschreitet mit 0,459 diesen Wert nur knapp. Vor dem Hintergrund der inhaltlichen Interpretierbarkeit wird die Vorgabe von zwei Faktoren gegenüber der Vierfaktoren-Lösung trotzdem vorgezogen. Die rotierte Komponentenmatrix zeigt eine Einfachstruktur. Alle Variablen laden auf nur einem Faktor, wobei die *Faktorladungen* zwischen 0,549 und 0,876 liegen, den Grenzwert von 0,4 somit deutlich überschreiten (vgl. vgl. Anhang A-4.6.1-3).

Im Anschluss erfolgte die Beurteilung der internen Konsistenz anhand von *Cronbachs Alpha*. Die Alpha-Werte können dabei mit 0,747 für den ersten Faktor sowie 0,693 für den zweiten Faktor als akzeptabel angesehen. Abschließend wird die *Item-to-Total-Korrelation*, als Korrelation jeder Variable mit der Summe aller dem Faktor zugewiesenen Variablen, berechnet (Homburg & Giering, 1996, S. 8 f.). In der Literatur existiert für die Item-to-Total-Korrelation kein kritischer Grenzwert. Churchill (1979) empfiehlt sie bei Bedarf heranzuziehen, um Cronbachs Alpha durch das Eliminieren von Indikatoren zu erhöhen. Eliminiert werden dabei Indikatoren mit einer sehr niedrigen Item-to-total-Korrelation. Für den ersten Faktor zeigen zwei Variablen mit 0,518 bzw. 0,474 niedrige Werte (vgl. Anhänge A-4.6.1-4 und A-4.6.1-5). Bei dem Item *Nützlichkeit3umkodiert* (Frage 6) mit dem Wert von 0,474 handelt es sich um das zum Teil-Fragebogen hinzugefügte Item, welches die Nützlichkeit des ergänzend eingesetzten Concept Mappings im Vergleich zur alleinigen Durchführung des Planspiels überprüfen sollte. Dieses Item wurde in keiner vorherigen Studie eingesetzt und dessen Reliabilität überprüft. Da die Item-to-Total-Korrelation doch deutlich unter 0,6 liegt, wird diese Variable eliminiert.

Anschließend wird eine Faktorenanalyse mit den verbleibenden neun Items durchgeführt. Zwei Kriterien zur Prüfung des Datenmaterials, *Bartlett-Test* (Signifikanzwert von 0,000) sowie *KMO-Kriterium* (0,616) sind erfüllt.

Tabelle 23: KMO-Kriterium und Bartlett-Test für die Zwei-Faktoren-Lösung mit neun Variablen

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		0,616
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	150,544
	df	36
	Signifikanz nach Bartlett	0,000

In der *Anti-Image-Kovarianz-Matrix* wird der Grenzwert von 25 Prozent, wie in den vorherigen Analysen, überschritten (44,444 Prozent), das Kriterium somit nicht erfüllt. Die *MSA-Werte* für drei Variablen unterschreiten mit 0,487, 0,484 bzw. 0,487 nur knapp den kritischen Wert von 0,5 und werden daher nicht ausgeschlossen. Alle anderen *MSA-Werte* liegen über dem Grenzwert (vgl. Anhang A-4.6.1-6).

Das als geeignet eingeschätzte Datenmaterial wird erneut auf Grundlage der *Hauptkomponentenanalyse* mit einer Rotation nach dem *Varimax-Verfahren* extrahiert, wobei die zwei resultierenden Faktoren 55,033 Prozent der *Varianz erklären* (vgl. Anhang A-4.6.1-7).

Die *Kommunalitäten* von drei Variablen unterschreiten die Grenze von 0,5 deutlich (0,283; 0,414; 0,385) eine Variable liegt mit 0,486 knapp unter diesem Wert (vgl. Tabelle 24). Vor dem Hintergrund der inhaltlichen Interpretierbarkeit wird diese Lösung trotzdem der Vier-Faktoren-Lösung vorgezogen.

Tabelle 24: Kommunalitäten für die Zwei-Faktoren-Lösung mit neun Variablen

	Anfänglich	Extraktion
Interessantheit1	1,000	0,758
Nützlichkeit1	1,000	0,283
Nützlichkeit2	1,000	0,643
Interessantheit2umkodiert	1,000	0,791
Interessantheit3	1,000	0,385
Handhabbarkeit1umkodiert	1,000	0,414
Handhabbarkeit2	1,000	0,486
Handhabbarkeit3	1,000	0,634
Handhabbarkeit4umkodiert	1,000	0,559

Extraktionsmethode: Hauptkomponentenanalyse.

Die *Faktorladungen* liegen zwischen 0,518 und 0,883, somit über dem kritischen Wert von 0,4 (vgl. Tabelle 25).

Tabelle 25: Rotierte Faktorladungen für die Zwei-Faktoren-Lösung mit neun Variablen

	Komponente	
	1	2
Interessantheit1	0,870	
Nützlichkeit1	0,532	
Nützlichkeit2	0,802	
Interessantheit2umkodiert	0,883	
Interessantheit3	0,621	
Handhabbarkeit1umkodiert		0,628
Handhabbarkeit2		0,691
Handhabbarkeit3umkodiert		0,785
Handhabbarkeit4umkodiert		0,747

Extraktionsmethode: Hauptkomponentenanalyse,
Rotationsmethode: Varimax mit Kaiser-Normalisierung^a

Cronbachs Alpha erhöht sich für den ersten Faktor durch das Ausschließen der Variable *Nützlichkeit3umkodiert* auf 0,778 (vgl. Tabelle 26). Die interne Konsistenz für den zweiten Faktor bleibt unverändert bei 0,693.

Tabelle 26: Interne Konsistenz für den Faktor 1 für die Zwei-Faktorenlösung mit neun Variablen

Reliabilitätsstatistiken	
Cronbachs Alpha	Anzahl der Items
0,778	5

Bezüglich der *Item-to-Total-Korrelation* unterschreitet nur noch eine Variable des ersten Faktors den Wert von 0,6 (vgl. Anhang A-4.6.1-8).

Die Güte der Analyse konnte durch Ausschluss der Variable *Nützlichkeit3umkodiert* erhöht werden. Der Anteil der erklärten Varianz erhöht sich von 50,885 auf 55,033 Prozent. Darüber hinaus verbessern sich die MSA-Werte sowie das Cronbachs Alpha für den ersten Faktor.

Letztlich muss festgestellt werden, dass die Faktorenanalyse mit der Extraktion von vier Faktoren eine bessere Güte aufweist. Jedoch sollte eine Faktorenanalyse auch immer vor dem Hintergrund der inhaltlichen Interpretierbarkeit betrachtet werden. Aufgrund der Doppelladungen mit ähnlichen Werten konnte dies nicht gewährleistet werden, was letztlich zu der Zwei-Faktoren-Lösung führte.

Nachfolgende Tabelle 27 stellt die Kriterien zur Prüfung des Datenmaterials sowie die Gütekriterien für die Zwei-Faktoren-Lösung mit neun Variablen noch einmal zusammenfassend dar.

Tabelle 27: Kriterien zur Prüfung des Datenmaterials und Gütekriterien für die Zwei-Faktoren-Lösung mit neun Variablen

Kriterium, Bedingung	Wert in Studie Erfüllt/nicht erfüllt
<i>Kriterien zur Prüfung des Datenmaterials</i>	
Bartlett-Test; Signifikanz $\leq 0,05$	0,000 erfüllt
Anti-Image-Kovarianzmatrix max. 25% der nichtdiagonalen Elemente betragsmäßig $\geq 0,09$	44,444% nicht erfüllt , aber kein Ausschlusskriterium
KMO-Kriterium $\geq 0,5$	0,616 erfüllt
MSA-Werte $\geq 0,5$	0,487, 0,484 bzw. 0,487; nur knapp unter 0,5; alle anderen darüber erfüllt
<i>Gütekriterien</i>	
Eigenwert-Kriterium ≥ 1	2 Faktoren, Eigenwert gesamt: 2,059 erfüllt
Erklärter Varianzanteil $\geq 50\%$	55,033% erfüllt
Kommunalitäten $\geq 0,5$	0,283; 0,414; 0,385; 0,486; alle anderen über 0,5 nicht erfüllt , Zwei-Faktorenlösung aufgrund besserer Interpretierbarkeit trotzdem bevorzugt
Faktorladungsstruktur Faktorladung $\geq 0,4$; keine Doppelladungen	Faktor 1 und 2: alle über 0,4; keine Doppelladungen erfüllt
Cronbachs Alpha $\geq 0,7$	Faktor 1: 0,778 Faktor 2: 0,693; nur knapp unter 0,7 erfüllt
Item-to-Total-Korrelation $\geq 0,6$	Faktor 1: 0,548, alle anderen über 0,6 Faktor 2: alle über 0 erfüllt

Auf dem ersten Faktor laden die Items des Bereiches Interessantheit und Nützlichkeit. Der Faktor wird für folgende Analysen als *Interessantheit-Nützlichkeit* bezeichnet. Die theoretische Zuordnung kann daher nicht beibehalten werden. Es zeigt sich, dass die Interessantheit mit der Nützlichkeit in Bezug auf den Umgang mit Concept Mapping zusammenhängt. Dem zweiten Faktor sind alle Variablen des Bereiches *Handhabbarkeit* zugewiesen. Somit konnte die theoretische Zuordnung bestätigt werden.

Abschließend werden die Mittelwerte und Standardabweichungen der Faktoren *Interessantheit-Nützlichkeit* und *Handhabbarkeit* dargestellt (vgl. Tabelle 28).

Tabelle 28: Mittelwerte und Standardabweichungen für die Faktoren Interessantheit-Nützlich-
 keit und Handhabbarkeit

Faktor	Mittelwert	Standardabweichung
Interessantheit-Nützlich- keit	2,694	0,633
Handhabbarkeit	2,438	0,587

trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4

Es zeigt sich, dass die Interessantheit-Nützlich-
 keit höher bewertet wird als die Handhab-
 barkeit. Die Probanden schätzen die Methode des Concept Mapping eher als interessant
 und nützlich ein. Das Concept Mapping wird hingegen als eher nicht handhabbar einge-
 schätzt.

4.6.2 Auswertung des Teil-Fragebogens zur Einschätzung der Nützlich- keit des Feedbacks

Die ausgewählten Antworten der Probanden wurden mit Zahlen kodiert (trifft nicht zu:
 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4) und in Excel erfasst. Die Items
 wurden entsprechend der Art des Feedbacks mit Abkürzungen (Einschätzung der Nützlich-
 keit des einfachen Feedbacks [EiFBN], Einschätzung der Nützlich-
 keit des elaborierten Feedbacks [ElFBN], gewünschtes Feedback [FBgew] und gewünschtes einfaches
 Feedback [ElFBgew]) kodiert. Für eine einheitliche Richtung bei der Auswertung wurden
 im nächsten Schritt die negativ gepolten Antworten umkodiert. Anschließend erfolgte
 eine Übertragung der Daten in SPSS.

Zunächst wurde die interne Konsistenz für die drei Items bezüglich EiFBN berechnet.
 Mit einem Wert von 0,324 ist diese inakzeptabel. In einem nächsten Schritt wurde das
 Item *EiFBN_3umkodiert* (Frage 13, vgl. Kapitel 4.3.3) ausgeschlossen. Die interne Kon-
 sistenz konnte so auf einen akzeptablen Wert von 0,767 erhöht werden. Das Item
EiFBN_3umkodiert passt mit den anderen Items zur Nützlich-
 keit des einfachen Feed-
 backs nicht zusammen. Dies kann mit der Fragerichtung zusammenhängen. Das Item
EiFBN_3 ist negativ gepolt. Es stellt sich folglich die Frage, inwieweit die Probanden die
 Frage richtig verstanden haben. Eine weitere Ursache kann in der Formulierung des Items
 begründet liegen. Die Frage ist sehr global formuliert. Die Probanden sollen allgemein
 einschätzen, inwieweit sie das einfache Feedback als überflüssig erachten. Die beiden
 anderen Items hinterfragen speziell die Möglichkeit, mithilfe des einfachen Feedbacks
 Wissenslücken aufzudecken bzw. Zusammenhänge zwischen Begriffen besser zu verste-
 hen.

Falls die Antwortkategorie „trifft eher zu“ oder „trifft voll zu“ bei der Frage 13 gewählt wurde, sollte ein Grund ausgewählt oder ein eigener Grund angegeben werden, warum die Probanden das einfache Feedback als überflüssig erachten. Von den 17 Probanden der Gruppe *EiFB* und *ElFB*, welche das einfache Feedback als überflüssig erachteten (d. h. „trifft eher zu“ oder „trifft voll zu“ ankreuzten), gaben den Grund „mir waren die Zusammenhänge auch schon nach dem Erstellen der Concept Maps klar“ (13a) fünf Schüler und „trotz des Vergleichens habe ich die Beziehungen zwischen den Begriffen nicht verstanden“ (13b) acht Schüler an. Als anderer Grund (13c) wurde einmal „ich hatte viel zu viele andere Wörter und dadurch andere Verknüpfungen“ genannt. Insgesamt hatten die Probanden Schwierigkeiten bei der Angabe der Gründe. Dies äußerte sich dadurch, dass einige Schüler keinen Grund nannten, teilweise mehrere Gründe angegeben wurden oder Gründe ausgewählt wurden, obwohl die Probanden das einfache Feedback gar nicht als überflüssig eingeschätzt hatten. Bei Aufgabe 13c wurde einmal der Grund „Wenn man es noch einmal korrigiert u. vergleicht bleibt mehr hängen“ und ein anderes Mal „fand es besser zu verstehen“ angegeben. Diese Antworten zeigen, dass die Probanden das einfache Feedback eigentlich als nützlich einschätzten und dass sie scheinbar mit der Richtung der Frageformulierung Probleme hatten.

Für weitergehende Analysen zur Einschätzung der Nützlichkeit des einfachen Feedbacks wurden daher lediglich die Items *EiFBN_1* und *EiFBN_2* einbezogen.

Anschließend wurde eine Faktorenanalyse durchgeführt, um die Skala statistisch zu validieren. Da der Teil-Fragebogen vorher noch nicht eingesetzt wurde, konnte auch noch nicht überprüft werden, inwieweit die Items die Skala angemessen repräsentieren. Des Weiteren musste für einen akzeptablen Wert für die interne Konsistenz (0,767) ein Item ausgeschlossen werden, so dass eine Überprüfung der Skala mittels einer Faktorenanalyse wichtig ist.

Mittels einer Faktorenanalyse können beide Items auf einen Faktor verdichtet werden. Die Kriterien zur Prüfung des Datenmaterials, *Bartlett-Test* (Signifikanzwert von 0,000), *KMO-Kriterium* (0,500) sowie *MSA-Werte* (0,500 für beide Variablen) erfüllt. In der *Anti-Image-Kovarianz-Matrix* sind mehr als 25 Prozent der nichtdiagonalen Elemente größer als 0,09, das Kriterium stellt jedoch kein Ausschlusskriterium dar. Das als geeignet eingeschätzte Datenmaterial wird mittels *Hauptkomponentenanalyse* auf einen Faktor mit einem *Varianzanteil* von 81,122 Prozent extrahiert. Die Werte für die Kommunalitäten (0,811; 0,811), die Faktorladungen (0,901; 0,901), für Cronbachs Alpha (0,767) sowie

für die Item-to-Total-Korrelation (0,900; 0,901) übersteigen die jeweilige Grenze, die Gütekriterien sind somit erfüllt. Die Ergebnisse sind zusammenfassend in den Anhängen 4.6.2-1 bis 4.6.2-6 dargestellt. Abschließend sind die Mittelwerte und Standardabweichungen für den Faktor *EiFBN* dargestellt (vgl. Tabelle 29). Mit einem Mittelwert von 2,776 schätzen die Probanden das einfache Feedback als eher nützlich ein.

Tabelle 29: Mittelwerte und Standardabweichungen für *EiFBN*

Faktor	Mittelwert	Standardabweichung
<i>EiFBN</i>	2,776	0,777

trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4

In einem nächsten Schritt wurde die interne Konsistenz für die Items bezüglich *EiFBN* berechnet (vgl. Tabelle 30).

Tabelle 30: Interne Konsistenz für *EiFBN*

Reliabilitätsstatistiken	
Cronbachs Alpha	Anzahl der Items
0,157	3

Mit einem Wert von 0,157 ist diese inakzeptabel. Cronbachs Alpha konnte auch durch den Ausschluss von Variablen nicht erhöht werden. Es zeigt sich, dass die Items den Bereich *EiFBN* nicht ausreichend repräsentieren. Bei genauerer Betrachtung der Variablen fällt auf, dass diese auf verschiedenen Ebenen formuliert sind. Das Item *EiFBN_2* ist global formuliert und fragt allgemein, inwieweit die Probanden das elaborierte Feedback als überflüssig erachten. Die beiden anderen Items beziehen sich darauf, was genau die Probanden an dem elaborierten Feedback als nützlich einschätzen. Das Item *EiFBN_1* zielt auf Fehlerursachen ab, das Item *EiFBN_3* hinterfragt, inwieweit das elaborierte Feedback das Verstehen der Zusammenhänge fördert. Inhaltlich werden hier folglich zwei völlig verschiedene Bereiche angesprochen. Hinsichtlich des Items *EiFBN_2* schätzen lediglich vier Probanden das elaborierte Feedback als überflüssig ein (Antwortkategorie trifft eher zu bzw. trifft voll zu). Bei der Angabe der Gründe hatten einige Schüler jedoch auch wie beim Item *EiFBN_3* Schwierigkeiten. Dreimal wurden z. B. Gründe angegeben, obwohl das elaborierte Feedback gar nicht als überflüssig eingeschätzt wurde. Eine Ursache hierfür kann, wie beim Item *EiFBN_3*, die Fragerichtung sein. Es stellt sich folglich die Frage, inwieweit die Probanden die Frage richtig verstanden haben.

Daher werden die Items *EIFBN_1* und *EIFBN_3* jeweils einzeln für weitere Analysen einbezogen und das Item *EIFBN_2* ausgeschlossen.

Nachfolgende Tabelle 31 zeigt die Standardabweichungen und Mittelwerte für die zwei Variablen zur Einschätzung der Nützlichkeit des elaborierten Feedbacks. Mit Mittelwerten von 3,100 bzw. 2,938 wird das elaborierte Feedback als eher nützlich eingeschätzt.

Tabelle 31: Standardabweichungen und Mittelwerte für EIFBN

Items zu EIFBN	Mittelwert	Standardabweichung
Durch das zusätzliche Besprechen der Fehler sowie die Erklärungen konnte ich die Ursachen meiner Fehler finden.	3,100	0,553
Die zusätzliche Fehlerbesprechung sowie die Erklärungen halfen mir dabei, die Beziehungen zwischen den Begriffen zu verstehen.	2,938	0,772

trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4

Die nachfolgende Tabelle 32 stellt den jeweiligen Mittelwert und die Standardabweichung für die Items bezüglich FBgew und EIFBgew dar. Die Probanden der Gruppe *kein FB* stimmen dem Wunsch nach einer Rückmeldung zu ihren Concept Maps mit einem Mittelwert von 2,833 eher zu. Die Probanden der Gruppe *EiFB* stimmen dem Wunsch nach einer zusätzlichen Besprechung der Fehler mit einem Mittelwert von 2,118 eher nicht zu.

Tabelle 32: Mittelwerte und Standardabweichungen zum gewünschten Feedback

Items zum gewünschten Feedback	Mittelwert	Standardabweichung
<i>FBgew (Gruppe kein FB)</i>		
Nach dem Erstellen der Concept Maps hätte ich mir noch eine Rückmeldung darüber gewünscht, was ich richtig und was ich falsch gemacht habe.	2,833	0,786
<i>EIFBgew (Gruppe EiFB)</i>		
Ich hätte mir eine zusätzliche Besprechung der Fehler gewünscht, da mir dies gezeigt hätte, was die Gründe für meine Fehler waren.	2,118	0,781

trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4

4.6.3 Auswertung des Teil-Fragebogens zur Einschätzung des Umgangs mit Fehlern

Die ausgewählten Antworten der Probanden wurden mit Zahlen kodiert (trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4) und in Excel erfasst. Anschließend erfolgte eine Übertragung der Daten in SPSS.

In einem ersten Schritt wurde die interne Konsistenz für alle neun Items der Skala *Einschätzung zum Umgang mit Fehlern* [UF] berechnet. Mit einem Wert von 0,741 kann diese als akzeptabel angesehen werden (vgl. Tabelle 33).

Tabelle 33: Interne Konsistenz für den Umgang mit Fehlern

Reliabilitätsstatistiken	
Cronbachs Alpha	Anzahl der Items
0,741	9

Der Faktor wird folglich durch die Variablen auch gut repräsentiert. Die Skala wurde durch Spychiger, Kuster und Oser (2006) bereits statistisch validiert. Des Weiteren ist kein Ausschluss einer Variablen vonnöten, um eine akzeptable interne Konsistenz zu erreichen. Alle neun Items repräsentieren die Skala gut. Daher ist es nicht notwendig, mithilfe einer Faktorenanalyse die Subskalen zu überprüfen bzw. neue Faktoren zu identifizieren.

Folgende Tabelle 34 zeigt den Mittelwert und die Standardabweichung für den Faktor *Einschätzung zum Umgang mit Fehlern*.

Tabelle 34: Mittelwerte und Standardabweichungen für den Umgang mit Fehlern

Faktor	Mittelwert	Standardabweichung
Umgang mit Fehlern	2,683	0,477

trifft nicht zu: 1, trifft eher nicht zu: 2, trifft eher zu: 3, trifft voll zu: 4

Mit dem Mittelwert von 2,4683 schätzen die Probanden ihren Umgang mit Fehlern somit als eher konstruktiv ein.

5 Empirische Untersuchung zur Lernwirksamkeit von Feedback zu Concept Maps – Analysen und Ergebnisse

Das folgende Kapitel beinhaltet Analysen und Ergebnisse der empirischen Untersuchung zur Lernwirksamkeit des Feedbacks zu Concept Maps. Die Lernwirksamkeit als abhängige Variable wird dabei über den Wissenserwerb sowie die Qualität der Maps operationalisiert. Im Kapitel 5.1 werden inferenzstatistische und deskriptive Analysen zum Einfluss der Art des Feedbacks auf den Wissenszuwachs (Kapitel 5.1.1) sowie zum Einfluss der Art des Feedbacks auf die Qualität der Maps (Kapitel 5.1.2) durchgeführt. Dabei werden auch personale und situative Einflussfaktoren untersucht. Des Weiteren wird der Einfluss der Qualität der Maps auf den Wissenserwerb untersucht (Kapitel 5.1.3). Die Ergebnisse werden jeweils zusammengefasst und im Hinblick auf die aufgestellten Hypothesen bewertet.

Kapitel 5.2 analysiert zunächst den Umgang mit dem Feedback (Kapitel 5.2.1). Die Hypothesen zum Einfluss des eingeschätzten Umgangs mit Feedback auf die Lernwirksamkeit des Feedbacks (H27-H32) werden erst hier untersucht, um quantitative und qualitative Ergebnisse zum Umgang mit dem Feedback gegenüberstellen zu können. Kapitel 5.2.2 beleuchtet die erstellten Concept Maps in Bezug auf ihre inhaltliche Richtigkeit, um auf diese Weise Fehlkonzepte und fehlende Propositionen identifizieren zu können. Kapitel 5.3 interpretiert die Ergebnisse unter Rückbezug auf die aufgestellten Hypothesen sowie die theoretischen Grundlagen der Kapitel 2 und 3.

5.1 Analysen zum Einfluss der Art des Feedbacks

5.1.1 Analysen zum Einfluss der Art des Feedbacks auf den Wissenserwerb

5.1.1.1 Analyse der Lernwirksamkeit des Concept Mapping

Bevor überprüft werden kann, inwieweit die Art des Feedbacks einen Einfluss auf den Wissenszuwachs hat, wird an dieser Stelle die Lernwirksamkeit des Concept Mapping analysiert. Auf diese Weise kann festgestellt werden, ob das Concept Mapping als ergänzende Lernstrategie einen Vorteil gegenüber der alleinigen Durchführung des Planspiels hat.

Hierzu werden zwei Varianzanalysen [ANOVA] mit Messwiederholung mit den Faktoren *Zeit* und *Gruppe* sowie der abhängigen Variable [AV] *Gruppenmittelwerte Wissens-*

test durchgeführt. Der Faktor *Gruppe* ist dabei zweifach gestuft (*Kontrollgruppe* vs. *Concept Mapping*). Für die Gruppe *Concept Mapping* [CM] werden die Probanden der Gruppen *kein FB*, *EiFB* und *ElFB* zusammengefasst. Der Faktor *Zeit* ist zwei- bzw. dreifach gestuft. Um den kurzfristigen Behaltenseffekt zu analysieren, werden die Ergebnisse von Vortest [VT] und 1. Nachtest [1. NT] verglichen. Zur Analyse des langfristigen Behaltenseffekts werden die Ergebnisse von Vortest, 1. Nachtest und 2. Nachtest [2. NT]⁸⁴ gegenübergestellt.

Zunächst gilt es, die Voraussetzungen zu prüfen. Die Messwerte in allen Teilstichproben sollten normalverteilt sein. Diese Voraussetzung ist erfüllt. Des Weiteren gilt es, die Varianzhomogenität zu prüfen. Der Levene-Test auf Gleichheit der Varianzen zeigt Varianzhomogenität für Vortest, 1. Nachtest und 2. Nachtest. Darüber hinaus sollten nicht nur die Varianzen der Messzeitpunkte, sondern auch die Zusammenhänge zwischen den Messzeitpunkten bei mehr als zwei Zeitpunkten gleich hoch ausfallen (Kovarianzen). Dies wird anhand des Mauchly-Tests auf Sphärizität geprüft (Bühner & Ziegler, 2009, S. 519). Die Sphärizitätsannahme ist bei der ANOVA mit Messwiederholung mit drei Zeitpunkten erfüllt.

Um sich zunächst einen Überblick über die Gruppenunterschiede zu verschaffen, werden in nachfolgender Tabelle 35 die Mittelwerte und Standardabweichungen für die Gruppe *Concept Mapping* und die *Kontrollgruppe* für den Vortest, 1. und 2. Nachtest angegeben.

Tabelle 35: Mittelwerte und Standardabweichungen für die Gruppe CM und die KG für den Vortest, 1. Nachtest und 2. Nachtest

		Mittelwerte	Standardabweichungen
Vortest	KG	4,68	2,30
	CM	5,82	2,50
1. Nachtest	KG	11,54	4,57
	CM	14,31	4,46
2. Nachtest	KG	11,08	2,62
	CM	10,99	4,69

Hierbei zeigt sich im 1. Nachtest ein geringer Vorteil der Gruppe *CM* gegenüber der *KG*, welcher jedoch im 2. Nachtest nicht mehr zu erkennen ist.

⁸⁴ Die Aufgaben 4b und 10b waren lediglich Teil des 2. Nachtests. Um den Wissenszuwachs analysieren zu können, werden diese Aufgaben daher für die ANOVAs mit Messwiederholung nicht mit einbezogen.

Nachfolgende Tabelle 36 stellt die Ergebnisse der zwei ANOVAs mit Messwiederholung dar.

Tabelle 36: Ergebnisse der ANOVAs mit Messwiederholung für die Lernwirksamkeit des Concept Mapping

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)</u>							
Zeit _{VT-1.NT}	2043,619	1	2043,619	248,785	0,000*	0,754	3,50*
Zeit _{VT-1.NT} X Gruppe	12,341	1	12,341	1,502	0,224	0,018	0,27
Zeit _{VT-1.NT-2.NT}	2117,934	2	1058,969	153,613	0,000*	0,681	2,92*
Zeit _{VT-1.NT-2.NT} X Gruppe	67,978	2	33,989	4,930	0,008*	0,064	0,52*

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt zeigt sich ein signifikanter Haupteffekt für den Faktor *Zeit* bei beiden Analysen mit großen Effekten von $d=3,50^{85}$ bzw. $d=2,92$. Demnach konnten die Versuchsgruppen einen signifikanten Wissenszuwachs vom Vortest zum 1. Nachttest ($F=248,785$; $p=0,000$) sowie über drei Zeitpunkte (VT-1. NT-2. NT; $F=153,613$; $p=0,000$) erzielen. Insgesamt kann kein signifikanter Interaktionseffekt für die Faktoren *Zeit_{VT-1.NT}* und *Gruppe* ($F=1,502$; $p=0,224$) nachgewiesen werden. Dementsprechend unterscheiden sich die Gruppen im Wissenszuwachs vom Vortest zum 1. Nachttest nicht signifikant voneinander. Hierbei wird ein lediglich kleiner Effekt von $d=0,27$ erzielt. Über die drei Zeitpunkte hinweg zeigt sich jedoch ein signifikanter Interaktionseffekt von *Zeit_{VT-1.NT-2.NT}* und *Gruppe* ($F=67,978$; $p=0,008$). Die *KG* ist der Gruppe *CM* insgesamt signifikant überlegen, mit einem mittleren Effekt von $d=0,52$. Dies ist auch aus der nachstehenden Abbildung 15 ablesbar. Der Wissenszuwachs vom Vortest zum 1. Nachttest ist bei der Gruppe *CM* mit einem Wert von 8,49 gegenüber der *KG* mit einem Wert von 6,86 zunächst höher (vgl. hierzu auch Anhang 5.1.1-1). Werden jetzt 1. Nachttest und 2. Nachttest verglichen, so zeigt sich jedoch bei der *KG* eine deutlich geringere Wissensabnahme gegenüber der Gruppe *CM* (-0,46 vs. -3,32). Der langfristige Behaltenseffekt scheint bei der *KG* somit deutlich höher zu sein.

⁸⁵ Die Berechnung der Effektstärke d erfolgte durch Umrechnung des Wertes 0,018 für das partielle Eta-Quadrat (η^2) (Cohen, 1988). Für alle nachfolgenden Angaben der Effektstärken wird ebenfalls die Umrechnung des partiellen Eta-Quadrats herangezogen. Zur Einschätzung der Höhe der Effekte wird im Folgenden die Systematik von Cohen (1988, S. 24 ff.) verwendet, nach der Effekte ab $d=0,2$ als klein, ab $d=0,5$ als mittelgroß und ab $d=0,8$ als groß angesehen werden.

Dieses Ergebnis muss jedoch kritisch hinterfragt werden. Die Durchführung des 2. Nachtests wurde bei der *KG* durch den Klassenlehrer betreut, wohingegen die anderen Versuchspersonen durch den Forscher selbst betreut wurden. Im Nachhinein bleibt somit fraglich, inwieweit der Klassenlehrer das Ergebnis verzerrt hat, indem er die Schüler z. B. inhaltlich auf den Test vorbereitet hat oder während der Bearbeitung Unterlagen verwendet werden durften. Das kann an dieser Stelle nicht abschließend beurteilt werden. Aus diesem Grund wird das Ergebnis über die drei Messzeitpunkte hinweg nicht weiter interpretiert. Hinsichtlich der Veränderungen vom Vortest zum 1. Nachtest zeigen sich zwar keine signifikanten Gruppenunterschiede zwischen der *KG* und der Gruppe *CM*. Dennoch ist von einem geringen Vorteil der Gruppe *CM* gegenüber der *KG* auszugehen ($d=0,27$).

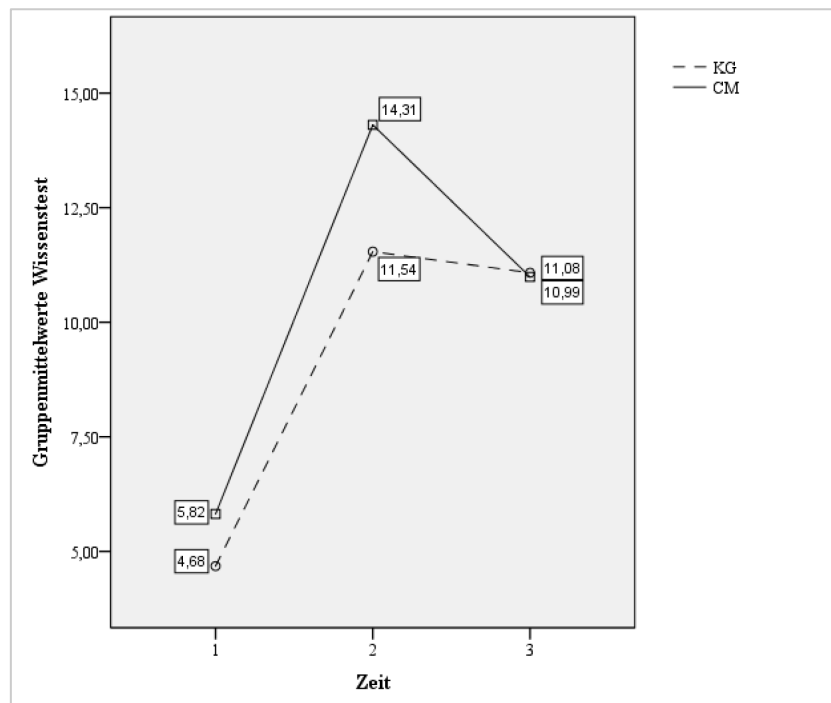


Abbildung 15: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die *KG* und die Gruppe *CM* über drei Messzeitpunkte

5.1.1.2 Gruppenunterschiede hinsichtlich des Wissenserwerbs

Um zu analysieren, inwieweit sich die drei Gruppen *kein FB*, *EiFB* und *EIFB* hinsichtlich des Wissenserwerbs unterscheiden, werden zunächst zwei einfaktorielle ANOVAs mit dem Faktor *Gruppe* sowie den AVs *1. Nachtest* bzw. *2. Nachtest* durchgeführt. Die Variable *Gruppe* ist dreifach gestuft (*kein FB*, *EiFB* und *EIFB*). Vor der Durchführung der Analysen werden die Voraussetzungen überprüft. Zunächst sollten die Messwerte in allen Teilstichproben eine Normalverteilung zeigen. Des Weiteren gilt es, die Varianzhomogenität zu prüfen. Beide Voraussetzungen sind gegeben.

Tabelle 37: Ergebnisse der einfaktoriellen ANOVAs mit der unabhängigen Variable [UV]
Gruppe und den AVs *1. Nachtest* bzw. *2. Nachtest*

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über alle Aufgaben)</u>							
Gruppe	108,262**	2	54,131	2,733	0,074	0,093	0,64
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über alle Aufgaben)</u>							
Gruppe	44,604***	2	22,302	0,844	0,436	0,035	0,38

* Korrigiertes R-Quadrat = 0,028

** Korrigiertes R-Quadrat = 0,059

*** Korrigiertes R-Quadrat = -0,007

Die Ergebnisse der Analysen in Tabelle 37 zeigen keine signifikanten Unterschiede zwischen den Gruppen *EIFB*, *EiFB* und *kein FB* hinsichtlich der Ergebnisse im 1. NT ($F=2,733$; $p=0,074$) und 2. NT ($F=0,844$; $p=0,436$). Hierbei zeigen sich mittlere Effekte für den 1. Nachtest mit $d=0,64$ sowie geringe Effekte für den 2. Nachtest mit $d=0,38$.

Um später Aussagen hinsichtlich der Hypothese H1 und H2 treffen zu können und somit den Einfluss der Art des Feedbacks auf den Wissenserwerb im 1. und 2. Nachtest zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs durchgeführt. Hierbei werden die beiden FB-Gruppen (*EiFB* und *EIFB*) zusammengefasst und der Gruppe *kein FB* gegenübergestellt. Darüber hinaus werden die Gruppen *EiFB* und *kein FB*, *EIFB* und *kein FB* sowie *EiFB* und *EIFB* jeweils paarweise verglichen. Dieses Vorgehen wurde bewusst gewählt, um nicht nur zu prüfen, ob Feedback allgemein lernwirksam ist, sondern auch zu analysieren, welches Feedback ggf. gegenüber keinem Feedback lernwirksamer ist. Nachfolgende Tabelle 38 zeigt die Effektstärken des 1. und 2. Nachtests für die Gruppenvergleiche, welche aus dem jeweiligen η^2 berechnet wurden (vgl. Anhang 5.1.1.-2).

Tabelle 38: Effektstärken für den Haupteffekt *Gruppe* der einfaktoriellen ANOVAs mit der UV *Gruppe* und der AV *1. Nachtest* bzw. *2. Nachtest*

	1. Nachtest	2. Nachtest
FB vs. kein FB	d=0,35 ($\eta^2=0,029$)	d=0,18 ($\eta^2=0,008$)
EiFB vs. kein FB	d=0,06 ($\eta^2=0,001$)	d=0,00 ($\eta^2=0,000$)
EIFB vs. kein FB	d=0,69* ($\eta^2=0,106$)	d=0,37 ($\eta^2=0,033$)
EIFB vs. EiFB	d=0,63 ($\eta^2=0,091$)	d=0,39 ($\eta^2=0,037$)

* signifikant auf dem Niveau $\alpha = 0,05$

Die Gruppe *FB* ist der Gruppe *kein FB* im 1. Nachtest ($d=0,35$) und 2. Nachtest ($d=0,18$) geringfügig überlegen. Dabei zeigen sich bei der Gruppe *EiFB* sowohl im 1. Nachtest ($d=0,06$) und 2. Nachtest ($d=0,00$) keine Effekte gegenüber der Gruppe *kein FB*. Dagegen ist ein mittlerer signifikanter Effekt der Gruppe *EIFB* gegenüber der Gruppe *kein FB* beim 1. Nachtest ($d=0,69$) und ein kleiner Effekt beim 2. Nachtest ($d=0,37$) erkennbar. Das *EIFB* ist dem *EiFB* mit einem mittleren Effekt beim 1. Nachtest ($d=0,63$) und einem kleinen Effekt beim 2. Nachtest ($d=0,39$) überlegen.

Die Gruppenmittelwerte sowie die Standardabweichungen des Vortests, 1. und 2. Nachtests sind in nachfolgender Tabelle 39 dargestellt.

Hierbei zeigt sich insbesondere für die Gruppe *EIFB* ein Vorteil im Vortest. Um diese Gruppenunterschiede hinsichtlich des Vortests zu kontrollieren, werden für die nachfolgenden Analysen die Wissenszuwächse vom Vor- zum 1. Nachtest und vom Vor- über den 1. zum 2. Nachtest herangezogen.

Tabelle 39: Mittelwerte und Standardabweichungen für die Gruppen *kein FB*, *EiFB* und *EIFB* für den Vortest, 1. und 2. Nachtest

		Mittelwerte	Standardabweichungen
Vortest	kein FB	4,97	2,39
	EiFB	5,27	2,54
	EIFB	7,03	2,19
1. Nachtest	kein FB	12,78	4,48
	EiFB	13,80	3,56
	EIFB	16,08	4,73
2. Nachtest	kein FB	10,50	4,25
	EiFB	10,07	4,42
	EIFB	12,19	5,24

Zur Analyse der Gruppenunterschiede hinsichtlich der Wissenszuwächse werden zwei zweifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit* und *Gruppe* sowie der AV *Gruppenmittelwerte Wissenstest* durchgeführt. Die Variable *Gruppe* ist dabei dreifach gestuft (*kein FB*, *EiFB* und *EIFB*). Der Faktor *Zeit* ist zwei- bzw. dreifach gestuft. Um den kurzfristigen Behaltenseffekt zu analysieren, werden die Gruppenmittelwerte von Vor- und 1. Nachtest verglichen. Zur Analyse des langfristigen Behaltenseffekts werden die Gruppenmittelwerte von Vor-, 1. und 2. Nachtest gegenübergestellt.

Vor der Durchführung der Analysen werden die Voraussetzungen überprüft. Normalverteilung, Varianzhomogenität sowie die Sphärizität sind gegeben. Nachfolgende Tabelle 40 fasst die Ergebnisse der beiden ANOVAs mit Messwiederholung zusammen.

Tabelle 40: Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs Zeit und Gruppe und der AV Gruppenmittelwerte Wissenstest

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partiel- les Eta- Quadrat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)</u>							
Zeit _{VT-1,NT}	1802,929	1	1802,929	236,170	0,000*	0,817	4,23*
Zeit _{VT-1,NT} x Gruppe	16,254	2	8,127	1,065	0,352	0,039	0,40
Zeit _{VT-1,NT-2,NT}	1774,727	2	887,364	123,884	0,000*	0,729	3,28*
Zeit _{VT-1,NT-2,NT} x Gruppe	14,307	4	3,577	0,499	0,736	0,021	0,29

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt zeigt sich ein signifikanter Haupteffekt für den Faktor *Zeit* bei beiden Analysen. Demnach konnten die Versuchsgruppen einen signifikanten Wissenszuwachs vom Vor- zum 1. Nachtest ($F=236,170$; $p=0,000$) sowie über drei Zeitpunkte (Vortest-1. Nachtest-2. Nachtest; $F=123,884$; $p=0,000$) erzielen. Die Effekte sind dabei mit $d=4,23$ und $d=3,28$ sehr groß. Jedoch kann kein signifikanter Interaktionseffekt für die Faktoren *Zeit_{VT-1,NT}* und *Gruppe* ($F=1,065$; $p=0,352$) sowie *Zeit_{VT-1,NT-2,NT} x Gruppe* ($F=0,499$; $p=0,736$) nachgewiesen werden. Hierbei zeigen sich nur geringe Effekte von $d=0,40$ und $d=0,29$. Zur genaueren Analyse der Gruppenunterschiede wird nachfolgende Abbildung 16 herangezogen.

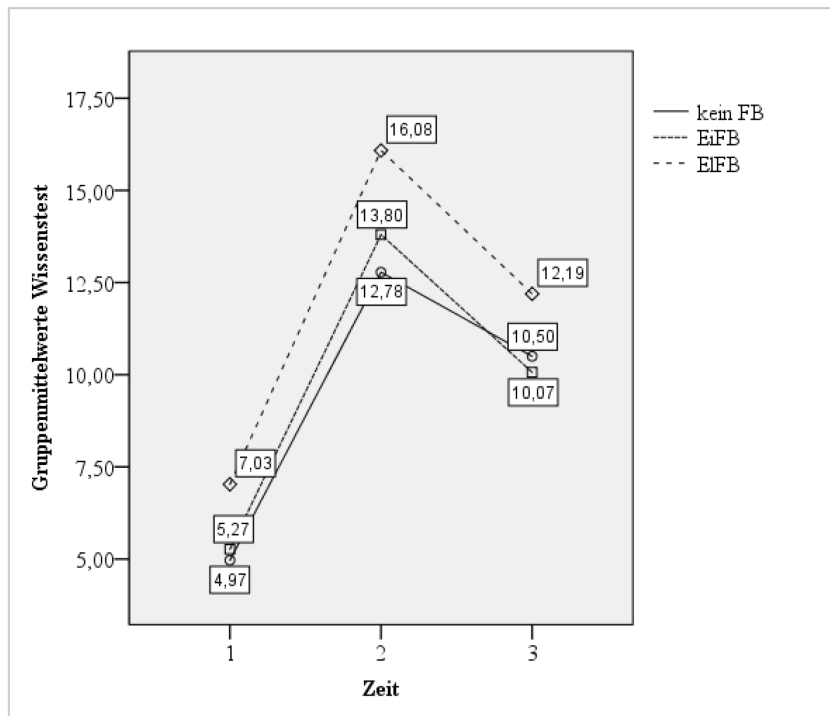


Abbildung 16: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen kein FB, EiFB und EIFB über drei Messzeitpunkte

Die Abbildung 16 zeigt einen ähnlichen Verlauf für alle drei Gruppen. Die Gruppe *EIFB* erzielt mit einem Wert von 9,06 den höchsten Wissenszuwachs zwischen Vor- und 1. Nachtest im Vergleich zur Gruppe *EiFB* (8,53) und *kein FB* (7,81). Vom 1. zum 2. Nachtest, zeigt sich jedoch bei der Gruppe *kein FB* mit einem Wert von -2,28 eine geringere Wissensabnahme gegenüber den Gruppen *EiFB* (-3,73) und *EIFB* (-3,89). Der langfristige Behaltenseffekt scheint bei der Gruppe *kein FB* somit am wenigsten zu sinken (vgl. hierzu auch Anhang 5.1.1-3).

Um später Aussagen hinsichtlich der Hypothese H1 und H2 treffen zu können und somit den Einfluss der Art des Feedbacks auf den Wissenszuwachs (VT-1. NT und VT-1. NT-2. NT) zu überprüfen, werden zweifaktorielle ANOVAs mit Messwiederholung durchgeführt. Die Ergebnisse der zweifaktorielle ANOVAs mit Messwiederholung für die paarweisen Vergleiche sind in Anhang 5.1.1-4 dargestellt. Nachfolgende Tabelle 41 zeigt die Effektstärken für die Gruppenvergleiche, welche aus dem jeweiligen η^2 berechnet wurden.

Tabelle 41: Effektstärken für den Interaktionseffekt *Zeit x Gruppe* der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Zeit* und *Gruppe* und der AV *Gruppenmittelwerte Wissenstest* über zwei bzw. drei Messzeitpunkte

	Vortest-1. Nachttest	Vortest-1. Nachttest-2. Nachttest
FB vs. kein FB	d=0,26 ($\eta^2=0,017$)	d=-0,29 ($\eta^2=0,020$)
EiFB vs. kein FB	d=0,09 ($\eta^2=0,002$)	d=-0,29 ($\eta^2=0,021$)
EIFB vs. kein FB	d=0,46 ($\eta^2=0,051$)	d=-0,33 ($\eta^2=0,026$)
EIFB vs. EiFB	d=0,35 ($\eta^2=0,029$)	d=-0,09 ($\eta^2=0,002$)

Kurzfristig scheint *FB* mit einer kleinen Effektstärke von $d=0,26$ etwas lernwirksamer als *kein FB* zu sein. Werden *EiFB* und *EiFB* jeweils separat mit *keinem FB* verglichen, so zeigt sich, dass *EiFB* mit einer mittleren Effektstärke von $d=0,46$ *keinem FB* überlegen ist. Das *EiFB* und *kein FB* unterscheiden sich dagegen nur sehr gering ($d=0,09$). Das *EiFB* ist dem *EiFB* mit einer kleinen Effektstärke von $d=0,35$ überlegen. Langfristig scheint *FB* nicht lernwirksamer zu sein sondern eher kontraproduktiv zu wirken, was die kleine negative Effektstärke von $d=-0,29$ verdeutlicht. Weder das *EiFB* ($d=-0,29$) noch das *EiFB* ($d=-0,33$) sind gegenüber *keinem FB* lernwirksamer. Das *EiFB* ist dem *EiFB* langfristig nicht mehr überlegen ($d=-0,09$). Dabei ist jedoch zu betonen, dass die Gruppenunterschiede insgesamt nicht signifikant sind.

5.1.1.3 Einfluss von personalen und weiteren Faktoren auf den Wissenserwerb

5.1.1.3.1 Einfluss von Vorwissen, Sprachfähigkeit, der Einschätzung zum Umgang mit Concept Mapping und zum Umgang mit Fehlern auf den Wissenserwerb

Nachfolgend wird der Einfluss des Vorwissens, der Sprachfähigkeit, der Einschätzung zum Umgang mit Fehlern und zum Umgang mit Concept Mapping auf den 1. und 2. Nachttest überprüft. Hierzu werden zwei einfaktorielle Kovarianzanalysen [ANCOVA] mit dem Faktor *Gruppe* und der AV *1. Nachttest* bzw. *2. Nachttest* durchgeführt.⁸⁶ Die

⁸⁶ Da die relevanten Faktoren nur einmalig erhoben wurden, ist ein Einbezug dieser als Kovariaten in eine ANOVA mit Messwiederholung nicht sinnvoll. Das Herausparsialisieren einer einmalig erhobenen Kontrollvariablen wirkt sich in diesem Fall lediglich auf den Gruppierungsfaktor nicht aber auf den Interakti-

Kovariaten sind dabei *Vorwissen [VW]*, *Sprachfähigkeit [SF]*, *eingeschätzte Interessantheit und Nützlichkeit [IN]* und *eingeschätzte Handhabbarkeit [H]* hinsichtlich des *Concept Mapping*⁸⁷ sowie *Einschätzung zum Umgang mit Fehlern [UF]*. Die Variable *Gruppe* ist dreifach gestuft (*kein FB*, *EiFB* und *ElFB*). Zur Prüfung der Voraussetzungen kann auf Kapitel 5.1.1.2 verwiesen werden.

Die nachfolgende Tabelle 42 zeigt die Ergebnisse der einfaktoriellen ANCOVA für die AV 1. *Nachtest*. Die einfaktorielle ANCOVA für die AV 1. *Nachtest* zeigt einen signifikanten Einfluss der Kovariaten *VW* ($F=8,960$; $p=0,004$) und *SF* ($F=6,984$; $p=0,011$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=1,168$; $p=0,320$). Durch das Herauspartialisieren der Kovariaten verkleinert sich der Gruppenunterschied im Vergleich zu der ANCOVA ohne Kovariaten (vgl. Kapitel 5.1.1.2). Die Effektstärke verringert sich von $d=0,64$ auf $d=0,44$. Die Unterschiede zwischen den Gruppen hinsichtlich der Ergebnisse im 1. *Nachtest* sind somit nicht nur auf die Art des Feedbacks sondern auch auf weitere Faktoren zurückzuführen. Neben dem *VW* ($d=0,86$) hat hierbei die *SF* ($d=0,76$) einen großen Effekt.

Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von $0,059$ ⁸⁸ auf $0,342$, d. h. 34,2 Prozent der Varianz der AV 1. *Nachtest* können durch das Modell erklärt werden.⁸⁹

onseffekt Zeit x Gruppe aus (Bortz & Schuster, 2016, S. 321). Aus diesem Grund werden univariate ANCOVAs mit der Variable 1. *Nachtest* und 2. *Nachtest* als AV durchgeführt. Der verzerrende Einfluss des Vortests wird dabei herauspartialisiert, indem das Vorwissen als Kovariate einbezogen wird.

⁸⁷ Die beiden Faktoren H und IN wurden im Rahmen der Auswertung des Fragebogens zur Einschätzung des Umgangs mit dem *Concept Mapping* extrahiert (vgl. Kapitel 4.6.1).

⁸⁸ vgl. Kapitel 5.1.1.2; ANOVA mit der UV Gruppe und der AV 1. *Nachtest*

⁸⁹ Als Maß für die Güte des Modells wird das korrigierte R^2 herangezogen. Im Vergleich zu R^2 berücksichtigt dieses Maß die Freiheitsgrade und verhindert somit, dass schon allein durch die Aufnahme weiterer UVs bzw. Kovariaten die Güte des Modells steigt (Brosius, 2011, S. 565 ff.).

Tabelle 42: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, der AV *1. Nachtest* sowie den Kovariaten *VW*, *SF*, *H*, *IN* und *UF*

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quad- rat	d
<u>Abh. Variable 1. Nachtest (über alle Aufgaben)**</u>							
VW	124,070	1	124,070	8,960	0,004*	0,157	0,86*
SF	96,710	1	96,710	6,984	0,011*	0,127	0,76*
H	30,848	1	30,848	2,228	0,142	0,044	0,43
IN	21,093	1	21,093	1,523	0,223	0,031	0,36
UF	6,914	1	6,914	0,510	0,499	0,010	0,20
Gruppe	32,342	2	16,171	1,168	0,320	0,046	0,44

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,342

Die einfaktorielle ANCOVA für die AV 2. Nachtest zeigt einen signifikanten Einfluss der Kovariaten *VW* ($F=12,736$; $p=0,00$) und *SF* ($F=5,114$; $p=0,029$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=0,169$; $p=0,845$). Das Neutralisieren der Kovariaten führt zu einer Verringerung der Gruppenunterschiede verglichen zu der ANCOVA ohne Kovariaten (vgl. Kapitel 5.1.1.2) mit einer ebenso kleineren Effektstärke von $d=0,18$ (vorher $d=0,38$). Die Gruppenunterschiede hinsichtlich des 2. Nachtests werden folglich nicht nur durch die Art des Feedbacks, sondern auch durch die Faktoren *VW* und *SF*, mit jeweils großen Effektstärken von $d=1,14$ und $d=0,71$, hervorgerufen (vgl. Tabelle 43).

Die Aufnahme der Kovariaten verbessert die Modellgüte von $-0,007$ ⁹⁰ auf $0,303$, d. h. 30,3 Prozent der Varianz der AV 2. Nachtest lassen sich durch das Modell erklären.⁹¹

⁹⁰ vgl. Kapitel 5.1.1.2; ANOVA mit der UV Gruppe und der AV 2. Nachtest

⁹¹ Als Maß für die Güte des Modells wird das korrigierte R^2 herangezogen.

Tabelle 43: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, der AV 2. Nachtest sowie den Kovariaten *VW*, *SF*, *H*, *IN* und *UF*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable 2. Nachtest (über alle Aufgaben)**</u>							
VW	232,777	1	232,777	12,736	0,001*	0,237	1,14*
SF	93,465	1	93,465	5,114	0,029*	0,111	0,71*
H	17,468	1	17,468	0,965	0,334	0,023	0,31
IN	1,862	1	1,862	0,102	0,751	0,002	0,09
UF	27,900	1	27,900	1,527	0,224	0,036	0,39
Gruppe	6,166	2	3,083	0,169	0,845	0,008	0,18

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,303

Nachfolgende Tabelle zeigt die Effektstärken des 1. und 2. Nachtests für die Gruppenvergleiche unter Berücksichtigung der Kovariaten *VW*, *SF*, *H*, *IN* und *UF*. Die Ergebnisse der Analysen sind in Anhang 5.1.1.-5 dargestellt.

Tabelle 44: Effektstärken für die einfaktoriellen ANCOVAs mit der UV *Gruppe* und der AV 1. *Nachtest* bzw. 2. *Nachtest* sowie den Kovariaten *VW*, *SF*, *H*, *IN* und *UF*

	1. Nachtest	2. Nachtest
FB vs. kein FB	d=0,39 ($\eta^2=0,036$)	d=-0,06 ($\eta^2=0,001$)
EiFB vs. kein FB	d=0,37 ($\eta^2=0,033$)	d=0,00 ($\eta^2=0,000$)
ElFB vs. kein FB	d=0,52 ($\eta^2=0,063$)	d=-0,13 ($\eta^2=0,004$)
ElFB vs. EiFB	d=0,13 ($\eta^2=0,004$)	d=-0,64 ($\eta^2=0,094$)

Bezogen auf den 1. Nachtest bleibt *FB* gegenüber *keinem FB* im Vergleich zu der Analyse ohne Kovariaten etwas lernwirksamer ($d=0,35$ vs. $d=0,39$), wobei sowohl *EiFB* als auch *ElFB* etwas lernwirksamer sind als *kein FB*. Der Effekt des *EiFB* vergrößert sich dabei deutlich ($d=0,06$ vs. $d=0,37$). Der Effekt der Gruppe *ElFB* verringert sich gegenüber der Gruppe *kein FB* (von $d=0,69$ auf $d=0,52$) und ist nicht länger signifikant. Darüber hinaus zeigen sich verringerte Effekte des *ElFB* gegenüber dem *EiFB* ($d=0,63$ vs. $d=0,13$).

Im 2. Nachtest verschwinden die Effekte bzw. kehren sich um. Das *FB* ist *keinem FB* nicht mehr überlegen ($0,18$ vs. $-0,06$). Dabei zeigen sich weiterhin keine Effekte der

Gruppe *EiFB* gegenüber *keinem FB* ($d=0,00$ unverändert). Darüber hinaus ist *EiFB* gegenüber *keinem FB* nicht länger lernwirksam ($d=0,37$ vs. $d=-0,13$). Insgesamt ist nun die Gruppe *EiFB* der Gruppe *EiFB* mit einem mittleren Effekt überlegen ($d=0,39$ vs. $d=-0,64$). Nachfolgende Abbildung 17 verdeutlicht den Einfluss der Kovariaten auf die Ergebnisse des 2. Nachtests, indem die Mittelwerte mit den adjustierten Mittelwerten verglichen sowie die Standardabweichungen dargestellt werden.

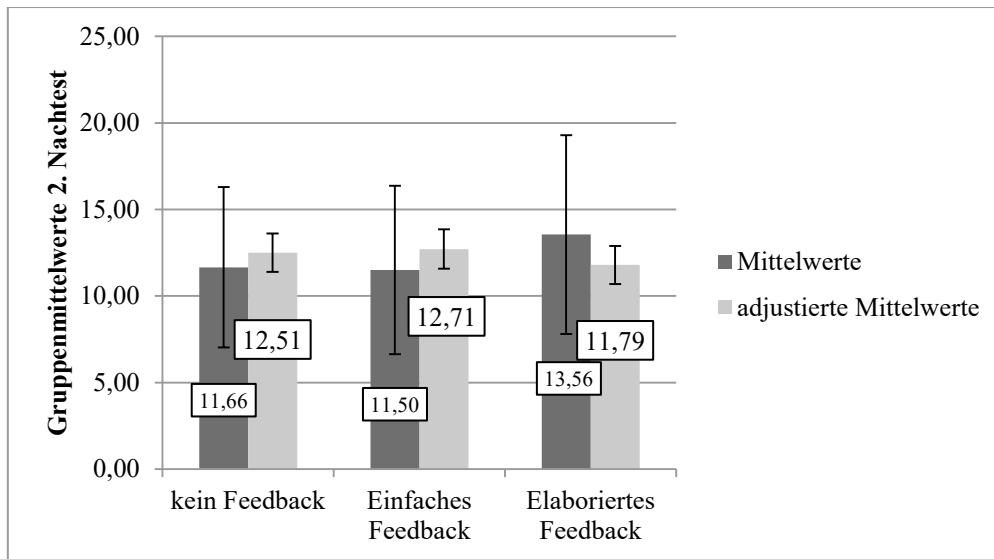


Abbildung 17: Graphische Darstellung der (adjustierten) Mittelwerte Standardabweichungen für den 2. Nachtest

Das Herauspartialisieren der Kovariaten, insbesondere der *SF* und des *VW*, führt folglich zu veränderten Gruppenunterschieden vor allem im 2. Nachtest. Um den Einfluss dieser Variablen genauer zu untersuchen, werden diese für weitere Analysen am Median dichotomisiert und jeweils zwei Gruppen, niedrig und hoch, gebildet. Hierbei wird der üblichen Verfahrensweise zur Dichotomisierung von Probanden gefolgt (Brünken & Leutner, 2005). Aufgrund dessen, dass sich bei den Variablen *H*, *IN* und *UF* kein signifikanter Einfluss auf die Ergebnisse des 1. und 2. Nachtests zeigte, werden hierzu keine weiteren Analysen durchgeführt.

5.1.1.3.2 Einfluss von niedrigem bzw. hohem Vorwissen auf den Wissenserwerb

Zur Analyse des Einflusses des Vorwissens werden zunächst zwei zweifaktorielle ANOVAs mit den UVs *Gruppe* und *niedriges bzw. hohes Vorwissen* [*VW_nh*⁹²] sowie der AV *1. bzw. 2. Nachtest* durchgeführt. Des Weiteren folgen zwei dreifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit*, *Gruppe* und *VW_nh* sowie der AV *Gruppenmittelwerte Wissenstest*. Der Faktor *Zeit* ist zwei- (VT-1. NT) bzw. dreifach (VT-1. NT-2. NT) gestuft. Die Voraussetzungen der Normalverteilung, der Varianzhomogenität sowie der Sphärizität sind gegeben.

Die Ergebnisse der zweifaktoriellen ANOVAs in Tabelle 45 zeigen keinen signifikanten Haupteffekt für den Faktor *Gruppe* hinsichtlich der Ergebnisse im 1. Nachtest ($F=1,367$; $p=0,264$) und 2. Nachtest ($F=0,318$; $p=0,730$). Hierbei zeigen sich mittlere Effekte für den 1. Nachtest mit $d=0,47$ sowie geringe Effekte für den 2. Nachtest mit $d=0,25$. Der Haupteffekt *VW_nh* ist dagegen sowohl für den 1. Nachtest ($F=150,537$; $p=0,005$) als auch für den 2. Nachtest ($F=174,265$; $p=0,007$) signifikant. Hierbei zeigen sich jeweils große Effektstärken von $d=0,83$ bzw. $d=0,86$. Der Interaktionseffekt zwischen *Gruppe* und *VW_nh* bleibt dagegen sowohl für den 1. Nachtest ($F=13,966$; $p=0,675$) als auch für den 2. Nachtest ($F=34,463$; $p=0,200$) nicht signifikant, wobei für den 1. Nachtest ein geringer Effekt von $d=0,26$ und für den 2. Nachtest ein mittlerer Effekt von $d=0,56$ erzielt wird.

Tabelle 45: Ergebnisse der zweifaktoriellen ANOVAs mit der UV *Gruppe* und *VW_nh* sowie der AV *1. Nachtest* bzw. *2. Nachtest*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über alle Aufgaben)</u>							
Gruppe	48,261**	2	24,13	1,367	0,264	0,052	0,47
VW_nh	150,537	1	150,537	8,527	0,005*	0,146	0,83
Gruppe x VW_nh	13,966	2	6,983	0,396	0,675	0,016	0,26
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über alle Aufgaben)</u>							
Gruppe	13,848***	2	6,924	0,318	0,730	0,015	0,25
VW_nh	174,265	1	174,265	7,993	0,007*	0,157	0,86
Gruppe x VW_nh	72,926	2	34,463	1,672	0,200	0,072	0,56

** korrigiertes R-Quadrat: 0,162

*** korrigiertes R-Quadrat: 0,169

⁹² *Niedriges VW* beinhaltet dabei alle Probanden mit einem $VW \leq 6,00$ (Median) und *hohes VW* umfasst alle Probanden mit einem $VW > 6,00$.

In der Abbildung 18 und Abbildung 19 sind die Ergebnisse graphisch dargestellt.

Um Aussagen hinsichtlich der Hypothese H3 bis H6 treffen zu können und somit den Einfluss des Vorwissens auf den Wissenserwerb des 1. und 2. Nachtests zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs jeweils separat für *niedriges* und *hohes VW* durchgeführt.

Die Ergebnisse der ANOVAs für die paarweisen Vergleiche sowie die Effektstärken können den Anhängen 5.1.1-6 und 5.1.1-7 entnommen werden. Insgesamt zeigt sich der signifikante Einfluss des VW deutlich. Probanden mit *niedrigem VW* erzielen sowohl im 1. Nachtest als auch im 2. Nachtest auch niedrige Ergebnisse und Probanden mit *hohem VW* erzielen hohe Ergebnisse. Im 1. Nachtest unterscheidet sich bei Probanden mit *niedrigem VW* das *FB* und *kein FB* ($d=0,09$) nicht. Dabei ist die Gruppe *EiFB* mit einem kleinen Effekt ($d=0,38$) und einem höheren Wert für den 1. Nachtest (13,89 Punkte) der Gruppe *kein FB* geringfügig überlegen. Das *EiFB* (12,19 Punkte) ist der Gruppe *kein FB* dagegen geringfügig unterlegen ($d=-0,14$; 11,62 Punkte). Das *EiFB* ist dem *EiFB* mit einem mittleren Effekt ($d=0,57$) überlegen.

Bei Probanden mit *hohem VW* ist das *FB* der Gruppe *kein FB* mit einem mittleren Effekt ($d=0,63$) überlegen, wobei sowohl das *EiFB* als auch das *EiFB* mit $d=0,74$ einen mittleren Effekt gegenüber *keinem FB* erzielen. Die Gruppe *kein FB* erreicht mit 14,40 Punkten den niedrigsten Wert für den 1. Nachtest. Das *EiFB* ist dem *EiFB* mit 17,36 Punkten (gegenüber 16,60 Punkten) geringfügig überlegen, wobei sich jedoch fast kein Effekt zeigt ($d=0,18$).

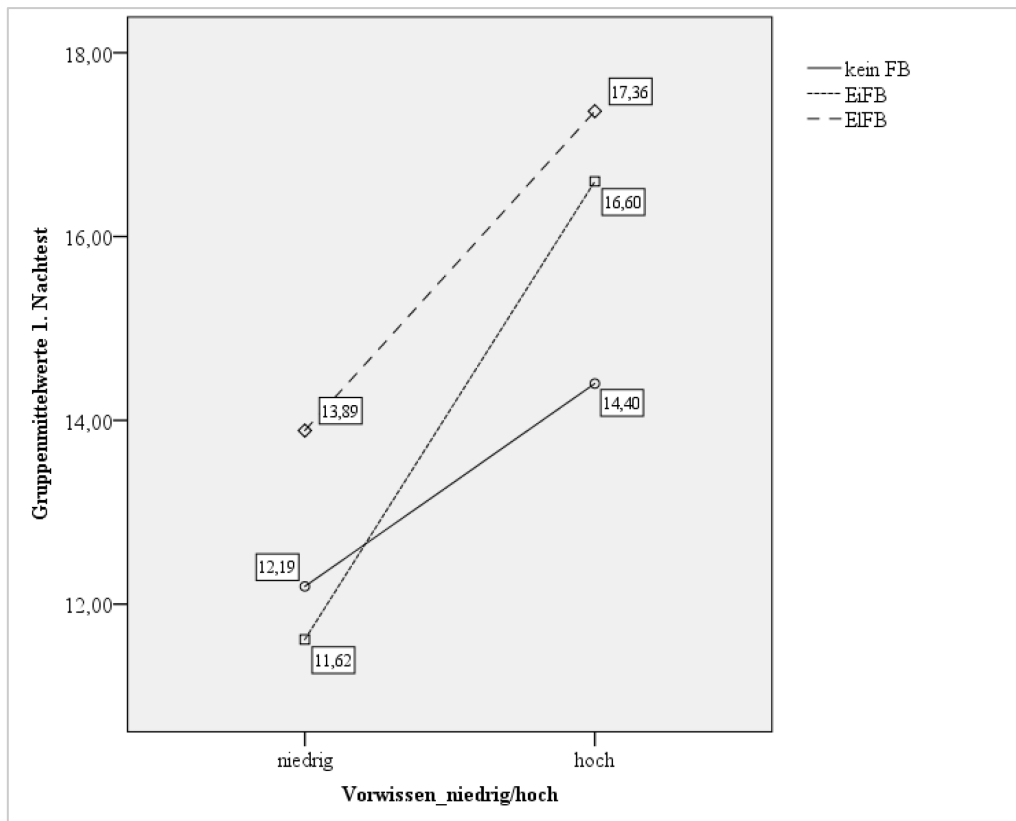


Abbildung 18: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit vom Vorwissensniveau

Im 2. Nachttest ist das *FB* bei Probanden mit *niedrigem VW* der Gruppe *kein FB* mit einem kleinen Effekt von $d=-0,36$ unterlegen, wobei sowohl das *EiFB* ($d=-0,50$) als auch das *EIFB* ($d=-0,17$) mit 9,36 Punkten und 10,71 Punkten gegenüber *keinem FB* mit 11,56 Punkten schlechtere Ergebnisse erzielen.

Bei Probanden mit *hohem VW* der Gruppe *kein FB* bleiben die Ergebnisse gegenüber niedrigem *VW* auf einem relativ konstanten Niveau (12,08 Punkte). Die Probanden der Feedbackgruppen erzielen deutlich bessere Ergebnisse (17,31 Punkte bzw. 15,36 Punkte) als die Gruppe *kein FB*. Das *FB* ist *keinem FB* mit einem mittleren Effekt von $d=0,66$ überlegen, wobei sich beim *EiFB* ein großer signifikanter Effekt ($d=3,23$), beim *EIFB* nur ein mittlerer Effekt zeigt. Das *EIFB* ist dem *EiFB* mit einem kleinen Effekt von $d=0,38$ unterlegen.

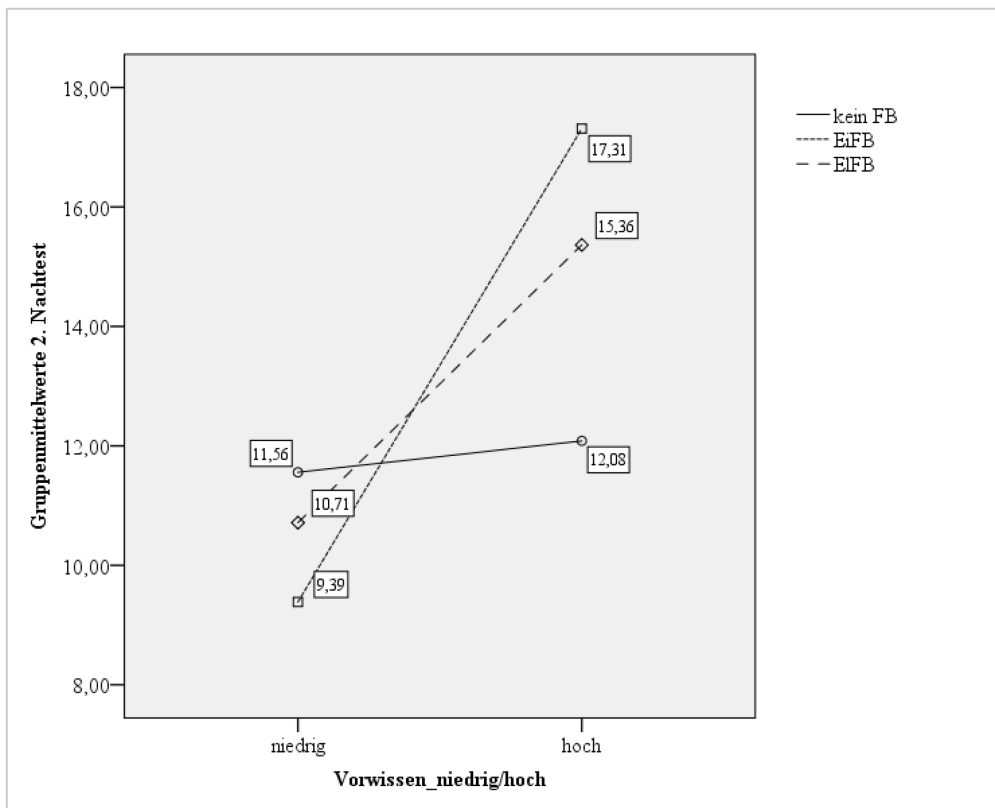


Abbildung 19: Graphische Darstellung der Gruppenmittelwerte des 2. Nachttests in Abhängigkeit vom Vorwissen

Um den Einfluss des VW auf den Wissenszuwachs zu analysieren, werden dreifaktorielle ANOVAs mit Messwiederholung durchgeführt (vgl. Tabelle 46).

Tabelle 46: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Gruppe* und *VW_nh* sowie der AV *Gruppenmittelwerte Wissenstest*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)</u>							
Zeit _{VT-1.NT}	1464,099	1	1464,099	186,386	0,000*	0,788	3,86*
Zeit _{VT-1.NT} x Gruppe	23,465	2	11,732	1,494	0,234	0,056	0,49
Zeit _{VT-1.NT} x VW_nh	5,706	1	5,706	0,726	0,398	0,014	0,24
Zeit _{VT-1.NT} x Gruppe x VW_nh	6,905	2	3,453	0,440	0,647	0,017	0,26
Zeit _{VT-1.NT-2.NT}	1264,957	2	632,478	87,454	0,000*	0,670	2,85*
Zeit _{VT-1.NT-2.NT} x Gruppe	10,668	4	2,667	0,369	0,830	0,017	0,26
Zeit _{VT-1.NT-2.NT} x VW_nh	3,411	2	1,706	0,236	0,790	0,005	0,14
Zeit _{VT-1.NT-2.NT} x Gruppe x VW_nh	33,258	4	8,315	1,150	0,339	0,051	0,46

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt können keine signifikanten Interaktionseffekte für die Faktoren $Zeit_{VT-1,NT}$ und VW_nh ($F=5,071$; $p=0,234$) sowie $Zeit_{VT-1,NT-2,NT}$ und VW_nh ($F=3,411$; $p=0,790$) nachgewiesen werden. Hierbei zeigen sich nur geringe Effekte von $d=0,24$ und $d=0,14$. Darüber hinaus sind die Interaktionseffekte zwischen den Faktoren $Zeit_{VT-1,NT}$, VW_nh und $Gruppe$ ($F=0,440$; $p=0,647$) sowie $Zeit_{VT-1,NT-2,NT}$, VW_nh und $Gruppe$ ($F=1,150$; $p=0,339$) nicht signifikant. Die Effekte sind mit $d=0,26$ bzw. $d=0,46$ gering bzw. mittel. Vorwissensstärkere Schüler unterscheiden sich nicht signifikant von den vorwissensschwächeren Schülern hinsichtlich des Wissenszuwachses, auch unabhängig davon, ob und welches Feedback sie erhielten. Diese Ergebnisse lassen sich aus der Abbildung 21 ablesen. Die Veränderungen zwischen Vor- und 1. Nachtest sowie zwischen 1. und 2. Nachtest verlaufen jeweils fast parallel.

Um Aussagen hinsichtlich der Hypothese H3 bis H6 treffen zu können und somit den Einfluss des Vorwissens auf den Wissenszuwachs (VT-1. NT; VT-1. NT-2. NT) zu überprüfen, wurden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung jeweils separat für *niedriges* und *hohes VW* durchgeführt. Die Ergebnisse der ANOVAs mit Messwiederholung für die paarweisen Vergleiche jeweils separat für *niedriges* und *hohes VW* sowie die Effektstärken können den Anhängen 5.1.1-8 und 5.1.1-9 entnommen werden.

Vom Vortest zum 1. Nachtest unterscheiden sich die Probanden mit *niedrigem VW* der Gruppen *FB* und *kein FB* nicht ($d=0,06$). Dabei zeigt sich beim *ElFB* gegenüber *keinem FB* ein geringer Effekt ($d=0,31$), beim *EiFB* gegenüber *keinem FB* ist dagegen kein Effekt erkennbar ($d=0,09$). Das *ElFB* ist dem *EiFB* mit einem geringen Effekt von $d=0,38$ geringfügig überlegen. Bei Probanden mit *hohem VW* ist das *FB* der Gruppe *kein FB* mit einem mittleren Effekt überlegen, wobei sich beim *EiFB* ein mittlerer Effekt von $d=0,64$ und beim *ElFB* ein hoher Effekt von $d=0,88$ gegenüber *keinem FB* zeigen. Das *ElFB* ist dem *EiFB* mit einem kleinen Effekt von $d=0,31$ überlegen.

Tendenziell lässt sich aus der Abbildung 20 und dem Anhang 5.1.1-10 ableiten, dass die Probanden der Gruppe *ElFB* den höchsten Wissenszuwachs vom Vor- zum 1. Nachtest erzielen, unabhängig davon, ob sie *niedriges* (9,21 Punkte) oder *hohes VW* (8,95 Punkte) haben (im Vergleich zu 8,03 und 6,83 bei der Gruppe *kein FB*; 8,82 und 7,75 bei der Gruppe *EiFB*).

Langfristig ist das *FB* bei Probanden mit *niedrigem VW* der Gruppe *kein FB* mit einem mittleren Effekt von $d=-0,54$ unterlegen, wobei sich sowohl beim *EiFB* ($d=-0,59$) als auch beim *ElFB* ($d=-0,50$) gegenüber *keinem FB* negative mittlere Effekte zeigen. Das *ElFB* ist dem *EiFB* nur geringfügig überlegen ($d=0,11$). Probanden mit *hohem VW* profitieren vom *FB* gegenüber *keinem FB* ($d=0,48$). Hierbei ist insbesondere das *EiFB* gegenüber *keinem FB* mit einem hohen Effekt von $d=1,06$ überlegen. Das *ElFB* zeigt gegenüber *keinem FB* einen mittleren Effekt ($d=0,50$). Das *ElFB* ist dem *EiFB* mit einem kleinen Effekt von $d=-0,39$ unterlegen.

Probanden mit *niedrigem VW* der Gruppe *kein FB* können das Wissensniveau im 2. Nachtest im Vergleich zum 1. Nachtest relativ konstant halten (geringe Wissensabnahme von -1,81 Punkten), das Wissen nimmt im Vergleich zu der Gruppe *EiFB* (-8,64 Punkte) und *ElFB* (-9,29 Punkte) über drei Zeitpunkte am geringsten ab (-5,96 Punkte). Werden Vortest, 1. und 2. Nachtest verglichen, so ist erkennbar, dass bei Probanden mit *hohem VW* das Wissen der Gruppe *EiFB* mit -10,00 Punkten gegenüber *keinem FB* (-12,83 Punkte) und *ElFB* (-11,95 Punkte) am geringsten abnimmt (vgl. Abbildung 20 und Anhang 5.1.1-10). Insgesamt zeigt sich, dass sich das Wissensniveau der Probanden mit niedrigem VW der Gruppe *kein FB* an das Niveau der Probanden mit hohem VW annähert. Bei den *FB*-Gruppen spiegelt ein niedriges Vorwissensniveau auch ein niedriges Wissensniveau im 2. Nachtest wider und umgekehrt.

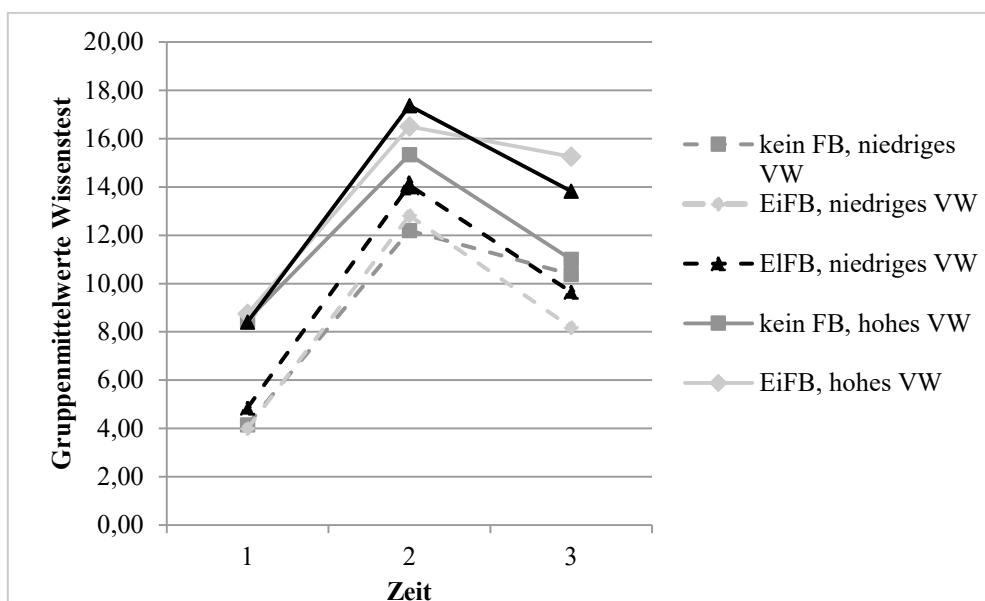


Abbildung 20: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *ElFB* über drei Messzeitpunkte in Abhängigkeit vom *VW*-Niveau

5.1.1.3.3 Einfluss von niedriger bzw. hoher Sprachfähigkeit auf den Wissenserwerb

Zur Analyse der Sprachfähigkeit werden zunächst zwei zweifaktorielle ANOVAs mit den UVs *Gruppe* und *niedrige bzw. hohe SF* [SF_nh^{93}] sowie der AV *1. bzw. 2. Nachtest* durchgeführt. Anschließend folgen zwei dreifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit*, *Gruppe* und *SF_nh* sowie der AV *Gruppenmittelwerte Wissenstest*. Der Faktor *Zeit* ist zwei- (VT-1.NT) bzw. dreifach (VT-1.NT-2.NT) gestuft. Die Voraussetzungen der Normalverteilung, der Varianzhomogenität sowie der Sphärizität sind gegeben.

Die Ergebnisse der zweifaktoriellen ANOVAs in Tabelle 47 zeigen einen signifikanten Haupteffekt für den Faktor *Gruppe* in Bezug auf die Ergebnisse im 1. Nachtest ($F=3,645$; $p=0,033$). Die Gruppen unterscheiden sich folglich signifikant hinsichtlich der Ergebnisse des 1. Nachtests. Hierbei zeigt sich ein großer Effekt von $d=0,76$. Der Haupteffekt für den Faktor *Gruppe* hinsichtlich der Ergebnisse des 2. Nachtests ist dagegen nicht signifikant ($F=0,431$; $p=0,653$). Der Effekt ist mit $d=0,29$ eher klein. Der Haupteffekt *SF_nh* ist sowohl für den 1. Nachtest ($F=122,801$; $p=0,005$) als auch für den 2. Nachtest ($F=129,508$; $p=0,027$) signifikant. Hierbei zeigt sich ein großer Effekt mit $d=0,88$ bzw. mittlerer Effekt mit $d=0,70$. Der Interaktionseffekt zwischen *Gruppe* und *SF_nh* bleibt dagegen sowohl für den 1. Nachtest ($F=48,610$; $p=0,246$) als auch für den 2. Nachtest ($F=12,769$; $p=0,775$) nicht signifikant, wobei für den 1. Nachtest ein mittlerer Effekt von $d=0,48$ und für den 2. Nachtest ein kleiner Effekt von $d=0,22$ erzielt wird.

Tabelle 47: Ergebnisse der zweifaktoriellen ANOVAs mit der UV *Gruppe* und *SF_nh* sowie der AV *1. Nachtest* sowie *2. Nachtest*

Quelle	Quadrat-summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über alle Aufgaben)</u>							
Gruppe	122,801**	2	61,401	3,645	0,033*	0,127	0,76*
SF_nh	162,664	1	162,664	9,656	0,003*	0,162	0,88*
Gruppe x SF_nh	48,610	2	24,305	1,443	0,246	0,055	0,48
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über alle Aufgaben)</u>							
Gruppe	21,413***	2	10,706	0,431	0,653	0,020	0,29
SF_nh	129,508	1	129,508	5,212	0,027*	0,108	0,70*
Gruppe x SF_nh	12,769	2	6,385	0,257	0,775	0,012	0,22

** korrigiertes R-Quadrat: 0,200

*** korrigiertes R-Quadrat: 0,053

⁹³ *Niedrige SF* beinhaltet dabei alle Probanden mit einer $SF \leq 92,75$ (Median) und *hoher SF* umfasst alle Probanden mit einer $SF > 92,75$.

Interessant an dieser Stelle ist der signifikante Haupteffekt des Faktors *Gruppe* für die AV 1. *Nachtest*. Durch den Einbezug der *SF* als dichotomisierte Variable in das Modell werden die Gruppenunterschiede signifikant. Darüber hinaus verbessert sich die Modellgüte von 0,059 auf 0,200. Aus diesem Grund werden zusätzlich die Effektstärken für die jeweiligen Gruppenvergleiche mit angegeben. Nachfolgende Tabelle 48 zeigt einen signifikanten mittleren Effekt der Gruppe *kein FB* gegenüber der Gruppe *FB* ($d=0,57$) sowie einen großen signifikanten Effekt der Gruppe *EiFB* gegenüber der Gruppe *kein FB* ($d=0,94$). Im Vergleich zur univariaten ANOVA ohne Einbezug weiterer Variablen wird folglich der Gruppenunterschied zwischen *FB* und *keinem FB* signifikant (vgl. Kapitel 5.1.1.2 sowie Anhang 5.1.1-11).⁹⁴

Tabelle 48: Effektstärken für den Haupteffekt *Gruppe* der zweifaktoriellen ANOVA mit den UVs *Gruppe* und *SF_nh* sowie der AV 1. *Nachtest*

	1. Nachtest
FB vs. kein FB	d=0,57* ($\eta^2=0,076$)
EiFB vs. kein FB	d=0,35 ($\eta^2=0,030$)
EiFB vs. kein FB	d=0,94* ($\eta^2=0,181$)
EiFB vs. EiFB	d=0,52 ($\eta^2=0,063$)

* signifikant auf dem Niveau $\alpha = 0,05$

Um Aussagen hinsichtlich der Hypothese H7 bis H10 treffen zu können und somit den Einfluss der *SF* auf den Wissenserwerb des 1. und 2. Nachtests überprüfen, werden jeweils weitere einfaktorielle ANOVAs jeweils separat für *niedrige* und *hohe SF* durchgeführt. In nachfolgender Abbildung 21 und Abbildung 22 sind die Ergebnisse der zweifaktoriellen ANOVAs graphisch dargestellt. Die Ergebnisse der ANOVAs für die paarweisen Vergleiche jeweils separat für *niedrige* und *hohe SF* sowie die Effektstärken können den Anhängen 5.1.1-13 und 5.1.1-14 entnommen werden.

Aus Abbildung 21 und Abbildung 22 ist deutlich erkennbar, dass die Probanden mit *niedriger SF* sowohl im 1. Nachtest als auch im 2. Nachtest auch niedrige Ergebnisse erzielen und umgekehrt. Im 1. Nachtest ist bei Probanden mit *niedriger SF* das *FB* signifikant besser als *kein FB* ($d=0,97$), wobei der Effekt aus einer signifikanten hohen Effektstärke

⁹⁴ Die Berechnung basiert auf den adjustierten Mittelwerten des 1. Nachtests: *kein FB* 12,11; *EiFB* 13,40; *EiFB* 15,70 (vgl. Anhang 5.1.1.-12).

des *EIFB* gegenüber *keinem FB* ($d=1,91$) und einer hohen Effektstärke des *EiFB* gegenüber *keinem FB* ($d=0,79$) resultiert. Das *EIFB* zeigt gegenüber *EiFB* einen mittleren Effekt von $d=0,64$. Bei Probanden mit *hoher SF* zeigt sich ein kleiner Effekt des *FB* gegenüber *keinem FB* ($d=0,17$), wobei dies durch einen kleinen Effekt des *EIFB* gegenüber *keinem FB* ($d=0,35$) hervorgerufen wird. Das *EiFB* zeigt gegenüber *keinem FB* einen kleinen negativen Effekt von $d=-0,20$. Das *EIFB* ist dem *EiFB* mit einem mittleren Effekt von $d=0,42$ überlegen. Aus der Abbildung 21 ist ebenfalls erkennbar, dass sich der Vorteil des *EIFB* insbesondere bei Probanden mit *niedriger SF* zeigt. Das *EIFB* ist mit 14,72 Punkten der Gruppe *EiFB* mit 12,21 Punkten und der Gruppe *kein FB* mit 9,00 Punkten deutlich überlegen.

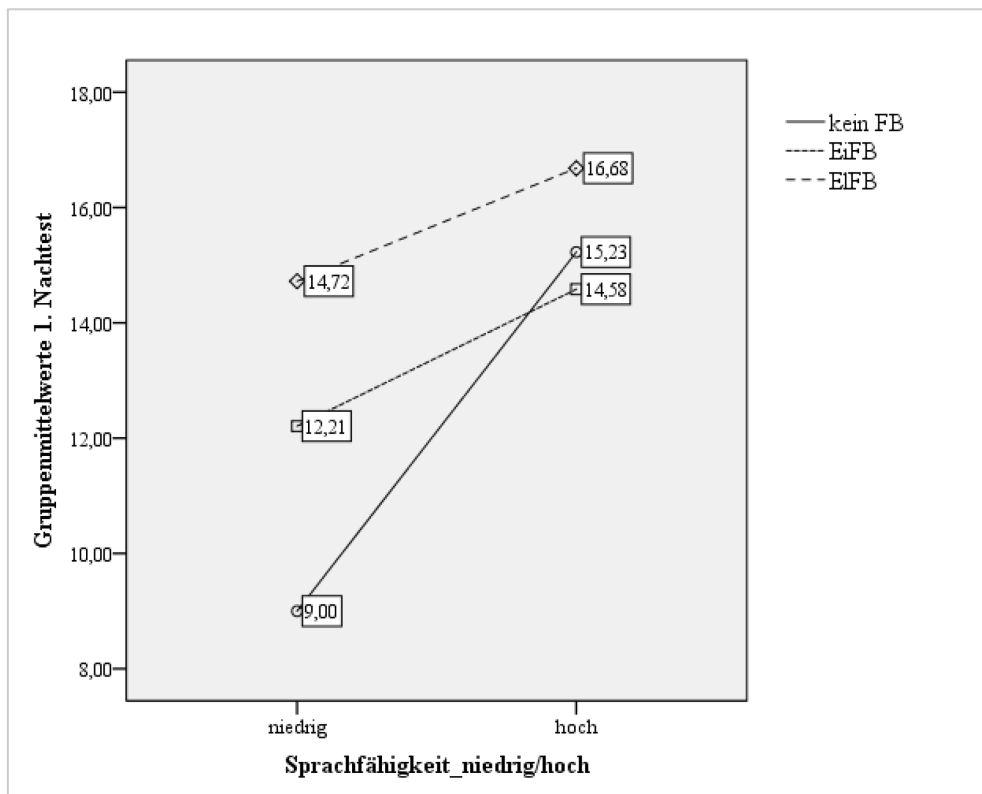


Abbildung 21: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit von der *SF*

Im 2. Nachttest zeigt sich bei Probanden mit *niedriger SF* kein Effekt des *FB* gegenüber *keinem FB* ($d=0,00$), wobei sich sowohl das *EiFB* ($d=0,00$) als auch das *EIFB* ($d=-0,06$) von der Gruppe *kein FB* nicht unterscheiden. Das *EIFB* zeigt gegenüber dem *EiFB* ebenfalls keinen Effekt ($d=-0,11$). Bei Probanden mit *hoher SF* zeigt sich ein kleiner Effekt des *FB* gegenüber *keinem FB* ($d=0,39$), wobei das *EiFB* mit einem kleinen Effekt von $d=0,18$ und das *EIFB* mit einem mittleren Effekt von $d=0,52$ der Gruppe *kein FB* überlegen ist. Das *EIFB* zeigt gegenüber dem *EiFB* einen kleinen Effekt von $d=0,38$. Im 2.

Nachtest stellt sich somit ein umgekehrtes Bild dar (vgl. Abbildung 22). Hier erzielen die Probanden mit *hoher SF* der Gruppe *ElFB* bessere Ergebnisse (15,36 Punkte) als die Probanden der Gruppe *EiFB* (13,38 Punkte) und *kein FB* (12,61 Punkte). Bei *niedriger SF* zeigen sich dagegen kaum Unterschiede (*ElFB*: 10,71 Punkte; *EiFB*: 10,25 Punkte; *kein FB*: 10,43 Punkte).

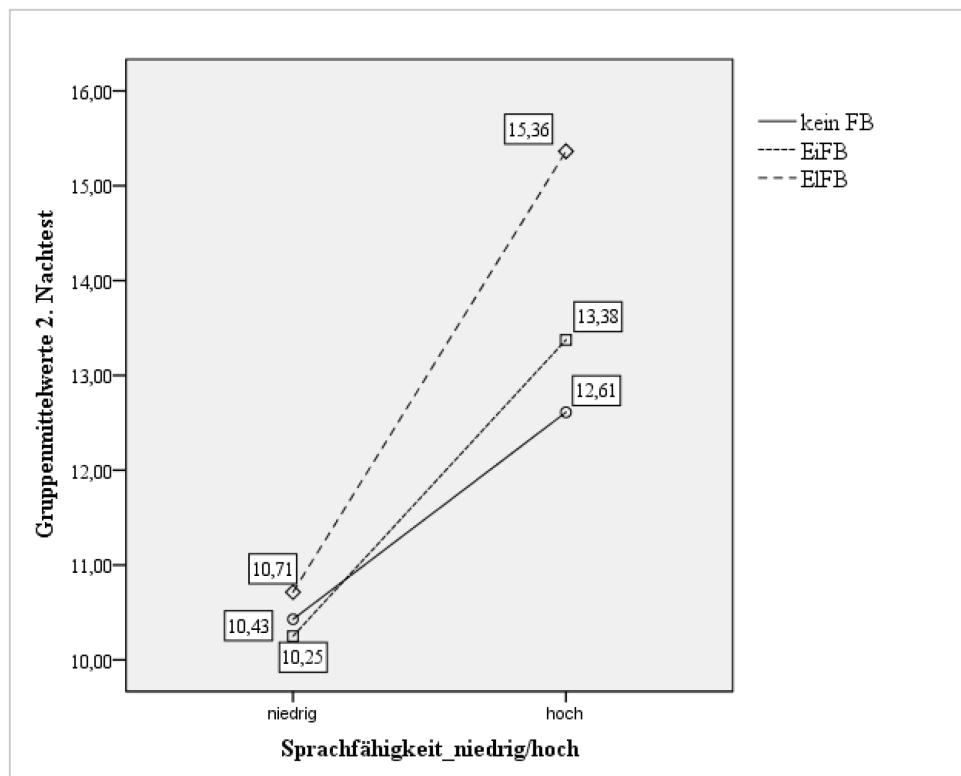


Abbildung 22: Graphische Darstellung der Gruppenmittelwerte des 2. Nachtests in Abhängigkeit von der *SF*

Um den Einfluss der *SF* auf den Wissenszuwachs zu analysieren, werden dreifaktorielle ANOVAs mit Messwiederholung durchgeführt (vgl. Tabelle 49).

Tabelle 49: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs
Gruppe und *SF_nh* sowie der AV *Gruppenmittelwerte Wissenstest*

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quad- rat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)</u>							
Zeit _{VT-1.NT}	1724,467	1	1724,467	233,775	0,000*	0,824	4,33*
Zeit _{VT-1.NT} x Gruppe	17,664	2	8,832	1,197	0,311	0,046	0,44
Zeit _{VT-1.NT} x SF_nh	32,430	1	32,430	4,396	0,041*	0,081	0,59*
Zeit _{VT-1.NT} x Gruppe x VW_nh	5,410	2	2,705	0,367	0,695	0,014	0,24
Zeit _{VT-1.NT-2.NT}	1696,539	2	848,270	131,745	0,000*	0,754	3,50*
Zeit _{VT-1.NT-2.NT} x Gruppe	20,265	4	5,066	0,787	0,537	0,035	0,38
Zeit _{VT-1.NT-2.NT} x SF_nh	49,7041	2	24,852	3,860	0,025*	0,082	0,60*
Zeit _{VT-1.NT-2.NT} x Gruppe x SF_nh	56,103	4	14,026	2,178	0,078	0,092	0,64

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt können signifikante Interaktionseffekte für die Faktoren *Zeit_{VT-1.NT}* und *SF_nh* ($F=4,396$; $p=0,041$) sowie *Zeit_{VT-1.NT-2.NT}* und *SF_nh* ($F=3,860$; $p=0,025$) nachgewiesen werden. Hierbei zeigen sich mittlere Effekte von $d=0,59$ und $d=0,60$. Die Interaktionseffekte zwischen den Faktoren *Zeit_{VT-1.NT}*, *SF_nh* und *Gruppe* ($F=0,367$; $p=0,695$) sowie *Zeit_{VT-1.NT-2.NT}*, *SF_nh* und *Gruppe* ($F=2,178$; $p=0,078$) sind dagegen nicht signifikant. Die Effekte sind mit $d=0,26$ und $d=0,64$ niedrig bzw. mittel. Schüler mit einer *niedrigen SF* unterscheiden sich folglich signifikant von den Schülern mit einer *höheren SF* hinsichtlich des Wissenszuwachses, jedoch unabhängig davon, ob und welches Feedback sie erhielten. Vom Vortest zum 1. Nachtest erzielen Probanden mit einer *höheren SF* auch einen höheren Wissenszuwachs und umgekehrt. Über drei Zeitpunkte nimmt das Wissen der Probanden mit *hoher SF* unabhängig vom Feedback geringer ab als das Wissen bei Probanden mit *niedriger SF*.

Um Aussagen hinsichtlich der Hypothese H7 bis H10 treffen zu können und somit den Einfluss der SF auf den Wissenszuwachs (VT-1. NT; VT-1. NT-2. NT) zu überprüfen, wurden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung jeweils separat für *niedrige* und *hohe SF* durchgeführt. Die Ergebnisse der ANOVAs mit Messwiederholung für die paarweisen Vergleiche jeweils separat für niedrige und hohe SF sowie die Effektstärken können den Anhängen 5.1.1-15 und 5.1.1-16 entnommen werden.

Bei Probanden mit *geringer SF* zeigt sich vom Vortest zum 1. Nachtest ein geringer Effekt von $d=0,42$ des *FB* gegenüber *keinem FB*. Hierbei ist beim *EIFB* ein großer Effekt ($d=0,94$) gegenüber *keinem FB* erkennbar. Das *EiFB* ist *keinem FB* mit einem geringen Effekt ($d=0,23$) überlegen. Beim *EIFB* zeigt sich gegenüber dem *EiFB* ein mittlerer Effekt von $d=0,57$. Bei Probanden mit *hoher SF* ist das *FB* der Gruppe *kein FB* ebenfalls mit einem geringen Effekt ($d=0,33$) überlegen, wobei sich sowohl beim *EiFB* als auch beim *EIFB* geringe Effekte gegenüber *keinem FB* zeigen. Bei der Gruppe *EIFB* ist gegenüber *EiFB* kein Effekt ($d=-0,06$) erkennbar.

Aus Abbildung 23 sowie Anhang 5.1.1-17 lässt sich ableiten, dass bei den Probanden mit *niedriger SF* die Gruppe *EIFB* auch den höchsten Wissenszuwachs vom Vor- zum 1. Nachtest erzielt (8,29 Punkte im Vergleich zu 7,67 Punkten bei der Gruppe *EiFB* und 5,86 Punkten bei der Gruppe *kein FB*). Bei *hoher SF* ist die Gruppe *EiFB* mit 9,83 Punkten der Gruppe *EIFB* mit 9,55 Punkten und der Gruppe *kein FB* mit 9,33 Punkten etwas überlegen.

Bei Probanden mit *niedriger SF* ist langfristig (VT-1. NT-2. NT) das *FB* der Gruppe *kein FB* deutlich unterlegen, was die negative Effektstärke von $d=-1,11$ verdeutlicht. Hierbei zeigen sich sowohl beim *EiFB* ($d=-1,08$) als auch beim *EIFB* ($d=-1,49$) signifikante negative Effekte. Das *EIFB* ist dabei dem *EiFB* mit einem geringen Effekt von $d=-0,20$ unterlegen. *Hohe SF* führt zu einem geringen Effekt des *FB* gegenüber *keinem FB* ($d=0,30$), wobei dies auf einen mittleren Effekt des *EiFB* von $d=0,45$ sowie einen geringen Effekt von $d=0,26$ des *EIFB* gegenüber *keinem FB* zurückzuführen ist. Das *EIFB* ist mit einem geringen negativen Effekt von $d=-0,14$ weniger lernwirksam als *EiFB*.

Werden 1. und 2. Nachtest verglichen (vgl. Anhang 5.1.1.-17 und Abbildung 24), so zeigt sich, dass niedrige *SF* bei den *FB*-Gruppen zu einer deutlichen Wissensabnahme führt (- 4,67 Punkte für *EiFB*, -5,50 Punkte bei *EIFB*), wohingegen *kein FB* das ohnehin niedrige Wissensniveau fast konstant halten kann (Wissensabnahme von -0,14 Punkten). Langfristig kann das hohe Wissensniveau der *FB*-Gruppen mit niedriger *SF* somit nicht gehalten werden (Wissensabnahme von -10,28 bzw. -12,36 Punkten). Die Gruppe *kein FB* hat mit -3,29 Punkten die geringste Wissensabnahme über drei Zeitpunkte. Bei den Probanden mit *hoher SF* nimmt das Wissen der Gruppe *EiFB* mit -7,08 Punkten (*EiFB*) gegenüber *EIFB* mit -10,00 Punkten und der Gruppe *kein FB* mit -10,33 Punkten langfristig am geringsten ab.

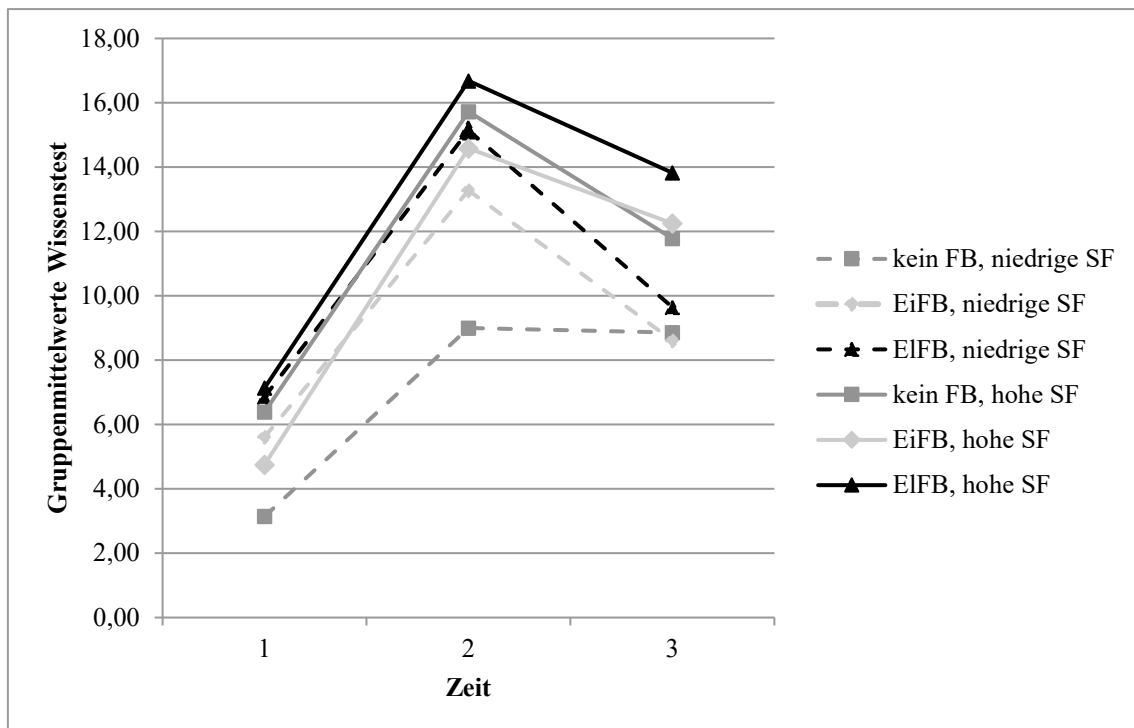


Abbildung 23: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *EIFB* über drei Messzeitpunkte in Abhängigkeit vom Niveau der *SF*

An dieser Stelle ist noch anzumerken, dass auf eine ANOVA mit gemeinsamer Betrachtung des VW und der SF als dichotomisierte Variable verzichtet wird. Eine solche Analyse hätte aufgrund zu kleiner Gruppengrößen bzw. einiger nichtbesetzter Zellen kein aussagekräftiges Bild vermittelt.

5.1.1.4 Einfluss des Lernzielniveaus auf den Wissenserwerb

5.1.1.4.1 Gruppenunterschiede hinsichtlich des Wissenserwerbs für einfache Aufgaben

In nachfolgendem Abschnitt wird zunächst der Einfluss für einfache Aufgaben auf den Wissenserwerb überprüft. Hierzu werden zwei einfaktorische ANOVAs mit der UV *Gruppe* sowie der AV *1. bzw. 2. Nachtest für einfache Aufgaben* durchgeführt. Es folgen zwei zweifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit* und *Gruppe* sowie der AV *Gruppenmittelwerte Wissenstest für einfache Aufgaben*. Der Faktor *Zeit* ist zwei- (VT-1.NT) bzw. dreifach (VT-1.NT-2.NT) gestuft. Die Voraussetzungen der Normalverteilung, der Varianzhomogenität sowie der Sphärizität sind gegeben.

Tabelle 50: Ergebnisse der einfaktoriellen ANOVAs mit der UV *Gruppe* und der AV 1. *Nachtest* bzw. *2. Nachtest* für einfache Aufgaben

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quad- rat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über einfache Aufgaben)</u>							
Gruppe	27,929*	2	13,964	2,224	0,118	0,077	0,58
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über einfache Aufgaben)</u>							
Gruppe	18,691**	2	9,346	1,467	0,241	0,060	0,51

* Korrigiertes R-Quadrat = 0,043

** Korrigiertes R-Quadrat = 0,019

Die Ergebnisse der Analysen in Tabelle 50 zeigen keine signifikanten Unterschiede zwischen den Gruppen *EIFB*, *EiFB* und *kein FB* hinsichtlich der Ergebnisse im 1. Nachtest ($F=2,224$; $p=0,118$) und 2. Nachtest ($F=2,224$; $p=0,241$) für einfache Aufgaben. Hierbei zeigen sich mittlere Effekte für den 1. Nachtest mit $d=0,58$ sowie für den 2. Nachtest mit $d=0,51$.

Um Aussagen hinsichtlich der Hypothese H23 und H24 treffen zu können und somit die Lernwirksamkeit des Feedbacks für einfache Aufgaben des 1. und 2. Nachtests zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs durchgeführt. Nachfolgende Tabelle 51 zeigt die Effektstärken für die Gruppenvergleiche hinsichtlich des 1. und 2. Nachtests für einfache Aufgaben, welche aus dem jeweiligen η^2 berechnet wurden (vgl. Anhang 5.1.1.-18).

Tabelle 51: Effektstärken für den Haupteffekt *Gruppe* der einfaktoriellen ANOVAs mit der UV *Gruppe* und der AV 1. *Nachtest* bzw. *2. Nachtest* für einfache Aufgaben

	1. Nachtest	2. Nachtest
FB vs. kein FB	d=0,38 ($\eta^2=0,034$)	d=0,09 ($\eta^2=0,002$)
EiFB vs. kein FB	d=0,14 ($\eta^2=0,005$)	d=-0,25 ($\eta^2=0,015$)
EIFB vs. kein FB	d=0,67 ($\eta^2=0,099$)	d=0,36 ($\eta^2=0,031$)
EIFB vs. EiFB	d=0,52 ($\eta^2=0,064$)	d=0,60 ($\eta^2=0,083$)

Hinsichtlich der *einfachen Aufgaben* ist die Gruppe *FB* der Gruppe *kein FB* im 1. Nachtest ($d=0,38$) und 2. Nachtest ($d=0,09$) geringfügig überlegen. Dabei zeigt sich bei der Gruppe *EiFB* im 1. Nachtest ($d=0,14$) ein sehr kleiner Effekt gegenüber der Gruppe *kein FB*. Im 2. Nachtest ist die Gruppe *kein FB* dem *EiFB* sogar geringfügig überlegen ($d=-0,25$).

Dagegen ist ein mittlerer Effekt der Gruppe *ElFB* gegenüber der Gruppe *kein FB* beim 1. Nachtest ($d=0,67$) und ein kleiner Effekt beim 2. Nachtest ($d=0,36$) erkennbar. Das *ElFB* ist dem *EiFB* mit einem mittleren Effekt beim 1. Nachtest ($d=0,52$) und 2. Nachtest ($d=0,60$) überlegen.

Zur Analyse der Lernwirksamkeit des Feedbacks auf den Wissenszuwachs einfacher Aufgaben wurden zweifaktorielle ANOVAs mit Messwiederholung durchgeführt (vgl. Tabelle 52). Es zeigen sich signifikante Haupteffekte für den Faktor *Zeit* bei beiden Analysen. Demnach konnten die Versuchsgruppen einen signifikanten Wissenszuwachs hinsichtlich der einfachen Aufgaben vom Vor- zum 1. Nachtest ($F=200,228$; $p=0,000$) sowie über drei Zeitpunkte (Vortest-1. Nachtest-2. Nachtest; $F=98,987$; $p=0,000$) erzielen. Die Effekte sind dabei mit $d=3,89$ und $d=2,94$ sehr groß. Jedoch kann kein signifikanter Interaktionseffekt für die Faktoren *Zeit*_{VT-1.NT} und *Gruppe* ($F=0,347$; $p=0,708$) sowie *Zeit*_{VT-1.NT-2.NT} und *Gruppe* ($F=0,801$; $p=0,527$) nachgewiesen werden. Hierbei zeigen sich nur geringe Effekte von $d=0,23$ und $d=0,38$.

Tabelle 52: Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Zeit* und *Gruppe* und der AV *Gruppenmittelwerte Wissenstest* für einfache Aufgaben

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über einfache Aufgaben)</u>							
<i>Zeit</i> _{VT-1.NT}	524,035	1	524,035	200,228	0,000*	0,791	3,89*
<i>Zeit</i> _{VT-1.NT} x <i>Gruppe</i>	1,816	2	0,908	0,347	0,708	0,013	0,23
<i>Zeit</i> _{VT-1.NT-2.NT}	501,536	2	250,768	98,987	0,000*	0,683	2,94*
<i>Zeit</i> _{VT-1.NT-2.NT} x <i>Gruppe</i>	8,121	4	2,030	0,801	0,527	0,034	0,38

* signifikant auf dem Niveau $\alpha = 0,05$

Zur genaueren Analyse der Gruppenunterschiede wird nachfolgende Abbildung 24 herangezogen.

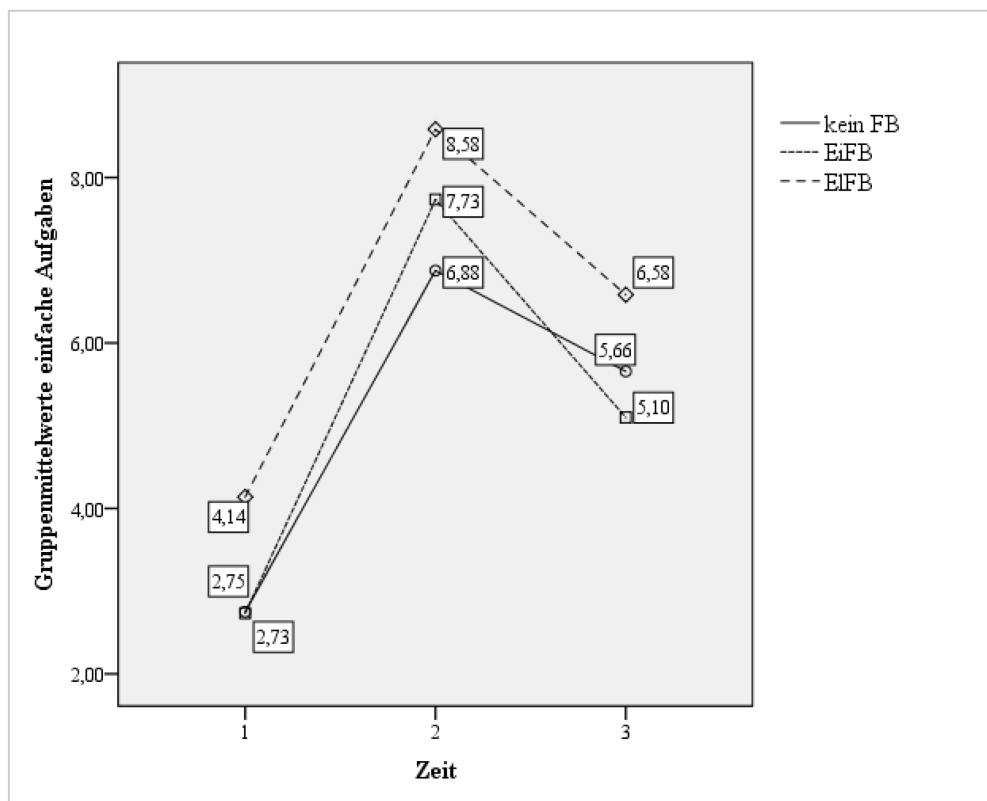


Abbildung 24: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *EIFB* über drei Messzeitpunkte für *einfache Aufgaben*

Die Abbildung 24 zeigt einen ähnlichen Verlauf für alle drei Gruppen. Die Gruppe *EiFB* erzielt mit einem Wert von 5,00 Punkten den höchsten Wissenszuwachs zwischen Vor- und 1. Nachtest im Vergleich zur Gruppe *EIFB* (4,44 Punkte) und *kein FB* (4,13 Punkte). Kurzfristig scheint somit das *EiFB* bei einfachen Aufgaben am lernwirksamsten. Vom 1. zum 2. Nachtest zeigt sich jedoch bei der Gruppe *kein FB* mit einem Wert von -1,22 Punkten eine geringere Wissensabnahme gegenüber den Gruppen *EIFB* (-2,00 Punkte) und *EiFB* (-2,63 Punkte). Langfristig kann die Gruppe *EiFB* das Wissensniveau nicht halten. Der langfristige Behaltenseffekt über drei Zeitpunkte hinsichtlich der einfachen Aufgaben scheint bei der Gruppe *kein FB* am höchsten zu sein, die Wissensabnahme ist hier mit -3,97 Punkten gegenüber -5,37 Punkten bei der Gruppe *EiFB* und -6,14 Punkten bei der Gruppe *EIFB* am geringsten (vgl. Anhang 5.1.1-19).

Um Aussagen hinsichtlich der Hypothese H23 und H24 treffen zu können und somit die Lernwirksamkeit des Feedbacks für den Wissenszuwachs in Bezug auf einfache Aufgaben zu überprüfen, werden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung durchgeführt. Nachfolgende Tabelle 53 zeigt die Effektstärken für die Gruppenvergleiche hinsichtlich des Wissenszuwachses für einfache Aufgaben, welche aus dem jeweiligen η^2 berechnet wurden (vgl. Anhang 5.1.1.-20).

Tabelle 53: Effektstärken für den Interaktionseffekt *Zeit x Gruppe* der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Zeit* und *Gruppe* und der AV *Gruppenmittelwerte Wissenstest* über zwei bzw. drei Messzeitpunkte für einfache Aufgaben

	Vortest-1. Nachtest	Vortest-1. Nachtest-2. Nachtest
FB vs. kein FB	d=0,23 ($\eta^2=0,013$)	d= -0,33 ($\eta^2=0,026$)
EiFB vs. kein FB	d=0,23 ($\eta^2=0,013$)	d=-0,49 ($\eta^2=0,057$)
EIFB vs. kein FB	d=0,26 ($\eta^2=0,016$)	d=-0,36 ($\eta^2=0,032$)
EIFB vs. EiFB	d=0,00 ($\eta^2=0,000$)	d= -0,22 ($\eta^2=0,012$)

Kurzfristig scheint *FB* mit einer kleinen Effektstärke von $d=0,23$ hinsichtlich der *einfachen Aufgaben* etwas lernwirksamer als *kein FB* zu sein. Werden *EIFB* und *EiFB* jeweils separat mit *keinem FB* verglichen, so zeigt sich, dass sowohl *EIFB* als auch *EiFB* mit einer kleinen Effektstärke von $d=0,23$ bzw. $d=0,26$ *keinem FB* überlegen ist. Das *EIFB* ist dem *EiFB* dagegen nicht überlegen ($d=0,00$). Langfristig scheint *FB* bei einfachen Aufgaben nicht lernwirksamer zu sein, sondern eher kontraproduktiv zu wirken ($d=-0,33$). Das *EiFB* ($d=-0,49$) sowie das *EIFB* ($d=-0,36$) sind gegenüber *keinem FB* mit einem geringen bzw. mittleren Effekt weniger lernwirksam. Das *EiFB* ist dem *EIFB* langfristig mit einem kleinen Effekt von $d=-0,22$ überlegen.

5.1.1.4.2 *Gruppenunterschiede hinsichtlich des Wissenserwerbs für komplexe Aufgaben*

Im nächsten Abschnitt wird der Einfluss für Lernziele mit einem hohen Anspruchsniveau auf den Wissenserwerb überprüft.

Hierzu werden zwei einfaktorielle ANOVAs mit der UV *Gruppe* sowie der AV 1. bzw. 2. Nachtest für komplexe Aufgaben durchgeführt. Es folgen zwei zweifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit* und *Gruppe* sowie der AV *Gruppenmittelwerte Wissenstest* für komplexe Aufgaben. Der Faktor *Zeit* ist zwei- (VT-1.NT) bzw. dreifach (VT-1.NT-2.NT) gestuft. Die Voraussetzungen der Normalverteilung sowie der Varianzhomogenität sind gegeben.

Tabelle 54: Ergebnisse der einfaktoriellen ANOVAs mit der UV *Gruppe* und der AV 1. *Nachtest* bzw. 2. *Nachtest* für komplexe Aufgaben

Quelle	Quadrat-summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über komplexe Aufgaben)</u>							
Gruppe	27,206*	2	13,603	2,382	0,102	0,082	0,60
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über komplexe Aufgaben)</u>							
Gruppe	8,143**	2	4,072	0,415	0,663	0,018	0,27

* Korrigiertes R-Quadrat = 0,048

** Korrigiertes R-Quadrat = -0,025

Die Ergebnisse der Analysen in Tabelle 54 zeigen keine signifikanten Unterschiede zwischen den Gruppen *ElFB*, *EiFB* und *kein FB* hinsichtlich der Ergebnisse im 1. Nachtest (F=2,382; p=0,102) und 2. Nachtest (F=0,415; p=0,663) für komplexe Aufgaben. Hierbei zeigt sich ein mittlerer Effekte für den 1. Nachtest mit d=0,60 bzw. ein geringer Effekt für den 2. Nachtest mit d=0,27.

Um Aussagen hinsichtlich der Hypothese H25 und H26 treffen zu können und somit die Lernwirksamkeit des Feedbacks bei komplexen Aufgaben des 1. und 2. Nachtests zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs durchgeführt. Nachfolgende Tabelle 55 zeigt die Effektstärken für die Gruppenvergleiche hinsichtlich des 1. und 2. Nachtests für komplexe Aufgaben, welche aus dem jeweiligen η^2 berechnet wurden (vgl. Anhang 5.1.1.-21).

Tabelle 55: Effektstärken für den Haupteffekt *Gruppe* der einfaktoriellen ANOVAs mit der UV *Gruppe* und der AV 1. *Nachtest* bzw. 2. *Nachtest* für komplexe Aufgaben

	1. Nachtest	2. Nachtest
FB vs. kein FB	d=0,26 ($\eta^2=0,016$)	d=0,22 ($\eta^2=0,012$)
EiFB vs. kein FB	d=-0,06 ($\eta^2=0,001$)	d=0,13 ($\eta^2=0,004$)
ElFB vs. kein FB	d=0,59 ($\eta^2=0,080$)	d=0,33 ($\eta^2=0,026$)
ElFB vs. EiFB	d=0,64 ($\eta^2=0,093$)	d=0,18 ($\eta^2=0,008$)

Hinsichtlich der komplexen Aufgaben ist die Gruppe *FB* der Gruppe *kein FB* im 1. Nachtest ($d=0,26$) und 2. Nachtest ($d=0,22$) geringfügig überlegen. Dabei zeigen sich beim *EiFB* keine Effekte im 1. und 2. Nachtest gegenüber der Gruppe *kein FB* ($d=-0,06$ bzw. $d=0,13$). Dagegen ist ein mittlerer Effekt der Gruppe *ElFB* gegenüber der Gruppe *kein FB* beim 1. Nachtest ($d=0,59$) und ein kleiner Effekt beim 2. Nachtest ($d=0,33$) erkennbar. Das *ElFB* ist dem *EiFB* mit einem mittleren Effekt beim 1. Nachtest ($d=0,59$) und einem kleinen Effekt im 2. Nachtest ($d=0,18$) überlegen.

Zur Analyse der Lernwirksamkeit des Feedbacks auf den Wissenszuwachs komplexer Aufgaben wurden zweifaktorielle ANOVAs mit Messwiederholung durchgeführt (vgl. Tabelle 56). Es zeigen sich signifikante Haupteffekte für den Faktor *Zeit* bei beiden Analysen. Demnach konnten die Versuchsgruppen einen signifikanten Wissenszuwachs hinsichtlich der komplexen Aufgaben vom Vor- zum 1. Nachtest ($F=147,208$; $p=0,000$) sowie über drei Zeitpunkte (Vortest-1. Nachtest-2. Nachtest; $F=73,798$; $p=0,000$) erzielen. Die Effekte sind dabei mit $d=3,33$ und $d=2,53$ sehr groß. Jedoch kann kein signifikanter Interaktionseffekt für die Faktoren *Zeit*_{VT-1.NT} und *Gruppe* ($F=1,896$; $p=0,160$) sowie *Zeit*_{VT-1.NT-2.NT} und *Gruppe* ($F=0,618$; $p=0,651$) nachgewiesen werden. Hierbei zeigen sich mittlere bzw. geringe Effekte von $d=0,54$ und $d=0,33$.

Tabelle 56: Ergebnisse der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Zeit* und *Gruppe* und der AV *Gruppenmittelwerte Wissenstest* für komplexe Aufgaben

Quelle	Quadrat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quad- rat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über komplexe Aufgaben)</u>							
Zeit _{VT-1.NT}	382,950	1	382,950	147,208	0,000*	0,735	3,33*
Zeit _{VT-1.NT} X Gruppe	9,865	2	4,933	1,896	0,160	0,067	0,54
Zeit _{VT-1.NT-2.NT}	391,481	2	195,740	73,798	0,000*	0,616	2,53*
Zeit _{VT-1.NT-2.NT} X Gruppe	6,559	4	1,640	0,618	0,651	0,026	0,33

* signifikant auf dem Niveau $\alpha = 0,05$

Die Unterschiede zwischen den Gruppen hinsichtlich des Wissenszuwachses für die komplexen Aufgaben werden anhand nachfolgender Abbildung 25 analysiert.

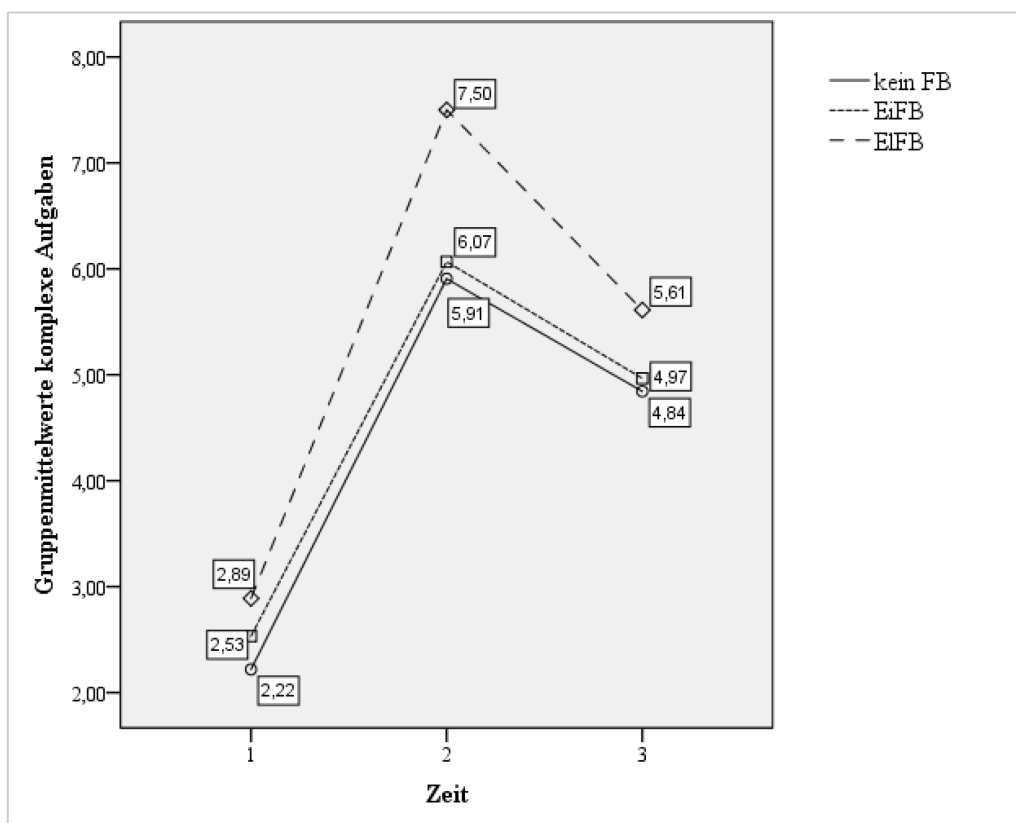


Abbildung 25: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *EIFB* über drei Messzeitpunkte für komplexe Aufgaben

Die Abbildung 25 zeigt für die Gruppe *EIFB* den höchsten Wissenszuwachs zwischen Vortest und 1. Nachttest mit 4,61 Punkten im Vergleich zur Gruppe *kein FB* (3,69 Punkte) und *EiFB* (3,53 Punkte). Kurzfristig scheint somit das *EIFB* bei den komplexen Aufgaben am lernwirksamsten. Vom 1. zum 2. Nachttest zeigt sich jedoch bei der Gruppe *kein FB*

mit einem Wert von -1,06 Punkten eine geringere Wissensabnahme gegenüber den Gruppen *EiFB* (-1,10 Punkte) und *EIFB* (-1,89 Punkte). Langfristig kann die Gruppe *EIFB* das Wissensniveau nicht halten. Der langfristige Behaltenseffekt über drei Zeitpunkte hinsichtlich der komplexen Aufgaben scheint bei der Gruppe *kein FB* am höchsten zu sein, das Wissen nimmt hier über drei Zeitpunkte mit -3,28 Punkten gegenüber -3,63 Punkten bei der Gruppe *EiFB* und -4,78 Punkten bei der Gruppe *EIFB* am geringsten ab (vgl. Anhang 5.1.1.-22).

Um Aussagen hinsichtlich der Hypothese H25 und H26 treffen zu können und somit die Lernwirksamkeit des Feedbacks auf den Wissenszuwachs in Bezug auf komplexe Aufgaben zu überprüfen, werden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung durchgeführt. In nachfolgender Tabelle 57 sind die Effektstärken für die Gruppenvergleiche hinsichtlich des Wissenszuwachses für komplexe Aufgaben, welche aus dem jeweiligen η^2 berechnet wurden, dargestellt (vgl. Anhang 5.1.1.-23).

Tabelle 57: Effektstärken für den Interaktionseffekt *Zeit x Gruppe* der zweifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Zeit* und *Gruppe* und der AV *Gruppenmittelwerte Wissenstest* über zwei bzw. drei Messzeitpunkte für *komplexe Aufgaben*

	Vortest-1. Nachtest	Vortest-1. Nachtest-2. Nachtest
FB vs. kein FB	d=0,22 ($\eta^2=0,012$)	d=-0,16 ($\eta^2=0,006$)
EiFB vs. kein FB	d=-0,06 ($\eta^2=0,001$)	d=-0,06 ($\eta^2=0,001$)
EIFB vs. kein FB	d=0,51 ($\eta^2=0,061$)	d=-0,32 ($\eta^2=0,025$)
EIFB vs. EiFB	d=0,56 ($\eta^2=0,073$)	d=-0,34 ($\eta^2=0,028$)

Kurzfristig scheint *FB* mit einer kleinen Effektstärke von $d=0,22$ hinsichtlich der komplexen Aufgaben etwas lernwirksamer als *kein FB* zu sein. Werden *EIFB* und *EiFB* jeweils separat mit *keinem FB* verglichen, so zeigt sich, dass *EIFB* mit einer mittleren Effektstärke von $d=0,51$ *keinem FB* überlegen ist. Das *EiFB* zeigt gegenüber *keinem FB* keine Effekte ($d=-0,06$). Das *EIFB* ist dem *EiFB* mit einer mittleren Effektstärke von $d=0,56$ überlegen. Langfristig scheint *kein FB* mit einem geringen Effekt von $d=-0,16$ lernwirksamer zu sein, wobei sich *EiFB* und *kein FB* nicht unterscheiden ($d=-0,06$). Das

EIFB ist der Gruppe *kein FB* mit einer geringen Effektstärke von $d=-0,32$ unterlegen. Das *EiFB* ist mit einem kleinen Effekt von $d=-0,34$ lernwirksamer als das *EIFB*.

5.1.1.5 Einfluss von personalen und weiteren Faktoren auf den Wissenserwerb einfacher bzw. komplexer Aufgaben

Nachfolgend wird der Einfluss der Variablen *VW*, *SF*, *UF*, *IN* und *H* auf einfache bzw. komplexe Aufgaben des 1. und 2. Nachtests überprüft.

Hierzu werden jeweils zwei einfaktorielle ANCOVAs mit dem Faktor *Gruppe* und der AV *1. Nachtest* bzw. *2. Nachtest* jeweils für einfache und komplexe Aufgaben durchgeführt. Die Kovariaten sind dabei *VW*, *SF*, *IN* und *H* sowie *UF*. Die Variable *Gruppe* ist dreifach gestuft (*kein FB*, *EiFB* und *EIFB*). Zur Prüfung der Voraussetzungen kann auf Kapitel 5.1.1.4 verwiesen werden. Die Ergebnisse sind im Anhang 5.1.1.-24 dargestellt. Hinsichtlich der einfachen Aufgaben zeigt sich ein signifikanter Einfluss der *SF* für den 1. Nachtest ($F=5,040$; $p=0,029$) und 2. Nachtest ($F=12,488$; $p=0,001$). Das *VW* hat lediglich einen signifikanten Einfluss auf die einfachen Aufgaben des 2. Nachtests ($F=10,920$; $p=0,002$). Bei den komplexen Aufgaben zeigt sich ein signifikanter Einfluss des *VW* auf den 1. Nachtest ($F=11,941$; $p=0,001$) und 2. Nachtest ($F=9,282$; $p=0,004$). Des Weiteren beeinflussen die *SF* ($F=5,821$; $p=0,020$) und die Kovariate *IN* ($F=4,685$; $p=0,035$) die komplexen Aufgaben des 1. Nachtests signifikant.

Um den Einfluss der Variablen *SF*, *VW* und *IN* genauer zu analysieren, werden zweifaktorielle ANOVAs bzw. dreifaktorielle ANOVAs mit Messwiederholung mit den dichotomisierten Variablen *SF_nh*, *VW_nh* sowie *niedrige bzw. hohe eingeschätzte Interessantheit und Nützlichkeit hinsichtlich des Concept Mapping [IN_nh]*⁹⁵ durchgeführt (vgl. Anhänge 5.1.1.25 bis 5.1.1.29). Dabei ist es von besonderem Interesse, inwieweit sich die Gruppen hinsichtlich der Ergebnisse im 1. bzw. 2. Nachtest sowie hinsichtlich des Wissenszuwachses für einfache bzw. komplexe Aufgaben in Abhängigkeit von ihrem *VW*, ihrer *SF* bzw. der *IN* unterscheiden. Aus diesem Grund liegt der Fokus insbesondere auf der Analyse der Interaktionseffekte *Gruppe* und *VW_nh*, *SF_nh* bzw. *IN_nh* der ANOVAs bzw. *Zeit*, *Gruppe* und *VW_nh*, *SF_nh* bzw. *IN_nh* der ANOVAs mit Messwiederholung. Nachfolgend werden lediglich signifikante Interaktionseffekte näher untersucht.

⁹⁵ *Niedriges IN* beinhaltet dabei alle Probanden mit einem Wert für $IN \leq 2,8$ (Median) und *hohes IN* umfasst alle Probanden mit einem Wert für $IN > 2,8$.

Insgesamt ist dabei lediglich bei der dreifaktoriellen ANOVA mit Messwiederholung mit den UVs Zeit $VT-1.NT-2.NT$, Gruppe und SF_nh und der AV Gruppenmittelwerte Wissenstest für komplexe Aufgaben ein signifikanter Interaktionseffekt für Zeit $VT-1.NT-2.NT$, Gruppe und SF_nh mit einem mittleren Effekt von $d=0,75$ ($\eta^2=0,123$) erkennbar (vgl. Anhang 5.1.1.29). Schüler mit einer *niedrigen SF* unterscheiden sich folglich signifikant von den Schülern mit einer *höheren SF* hinsichtlich des Wissenszuwachses für komplexe Aufgaben, in Abhängigkeit davon, ob und welches Feedback sie erhielten.

Die Ergebnisse der ANOVAs mit Messwiederholung für die paarweisen Vergleiche jeweils separat für niedrige und hohe SF sowie die Effektstärken können den Anhängen 5.1.1-30 und 5.1.1-31 entnommen werden.

Bei den Probanden mit *niedriger SF* ist das *FB* der Gruppe *kein FB* mit einem hohen negativen Effekt von $d=-0,78$ signifikant unterlegen. Dabei ist sowohl das *EiFB* mit einem mittleren negativen Effekt von $d=-0,69$ als auch das *ElFB* mit einem hohen signifikanten negativen Effekt von $d=-1,19$ *keinem FB* unterlegen. Das *ElFB* ist gegenüber *EiFB* mit einem mittleren negativen Effekt von $d=-0,54$ weniger lernwirksam.

Bei den Probanden mit *hoher SF* zeigt sich ein gegensätzliches Bild. Das *FB* ist *keinem FB* mit einer mittleren Effektstärke von $d=0,69$ überlegen, wobei sowohl das *EiFB* mit einer hohen signifikanten Effektstärke von $d=1,05$ als auch das *ElFB* mit einer mittleren Effektstärke von $d=0,60$ lernwirksamer als *kein FB* sind. Das *ElFB* ist mit einer negativen kleinen Effektstärke von $d=-0,23$ dem *EiFB* unterlegen.

Nachfolgende Abbildung 26 sowie Anhang 5.1.1.-32 verdeutlichen noch einmal den Einfluss der SF auf die Gruppenunterschiede. Bei *niedriger SF* nimmt das Wissen über drei Zeitpunkte der Gruppe *kein FB* weniger ab ($-0,93$ Punkte) als bei *EiFB* ($-4,72$) und *ElFB* ($-5,21$). Zwischen 1. und 2. Nachtest ist bei der Gruppe *kein FB* sogar einen Wissenszuwachs erkennbar ($0,36$ Punkte). *Hohe SF* führt dagegen bei der Gruppe *kein FB* zur höchsten Wissensabnahme über drei Zeitpunkte ($-5,11$ Punkte), wohingegen das Wissen der Gruppe *ElFB* ($-4,50$ Punkte) und der Gruppe *EiFB* ($-2,00$ Punkte) weniger abnimmt. An dieser Stelle ist insbesondere der Vorteil des *EiFB* erkennbar. Vom 1. zum 2. Nachtest kann diese Gruppe ihr Wissensniveau mit $-0,25$ Punkten fast konstant halten.

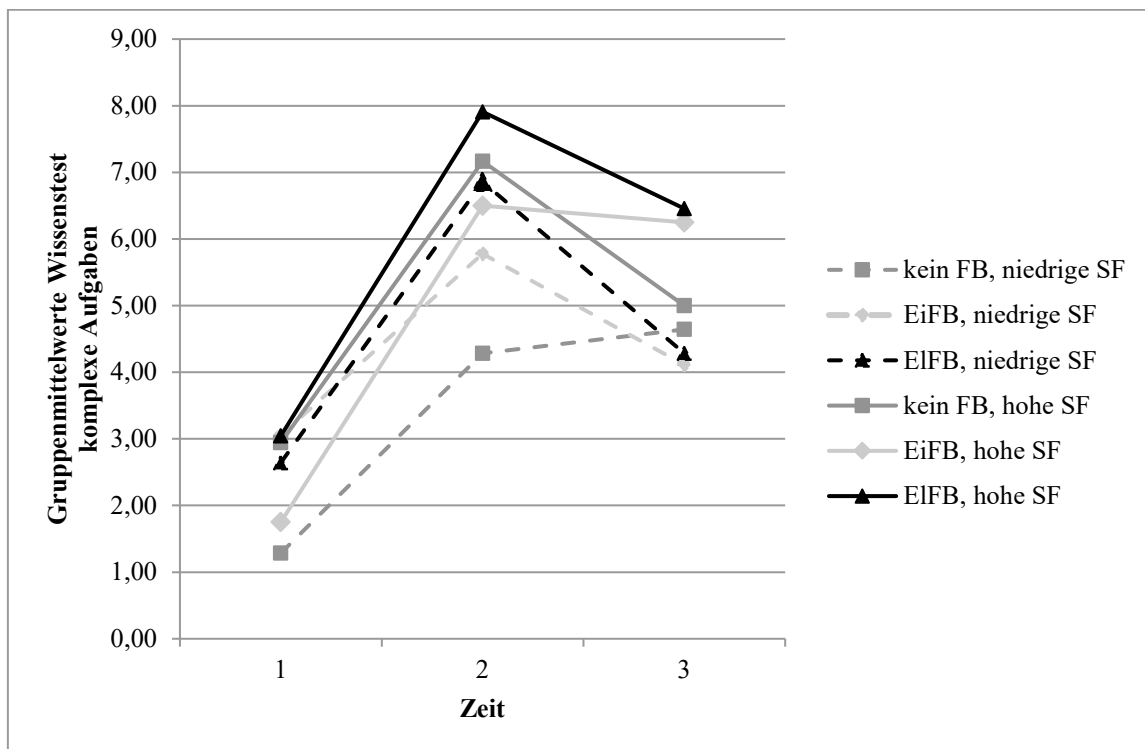


Abbildung 26: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen kein *FB*, *EiFB* und *EIFB* über drei Messzeitpunkte für komplexe Aufgaben in Abhängigkeit vom Niveau der *SF*

5.1.1.6 Zusammenfassung der Ergebnisse und Bewertung der Hypothesen hinsichtlich des Wissenserwerbs

Gruppenunterschiede hinsichtlich des Wissenserwerbs (H1 und H2)

Nachfolgend werden die Hypothesen (vgl. Kapitel 4.1) anhand der in Kapitel 5.1.1 durchgeführten Analysen hinsichtlich des Wissenserwerbs bewertet. Hierbei werden die Ergebnisse des 1. und 2. Nachttests sowie der Wissenszuwachs vom Vor- zum 1. Nachttest und über drei Zeitpunkte (Vortest-1. Nachttest-2. Nachttest) herangezogen.

Zunächst wurde die Lernwirksamkeit des CMs analysiert (vgl. Kapitel 5.1.1.1). Hierbei zeigt sich kurzfristig ein geringer Vorteil des CM gegenüber der KG ($d=0,27$). Langfristig ist das CM der KG jedoch mit einem mittleren Effekt von $d=0,52$ unterlegen. Dieses Ergebnis ist jedoch kritisch zu hinterfragen. An dieser Stelle kann nicht geklärt werden, inwieweit das Ergebnis des 2. Nachttests durch zusätzliche Hilfestellungen bzw. Hilfsmittel bei der Bearbeitung der Tests verzerrt wurde und somit insbesondere die Durchführungsobjektivität nicht gegeben ist. Das Ergebnis wird somit nicht weiter interpretiert.

Weiterhin wurde überprüft, inwieweit das zusätzliche Feedback einen Einfluss auf den Wissenserwerb hat. Hierzu gilt es nun, die Ergebnisse im Hinblick auf die Hypothesen H1, H1a und H1b sowie H2 zu bewerten.

Hierbei zeigten sich keine signifikanten Unterschiede hinsichtlich der Ergebnisse im 1. Nachtest und 2. Nachtest zwischen den drei Gruppen insgesamt. Werden die Gruppen paarweise gegenübergestellt, so ist das *EIFB* der Gruppe *kein FB* hinsichtlich der Ergebnisse des 1. Nachtests signifikant überlegen. Alle anderen Paarvergleiche zeigen keine signifikanten Unterschiede. Des Weiteren unterscheiden sich die Gruppen nicht hinsichtlich des Wissenszuwachses vom Vor- zum 1. Nachtest sowie vom Vor- über den 1. Nachtest zum 2. Nachtest. Die Hypothesen H1, H1a und H2 müssen demnach abgelehnt werden. Das *FB* ist insgesamt nicht signifikant besser als *kein FB*, wobei sich das *EiFB* und *kein FB* dabei nicht signifikant unterscheiden. Das *EIFB* ist nicht signifikant besser als *EiFB*. Hypothese 1b kann bezüglich des 1. Nachtests angenommen, muss jedoch für den 2. Nachtest sowie für die Wissenszuwächse über zwei und drei Zeitpunkte abgelehnt werden. Das *EIFB* ist für den 1. Nachtest signifikant besser als *kein FB*. Hinsichtlich der Ergebnisse im 2. Nachtest sowie der Wissenszuwächse über zwei bzw. drei Zeitpunkte unterscheiden sich das *EIFB* und *kein FB* nicht (vgl. Kapitel 5.1.1.2).

Beim 1. und 2. Nachtest lassen sich Tendenzen ableiten, die für einen kleinen Vorteil des *FB* gegenüber *keinem FB* sprechen ($d=0,35$; $d=0,18$). Dabei ist der Unterschied jedoch auf die Gruppe *EIFB* zurückzuführen. Hier zeigt sich ein mittlerer signifikanter Effekt beim 1. Nachtest ($d=0,69$) sowie ein kleiner Effekt beim 2. Nachtest ($d=0,37$). Die Gruppen *EiFB* und *kein FB* unterscheiden sich auch tendenziell nicht ($d=0,06$ bzw. $d=0,00$).

Kurzfristig, d. h. hinsichtlich des Wissenszuwachses vom Vor- zum 1. Nachtest, scheint *FB* tendenziell etwas lernwirksamer als *kein FB* zu sein ($d=0,26$), wobei der Unterschied auch hier durch eine mittlere Effektstärke von $d=0,46$ des *EIFB* gegenüber *keinem FB* zum Ausdruck kommt, das *EiFB* gegenüber *keinem FB* dagegen keine Effekte zeigt ($d=0,09$).

Langfristig (Vortest-1. Nachtest-2. Nachtest) ist dagegen *kein FB* mit einem mittleren Effekt sowohl dem *FB* ($d=-0,29$) als auch dem *EiFB* ($d=-0,29$) sowie dem *EIFB* ($d=-0,33$) überlegen.

In Bezug auf die Hypothese 2 ist das *EIFB* dem *EiFB* beim 1. Nachtest mit einem mittleren Effekt von $d=0,63$ und beim 2. Nachtest mit einem kleinen Effekt von $d=0,39$ überlegen. Hinsichtlich des Wissenszuwachses zeigt sich kurzfristig ein kleiner Effekt von $d=0,35$ des *EIFB* gegenüber *EiFB*, langfristig unterscheiden sich *EIFB* und *EiFB* nicht ($d=-0,09$).

Einfluss der individuellen Eingangsvoraussetzungen und weiteren Faktoren auf den Wissenserwerb (H3-H22)

Insgesamt zeigt sich ein signifikanter Einfluss der Faktoren *VW* und *SF* auf die Ergebnisse des 1. sowie 2. Nachtests. Die weiteren Faktoren *UF*, *IN* und *H* haben dagegen keinen signifikanten Einfluss auf die Ergebnisse des 1. und 2. Nachtests. Aus diesem Grund wird die Analyse der Hypothesen zu den Variablen *UF* (H11-H14), *IN* (H15-H18), und *H* (H19-H22) nicht weiter verfolgt.

Die paarweisen Vergleiche der Gruppen zeigen, dass sich insbesondere die Effekte des *EIFB* durch das Herauspartialisieren der Faktoren beim 1. Nachtest verringern, während sich die Effekte des *EiFB* erhöhen. Beim 2. Nachtest verschwinden die Effekte, so dass das *FB*, insbesondere das *EIFB*, gegenüber *keinem FB* auch tendenziell nicht mehr länger lernwirksam sind (vgl. Kapitel 5.1.1.3.1). Demnach müssen unter Einbezug des Einflusses der Faktoren *VW*, *SF*, *UF*, *IN* und *H* die Hypothesen 1, 1a, 1b und 2 abgelehnt werden. Für weitere Analysen wurde lediglich das *VW* und die *SF* näher betrachtet, da diese Variablen einen signifikanten Einfluss zeigen.

Um genauere Aussagen hinsichtlich der Hypothesen H3 bis H10 treffen zu können und somit den Einfluss des *VW* und der *SF* auf die Lernwirksamkeit des Feedbacks zu überprüfen, wurden die signifikanten Einflüsse des *VW* und der *SF* mittels Dichotomisierung dieser Variablen am Median (Einteilung in niedrig und hoch) genauer untersucht.

Es zeigt sich, dass Probanden mit einem *geringen VW* (*hohem VW*) auch signifikant niedrigere (höhere) Ergebnisse im 1. und 2. Nachtest erzielen, aber unabhängig davon, ob und welches Feedback sie erhielten. *VW*-stärkere Schüler unterscheiden sich nicht signifikant von *VW*-schwächeren Schülern hinsichtlich des Wissenszuwachses, auch unabhängig davon, ob und welches Feedback sie erhielten.

Bei *niedrigem VW* ist das *FB* nicht signifikant besser als *kein FB*, wobei sich sowohl das *EiFB* als auch das *EIFB* nicht signifikant von *keinem FB* unterscheiden. Das *EIFB* ist nicht signifikant besser als *EiFB*. Aus diesem Grund müssen die Hypothesen H3, H3a,

H3b und H4 abgelehnt werden. Bei *hohem VW* unterscheiden sich *FB* und *kein FB* nicht signifikant voneinander. Hierbei ist das *EIFB* nicht signifikant besser als *kein FB*. Die Hypothesen H5 sowie H5b können daher angenommen werden. Das *EiFB* ist für den 1. Nachtest sowie den Wissenszuwachs über zwei bzw. drei Zeitpunkte nicht signifikant besser als *kein FB*. Im 2. Nachtest ist das *EiFB* *keinem FB* signifikant überlegen. Die Hypothese H5a muss daher für den 2. Nachtest abgelehnt, kann für den 1. Nachtest sowie den Wissenszuwachs über zwei bzw. drei Zeitpunkte angenommen werden. Das *EIFB* ist nicht signifikant besser als *EiFB*. Die Hypothese H6 kann angenommen werden (vgl. Kapitel 5.1.1.3.2).

Im 1. Nachtest zeigt das *FB* gegenüber *keinem FB* bei Probanden mit *niedrigem VW* keinen Effekt ($d=0,09$). Bei Probanden mit *hohem VW* ist das *FB* der Gruppe *kein FB* mit einem mittleren Effekt ($d=0,63$) überlegen. Das *EIFB* ist sowohl *keinem FB* ($d=0,38$; $d=0,74$) als auch dem *EiFB* ($d=0,57$; $d=0,18$) tendenziell überlegen, unabhängig davon ob die Probanden *niedriges* oder *hohes VW* besitzen. Probanden der Gruppe *EiFB* sind der Gruppe *kein FB* mit einem mittleren Effekt lediglich überlegen, wenn diese *hohes VW* besitzen ($d=0,74$; niedriges *VW*: $d=-0,14$).

Schüler mit *geringem VW* profitieren im 2. Nachtest nicht vom Feedback und sind der Gruppe *kein FB* unterlegen (*FB* vs. *kein FB*: $d=-0,36$; *EiFB* vs. *kein FB*: $d=-0,50$; *EIFB* vs. *kein FB*: $d=-0,17$). Das *EIFB* ist dem *EiFB* dabei geringfügig überlegen ($d=0,31$). Probanden mit *hohem VW* der beiden *FB-Gruppen* erzielen tendenziell bessere Ergebnisse als die Gruppe *kein FB* ($d=0,66$), wobei sich beim *EiFB* ein großer signifikanter Effekt gegenüber *keinem FB* zeigt ($d=3,23$), der Effekt des *EIFB* gegenüber *keinem FB* mit $d=0,57$ lediglich mittel ist. Das *EIFB* ist dem *EiFB* geringfügig unterlegen ($d=-0,38$). Es lässt sich also vermuten, dass im 2. Nachtest insbesondere Probanden mit *hohem VW* vom Feedback profitieren.

Vom Vor- zum 1. Nachtest zeigt das *FB* bei Probanden mit *niedrigem VW* gegenüber *keinem FB* keinen Effekt ($d=0,06$), bei Probanden mit *hohem VW* ist es *keinem FB* mit einem mittleren Effekt ($d=0,68$) tendenziell überlegen. Das *EIFB* ist tendenziell sowohl *keinem FB* ($d=0,31$; $d=0,88$) als auch dem *EiFB* ($d=0,38$; $d=0,31$) überlegen, unabhängig vom *VW*-Stand der Schüler. Das *EiFB* scheint tendenziell lediglich bei Probanden mit *hohem VW* mit einem mittleren Effekt zu wirken ($d=0,64$; niedriges *VW*: $d=0,09$).

Langfristig scheinen die Probanden der Gruppe *EiFB* mit *hohem VW* mit einer großen Effektstärke von $d=1,06$ am ehesten vom Feedback zu profitieren (*FB* vs. *kein FB*: $d=0,48$; *ElFB* vs. *kein FB*: $d=0,50$). Bei Schülern mit *niedrigem VW* wirkt das Feedback dagegen scheinbar kontraproduktiv, da hier bei den Probanden der Gruppe *kein FB* das Wissen langfristig am geringsten abnimmt (*FB* vs. *kein FB*: $d=-0,54$; *EiFB* vs. *kein FB*: $d=-0,59$; *ElFB* vs. *kein FB*: $d=-0,50$). Das *ElFB* zeigt gegenüber dem *EiFB* bei *niedrigem VW* keinen Effekt ($d=0,11$), bei *hohem VW* ist es dem *EiFB* dagegen geringfügig unterlegen ($d=-0,39$).

Probanden mit einer *niedrigen SF* erzielen unabhängig vom Einfluss des Feedbacks signifikant niedrigere Ergebnisse in beiden Nachttests sowie einen niedrigeren Wissenszuwachs über zwei Zeitpunkte und umgekehrt. Über drei Zeitpunkte zeigt sich jedoch ein gegensätzliches Bild. Das Wissen der Probanden mit *niedriger SF* nimmt langfristig am geringsten ab und umgekehrt, unabhängig davon, ob und welches Feedback sie erhielten.

Bei Probanden mit *niedriger SF* ist das *FB* im 1. Nachttest der Gruppe *kein FB* sowie das *ElFB* der Gruppe *kein FB* signifikant überlegen. Im 2. Nachttest bzw. in Bezug auf die Wissenszuwächse über zwei bzw. drei Zeitpunkte unterscheiden sich *FB* und *kein FB* bzw. *ElFB* und *kein FB* nicht. Daher können die Hypothesen H7 und H7b für den 1. Nachttest angenommen, müssen für den 2. Nachttest sowie die Wissenszuwächse über zwei bzw. drei Zeitpunkte abgelehnt werden. Das *EiFB* und *kein FB* unterscheiden sich nicht signifikant. Die Hypothese H7a muss insgesamt abgelehnt werden. Bei *hoher SF* unterscheiden sich *FB* und *kein FB* nicht signifikant voneinander. Hierbei sind das *ElFB* sowie das *EiFB* nicht signifikant besser als *kein FB*. Die Hypothesen H9 sowie H9a und H9b können daher angenommen werden. Das *ElFB* ist nicht signifikant besser als *EiFB*. Die Hypothese H10 kann angenommen werden (vgl. Kapitel 5.1.1.3.3).

Im 1. Nachttest scheint das Feedback insbesondere bei Probanden mit *niedriger SF* zu wirken. Hier ist das *FB* mit einer hohen signifikanten Effektstärke von $d=0,97$ *keinem FB* überlegen, wobei der Effekt durch den sehr hohen signifikanten Effekt des *ElFB* gegenüber *keinem FB* zum Tragen kommt ($d=1,91$). Das *EiFB* ist *keinem FB* mit einem hohen Effekt von $d=0,79$ tendenziell überlegen. Das *ElFB* zeigt einen mittleren Effekt von $d=0,64$ gegenüber *EiFB*. Bei Probanden mit *hoher SF* ist das *FB* der Gruppe *kein FB*

lediglich mit einem kleinen Effekt von $d=0,17$ tendenziell überlegen, wobei dies aus einem kleinen Effekt des *EIFB* gegenüber *keinem FB* resultiert ($d=0,35$). Das *EiFB* ist *keinem FB* mit einer kleinen negativen Effektstärke von $d=-0,20$ tendenziell unterlegen. Das *EIFB* zeigt gegenüber dem *EiFB* einen kleinen Effekt ($d=0,42$).

Im 2. Nachtest scheinen dagegen Probanden mit *hoher SF* vom Feedback zu profitieren. Hier erzielten die Probanden der Gruppe *FB* tendenziell bessere Ergebnisse als *kein FB* ($d=0,39$), wobei dies insbesondere aus einem mittleren Effekt des *EIFB* gegenüber *keinem FB* ($d=0,52$) resultiert. Das *EiFB* zeigt lediglich einen kleinen Effekt von $d=0,18$ gegenüber *keinem FB*. Das *EIFB* ist dem *EiFB* mit einem kleinen Effekt tendenziell überlegen ($d=0,38$). Bei *niedriger SF* zeigen sich keine Unterschiede (*FB insgesamt* vs. *kein FB*: $d=0,00$; *EiFB* vs. *kein FB*: $d=0,00$; *EIFB* vs. *kein FB*: $d=-0,06$; *EIFB* vs. *EiFB*: $d=0,11$).

Hinsichtlich des Wissenszuwachses vom Vor- zum 1. Nachtest profitieren Probanden mit *niedriger SF* insbesondere vom *EIFB* ($d=0,94$), das *EiFB* zeigt gegenüber *keinem FB* lediglich einen kleinen Effekt ($d=0,23$), so dass das *FB* insgesamt der Gruppe *kein FB* mit einem kleinen Effekt von $d=0,42$ tendenziell überlegen ist. Das *EIFB* ist dem *EiFB* mit einem mittleren Effekt tendenziell überlegen ($d=0,57$). Das Feedback hat bei *hoher SF* einen geringen Einfluss (*FB* vs. *kein FB*: $d=0,33$; *EiFB* vs. *kein FB*: $d=0,42$; *EIFB* vs. *kein FB*: $d=0,31$), wobei sich *EIFB* und *EiFB* nicht unterscheiden ($d=-0,06$).

Langfristig können die Probanden mit *niedriger SF* der *FB*-Gruppen das Wissensniveau jedoch nicht halten, das Feedback scheint hier eher kontraproduktiv zu wirken. Das Wissen der Gruppe *kein FB* nimmt am geringsten ab, weshalb sich bei dem *FB* ($d=-1,11$) sowie bei dem *EiFB* ($d=-1,08$) und dem *EIFB* ($d=-1,49$) negative Effektstärken gegenüber *keinem FB* zeigen. Das *EIFB* ist dem *EiFB* dabei geringfügig unterlegen ($d=-0,20$). Bei Probanden mit *hoher SF* ist ein kleiner Vorteil des *FB* gegenüber *keinem FB* erkennbar ($d=0,39$), wobei das *EIFB* gegenüber *keinem FB* einen mittleren Effekt ($d=0,52$) und das *EiFB* gegenüber *keinem FB* lediglich einen geringen Effekt zeigt ($d=0,18$). Das *EIFB* zeigt gegenüber dem *EiFB* einen geringen Vorteil ($d=0,38$).

Für den 1. Nachtest zeigt sich zusätzlich ein signifikanter Haupteffekt für den Faktor *Gruppe* unter Berücksichtigung der *SF* als dichotomisierte Variable. Der Einfluss der *SF* scheint somit die Wirkung des Feedbacks zu reduzieren. Das Herausparsialisieren der *SF_nh* führt folglich dazu, dass sich die Gruppen signifikant voneinander unterscheiden sowie die Güte des Modells steigt. Aus diesem Grund wurden die Gruppen zusätzlich

paarweise bewertet, um Rückschlüsse auf die H1 ziehen zu können. Bei genauerer Betrachtung der Paarvergleiche lässt sich neben einem hohen signifikanten Effekt der Gruppe *EIFB* gegenüber *keinem FB* ($d=0,94$) zusätzlich ein signifikanter mittlerer Effekt des *FB* gegenüber *keinem FB* von $d=0,57$ nachweisen (vgl. Kapitel 5.1.1.3.3; Tabelle 48). An dieser Stelle lässt sich folglich feststellen, dass die Hypothese H1 für den 1. Nachtest ebenfalls angenommen werden kann, sofern der Einfluss der SF neutralisiert wird, d. h. wenn von einer gleichen SF der Gruppen ausgegangen werden kann.

Einfluss der Aufgabenanforderungen auf den Wissenserwerb (H23-H26)

Die Gruppen *kein FB*, *EiFB* und *EIFB* unterscheiden sich nicht hinsichtlich der Ergebnisse für einfache und komplexe Aufgaben des 1. und 2. Nachtests bzw. hinsichtlich des Wissenszuwachses über zwei bzw. drei Zeitpunkte. Die Paarvergleiche zeigen ebenfalls keine signifikanten Effekte. Demnach können die Hypothesen H23, H23a, H23b und H24 angenommen werden. Bei einfachen Aufgaben hat das Feedback keinen signifikanten Einfluss auf den Wissenserwerb, wobei sich *EiFB* und *EIFB* nicht signifikant unterscheiden. Die Hypothesen H25, H25a, H25b und H26 müssen abgelehnt werden. Bei komplexen Aufgaben hat das Feedback ebenfalls keinen signifikanten Einfluss auf den Wissenserwerb, wobei sich *EIFB* und *EiFB* nicht signifikant unterscheiden (vgl. Kapitel 5.1.1.4).

Es lassen sich Tendenzen ableiten, die bei einfachen Aufgaben des 1. Nachtests für einen geringen Vorteil des *FB* ($d=0,38$) sowie für einen mittleren Vorteil des *EIFB* gegenüber *keinem Feedback* ($d=0,67$) sprechen. Das *EiFB* zeigt gegenüber *keinem FB* keine Effekte ($d=0,14$). Beim 2. Nachtest unterscheiden sich *FB* und *kein FB* nicht ($d=0,09$). Dieses Ergebnis zeigt sich, da zum einen *EiFB* der Gruppe *kein FB* mit einem kleinen Effekt tendenziell unterlegen ($d=-0,25$) und zum anderen *EIFB* der Gruppe *kein FB* mit einem kleinen Effekt von $d=0,36$ tendenziell überlegen ist.

Hinsichtlich des Wissenszuwachses vom Vortest zum 1. Nachtest für einfache Aufgaben ist das *FB* insgesamt ($d=0,23$), und dabei sowohl das *EiFB* ($d=0,23$) als auch das *EIFB* ($d=0,26$), der Gruppe *kein FB* mit einem kleinen Effekt tendenziell überlegen. Langfristig scheint *FB* bei einfachen Aufgaben nicht lernwirksamer zu sein, sondern eher kontraproduktiv zu wirken ($d=-0,33$). Das *EiFB* ($d=-0,49$) sowie das *EIFB* ($d=-0,36$) sind gegenüber *keinem FB* mit einem geringen bzw. mittleren Effekt tendenziell weniger lernwirksam. Für einfache Aufgaben scheint das Feedback somit kurzfristig wirksam zu sein, langfristig dagegen nicht.

Das *EIFB* ist dem *EiFB* im 1. und 2. Nachtest bei einfachen Aufgaben mit einem jeweils mittleren Effekt ($d=0,52$; $d=0,60$) tendenziell überlegen. Im Wissenszuwachs über zwei Zeitpunkte unterscheiden sich die Gruppen nicht. Das *EiFB* ist dem *EIFB* hinsichtlich des Wissenszuwachses über drei Zeitpunkte mit einem kleinen Effekt von $d=-0,22$ sogar überlegen.

Bei den komplexen Aufgaben des 1. und 2. Nachtests ist das *FB* der Gruppe *kein FB* mit einem geringen Effekt tendenziell überlegen ($d=0,26$), wobei sich *EiFB* und *kein FB* nicht unterscheiden ($d=-0,06$ bzw. $d=0,13$), das *EIFB* gegenüber *keinem FB* dagegen einen mittleren Effekt beim 1. Nachtest ($d=0,59$) bzw. einen geringen Effekt beim 2. Nachtest mit $d=0,33$ zeigt.

Hinsichtlich des Wissenszuwachses für komplexe Aufgaben vom Vor zum 1. Nachtest scheint das *FB* mit einem kleinen Effekt etwas lernwirksamer zu sein ($d=0,22$), wobei dies auf einen mittleren Effekt des *EIFB* gegenüber *keinem FB* ($d=0,51$) zurückzuführen ist. Das *EiFB* zeigt gegenüber *keinem FB* dagegen keine Effekte ($d=-0,06$). Über drei Zeitpunkte zeigt das *FB* insgesamt, und dabei das *EiFB* bzw. das *EIFB* jeweils separat, gegenüber *keinem FB* keine Wirkung. An dieser Stelle ist das *FB* der Gruppe *kein FB* sogar geringfügig unterlegen ($d=-0,16$), wobei sich dies durch einen kleinen Effekt der Gruppe *kein FB* gegenüber *EIFB* widerspiegelt ($d=-0,32$), sich *EiFB* und *kein FB* dagegen nicht unterscheiden ($d=-0,06$).

Das *EIFB* ist im 1. Nachtest sowie beim Wissenszuwachs über zwei Zeitpunkte mit einem mittleren Effekt ($d=0,64$ bzw. $d=0,56$), für den 2. Nachtest mit einem kleinen Effekt ($d=0,18$) tendenziell überlegen. Langfristig ist das *EIFB* dagegen dem *EiFB* mit einem kleinen Effekt von ($d=-0,34$) tendenziell unterlegen.

Werden die Ergebnisse der Analysen, insbesondere die Effektstärken, für einfache und komplexe Aufgaben in Bezug auf den Einfluss des Feedbacks mit den Ergebnissen über alle Aufgaben verglichen, so zeigen sich insgesamt kaum Unterschiede.

Abschließend wurde der Einfluss der Variablen *VW*, *SF*, *UF*, *IN* und *H* auf einfache bzw. komplexe Aufgaben des 1. und 2. Nachtests überprüft. Hierbei zeigen sich signifikante Einflüsse der *SF*, des *VW* sowie der Kovariate *IN*. Analysen mit den jeweils dichotomisierten Variablen zeigen lediglich einen signifikanten Effekt der *SF* auf den Wissenszuwachs komplexer Aufgaben über drei Zeitpunkte in Abhängigkeit vom Feedback. Dabei zeigen sich bei Probanden mit *niedriger SF* negative Effekte des *FB* insgesamt ($d=-0,78$;

signifikant) sowie des *EiFB* ($d=-0,69$) bzw. *EIFB* ($d=-1,19$; signifikant) jeweils separat gegenüber *keinem FB*. Das *EIFB* ist dem *EiFB* tendenziell unterlegen ($d=-0,54$). Die Effekte verringern sich somit im Vergleich zu den Ergebnissen über alle Aufgaben. *Hohe SF* führt zu mittleren bzw. hohen Effekten des *FB* insgesamt ($d=0,69$), des *EiFB* ($d=1,05$; signifikant) sowie des *EIFB* ($d=0,60$) gegenüber *keinem FB*. Das *EIFB* ist dem *EiFB* tendenziell unterlegen ($d=-0,23$). Im Vergleich zu den Ergebnissen über alle Aufgaben vergrößern sich die Effekte. Dies deutet darauf hin, dass Probanden mit hoher *SF* langfristig vor allem bei komplexen Aufgaben einen Vorteil erzielen. Die Ergebnisse unter Berücksichtigung des Einflussfaktors *SF* auf einfache bzw. komplexe Aufgaben sind folglich mit Ausnahme der Analyse des Einflusses der *SF* auf den Wissenszuwachs komplexer Aufgaben über drei Zeitpunkte mit den Ergebnissen über alle Aufgaben vergleichbar. Demnach scheint eine weitere Differenzierung der Aufgaben entsprechend des Lernzielniveaus in einfache und komplexe Aufgaben keine veränderten Einflüsse der *SF* hervorzurufen (vgl. Kapitel 5.1.1.5 sowie Anhang 5.1.1.-31).

Die Ergebnisse in Bezug auf die Hypothesen werden in nachfolgender Tabelle 58 noch einmal zusammenfassend veranschaulicht. Hierbei werden die Ergebnisse der in Kapitel 5.1.1 durchgeführten Hypothesenprüfung vorgestellt. Des Weiteren werden die Ergebnisse der paarweisen Vergleiche der Gruppen für den 1. Nachtest, 2. Nachtest, den Wissenszuwachs über zwei und drei Zeitpunkte anhand der Effektstärken dargestellt, um Tendenzen hinsichtlich der Hypothesen ableiten zu können. Dabei werden Effektstärken in Anlehnung an Cohen (1988, S. 24 ff.) von $d=0,16$ bis $d=0,45$ als klein, Effektstärken von $d=0,46$ bis $d=0,75$ als mittel und Effektstärken ab $d=0,76$ als groß angesehen.

Tabelle 58: Ergebnisse der Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs -
Prüfung der Hypothesen

Hypothese	Inhalt	Ergebnis der Prüfung	Effektstärken			
			1.NT	2.NT	WZ Vor-1.NT	WZ Vor-1.NT-2.NT
H1	FB>kein FB	X	0,35	0,18	0,26	- 0,29
H1a	EiFB>kein FB	X	0,06	0,00	0,09	- 0,29
H1b	ElFB>kein FB	✓ (1.NT) bzw. X	0,69*	0,37	0,46	- 0,33
H2	ElFB>EiFB	X	0,63	0,39	0,35	- 0,09
Prüfung der Hypothesen unter Herauspriorisierung der Einflussfaktoren gesamt (SF, VW, H, IN, UF)						
H1	FB>kein FB	X	0,39	- 0,06		
H1a	EiFB>kein FB	X	0,37	0,00		
H1b	ElFB>kein FB	X	0,52	- 0,13		
H2	ElFB>EiFB	X	0,13	- 0,64		
Einfluss des VW						
Niedriges VW						
H3	FB>kein FB	X	0,09	- 0,36	0,06	- 0,54
H3a	EiFB>kein FB	X	- 0,14	- 0,50	0,09	- 0,59
H3b	ElFB>kein FB	X	0,38	- 0,17	0,31	- 0,50
H4	ElFB>EiFB	X	0,57	0,31	0,38	0,11
Hohes VW						
H5	FB=kein FB	✓	0,63	0,66	0,68	0,48
H5a	EiFB=kein FB	✓ bzw. X (2.NT)	0,74	3,23*	0,64	1,06
H5b	ElFB=kein FB	✓	0,74	0,57	0,88	0,50
H6	ElFB=EiFB	✓	0,18	- 0,38	0,31	- 0,39
Einfluss der SF						
Niedrige SF						
H7	FB>kein FB	✓ (1.NT) bzw. X	0,97*	0,00	0,42	- 1,11*
H7a	EiFB>kein FB	X	0,79	0,00	0,23	- 1,08*
H7b	ElFB>kein FB	✓ (1.NT) bzw. X	1,91*	0,06	0,94	- 1,49*
H8	ElFB>EiFB	X	0,64	0,11	0,57	- 0,20
Hohe SF						
H9	FB=kein FB	✓	0,17	0,39	0,33	0,30
H9a	EiFB=kein FB	✓	- 0,20	0,18	0,42	0,45
H9b	ElFB=kein FB	✓	0,35	0,52	0,31	0,26
H10	ElFB=EiFB	✓	0,42	0,38	- 0,06	- 0,14
Einfluss des Lernzielniveaus						
Einfache Aufgaben						
H23	FB=kein FB	✓	0,38	0,09	0,23	- 0,33
H23a	EiFB=kein FB	✓	0,14	- 0,25	0,23	- 0,49
H23b	ElFB=kein FB	✓	0,67	0,36	0,26	- 0,36
H24	ElFB=EiFB	✓	0,52	0,60	0,00	- 0,22
Komplexe Aufgaben						
H25	FB>kein FB	X	0,26	0,22	0,22	- 0,16
H25a	EiFB>kein FB	X	- 0,06	0,13	- 0,06	- 0,06
H25b	ElFB>kein FB	X	0,59	0,33	0,51	- 0,32
H26	ElFB>EiFB	X	0,64	0,18	0,56	- 0,34

* signifikant auf dem Niveau $\alpha = 0,05$

✓ Hypothese angenommen

X Hypothese abgelehnt

kein Effekt, kleiner Effekt, mittlerer Effekt, großer Effekt

5.1.2 Einfluss der Art des Feedbacks auf die Qualität der Maps

5.1.2.1 Gruppenunterschiede hinsichtlich der Map-Qualität

Um zu analysieren, inwieweit sich die Gruppen *kein FB*, *EiFB* und *EIFB* hinsichtlich der Qualität der Maps unterscheiden, wird eine multivariate Varianzanalyse [MANOVA] mit der UV *Gruppe* und den AVs Scoringwerte für das Map 1 [*ScorM1*], das Map 2 [*ScorM2*] und das Map 3 [*ScorM3*] durchgeführt. Zunächst gilt es, die Voraussetzungen zu prüfen. Die Messwerte in allen Teilstichproben sollten normalverteilt sein. Diese Voraussetzung ist für den Wert *ScorM3* bei der Gruppe *kein FB* und *EIFB* nicht erfüllt. Dennoch kann das Verfahren durchgeführt werden, da es relativ robust bei annähernd gleich großen Stichproben ist (Bühner & Ziegler, 2009, S. 518). Als weitere Voraussetzung gilt es, die Varianzhomogenität zu prüfen. Der Levene-Test auf Gleichheit der Varianzen sowie der Box-M Test zeigen eine Varianzhomogenität für alle drei Scoringwerte.

Um sich zunächst einen Überblick über die Gruppenunterschiede zu verschaffen, werden in nachfolgender Tabelle die Mittelwerte und Standardabweichungen für *ScorM1*, *ScorM2* und *ScorM3* angegeben. Hierbei ist erkennbar, dass sich hinsichtlich *ScorM1* und *ScorM3* Gruppenunterschiede zeigen, wohingegen bei *ScorM2* kaum Unterschiede ersichtlich sind.

Tabelle 59: Mittelwerte und Standardabweichungen für die Gruppen *kein FB*, *EiFB* und *EIFB* *ScorM1*, *ScorM2* und *ScorM3*

		Mittelwerte	Standardabweichungen
<i>ScorM1</i>	<i>kein FB</i>	0,22	0,08
	<i>EiFB</i>	0,17	0,07
	<i>EIFB</i>	0,26	0,07
<i>ScorM2</i>	<i>kein FB</i>	0,19	0,09
	<i>EiFB</i>	0,14	0,09
	<i>EIFB</i>	0,16	0,09
<i>ScorM3</i>	<i>kein FB</i>	0,04	0,05
	<i>EiFB</i>	0,06	0,05
	<i>EIFB</i>	0,14	0,08

Die Ergebnisse der Analyse in Tabelle 60 zeigen signifikante Unterschiede zwischen den Gruppen *EIFB*, *EiFB* und *kein FB* hinsichtlich der Scoringwerte insgesamt (F=6,308; p=0,000) mit einem großen Effekt von d=1,21. Dabei unterscheiden sich die Gruppen hinsichtlich *ScorM1* (F=6,103; p=0,0041) und *ScorM3* (F=13,489; p=0,000) mit jeweils großen Effekten von d=0,96 und d=1,43 signifikant voneinander. In Bezug auf *ScorM2*

sind jedoch keine signifikanten Gruppenunterschiede erkennbar ($F=1,861$; $p=0,166$). Hierbei zeigt sich ein mittlerer Effekt mit $d=0,53$.

Tabelle 60: Ergebnisse der MANOVA mit der UV Gruppe und den AVs ScorM1, ScorM2 und ScorM3

Quelle	Quad- rat- summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
UV: Gruppe, AV: ScorMges.		6		6,308	0,000*	0,267	1,21
<u>Abh. Variable Scoringwerte (ScorM1, ScorM2, ScorM3)</u>							
ScorM1	0,063**	2	0,032	6,103	0,004*	0,187	0,96
ScorM2	0,030***	2	0,015	1,861	0,166	0,066	0,53
ScorM3	0,110****	2	0,055	13,489	0,000*	0,337	1,43

- * signifikant auf dem Niveau $\alpha = 0,05$
- ** Korrigiertes R-Quadrat = 0,157
- *** Korrigiertes R-Quadrat = 0,030
- **** Korrigiertes R-Quadrat = 0,312

Die Gruppenunterschiede beim ersten Map müssen relativiert werden, da zu diesem Zeitpunkt noch keine Intervention stattgefunden hatte. Eventuelle Gruppenunterschiede müssen an dieser Stelle andere Ursachen haben, weshalb eine multivariate Kovarianzanalyse [MANCOVA] mit der UV *Gruppe*, der Kovariaten *ScorM1* und den Scoringwerten als AVs (ScorM2 und ScorM3) durchgeführt wird. Die Ergebnisse sind in nachfolgender Tabelle 61 dargestellt.

Tabelle 61: Ergebnisse der MANCOVA mit der UV Gruppe, der Kovariaten ScorM1 und den AVs ScorM2 sowie ScorM3

Quelle	Abhängige Variable	Quadrat- summe	df	Mittel der Qua- drate	F-Wert	p	Partielles Eta- Quadrat	d
Kovariate: ScorM1	ScorMges.		2		5,845	0,005*	0,186	0,96*
UV: Gruppe	ScorMges.		4		5,636	0,000*	0,178	0,93*
Kovariate: ScorM1	ScorM2	0,008**	1	0,008	1,024	0,316	0,019	0,28
	ScorM3	0,040***	1	0,040	11,690	0,001*	0,184	0,95*
UV: Gruppe	ScorM2	0,024	2	0,012	1,496	0,234	0,054	0,48
	ScorM3	0,074	2	0,037	10,924	0,000*	0,296	1,30*

- * signifikant auf dem Niveau $\alpha = 0,05$
- ** Korrigiertes R-Quadrat = 0,031
- *** Korrigiertes R-Quadrat = 0,428

Die Kovariate *ScorM1* hat einen signifikanten Einfluss auf die Scoringwerte insgesamt ($F=5,845$; $p=0,005$) mit einem großen Effekt von $d=0,96$. Die Gruppen unterscheiden

sich dabei hinsichtlich der Scoringwerte insgesamt ebenfalls signifikant voneinander ($F=5,636$; $p=0,000$) mit einem großen Effekt von $d=0,93$. Dabei hat die Kovariate keinen signifikanten Einfluss auf *ScorM2* ($F=1,024$; $p=0,316$), der Effekt ist mit $d=0,28$ gering. Der Einfluss auf *ScorM3* ist dagegen signifikant ($F=11,690$; $p=0,001$) mit einem großen Effekt von $d=0,95$. Hinsichtlich *ScorM2* unterscheiden sich die Gruppen nicht signifikant voneinander ($F=1,496$; $p=0,234$) mit einem mittleren Effekt von $d=0,48$, bezüglich *ScorM3* sind die Gruppenunterschiede signifikant ($F=10,924$; $p=0,000$) mit einem großen Effekt von $d=1,30$. Die Aufnahme der Kovariate *ScorM1* verbessert die Modellgüte von 0,030 auf 0,031 für *ScorM2* und von 0,312 auf 0,428 für *ScorM3*.

Um Aussagen hinsichtlich der Hypothesen H1 und H2 treffen zu können und somit die Lernwirksamkeit des Feedbacks zu überprüfen, werden jeweils weitere einfaktorielle MANCOVAs durchgeführt. In nachfolgender Tabelle 62 sind die Effektstärken der Gruppenvergleiche für die Scoringwerte *ScorM2* und *ScorM3* unter Berücksichtigung der Kovariaten *ScorM1* dargestellt (vgl. Anhang 5.1.2.-1).

Tabelle 62: Effektstärken für die MANCOVA mit der UV Gruppe, der Kovariaten *ScorM1* und den AVs *ScorM2* sowie *ScorM3*

	ScorM2	ScorM3
FB vs. kein FB	d=-0,47 ($\eta^2=0,053$)	d=1,08* ($\eta^2=0,225$)
EiFB vs. kein FB	d=-0,55 ($\eta^2=0,071$)	d=0,85* ($\eta^2=0,152$)
EIFB vs. kein FB	d=-0,42 ($\eta^2=0,043$)	d=1,40* ($\eta^2=0,328$)
EIFB vs. EiFB	d=0,00 ($\eta^2=0,000$)	d=0,66 ($\eta^2=0,098$)

* signifikant auf dem Niveau $\alpha = 0,05$

Beim 2. Mapping zeigen sich mittlere negative Effekte des *FB* gegenüber *keinem FB* ($d=-0,47$), wobei sowohl das *EiFB* als auch *EIFB* mit einem mittleren ($d=-0,55$) bzw. kleinen Effekt ($d=-0,42$) *keinem FB* unterlegen sind. Das *EIFB* und das *EiFB* unterscheiden sich nicht ($d=0,00$).

Für das 3. Mapping zeigt sich ein großer signifikanter Effekt von $d=1,08$ des *FB* gegenüber *keinem FB*. Dies resultiert insbesondere durch den großen signifikanten Effekt des *EIFB* gegenüber *keinem FB* ($d=1,40$). Das *EiFB* ist *keinem FB* ebenfalls mit einem großen

Effekt überlegen ($d=0,85$). Der zunächst negative Effekt kehrt sich folglich beim 3. Mapping um. Das *EIFB* ist dem *EiFB* mit einem mittleren Effekt überlegen ($d=0,66$).

Nachfolgend wird überprüft, inwieweit sich die Qualität der Maps, operationalisiert durch die Scoring-Werte, zwischen den verschiedenen Messzeitpunkten verändert und ob die Veränderung über die Zeit von der Art des Feedbacks abhängt. Aufgrund dessen, dass bis zum 1. Mapping keine Intervention stattfand, können so die Gruppenunterschiede hinsichtlich des ersten selbsterstellten Concept Maps ebenfalls relativiert werden (vgl. Tabelle 60). Hierfür wird eine univariate zweifaktorielle ANOVA mit Messwiederholung mit den Faktoren *Zeit* und *Gruppe* sowie der AV „Scoringwerte“ durchgeführt. Insgesamt lassen sich drei Messzeitpunkte definieren: 1. Mapping, 2. Mapping und 3. Mapping. Demzufolge ergeben sich für die ANOVA mit Messwiederholung drei Stufen für den Faktor *Zeit* mit drei Gruppen.

Für die Durchführung der ANOVAs gilt es, zunächst die Voraussetzungen zu prüfen. Zur Prüfung der Normalverteilung kann auf die entsprechenden Ausführungen zur MANOVA verwiesen werden. Als weitere Voraussetzung gilt es, die Varianzhomogenität zu prüfen. Der Levene-Test auf Gleichheit der Varianzen zeigt Varianzhomogenität für alle drei Zeitpunkte. Die Sphärizitätsannahme ist jedoch verletzt. Aus diesem Grund müssen die Freiheitsgrade bei den Tests des Messwiederholungseffekts und der Wechselwirkung korrigiert werden.

In nachfolgender Tabelle 63 sind die Ergebnisse der ANOVA mit Messwiederholung über drei Gruppen, unter Berücksichtigung der Korrekturverfahren nach Greenhouse-Geisser, dargestellt.⁹⁶

⁹⁶ Die Korrektur nach Greenhouse-Geisser stellt dabei das strengere, d. h. das konservativere Verfahren gegenüber dem Verfahren nach Huynh-Feldt dar und wird daher für weitere Analysen herangezogen. Die Freiheitsgrade aus dem F-Test werden mit dem Korrekturfaktor Epsilon multipliziert. Die Freiheitsgrade reduzieren sich, da der Faktor kleiner als 1 ist. Somit wird eine strengere Prüfgrößenverteilung herangezogen. Um ein signifikantes Ergebnis zu erzielen, muss ein größerer F-Wert erreicht werden (Bühner & Ziegler, 2009, S. 462 f.).

Tabelle 63: Ergebnisse der zweifaktoriellen ANOVA mit Messwiederholung mit den UVs Zeit und Gruppe und der AV Scoringwerte

Quelle	Quadrat-summe	df	Mittel der Qua-drate	F-Wert	p	Partielles Eta-Quadrat	d
Zeit _{ScorM1-ScorM2-ScorM3}	0,546	1,705	0,273	61,213	0,000*	0,536	2,15
Zeit _{ScorM1-ScorM2-ScorM3} XGruppe	0,097	3,410	0,028	5,434	0,001*	0,170	0,91

* signifikant auf dem Niveau $\alpha = 0,05$

Die ANOVA mit Messwiederholung über drei Gruppen zeigt einen signifikanten Haupteffekt für den Faktor *Zeit* ($F=61,213$; $p=0,000$) und einen signifikanten Interaktionseffekt für die Faktoren *Zeit* und *Gruppe* ($F=5,434$; $p=0,001$). Die Mapqualität unterscheidet sich folglich zwischen den drei Messzeitpunkten signifikant voneinander. Des Weiteren hängt die Veränderung über die *Zeit* signifikant von der Art des Feedbacks ab.

Um Aussagen hinsichtlich der Hypothesen H1 und H2 treffen zu können und somit die Lernwirksamkeit des Feedbacks zu überprüfen, werden weitere zweifaktorielle ANOVAs mit Messwiederholung durchgeführt. Nachfolgende Tabelle 64 zeigt die Effektstärken der Gruppenvergleiche für die Entwicklung der Scoringwerte über die drei Zeitpunkte (vgl. Anhang 5.1.2-2).

Tabelle 64: Effektstärken für die zweifaktorielle ANOVA mit Messwiederholung mit den UVs Zeit und Gruppe und der AV Scoringwerte

	ScorM1-ScorM2-ScorM3
FB vs. kein FB	d=0,79* ($\eta^2=0,134$)
EiFB vs. kein FB	d=0,69* ($\eta^2=0,107$)
ElFB vs. kein FB	d=1,03* ($\eta^2=0,211$)
ElFB vs. EiFB	d= 0,51 ($\eta^2=0,061$)

* signifikant auf dem Niveau $\alpha = 0,05$

Das *FB* ist *keinem FB* mit einem großen signifikanten Effekt von $d=0,79$ überlegen, wobei der Vorteil des *ElFB* mit einem großen signifikanten Effekt von $d=1,03$ gegenüber *keinem FB* größer ist als der Vorteil des *EiFB* gegenüber *keinem FB* (mittlerer signifikanter Effekt $d=0,69$). Das *ElFB* ist dem *EiFB* mit einem mittleren Effekt ($d=0,51$) überlegen.

Aus nachfolgender Abbildung 27 ist ableitbar, dass die Qualität der Maps insgesamt über die Zeit abnimmt. Dabei nimmt die Qualität der Maps der Gruppen *EiFB* (-0,25) und *EIFB* (-0,28) im Vergleich zur Gruppe *kein FB* (-0,38) insgesamt nicht so stark ab.

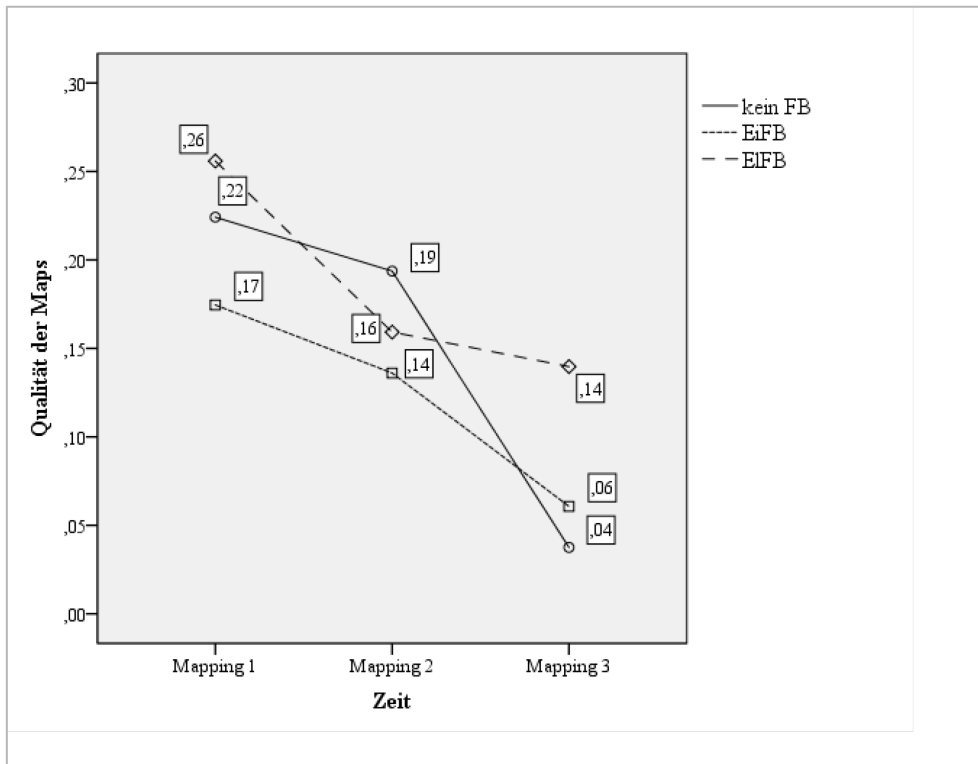


Abbildung 27: Graphische Darstellung der Scoringwerte für die Gruppen kein FB, EiFB und EIFB über drei Messzeitpunkte

5.1.2.2 Einfluss des Vorwissens, der Sprachfähigkeit, der Einschätzung zum Umgang mit Concept Mapping und zum Umgang mit Fehlern auf die Qualität der Maps

Nachfolgend wird der Einfluss der Faktoren *VW*, *SF*, *H*, *IN* und *UF* auf die Qualität der Maps überprüft. Hierzu wird eine einfaktorielle MANCOVA mit dem Faktor *Gruppe* und den AVs *ScorM2* sowie *ScorM3* durchgeführt. *ScorM1* wird als Kovariate aufgenommen, um den Einfluss herauszupartialisieren (vgl. Kapitel 5.1.2.1) Weitere Kovariaten sind *VW*, *SF*, *H*, *IN* und *UF*. Zur Prüfung der Voraussetzungen kann auf Kapitel 5.1.2.1 verwiesen werden. Die Tabelle 65 und die Tabelle 66 zeigen die Ergebnisse der einfaktoriellen MANCOVA.

Tabelle 65: Ergebnisse der einfaktoriellen MANCOVA mit der UV Gruppe, den Kovariaten ScorM1, VW, SF, H, IN, UF und der AV Scoringwerte

Effekt	Wert	F-Wert	df	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Scoringwerte (ScorM2 und ScorM3)</u>						
ScorM1	0,145	3,898	2	0,027*	0,145	0,82*
VW	0,052	1,270	2	0,291	0,052	0,47
SF	0,087	2,203	2	0,122	0,087	0,62
H	0,105	2,685	2	0,079	0,105	0,69
IN	0,126	3,322	2	0,045*	0,126	0,76*
UF	0,006	0,134	2	0,875	0,006	0,16
Gruppe	0,351	5,000	4	0,001*	0,175	0,92*

* signifikant auf dem Niveau $\alpha = 0,05$

Die Kovariaten *ScorM1* und *IN* haben insgesamt einen signifikanten Einfluss auf die Qualität der Maps ($F=3,898$; $p=0,027$; $F=3,322$; $p=0,045$) mit hohen Effekten ($d=0,82$ bzw. $d=0,76$). Die Gruppenunterschiede bleiben dabei im Vergleich zur Analyse ohne Kovariaten mit einem hohen Effekt von $d=0,82$ signifikant ($F=5,000$; $p=0,001$).

Tabelle 66: Ergebnisse der MANCOVA mit der UV Gruppe, den Kovariaten ScorM1, VW, SF, H, IN, UF und der AV ScorM2 bzw. ScorM3

Quelle	Abhängige Variable	Quadratsumme	df	Mittel der Quadratrate	F	p	Partielles Eta-Quadrat	d
ScorM1	ScorM2	0,001	1	0,001	0,179	0,674	0,004	0,13
	ScorM3	0,028	1	0,028	7,948	0,007*	0,145	0,82*
VW	ScorM2	0,016	1	0,016	2,405	0,128	0,049	0,46
	ScorM3	0,001	1	0,001	0,349	0,557	0,007	0,17
SF	ScorM2	0,019	1	0,019	2,896	0,095	0,058	0,50
	ScorM3	0,007	1	0,007	2,055	0,158	0,042	0,42
H	ScorM2	0,034	1	0,034	5,032	0,030*	0,097	0,66*
	ScorM3	0,001	1	0,001	0,196	0,660	0,004	0,13
IN	ScorM2	0,044	1	0,044	6,518	0,014*	0,122	0,75*
	ScorM3	0,002	1	0,002	0,604	0,441	0,013	0,23
UF	ScorM2	0,001	1	0,001	0,136	0,714	0,003	0,11
	ScorM3	0,000	1	0,000	0,111	0,741	0,002	0,09
Gruppe	ScorM2**	0,033	2	0,016	2,432	0,099	0,094	0,64
	ScorM3***	0,063	2	0,032	9,072	0,000*	0,279	1,24*

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,203

*** korrigiertes R-Quadrat: 0,413

Werden die Scoringwerte einzeln betrachtet, so zeigt sich, dass *ScorM1* einen signifikanten Einfluss mit einem hohen Effekt von $d=0,82$ auf *ScorM3* hat ($F=7,948$; $p=0,007$). Des

Weiteren zeigt sich ein signifikanter mittlerer Einfluss von $d=0,66$ der Kovariaten *H* auf *ScorM2* ($F=5,032$; $p=0,030$) sowie ein hoher Einfluss von $d=0,75$ der Kovariaten *IN* auf *ScorM2* ($F=6,518$; $p=0,014$). Die Kovariaten *VW* ($d=0,46$; $d=0,17$) und *SF* ($d=0,50$; $d=0,42$) haben keinen signifikanten Einfluss auf die Qualität der Maps, wobei sich kleine bzw. mittlere Effekte zeigen. Es zeigt sich kein Einfluss der Kovariate *UF* ($d=0,11$; $d=0,09$). Die Gruppenunterschiede hinsichtlich *ScorM2* bleiben dabei nicht signifikant, wobei sich ein mittlerer Effekt von $d=0,64$ zeigt. Die Gruppen unterscheiden sich dagegen signifikant unter Herauspriorisierung der Kovariaten hinsichtlich *ScorM3* mit einem hohen Effekt von $d=1,24$ ($F=9,072$; $p=0,000$). Die Gruppenunterschiede im Vergleich zu der Analyse nur mit *ScorM1* als Kovariate (vgl. Tabelle 61) vergrößern sich hinsichtlich *ScorM2* (von $d=0,48$ auf $d=0,64$) und bleiben hinsichtlich *ScorM3* fast konstant ($d=1,30$ vs. $d=1,24$). Die Aufnahme weiterer Kovariaten (neben *ScorM1*) führt zu einer Verbesserung der Modellgüte von 0,031 auf 0,203 für *ScorM2*. Für *ScorM3* verringert sich die Güte des Modells geringfügig von 0,428 auf 0,413 (vgl. Kapitel 5.1.2.1).

Um Aussagen hinsichtlich der Hypothesen H1 und H2 unter Herauspriorisierung der Einflussfaktoren insgesamt treffen zu können, werden weitere einfaktorielle MANCOVAs durchgeführt. Nachfolgende Tabelle 67 zeigt die Effektstärken der Gruppenvergleiche hinsichtlich der Scoringwerte *ScorM2* und *ScorM3* unter Berücksichtigung der Kovariaten *ScorM1*, *VW*, *SF*, *H*, *IN* und *UF* (vgl. Anhang 5.1.2-3).

Tabelle 67: Effektstärken für die einfaktorielle MANCOVA mit der UV Gruppe, den Kovariaten *ScorM1*, *VW*, *SF*, *H*, *IN*, *UF* und der AV *ScorM2* bzw. *ScorM3*

	ScorM2	ScorM3
FB vs. kein FB	$d=-0,63^*$ ($\eta^2=0,090$)	$d=1,08^*$ ($\eta^2=0,227$)
EiFB vs. kein FB	$d=-0,61$ ($\eta^2=0,085$)	$d=0,73$ ($\eta^2=0,117$)
EiFB vs. kein FB	$d=-0,84^*$ ($\eta^2=0,150$)	$d=1,38^*$ ($\eta^2=0,321$)
EiFB vs. EiFB	$d=-0,17$ ($\eta^2=0,007$)	$d=0,44$ ($\eta^2=0,047$)

* signifikant auf dem Niveau $\alpha = 0,05$

Die Effekte haben sich durch das Herauspriorisieren der Kovariaten hinsichtlich *ScorM2* vergrößert (vgl. Tabelle 62). Das *FB* bleibt *keinem FB* mit einem mittleren Effekt unterlegen, wobei der Effekt signifikant wird ($d=-0,47$ vs. $d=-0,63$). Hierbei vergrößert

sich insbesondere der Effekt von *keinem FB* gegenüber *EiFB*, der jetzt signifikant wird ($d=-0,42$ vs. $d=-0,84$). *Kein FB* bleibt mit einem mittleren Effekt lernwirksamer als *EiFB* ($d=-0,55$ vs. $d=-0,61$). Das *EiFB* ist dem *EiFB* nun geringfügig überlegen ($d=0,00$ vs. $d=-0,17$). Für das 3. Mapping lassen sich nur geringfügige Veränderungen ableiten. Das *FB* bleibt *keinem FB* mit einem unveränderten großen Effekt von $d=1,08$ überlegen, wobei sich beim *EiFB* unverändert große signifikante Effekte gegenüber *keinem FB* zeigen ($d=1,40$ vs. $d=1,38$). Das *EiFB* ist *keinem FB* nicht mehr signifikant überlegen, wobei sich der Effekt etwas verringert ($d=0,85$ vs. $d=0,73$). Das *EiFB* bleibt dem *EiFB* überlegen, wobei auch hier eine Verringerung des Effekts erkennbar ist ($d=0,66$ vs. $d=0,44$).

Der Einfluss der Kovariaten *IN* und *H* führen folglich zu veränderten Gruppenunterschieden insbesondere beim 2. Mapping. Um den Einfluss dieser Kovariaten genauer zu analysieren, werden diese für weitere Analysen am Median dichotomisiert und jeweils zwei Gruppen, niedrig und hoch gebildet. Dabei wird lediglich das 2. Mapping betrachtet, da hier die Kovariaten *IN* und *H* einen signifikanten Einfluss haben.

Zur Analyse der dichotomisierten Kovariate *IN_nh* wird eine zweifaktorielle ANCOVA mit den UVs *Gruppe* und *IN_nh*, der Kovariaten *ScorM1* und der AV *ScorM2* durchgeführt. Die Kovariate *ScorM1* wird an dieser Stelle mit aufgenommen, um den Einfluss des ersten Maps herauszupartialisieren. Die Ergebnisse sind in nachfolgender Tabelle 68 dargestellt.

Tabelle 68: Ergebnisse der zweifaktoriellen ANCOVA mit den UVs *Gruppe* und *IN_nh*, den Kovariaten *ScorM1* sowie der AV *ScorM2*

Quelle	Abhängige Variable	Quadratsumme	df	Mittel der Quadrate	F	p	Partielles Eta-Quadrat	d
ScorM1	ScorM2	0,006*	1	0,006	0,764	0,386	0,015	0,25
Gruppe	ScorM2	0,032	2	0,016	2,027	0,143	0,076	0,57
IN_nh	ScorM2	0,005	1	0,005	0,649	0,424	0,013	0,23
Gruppe * IN_nh	ScorM2	0,029	2	0,015	1,860	0,167	0,071	0,55

* korrigiertes R-Quadrat: 0,064

Der Haupteffekt *Gruppe* ist unter Neutralisierung der Kovariate *ScorM1* für das 2. Mapping mit einem mittleren Effekt von $d=0,57$ nicht signifikant, ($F=2,027$; $p=0,143$). Der Haupteffekt *IN_nh* bleibt mit einem geringen Effekt von $d=0,31$ nicht signifikant ($F=0,649$; $p=0,424$). Der Interaktionseffekt zwischen *Gruppe* und *IN_nh* bleibt mit einem

mittleren Effekt von $d=0,51$ nicht signifikant ($F=1,860$; $p=0,167$). Die Gruppen unterscheiden sich hinsichtlich *ScorM2* nicht signifikant voneinander, auch unabhängig davon, ob sie das Concept Mapping als interessant bzw. nützlich einschätzen oder nicht.

Um Aussagen hinsichtlich der Hypothese H15 bis H18 treffen zu können und somit den Einfluss der Variablen *IN* auf die Lernwirksamkeit des Feedbacks, operationalisiert durch die Qualität des 2. Maps, zu überprüfen, werden jeweils weitere einfaktorielle ANCOVAs jeweils separat für *IN niedrig* und *IN hoch* durchgeführt.

Die Ergebnisse der ANCOVAs für die paarweisen Vergleiche sowie die Effektstärken können den Anhängen 5.1.2.-4 und 5.1.2.-5 entnommen werden.

Für *IN niedrig* zeigt sich ein tendenzieller Vorteil von *keinem FB* gegenüber dem *FB* mit einem mittleren Effekt ($d=-0,68$), was insbesondere auf einen großen signifikanten Effekt von *keinem FB* gegenüber dem *EIFB* ($d=-1,29$) zurückzuführen ist. *Kein FB* zeigt gegenüber dem *EiFB* lediglich einen kleinen Effekt ($d=-0,39$). Das *EIFB* ist dem *EiFB* mit einem mittleren Effekt von $d=-0,54$ unterlegen.

Für *IN hoch* ist *kein FB* dem *FB* ebenfalls tendenziell überlegen ($d=-0,38$), wobei dies hier aus einem großen Effekt von *keinem FB* gegenüber *EiFB* ($d=-0,85$) resultiert. Das *EIFB* zeigt gegenüber *keinem FB* keinen Effekt ($d=-0,11$). Des Weiteren ist das *EIFB* dem *EiFB* mit einem kleinen Effekt von $d=0,38$ überlegen.

Nachfolgende Abbildung 28 zeigt, dass bei der Gruppe *kein FB* unabhängig von *IN_nh* die Qualität des 2. Maps am höchsten ist (*ScorM2*=0,19 bzw. 0,20). Bei der Gruppe *EiFB* ist die Qualität des 2. Maps bei den Probanden mit *IN hoch* geringer (*ScorM2*=0,12) als bei den Probanden mit *IN niedrig* (*ScorM2*=0,15). Am deutlichsten ist der Einfluss von *IN_nh* bei dem *EIFB* erkennbar. Probanden mit *IN niedrig* erstellen auch das 2. Map mit einer deutlich niedrigeren Qualität (*ScorM2*=0,10) als Probanden mit *IN hoch* (*ScorM2*=0,19). Ein zusätzlicher Paarvergleich zeigt hier eine große Effektstärke von 1,02.

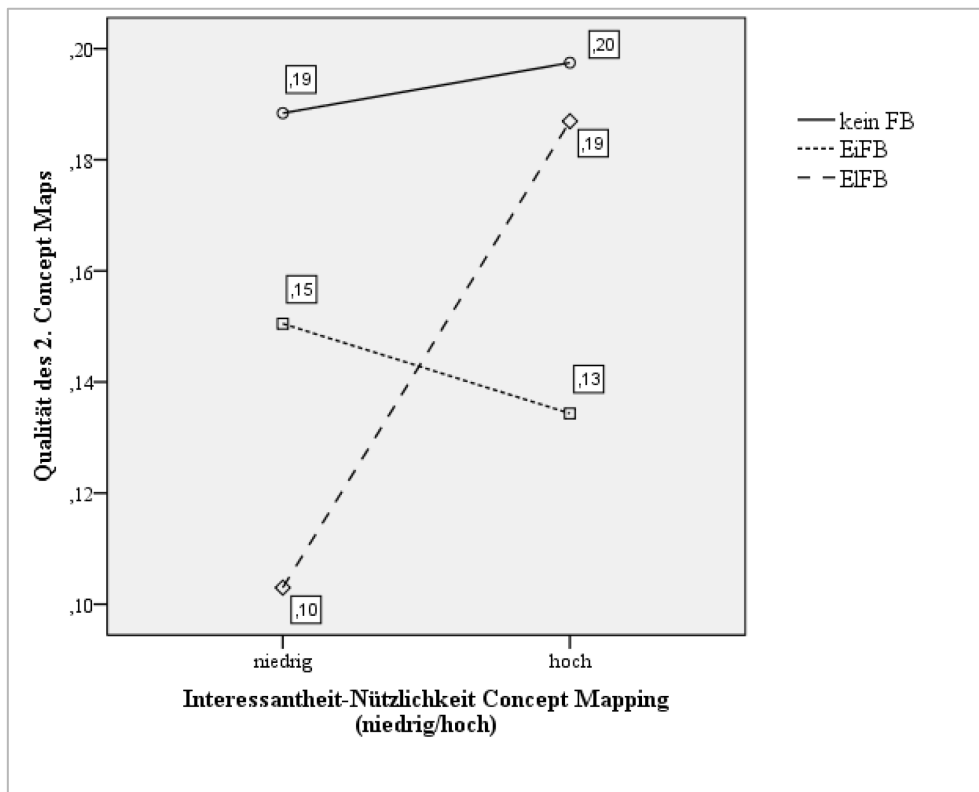


Abbildung 28: Graphische Darstellung der Werte für ScorM2 in Abhängigkeit vom Niveau für IN_nh

Zur Analyse der dichotomisierten Kovariaten *niedrig* bzw. *hohe eingeschätzte Handhabbarkeit hinsichtlich des Concept Mapping* [H_nh] wird eine zweifaktorielle ANCOVA mit den UVs *Gruppe* und *H_nh*⁹⁷, der Kovariaten *ScorM1* und der AV *ScorM2* durchgeführt. Die Kovariaten *ScorM1* wird an dieser Stelle mit aufgenommen, um den Einfluss des ersten Maps herauszupartialisieren. Die Ergebnisse sind in nachfolgender Tabelle 69 dargestellt.

Tabelle 69: Ergebnisse der zweifaktoriellen ANCOVA mit den UVs *Gruppe* und *H_nh*, den Kovariaten *ScorM1* sowie der AV *ScorM2*

Quelle	Abhängige Variable	Quadratsumme	df	Mittel der Quadrate	F	p	Partielles Eta-Quadrat	d
ScorM1	ScorM2	0,014**	1	0,014	1,827	0,183	0,036	0,39
Gruppe	ScorM2	0,024	2	0,012	1,520	0,229	0,058	0,50
H_nh	ScorM2	0,042	1	0,042	5,468	0,023*	0,100	0,67
Gruppe * H_nh	ScorM2	0,001	2	0,001	0,093	0,911	0,004	0,13

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,078

⁹⁷ *Niedrige H* beinhaltet dabei alle Probanden mit einem Wert für $H \leq 2,375$ (Median) und *hohe H* umfasst alle Probanden mit einem Wert für $H > 2,375$.

Der Haupteffekt *Gruppe* ist unter Neutralisierung der Kovariate *ScorM1* für das 2. Map mit einem mittleren Effekt von $d=0,50$ nicht signifikant ($F=1,520$; $p=0,229$). Der Haupteffekt *H_nh* wird mit einem mittleren Effekt von $d=0,67$ signifikant ($F=5,468$; $p=0,023$). Probanden, die das Concept Mapping als weniger handhabbar einschätzen, erzielen eine höhere Qualität im 2. Map als die Probanden, welche das Concept Mapping als handhabbar einschätzen. Zwischen *Gruppe* und *H_nh* zeigt sich kein Interaktionseffekt ($F=0,093$; $p=0,911$; $d=0,13$). Die Gruppen unterscheiden sich hinsichtlich *ScorM2* nicht signifikant voneinander, auch unabhängig davon, ob sie das Concept Mapping als handhabbar einschätzen oder nicht.

Um Aussagen hinsichtlich der Hypothese H19 bis H22 treffen zu können und somit den Einfluss der Variablen *H* auf die Lernwirksamkeit des Feedbacks, operationalisiert durch die Qualität des 2. Maps, zu überprüfen, werden jeweils weitere einfaktorielle ANCOVAs jeweils separat für *IN niedrig* und *IN hoch* durchgeführt.

Die Ergebnisse der ANCOVAs für die paarweisen Vergleiche sowie die Effektstärken können den Anhängen 5.1.2-6 und 5.1.2-7 entnommen werden.

Für *H niedrig* zeigt sich ein tendenzieller Vorteil von *keinem FB* gegenüber dem *FB* mit einem mittleren Effekt ($d=-0,55$), was auf mittlere Effekte von *keinem FB* gegenüber dem *EiFB* ($d=-0,64$) sowie gegenüber dem *ElFB* ($d=-0,48$) zurückzuführen ist.

Für *H hoch* ist *kein FB* dem *FB* ebenfalls tendenziell überlegen ($d=-0,44$), wobei dies aus einem niedrigen Effekt von *keinem FB* gegenüber *EiFB* ($d=-0,43$) und aus einem mittleren Effekt gegenüber *ElFB* ($d=-0,48$) resultiert.

Das *ElFB* zeigt gegenüber dem *EiFB* unabhängig vom Niveau für *H* keine Effekte ($d=0,11$; $d=-0,06$).

Nachfolgende Abbildung 29 zeigt, dass bei der Gruppe *kein FB* unabhängig von *H_nh* die Qualität des 2. Maps am höchsten ist (0,23 bzw. 0,16). Die Gruppen *EiFB* und *ElFB* unterschieden sich nur geringfügig, der Wert liegt bei 0,17 bzw. 0,18 für *H niedrig* und 0,12 bzw. 0,13 für *H hoch*. Insgesamt erstellen Probanden, welche die *H* als niedrig einschätzen, im 2. Mapping qualitativ hochwertigere Maps als Probanden, welche die *H* als hoch einschätzen.

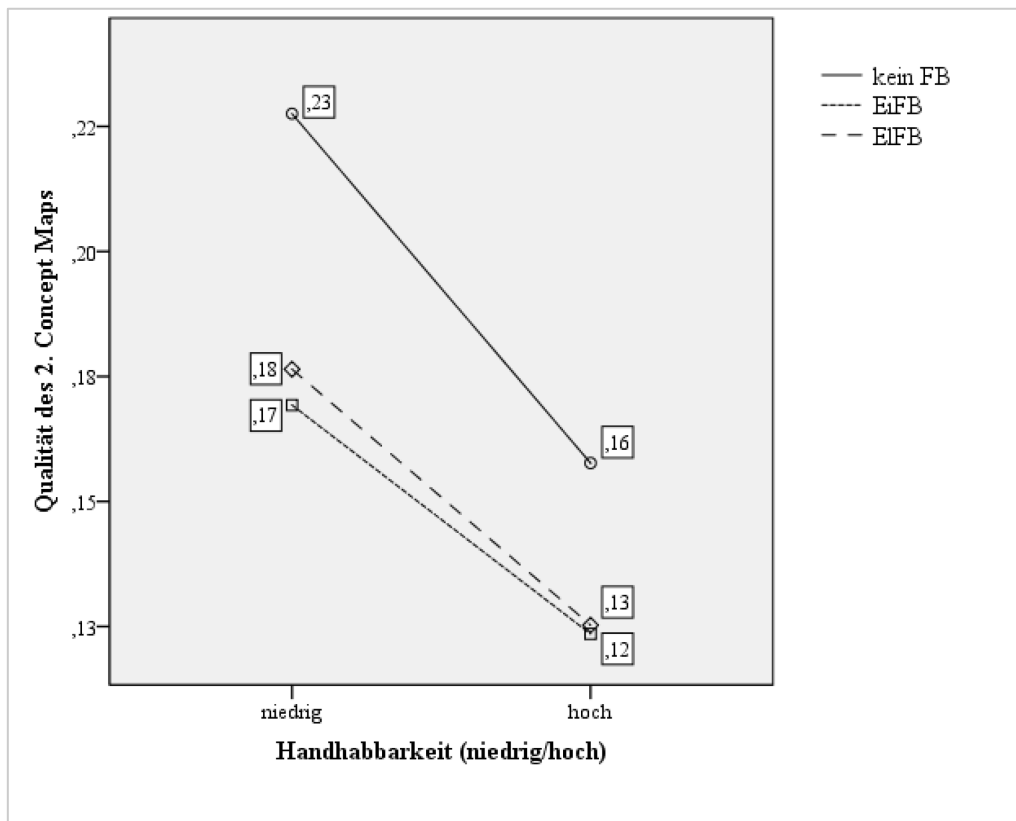


Abbildung 29: Graphische Darstellung der Werte für ScorM2 in Abhängigkeit vom Niveau für H_nh

5.1.2.3 Zusammenfassung der Ergebnisse und Bewertung der Hypothesen hinsichtlich der Qualität der Maps

Gruppenunterschiede hinsichtlich der Qualität der Maps (H1 und H2)

Nachfolgend werden die Hypothesen (vgl. Kapitel 4.1) anhand der in Kapitel 5.1.2 durchgeführten Analysen hinsichtlich der Qualität der Maps bewertet. Hierbei werden die Scoringwerte für die drei Mappings, ScorM1, ScorM2 und ScorM3 sowie die Entwicklung der Mapqualität über die drei Mappings herangezogen.

Zunächst wurde überprüft, inwieweit das Feedback einen Einfluss auf die Qualität der Maps hat. Hierzu gilt es, die Ergebnisse im Hinblick auf die Hypothesen H1, H1a und H1b sowie H2 zu bewerten.

Hinsichtlich ScorM1 zeigen sich signifikante Gruppenunterschiede, welche jedoch nicht auf die Intervention zurückgeführt werden können, sondern andere Ursachen haben müssen. Aus diesem Grund muss der Einfluss dieser Variable kontrolliert werden. ScorM1 wurde demzufolge als Kovariate in alle weiteren Analysen einbezogen, um den Einfluss herauszupartialisieren.

Es zeigen sich signifikante Gruppenunterschiede hinsichtlich ScorM3. In Bezug auf ScorM2 unterscheiden sich die Gruppen nicht signifikant. Werden die Gruppen paarweise gegenübergestellt, so zeigen sich hinsichtlich ScorM3 signifikante Vorteile des *FB* gegenüber *keinem FB*, des *EiFB* gegenüber *keinem FB* sowie des *ElFB* gegenüber *keinem FB*. Das *ElFB* ist dem *EiFB* nicht signifikant überlegen. In Bezug auf ScorM2 zeigen die Paarvergleiche auch keine signifikanten Unterschiede. Die Qualität der Maps nimmt insgesamt über die drei Zeitpunkte (ScorM1, ScorM2 und ScorM3) ab. Dabei hängt die Veränderung über die Zeit signifikant von der Art des Feedbacks ab. Die Paarvergleiche zeigen ebenfalls signifikante Vorteile des *FB* gegenüber *keinem FB*, des *EiFB* gegenüber *keinem FB* sowie des *ElFB* gegenüber *keinem FB*. Das *ElFB* ist dem *EiFB* nicht signifikant überlegen.

Die Hypothese H1 sowie H1a und H1b kann daher für ScorM3 sowie die Entwicklung der Mapqualität angenommen werden. Das *FB* ist in Bezug auf die Qualität des 3. Maps sowie die Entwicklung der Mapqualität über die Zeit signifikant besser als *kein FB*, wobei sowohl das *EiFB* als auch das *ElFB* *keinem FB* signifikant überlegen sind. Die Hypothesen H1 sowie H1a und H1b müssen für ScorM2 abgelehnt werden. Des Weiteren ist die Hypothese H2 abzulehnen. Das *ElFB* ist weder hinsichtlich ScorM2 bzw. ScorM3 noch die Entwicklung der Mapqualität signifikant besser als das *EiFB*.

In Bezug auf *ScorM2* ist das *FB* *keinem FB* mit einem mittleren Effekt ($d=-0,47$) tendenziell unterlegen, wobei dies durch einen mittleren negativen Effekt des *EiFB* ($d=-0,55$) sowie einen kleinen negativen Effekt des *ElFB* ($d=-0,42$) gegenüber *keinem FB* zum Ausdruck kommt. Das *ElFB* zeigt gegenüber dem *EiFB* keinen Effekt ($d=0,00$). Im 3. Mapping kehren sich die Unterschiede um. Das *FB* ist *keinem FB* mit einem großen signifikanten Effekt ($d=1,08$) überlegen, wobei dies aus großen Effekten des *EiFB* ($d=0,85$) sowie des *ElFB* ($d=1,40$) gegenüber *keinem FB* resultiert. Das *ElFB* ist dem *EiFB* mit einem mittleren Effekt von $d=0,66$ tendenziell überlegen. Für die Entwicklung der Mapqualität über die Zeit zeigt sich ein großer signifikanter Effekt des *FB* gegenüber *keinem FB* ($d=0,79$), wobei dies insbesondere auf einen großen signifikanten Effekt des *ElFB* gegenüber *keinem FB* ($d=1,03$) zurückzuführen ist. Das *EiFB* ist *keinem FB* mit einem mittleren signifikanten Effekt überlegen ($d=0,69$). Des Weiteren zeigt sich ein tendenzieller Vorteil des *ElFB* gegenüber *EiFB* mit einem mittleren Effekt von $d=0,51$ (vgl. Kapitel 5.1.2.1).

Einfluss der individuellen Eingangsvoraussetzungen und weiteren Faktoren auf die Qualität der Maps (H3-H22)

Insgesamt zeigt sich ein signifikanter Einfluss der Faktoren *IN* und *H* auf die Qualität des 2. Maps, auf die Qualität des 3. Maps haben sie keinen Einfluss. Die weiteren Faktoren *VW*, *SF*, und *UF* haben dagegen keinen signifikanten Einfluss auf die Qualität des 2. und 3. Maps. Aus diesem Grund wird die Analyse der Hypothesen zu den Variablen *VW* (H3-H6), *SF* (H7-H10) und *UF* (H11-H14) nicht weiter verfolgt.

Die paarweisen Vergleiche der Gruppen zeigen, dass sich die Effekte durch das Herausparsialisieren der Faktoren insgesamt hinsichtlich *ScorM2* vergrößern. Hierbei vergrößert sich insbesondere der Effekt von *keinem FB* gegenüber *EiFB*, der jetzt signifikant wird. Des Weiteren ist das *EiFB* dem *EiFB* nun geringfügig überlegen. Für *ScorM3* lassen sich nur geringfügige Veränderungen ableiten. Der Effekt des *EiFB* gegenüber *keinem FB* ist nicht mehr länger signifikant und verringert sich etwas.

Demnach müssen unter Einbezug des Einflusses der Faktoren *VW*, *SF*, *UF*, *IN* und *H* insgesamt die Hypothesen 1a für *ScorM3* abgelehnt werden. Für weitere Analysen wurde lediglich der Einfluss der Faktoren *IN* und *H* auf die Qualität des 2. Maps näher betrachtet, da sich hier ein signifikanter Einfluss zeigt.

Um genauere Aussagen hinsichtlich der Hypothesen H15 bis H22 treffen zu können und somit den Einfluss der Faktoren *IN* und *H* auf die Lernwirksamkeit des Feedbacks zu überprüfen, wurden die signifikanten Einflüsse von *IN* und *H* auf *ScorM2* mittels Dichotomisierung dieser Variablen am Median (Einteilung in niedrig und hoch) genauer untersucht.

Bei *IN niedrig* ist *kein FB* nicht signifikant besser als *FB*, wobei sich das *EiFB* und *kein FB* nicht signifikant voneinander unterscheiden, das *EiFB* der Gruppe *kein FB* jedoch signifikant unterlegen ist. Das *EiFB* und das *EiFB* unterscheiden sich nicht signifikant. Demnach müssen die Hypothesen H15, H15a sowie H16 abgelehnt werden. H15b kann für *ScorM2* angenommen werden. Probanden, welche das Concept Mapping als weniger interessant und nützlich erachten und *kein FB* erhalten, sind signifikant besser als Probanden der Gruppe *EiFB* mit *IN niedrig*.

Hinsichtlich *IN hoch* zeigen die Paarvergleiche keine signifikanten Gruppenunterschiede. Daher müssen die Hypothesen H17, H17a, H17b und H18 abgelehnt werden. Das *FB* ist nicht signifikant besser als *kein FB*, wobei weder das *EiFB* noch das *EiFB* der Gruppe

kein FB signifikant überlegen sind. *ElFB* und *EiFB* unterscheiden sich ebenfalls nicht signifikant.

Tendenziell ist das *FB* bei *IN niedrig* *keinem FB* mit einem mittleren Effekt unterlegen ($d=-0,68$), was sich insbesondere durch einen großen signifikanten negativen Effekt des *ElFB* gegenüber *keinem FB* zeigt ($d=-1,29$). Das *EiFB* ist *keinem FB* mit einem kleinen Effekt tendenziell unterlegen ($d=-0,39$). Das *ElFB* zeigt gegenüber dem *EiFB* einen mittleren negativen Effekt ($d=-0,54$). Bei *IN hoch* zeigt sich ein ähnliches Bild. Das *FB* ist *keinem FB* tendenziell mit einem kleinen Effekt unterlegen ($d=-0,38$), wobei dies hier auf einen großen negativen Effekt des *EiFB* gegenüber *keinem FB* zurückzuführen ist ($d=-0,85$). Das *ElFB* zeigt gegenüber *keinem FB* keinen Effekt ($d=-0,11$). Das *ElFB* ist dem *EiFB* mit einem kleinen Effekt tendenziell überlegen ($d=0,38$).

Es zeigt sich weiterhin, dass bei *keinem FB* unabhängig vom Niveau für *IN* die Qualität des 2. Maps am höchsten ist. Beim *ElFB* zeigt sich der Einfluss von *IN_nh* deutlich ($d=1,02$). Probanden mit *IN niedrig* erzielen eine deutlich geringere Qualität im 2. Map als Probanden mit *IN hoch*. Beim *EiFB* zeigen sich nur geringfügige Qualitätsunterschiede in Abhängigkeit vom Niveau für *IN*.

Hinsichtlich *H niedrig* und *H hoch* zeigen die Paarvergleiche keine signifikanten Gruppenunterschiede. Daher müssen die Hypothesen H19, H19a, H19b, H20; H21, H21a, H21b und H22 abgelehnt werden. Für *H niedrig* ist *kein FB* nicht signifikant besser als *FB*, wobei *kein FB* weder dem *EiFB* noch dem *ElFB* signifikant überlegen ist. Das *EiFB* ist nicht signifikant besser als das *ElFB*. Für *H hoch* ist das *FB* nicht signifikant besser als *kein FB*, wobei weder das *EiFB* noch das *ElFB* *keinem FB* überlegen sind. Das *ElFB* ist nicht signifikant besser als *EiFB*.

Für *H niedrig* und *H hoch* zeigen sich tendenzielle Vorteile von *keinem FB* gegenüber dem *FB* ($d=-0,55$; $d=-0,44$), sowie gegenüber dem *EiFB* ($d=-0,64$; $d=-0,43$) und dem *ElFB* ($d=-0,48$; $d=-0,48$). Das *ElFB* zeigt gegenüber dem *EiFB* unabhängig vom Niveau für *H* keine Effekte ($d=0,11$; $d=-0,06$).

Es zeigt sich auch hier, dass bei *keinem FB* unabhängig vom Niveau für *H* die Qualität des 2. Maps am höchsten ist. Insgesamt erstellen die Probanden der Gruppen *kein FB*, *EiFB* und *ElFB* mit *H niedrig* qualitativ schlechtere Maps im 2. Mapping als Probanden mit *H hoch* (vgl. Kapitel 5.1.2.2).

Die Ergebnisse in Bezug auf die Hypothesen werden in nachfolgender Tabelle 70 noch einmal zusammenfassend veranschaulicht.

Hierbei werden die Hypothesen anhand der in Kapitel 5.1.2 durchgeführten Analysen bewertet. Des Weiteren werden die Ergebnisse der paarweisen Vergleiche der Gruppen für ScorM2, ScorM3 unter Berücksichtigung von ScorM1 als Kovariate sowie die Entwicklung der Mapqualität anhand der Effektstärken dargestellt, um Tendenzen hinsichtlich der Hypothesen ableiten zu können.

Tabelle 70: Ergebnisse der Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps - Prüfung der Hypothesen

Hypothese	Inhalt	Prüfung der Hypothesen	Effektstärken		
			ScorM2	ScorM3	Entwicklung Map-Qualität
H1	FB>kein FB	✓ bzw. X (2. CM)	- 0,47	1,08*	0,79*
H1a	EiFB>kein FB	✓ bzw. X (2. CM)	- 0,55	0,85*	0,69*
H1b	EIFB>kein FB	✓ bzw. X (2. CM)	- 0,42	1,40*	1,03*
H2	EIFB>EiFB	X	0,00	0,66	0,51
Bewertung der Hypothesen unter Herauspartialisierung der Einflussfaktoren gesamt (ScorM1, SF, VW, H, IN, UF)					
H1	FB>kein FB	X bzw. ✓ (3. CM)	- 0,63	1,08*	
H1a	EiFB>kein FB	X	- 0,61	0,73	
H1b	EIFB>kein FB	X bzw. ✓ (3. CM)	- 0,84*	1,38*	
H2	EIFB>EiFB	X	- 0,17	0,44	
Einfluss des IN					
IN niedrig					
H15	FB<kein FB	X	- 0,68		
H15a	EiFB<kein FB	X	- 0,39		
H15b	EIFB<kein FB	✓	- 1,29*		
H16	EIFB<EiFB	X	- 0,54		
IN hoch					
H17	FB>kein FB	X	- 0,38		
H17a	EiFB>kein FB	X	- 0,85		
H17b	EIFB>kein FB	X	- 0,11		
H18	EIFB>EiFB	X	0,38		
Einfluss von H					
H niedrig					
H19	FB<kein FB	X	- 0,55		
H19a	EiFB<kein FB	X	- 0,64		
H19b	EIFB<kein FB	X	- 0,48		
H20	EIFB<EiFB	X	0,11		
H hoch					
H21	FB>kein FB	X	- 0,44		
H21a	EiFB>kein FB	X	- 0,43		
H21b	EIFB>kein FB	X	- 0,48		
H22	EIFB>EiFB	X	- 0,06		

* signifikant auf dem Niveau $\alpha = 0,05$

✓ Hypothese angenommen

X Hypothese abgelehnt

kein Effekt, kleiner Effekt, *mittlerer Effekt*, **großer Effekt**

5.1.3 Einfluss der Qualität der Maps auf den Wissenserwerb

5.1.3.1 Einfluss der Qualität des 1., 2. und 3. Maps auf den 1. bzw. 2. Nachtest

Um zu analysieren, inwieweit die Qualität der Maps einen Einfluss auf den Wissenserwerb hat, werden zunächst zwei univariate einfaktorielle ANCOVAs mit der UV Gruppe, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 1. bzw. 2. *Nachtest* durchgeführt. Zunächst gilt es, die Voraussetzungen zu prüfen. Die Messwerte sollten in allen Teilstichproben normalverteilt sein. Des Weiteren gilt es, die Varianzhomogenität zu prüfen. Diese Voraussetzungen sind erfüllt.

Tabelle 71: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 1. *Nachtest*

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable 1. Nachtest (über alle Aufgaben)**</u>							
ScorM1	12,016	1	12,016	0,882	0,352	0,017	0,26
ScorM2	153,689	1	153,689	11,276	0,002*	0,184	0,95*
ScorM3	77,145	1	77,145	5,660	0,021*	0,102	0,67*
Gruppe	12,935	2	6,468	0,475	0,625	0,019	0,28

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,353

Die Ergebnisse der einfaktoriellen ANCOVA für den 1. Nachtest sind in Tabelle 71 dargestellt. Die einfaktorielle ANCOVA für den 1. Nachtest zeigt einen signifikanten Einfluss der Kovariaten *ScorM2* ($F=11,276$; $p=0,002$) und *ScorM3* ($F=5,660$; $p=0,021$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=0,475$; $p=0,625$). Das Neutralisieren der Kovariaten führt zu einer Verringerung der Gruppenunterschiede von $d=0,64$ auf $d=0,28$ (vgl. Kapitel 5.1.1.2, Tabelle 37). Die Unterschiede zwischen den Gruppen hinsichtlich des 1. Nachtests sind somit weniger auf die Art des Feedbacks, sondern auf den Einfluss der Qualität des 2. Maps ($d=0,95$) und 3. Maps ($d=0,67$) zurückzuführen. Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von $0,059^{98}$ auf 0,353.

⁹⁸ vgl. Kapitel 5.1.1.2, Tabelle 37; ANOVA mit der UV Gruppe und der AV 1. Nachtest

Tabelle 72: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV *2. Nachtest*

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable 2. Nachtest (über alle Aufgaben)**</u>							
ScorM1	6,327	1	6,327	0,274	0,603	0,006	0,16
ScorM2	9,047	1	9,047	0,392	0,534	0,009	0,19
ScorM3	122,031	1	122,031	5,293	0,026*	0,110	0,70*
Gruppe	9,287	2	4,644	0,201	0,818	0,009	0,19

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat = 0,121

Die Ergebnisse der einfaktoriellen ANCOVA für den 2. Nachtest sind in Tabelle 72 dargestellt. Die einfaktorielle ANCOVA für den 2. Nachtest zeigt einen signifikanten Einfluss der Kovariate *ScorM3* ($F=5,293$; $p=0,026$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=0,201$; $p=0,818$). Das Neutralisieren der Kovariaten führt zu einer Verringerung der Gruppenunterschiede von $d=0,38$ auf $d=0,19$ (vgl. Kapitel 5.1.1.2, Tabelle 37). Die Unterschiede zwischen den Gruppen hinsichtlich des 2. Nachtests sind somit weniger auf die Art des Feedbacks, sondern auf den Einfluss der Qualität des 3. Maps ($d=0,70$) zurückzuführen. Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von $-0,007^{99}$ auf $0,121$.

Um den Einfluss der signifikanten Kovariaten näher zu untersuchen, werden *ScorM2* und *ScorM3* für weitere Analysen am Median dichotomisiert und zwei Gruppen, niedrig und hoch, gebildet.

5.1.3.2 Einfluss von niedriger bzw. hoher Qualität des 2. Maps auf den Wissenserwerb

Zur Analyse von *ScorM2* werden zunächst zwei univariate zweifaktorielle ANOVAs mit den UVs *Gruppe* und *ScorM2 niedrig bzw. hoch* (*ScorM2_nh*¹⁰⁰) sowie der AV *1.* bzw. *2. Nachtest* durchgeführt. Anschließend folgen zwei dreifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit* (zwei- und dreifach gestuft), *Gruppe* und *ScorM2_nh* sowie der AV *Gruppenmittelwerte Wissenstest*. Die Voraussetzungen der Normalverteilung, der Varianzhomogenität sowie der Sphärizität sind gegeben.

⁹⁹ vgl. Kapitel 5.1.1.2, Tabelle 37; ANOVA mit der UV *Gruppe* und der AV *2. Nachtest*

¹⁰⁰ *ScorM2 niedrig* beinhaltet dabei alle Probanden mit einem Wert für *ScorM2* $\leq 0,1538840025$ (Median) und *ScorM2 hoch* umfasst alle Probanden mit einem Wert für *ScorM2* $> 0,1538840025$.

Die Ergebnisse der zweifaktoriellen Analysen in Tabelle 73 zeigen keine signifikanten Haupteffekte für den Faktor *Gruppe* hinsichtlich der Ergebnisse im 1. Nachtest ($F=2,776$; $p=0,072$) und 2. Nachtest ($F=0,832$; $p=0,442$). Hierbei zeigt sich ein mittlerer Effekt für den 1. Nachtest ($d=0,67$) und ein kleiner Effekt für den 2. Nachtest ($d=0,39$). Der Haupteffekt für den Faktor *ScorM2_nh* ist für den 1. Nachtest signifikant ($F=4,637$ $p=0,036$), für den 2. Nachtest dagegen nicht ($F=2,671$; $p=0,759$). Hierbei zeigt sich eine mittlere Effektstärke für den 1. Nachtest ($d=0,61$), für den 2. Nachtest ist kein Effekt erkennbar ($d=0,09$). Der Interaktionseffekt zwischen *ScorM2_nh* und *Gruppe* bleibt sowohl für den 1. Nachtest ($F=0,854$; $p=0,432$) als auch für den 2. Nachtest ($F=0,157$; $p=0,855$) mit kleinen Effekten nicht signifikant ($d=0,37$; $d=0,17$).

Tabelle 73: Ergebnisse der zweifaktoriellen ANOVAs mit der UV *Gruppe* und *ScorM2_nh* sowie der AV 1. *Nachtest* bzw. 2. *Nachtest*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über alle Aufgaben)</u>							
Gruppe	103,274**	2	51,637	2,776	0,072	0,100	0,67
ScorM2_nh	86,242	1	86,242	4,637	0,036*	0,085	0,61*
Gruppe x ScorM2_nh	31,769	2	15,885	0,854	0,432	0,033	0,37
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über alle Aufgaben)</u>							
Gruppe	46,594***	2	23,297	0,832	0,442	0,037	0,39
ScorM2_nh	2,671	1	2,671	0,095	0,759	0,002	0,09
Gruppe x ScorM2_nh	8,813	2	4,406	0,157	0,855	0,007	0,17

** korrigiertes R-Quadrat: 0,117

*** korrigiertes R-Quadrat: -0,067

In der Abbildung 30 und der Abbildung 31 sind die Ergebnisse graphisch dargestellt.

Um Aussagen hinsichtlich der Hypothese H33 bis H39 treffen zu können und somit den Einfluss der Qualität der Maps auf die Lernwirksamkeit des Feedbacks zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs jeweils separat für *ScorM2 niedrig* und *hoch* durchgeführt.

Die Ergebnisse der ANOVAs für die paarweisen Vergleiche sowie die Effektstärken können den Anhängen 5.1.3.-1 und 5.1.3.-2 entnommen werden.

Insgesamt zeigt sich der Einfluss der Qualität des 2. Maps auf den 1. Nachtest deutlich. Probanden, welche Concept Maps mit einer niedrigen Qualität erstellen, erzielen im

1. Nachtest deutlich schlechtere Ergebnisse als Probanden, welche qualitativ bessere Maps erstellen. Im 2. Nachtest sind die Unterschiede nur gering ausgeprägt.

Im 1. Nachtest ist das *FB* gegenüber *keinem FB* bei Probanden mit *ScorM2_niedrig* mit einem geringen Effekt ($d=0,16$) überlegen. Hierbei ist das *EiFB* *keinem FB* geringfügig unterlegen ($d=-0,26$), wohingegen das *EIFB* *keinem FB* mit einem mittleren Effekt ($d=0,63$) überlegen ist. Das *EIFB* zeigt einen großen signifikanten Effekt gegenüber dem *EiFB* ($d=1,03$). Eine hohe Qualität der 2. Maps (*ScorM2 hoch*) führt im 1. Nachtest zu einem großen Effekt des *FB* gegenüber *keinem FB* ($d=0,77$), wobei sowohl das *EiFB* ($d=0,72$) als auch das *EIFB* ($d=0,85$) *keinem FB* mit einem mittleren bzw. hohen Effekt überlegen sind. Das *EIFB* zeigt gegenüber dem *EiFB* einen kleinen Effekt ($d=0,24$).

Der Einfluss von *ScorM2_nh* auf den 1. Nachtest ist insbesondere bei den Gruppen *EiFB* und *EIFB* erkennbar. Die Probanden dieser Gruppen, welche Maps mit einer niedrigeren Qualität erstellen, erzielen deutlich geringere Ergebnisse im 1. Nachtest (11,18 bzw. 14,82 Punkte) als Probanden, welche Maps mit einer höheren Qualität erstellen (15,86 bzw. 17,00 Punkte). Für die Gruppe *EiFB* zeigt sich dabei eine signifikante hohe Effektstärke von $d=1,25$ für die Gruppe *EIFB* eine mittlere Effektstärke von $d=0,49$. Probanden der Gruppe *kein FB* erzielen ähnliche Ergebnisse im 1. Nachtest, unabhängig vom Niveau für *ScorM2* (12,25 bzw. 13,08 Punkte) mit einer geringen Effektstärke von $d=0,35$.

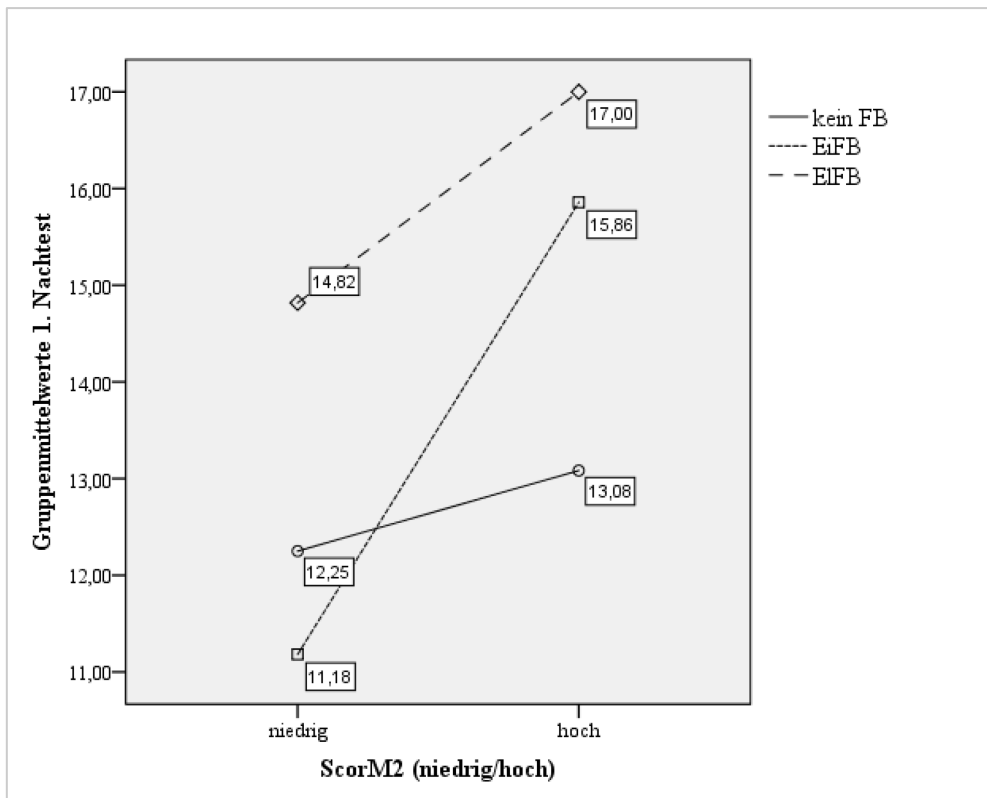


Abbildung 30: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit vom Niveau für ScorM2

Im 2. Nachttest führt *ScorM2 niedrig* zu einem geringen Vorteil des *FB* gegenüber *keinem FB* ($d=0,44$), wobei dies insbesondere auf einen mittleren Effekt des *EIFB* gegenüber *keinem FB* ($d=0,71$) zurückzuführen ist, das *EiFB* der Gruppe *kein FB* nur mit einem kleinen Effekt ($d=0,24$) überlegen ist. Das *EIFB* zeigt gegenüber dem *EiFB* einen mittleren Effekt ($d=0,5$). Bei Probanden mit *ScorM2 hoch* ist kein Vorteil des *FB* insgesamt bzw. des *EiFB* gegenüber *keinem FB* erkennbar ($d=0,06$ bzw. $d=-0,11$). Es zeigt sich lediglich ein kleiner Effekt des *EIFB* gegenüber *keinem FB* ($d=0,18$). Das *EIFB* ist dem *EiFB* mit einem kleinen Effekt überlegen ($d=0,26$).

Im 2. Nachttest zeigt sich bei allen drei Gruppen ein nur geringer Einfluss des Niveaus für *ScorM2* (*kein FB*: $d=0,19$; *EiFB*: $d=0,06$; *EIFB*: $d=-0,09$). Probanden mit *ScorM2 niedrig* der Gruppe *kein FB* erzielen geringfügig geringere Ergebnisse als Probanden mit *ScorM2 hoch* (10,67 vs. 12,25 Punkte). Bei den Gruppen *EiFB* und *EIFB* sind die Ergebnisse für beide Faktorstufen vergleichbar (*EiFB*: 11,34 bzw. 11,68 Punkte; *EIFB*: 13,78 bzw. 13,28 Punkte). Interessant ist an dieser Stelle, dass die Probanden der Gruppe *EIFB* mit *ScorM2 niedrig* im 2. Nachttest sogar geringfügig bessere Ergebnisse erzielen (13,78 Punkte) als die Probanden der Gruppe *EIFB* mit *ScorM2 hoch* (13,28 Punkte).

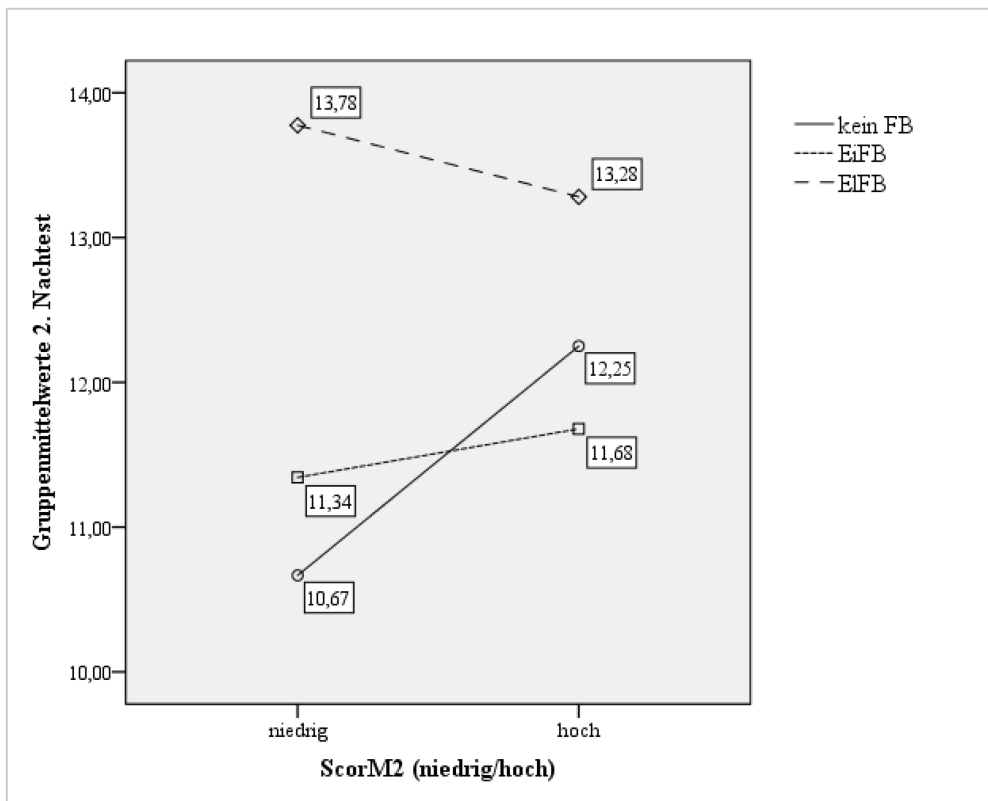


Abbildung 31: Graphische Darstellung der Gruppenmittelwerte des 2. Nachttests in Abhängigkeit vom Niveau für ScorM2

Die Ergebnisse der dreifaktoriellen ANOVA mit Messwiederholung sind in nachfolgender Tabelle 74 dargestellt.

Tabelle 74: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs *Gruppe* und *ScorM2_nh* sowie der AV *Gruppenmittelwerte Wissenstest*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)							
Zeit _{VT-1.NT}	1792,864	1	1792,864	251,222	0,000*	0,834	4,48*
Zeit _{VT-1.NT} x Gruppe	13,905	2	6,953	0,974	0,385	0,038	0,40
Zeit _{VT-1.NT} x ScorM2_nh	20,604	1	20,604	2,887	0,096	0,055	0,48
Zeit _{VT-1.NT} x Gruppe x ScorM2_nh	24,219	2	12,110	1,697	0,194	0,064	0,52
Zeit _{VT-1.NT-2.NT}	1757,691	2	878,845	122,108	0,000*	0,740	3,37*
Zeit _{VT-1.NT-2.NT} x Gruppe	15,339	4	3,835	0,533	0,712	0,024	0,31
Zeit _{VT-1.NT-2.NT} x ScorM2_nh	21,727	2	10,864	1,509	0,227	0,034	0,38
Zeit _{VT-1.NT-2.NT} x Gruppe x ScorM2_nh	17,186	4	4,297	0,597	0,666	0,027	0,33

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt können keine signifikanten Interaktionseffekte für die Faktoren $Zeit_{VT-1.NT}$ und $ScorM2_nh$ ($F=0,974$; $p=0,385$) sowie $Zeit_{VT-1.NT-2.NT}$ und $ScorM2_nh$ ($F=1,509$; $p=0,227$) nachgewiesen werden. Hierbei zeigt sich ein mittlerer bzw. geringer Effekt von $d=0,48$ und $d=0,31$. Darüber hinaus sind die Interaktionseffekte zwischen den Faktoren $Zeit_{VT-1.NT}$, $ScorM2_nh$ und $Gruppe$ ($F=1,697$; $p=0,194$) sowie $Zeit_{VT-1.NT-2.NT}$, $ScorM2_nh$ und $Gruppe$ ($F=0,597$; $p=0,666$) nicht signifikant. Die Effekte sind mit $d=0,52$ bzw. $d=0,33$ mittel bzw. gering. Schüler mit einer hohen Qualität für das 2. Map unterscheiden sich nicht signifikant von den Schülern mit einer niedrigen Qualität für das 2. Map hinsichtlich des Wissenszuwachses, auch unabhängig davon, ob und welches Feedback sie erhielten. Diese Ergebnisse lassen sich aus der Abbildung 32 ablesen. Die Veränderungen zwischen Vor- und 1. Nachtest sowie zwischen 1. und 2. Nachtest verlaufen jeweils fast parallel.

Um Aussagen hinsichtlich der Hypothese H33 bis H39 treffen zu können, werden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung jeweils separat für $ScorM2$ *niedrig* und *hoch* durchgeführt. Die Ergebnisse der ANOVAs mit Messwiederholung für die paarweisen Vergleiche jeweils separat für $ScorM2$ *niedrig* und $ScorM2$ *hoch* sowie die Effektstärken können den Anhängen 5.1.3-3 und 5.1.3-4 entnommen werden.

Vom Vortest zum 1. Nachtest zeigt FB gegenüber *keinem* FB bei Probanden mit $ScorM2$ *niedrig* einen kleinen negativen Effekt ($d=-0,23$), wobei sich $EiFB$ und *kein* FB nicht unterscheiden ($d=-0,06$), das $EiFB$ *keinem* FB jedoch mit einem mittleren Effekt unterlegen ist ($d=-0,47$). Das $EiFB$ ist dem $EiFB$ mit einem mittleren Effekt ($d=0,45$) überlegen. Bei Probanden mit $ScorM2$ *hoch* zeigt sich ein gegensätzliches Bild. Das FB ist *keinem* FB mit einem signifikanten hohen Effekt ($d=0,84$) überlegen, wobei sowohl das $EiFB$ ($d=0,76$) als auch das $EiFB$ ($d=0,94$) gegenüber *keinem* FB hohe Effekte zeigen. Das $EiFB$ ist dem $EiFB$ mit einem kleinen Effekt überlegen ($d=0,20$).

Langfristig zeigt sich bei Probanden mit $ScorM2$ *niedrig* kein Einfluss des FB gegenüber *keinem* FB ($d=0,06$), wobei das $EiFB$ einen kleinen Effekt ($d=0,23$) und das $EiFB$ keinen Effekt ($d=-0,09$) gegenüber *keinem* FB zeigen. Das $EiFB$ ist dem $EiFB$ geringfügig unterlegen ($d=-0,16$). Bei Probanden mit $ScorM2$ *hoch* ist das FB *keinem* FB mit einem mittleren Effekt unterlegen ($d=-0,61$), wobei sowohl das $EiFB$ ($d=-0,56$) als auch das $EiFB$ ($d=-0,63$) gegenüber *keinem* FB negative mittlere Effekte zeigen. Das $EiFB$ ist dem $EiFB$ mit einem kleinen negativen Effekt unterlegen ($d=-0,20$).

Beim separaten Vergleich der Gruppen in Abhängigkeit von *ScorM2_nh* zeigt sich, dass die Probanden der Gruppen *EiFB* und *EIFB* mit *ScorM2 niedrig* kurzfristig den Probanden mit *ScorM2 hoch* unterlegen ($d=0,96$; $d=0,76$), langfristig jedoch überlegen sind ($d=-0,70$; $d=-0,50$). Das Wissensniveau nähert sich bei den Probanden mit *ScorM2 niedrig* und *ScorM2 hoch* beim 2. Nachtest wieder an. Bei der Gruppe *kein FB* zeigt sich kurzfristig ein Vorteil der Probanden mit *ScorM2 niedrig* gegenüber den Probanden mit *ScorM2 hoch* ($d=-0,28$). Langfristig hat das Niveau von *ScorM2_nh* keinen Einfluss ($d=-0,09$).

Aus Abbildung 32 sowie Anhang 5.1.3-5 ist erkennbar, dass das Wissen der Probanden mit *ScorM2 niedrig* der Gruppe *EiFB* über drei Zeitpunkte am geringsten abnimmt (-6,69 Punkte; *kein FB*: -6,75 Punkte; *EIFB*: -10,25 Punkte). Bei Probanden mit *ScorM2 hoch* ist der Vorteil des *EiFB* und *EIFB* gegenüber *keinem FB* vom Vor- zum 1. Nachtest erkennbar. Langfristig können diese Gruppen das Wissensniveau jedoch nicht halten. Das Wissen der Gruppe *kein FB* nimmt über drei Zeitpunkte am geringsten ab (-7,55 Punkte; *EiFB*: 11,64 Punkte; *EIFB*: 11,75 Punkte).

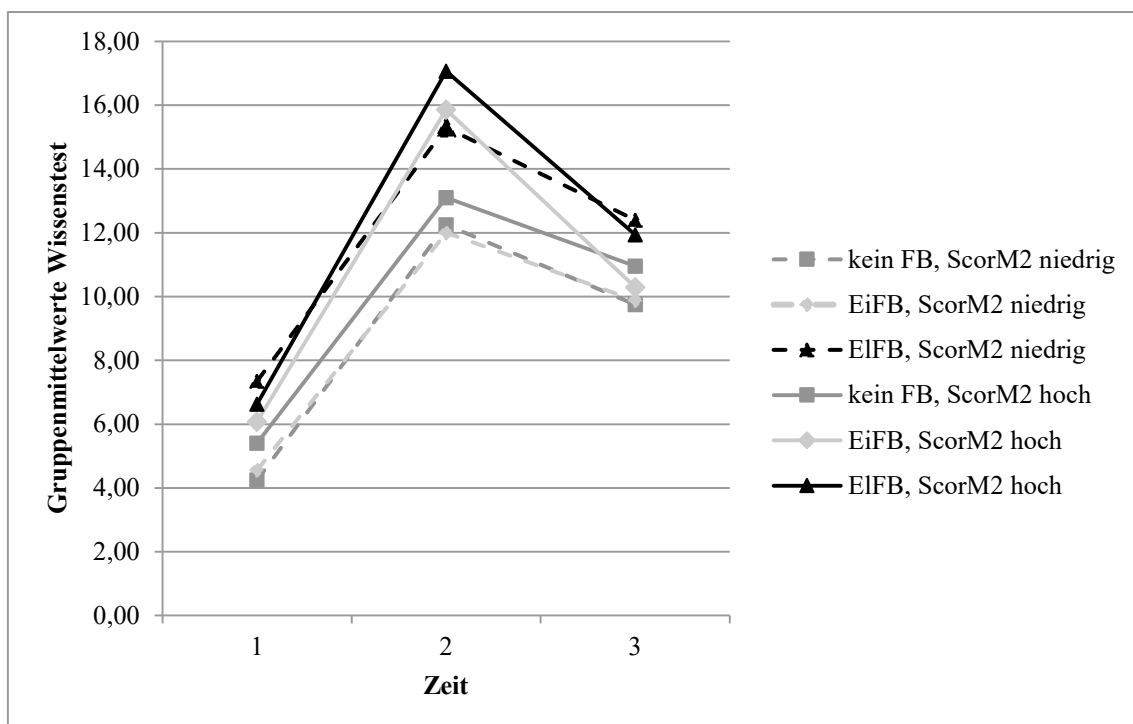


Abbildung 32: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *EIFB* über drei Messzeitpunkte in Abhängigkeit vom Niveau für *ScorM2*

5.1.3.3 Einfluss von niedriger bzw. hoher Qualität des 3. Maps auf den Wissenserwerb

Zur Analyse von ScorM3 werden zunächst zwei univariate zweifaktorielle ANOVAs mit den UVs *Gruppe* und *ScorM3 niedrig bzw. hoch* (*ScorM3_nh*¹⁰¹) sowie der AV 1. bzw. 2. *Nachtest* durchgeführt. Anschließend folgen zwei dreifaktorielle ANOVAs mit Messwiederholung mit den Faktoren *Zeit* (zwei- und dreifach gestuft), *Gruppe* und *ScorM3_nh* sowie der AV *Gruppenmittelwerte Wissenstest*. Die Voraussetzungen der Normalverteilung, der Varianzhomogenität sowie der Sphärizität sind gegeben.

Tabelle 75: Ergebnisse der zweifaktoriellen ANOVAs mit der UV *Gruppe* und *ScorM3_nh* sowie der AV 1. *Nachtest* bzw. 2. *Nachtest*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte 1. Nachtest (über alle Aufgaben)</u>							
Gruppe	22,842**	2	11,421	0,582	0,563	0,023	0,31
ScorM3_nh	57,360	1	57,360	2,922	0,094	0,055	0,48
Gruppe x ScorM3_nh	17,098	2	8,549	0,436	0,649	0,017	0,26
<u>Abh. Variable Gruppenmittelwerte 2. Nachtest (über alle Aufgaben)</u>							
Gruppe	46,878***	2	23,439	0,923	0,405	0,041	0,41
ScorM3_nh	21,576	1	21,576	0,849	0,362	0,019	0,28
Gruppe x ScorM3_nh	101,607	2	50,803	2,000	0,148	0,085	0,61

** korrigiertes R-Quadrat: 0,068

*** korrigiertes R-Quadrat: 0,032

Die Ergebnisse der zweifaktoriellen Analysen in Tabelle 75 zeigen keine signifikanten Haupteffekte für den Faktor *Gruppe* hinsichtlich der Ergebnisse im 1. *Nachtest* ($F=0,582$; $p=0,563$) und 2. *Nachtest* ($F=0,923$; $p=0,405$). Hierbei zeigen sich kleine Effekte für den 1. *Nachtest* ($d=0,31$) und 2. *Nachtest* ($d=0,41$). Der Haupteffekt für den Faktor *ScorM3_nh* ist für den 1. *Nachtest* ($F=2,922$; $p=0,094$) und 2. *Nachtest* nicht signifikant ($F=0,849$; $p=0,362$). Hierbei zeigt sich eine mittlere Effektstärke für den 1. *Nachtest* ($d=0,48$), für den 2. *Nachtest* ist ein kleiner Effekt erkennbar ($d=0,28$). Der Interaktionseffekt zwischen *ScorM3_nh* und *Gruppe* bleibt sowohl für den 1. *Nachtest* ($F=0,436$; $p=0,649$) als auch für den 2. *Nachtest* ($F=2,000$; $p=0,085$) nicht signifikant mit einem

¹⁰¹ *ScorM3 niedrig* beinhaltet dabei alle Probanden mit einem Wert für *ScorM3* $\leq 0,0731150040$ (Median) und *ScorM3 hoch* umfasst alle Probanden mit einem Wert für *ScorM3* $> 0,0731150040$.

kleinen Effekt ($d=0,26$) für den 1. Nachtest und einem mittleren Effekt für den 2. Nachtest ($d=0,61$).

In der Abbildung 33 und Abbildung 34 sind die Ergebnisse graphisch dargestellt.

Um Aussagen hinsichtlich der Hypothese H33 bis H39 treffen zu können und somit den Einfluss der Qualität der Maps auf die Lernwirksamkeit des Feedbacks zu überprüfen, werden jeweils weitere einfaktorielle ANOVAs jeweils separat für *ScorM3 niedrig* und *hoch* durchgeführt. Die Ergebnisse der ANOVAs für die paarweisen Vergleiche sowie die Effektstärken können den Anhängen 5.1.3.-6 und 5.1.3.-7 entnommen werden.

Im 1. Nachtest ist das *FB* gegenüber *keinem FB* bei Probanden mit *ScorM3 niedrig* mit einem geringen Effekt ($d=0,21$) überlegen, wobei dies auf geringe Effekte des *EiFB* ($d=0,17$) sowie des *EIFB* ($d=0,29$) zurückzuführen ist. Das *EIFB* zeigt gegenüber dem *EiFB* keinen Effekt ($d=0,11$). Eine hohe Qualität der 3. Maps (*ScorM3 hoch*) führt im 1. Nachtest zu keinem Effekt des *FB* gegenüber *keinem FB* ($d=0,00$), wobei das *EiFB* *keinem FB* mit einem mittleren Effekt unterlegen ($d=-0,74$), das *EIFB* jedoch mit einem kleinen Effekt ($d=0,17$) überlegen ist. Das *EIFB* zeigt gegenüber dem *EiFB* einen mittleren Effekt ($d=0,61$).

Der Einfluss von *ScorM3_nh* auf den 1. Nachtest ist insbesondere bei den Gruppen *kein FB* und *EIFB* erkennbar. Die Probanden dieser Gruppen, welche Maps mit einer niedrigeren Qualität erstellen, erzielen deutlich geringere Ergebnisse im 1. Nachtest (12,04 bzw. 13,38 Punkte) als Probanden, welche Maps mit einer höheren Qualität erstellen (15,50 bzw. 16,41 Punkte). Hierbei werden Effektstärken von $d=0,75$ bzw. $d=0,55$ erzielt. Probanden der Gruppe *EiFB* haben ähnliche Ergebnisse im 1. Nachtest erreicht, unabhängig vom Niveau für *ScorM2* (12,79 bzw. 13,42 Punkte).

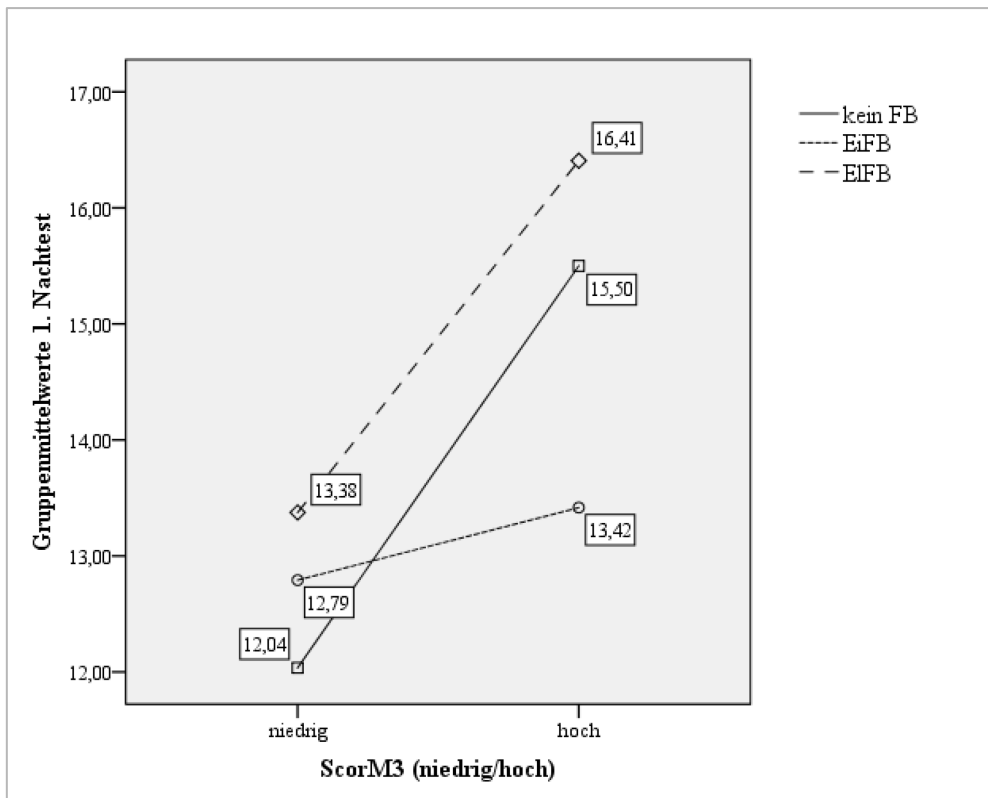


Abbildung 33: Graphische Darstellung der Gruppenmittelwerte des 1. Nachttests in Abhängigkeit vom Niveau für ScorM3

Im 2. Nachttest führt *ScorM3 niedrig* zu einem mittleren Effekt des *FB* gegenüber *keinem FB* ($d=0,44$), wobei dies insbesondere auf einen hohen Effekt des *EIFB* gegenüber *keinem FB* ($d=0,84$) zurückzuführen ist, das *EiFB* der Gruppe *kein FB* nur mit einem kleinen Effekt ($d=0,33$) überlegen ist. Das *EIFB* zeigt gegenüber dem *EiFB* einen kleinen Effekt ($d=0,40$). Bei Probanden mit *ScorM3 hoch* ist das *FB* der Gruppe *kein FB* mit einem mittleren Effekt unterlegen ($d=-0,56$). Hierbei ist sowohl das *EiFB* mit einem sehr großen Effekt ($d=-1,71$) sowie das *EIFB* mit einem mittleren Effekt ($d=-0,50$) *keinem FB* unterlegen. Das *EIFB* zeigt gegenüber dem *EiFB* einen kleinen Effekt ($d=0,40$).

Im 2. Nachttest zeigt sich bei der Gruppe *kein FB* ein großer Einfluss des Niveaus für *ScorM3*, Probanden mit *ScorM3 niedrig* erzielen signifikant geringere Ergebnisse als Probanden mit *ScorM3 hoch* (10,35 vs. 17,33 Punkte; $d=1,53$). Bei den Gruppen *EiFB* und *EIFB* sind die Ergebnisse für beide Faktorstufen vergleichbar (*EiFB*: 11,78 bzw. 10,95 Punkte mit $d=-0,17$; *EIFB*: 14,37 bzw. 13,45 Punkte mit $d=-0,11$). Interessant ist an dieser Stelle, dass die Probanden der beiden *FB*-Gruppen mit *ScorM3 niedrig* im 2. Nachttest sogar geringfügig bessere Ergebnisse erzielen als die Probanden mit *ScorM3 hoch*.

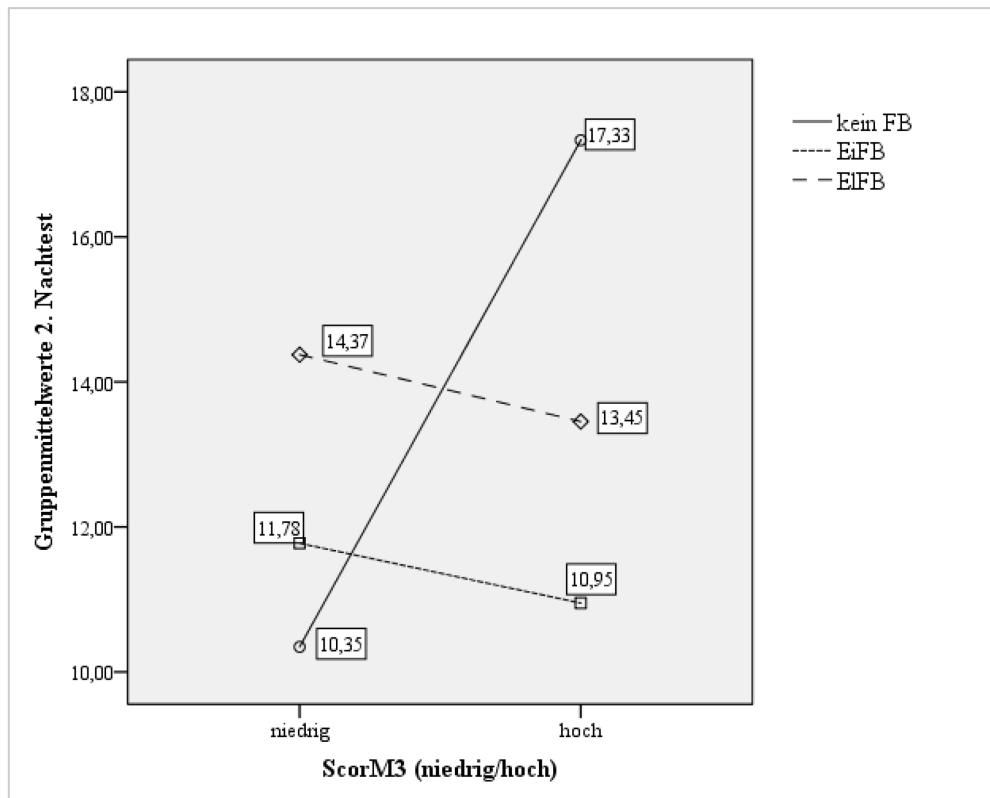


Abbildung 34: Graphische Darstellung der Gruppenmittelwerte des 2. Nachttests in Abhängigkeit vom Niveau für ScorM3

Die Ergebnisse der dreifaktoriellen ANOVA mit Messwiederholung sind in nachfolgender Tabelle 76 dargestellt.

Tabelle 76: Ergebnisse der dreifaktoriellen ANOVAs mit Messwiederholung mit den UVs Gruppe und ScorM3_nh sowie der AV Gruppenmittelwerte Wissenstest

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Gruppenmittelwerte Wissenstest (über alle Aufgaben)</u>							
Zeit _{VT-1.NT}	1293,996	1	1293,996	165,093	0,000*	0,768	3,64*
Zeit _{VT-1.NT} x Gruppe	1,783	2	0,891	0,114	0,893	0,005	0,14
Zeit _{VT-1.NT} x ScorM3_nh	9,743	1	9,743	1,243	0,270	0,024	0,31
Zeit _{VT-1.NT} x Gruppe x ScorM3_nh	4,126	2	2,063	0,263	0,770	0,010	0,20
Zeit _{VT-1.NT-2.NT}	987,122	2	493,561	69,905	0,000*	0,619	2,55*
Zeit _{VT-1.NT-2.NT} x Gruppe	21,807	4	5,452	0,772	0,546	0,035	0,38
Zeit _{VT-1.NT-2.NT} x ScorM3_nh	19,664	2	9,832	1,393	0,254	0,031	0,36
Zeit _{VT-1.NT-2.NT} x Gruppe x ScorM3_nh	36,946	4	9,237	1,308	0,273	0,057	0,49

* signifikant auf dem Niveau $\alpha = 0,05$

Insgesamt können keine signifikanten Interaktionseffekte für die Faktoren $Zeit_{VT-1,NT}$ und $ScorM3_nh$ ($F=1,243$; $p=0,270$) sowie $Zeit_{VT-1,NT-2,NT}$ und $ScorM3_nh$ ($F=1,393$; $p=0,254$) nachgewiesen werden. Hierbei zeigen sich kleine Effekte von $d=0,31$ und $d=0,36$. Darüber hinaus sind die Interaktionseffekte zwischen den Faktoren $Zeit_{VT-1,NT}$, $ScorM3_nh$ und $Gruppe$ ($F=0,263$; $p=0,770$) sowie $Zeit_{VT-1,NT-2,NT}$, $ScorM3_nh$ und $Gruppe$ ($F=1,308$; $p=0,273$) nicht signifikant. Die Effekte sind mit $d=0,20$ bzw. $d=0,49$ klein bzw. mittel. Schüler mit einer hohen Qualität für das 3. Map unterscheiden sich nicht signifikant von den Schülern mit einer niedrigen Qualität für das 3. Map hinsichtlich des Wissenszuwachses, auch unabhängig davon, ob und welches Feedback sie erhielten.

Um Aussagen hinsichtlich der Hypothese H33 bis H39 treffen zu können, werden jeweils weitere zweifaktorielle ANOVAs mit Messwiederholung jeweils separat für $ScorM3$ *niedrig* und *hoch* durchgeführt. Die Ergebnisse der ANOVAs mit Messwiederholung für die paarweisen Vergleiche jeweils separat für $ScorM3$ *niedrig* und $ScorM3$ *hoch* sowie die Effektstärken können den Anhängen 5.1.3-8 und 5.1.3-9 entnommen werden.

Vom Vortest zum 1. Nachtest zeigt FB gegenüber *keinem* FB bei Probanden mit $ScorM3$ *niedrig* einen kleinen Effekt ($d=0,16$), wobei sich $EiFB$ und *kein* FB nicht unterscheiden ($d=-0,06$), das $EiFB$ *keinem* FB jedoch mit einem kleinen Effekt überlegen ist ($d=0,18$). Das $EiFB$ erzielt gegenüber dem $EiFB$ keinen Effekt ($d=-0,09$). Bei Probanden mit $ScorM3$ *hoch* unterscheiden sich FB und *kein* FB nicht ($d=-0,09$), wobei sowohl das $EiFB$ ($d=-0,22$) als auch das $EiFB$ ($d=-0,18$) gegenüber *keinem* FB geringfügig unterlegen sind. Das $EiFB$ ist dem $EiFB$ mit einem kleinen Effekt überlegen ($d=0,38$).

Langfristig ist bei Probanden mit $ScorM3$ *niedrig* ein geringer Vorteil von *keinem* FB gegenüber FB erkennbar ($d=-0,19$), wobei sowohl das $EiFB$ ($d=-0,31$) als auch das $EiFB$ ($d=0,39$) mit einem kleinen Effekt unterlegen sind. Das $EiFB$ zeigt einen mittleren Effekt mit $d=0,57$ gegenüber dem $EiFB$. Bei Probanden mit $ScorM3$ *hoch* ist das FB *keinem* FB mit einem mittleren Effekt unterlegen ($d=-0,61$), wobei sowohl das $EiFB$ ($d=-1,00$) als auch das $EiFB$ ($d=-0,64$) gegenüber *keinem* FB negative mittlere Effekte zeigen. Das $EiFB$ ist dem $EiFB$ mit einem kleinen Effekt unterlegen ($d=-0,25$).

Beim separaten Vergleich der Gruppen in Abhängigkeit von $ScorM3_nh$ zeigt sich, dass die Probanden der Gruppen *kein* FB mit $ScorM3$ *niedrig* sowohl kurz- als auch langfristig den Probanden mit $ScorM3$ *hoch* unterlegen sind ($d=0,41$; $d=0,96$). Beim $EiFB$ hat das Niveau von $ScorM3_nh$ kurzfristig keinen Einfluss ($d=-0,06$), langfristig zeigt sich ein

geringer Vorteil der Probanden mit *ScorM3 hoch* gegenüber *ScorM3 niedrig* ($d=0,17$). Probanden der Gruppe *EIFB* mit *ScorM3 hoch* erzielen kurzfristig einen mittleren positiven Effekt ($d=0,46$), langfristig einen negativen mittleren Effekt ($d=-0,54$) gegenüber Probanden mit *ScorM3 niedrig*.

Aus Abbildung 35 sowie Anhang 5.1.3-10 ist erkennbar, dass der Wissenszuwachs der Probanden mit *ScorM3 niedrig* der Gruppe *EiFB* vom Vor- zum 1. Nachtest am höchsten ist (8,80 Punkte; *kein FB*: 7,27 Punkte; *EIFB*: 5,25 Punkte). Bei Probanden mit *ScorM3 hoch* zeigt sich bei der Gruppe *kein FB* kurzfristig der höchste Wissenszuwachs (10,17 Punkte; *EiFB*: 8,00 Punkte; *EIFB*: 9,53 Punkte). Langfristig nimmt das Wissen der Gruppe *kein FB* am geringsten ab, dies gilt sowohl für Probanden mit *ScorM3 niedrig* (-7,46 Punkte) als auch für Probanden mit *ScorM3 hoch* (-6,33 Punkte). Zu erkennen ist weiterhin, dass die Gruppe *kein FB* mit *ScorM3 hoch* ihr Wissensniveau vom 1. zum 2. Nachtest fast konstant halten kann, wohingegen das Wissen der Probanden mit *ScorM3 niedrig* vom 1. zum 2. Nachtest weiter abnimmt. Die Gruppe *EiFB* zeigt einen ähnlichen Verlauf hinsichtlich der Entwicklung des Wissens über die drei Zeitpunkte hinweg, sowohl für *ScorM3_niedrig* als auch für *ScorM3_hoch*. Beim *EIFB* profitieren kurzfristig die Probanden mit *ScorM3 hoch*, langfristig sind die Probanden mit *ScorM3 niedrig* den Probanden mit *ScorM3 hoch* überlegen, so dass sich beide Gruppen im 2. Nachtest insgesamt wieder annähern.

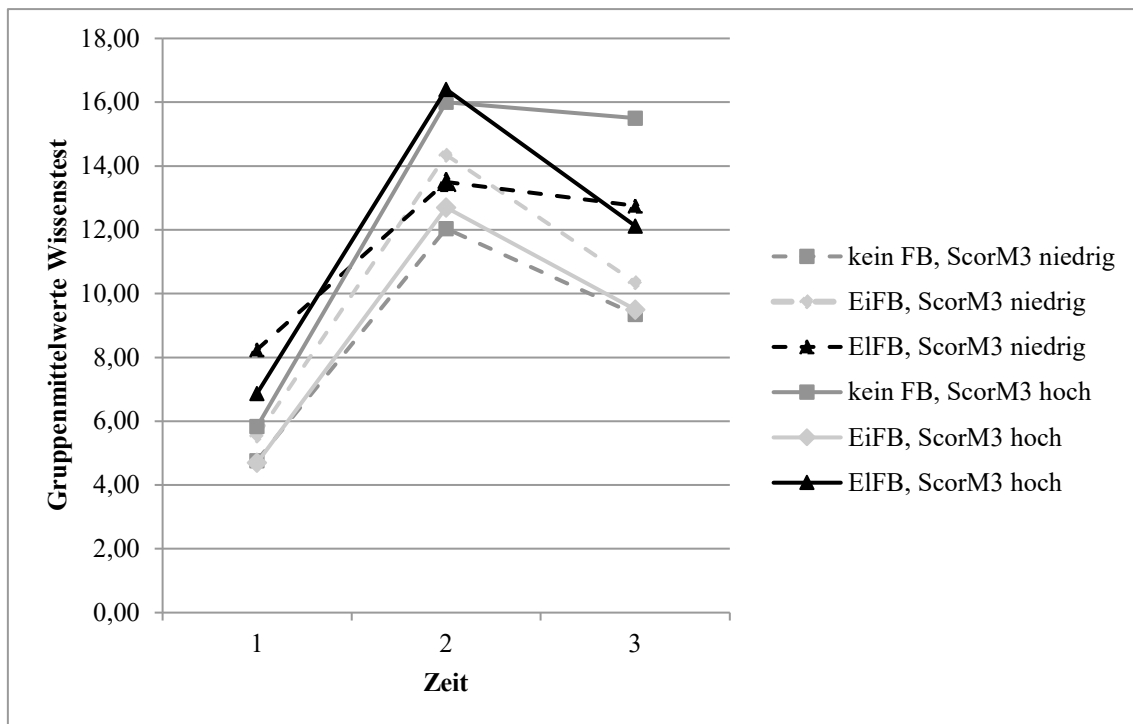


Abbildung 35: Graphische Darstellung der Gruppenmittelwerte des Wissenstests für die Gruppen *kein FB*, *EiFB* und *EiFB* über drei Messzeitpunkte in Abhängigkeit vom Niveau für *ScorM3*

5.1.3.4 Einfluss der Qualität des 1., 2. und 3. Maps auf den Wissenserwerb einfacher und komplexer Aufgaben

Nachfolgend wird der Einfluss der Qualität der Maps auf den Wissenserwerb einfacher bzw. komplexer Aufgaben untersucht. Hierzu werden je zwei univariate einfaktorielle ANCOVAs mit der UV Gruppe, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 1. bzw. 2. *Nachtest* für einfache bzw. komplexe Aufgaben durchgeführt. Zunächst gilt es, die Voraussetzungen zu prüfen. Die Messwerte sollten in allen Teilstichproben normalverteilt sein. Des Weiteren gilt es, die Varianzhomogenität zu prüfen. Diese Voraussetzungen sind erfüllt.

Die Ergebnisse der einfaktoriellen ANCOVA für die einfachen Aufgaben des 1. *Nachtests* sind in Tabelle 77 dargestellt. Die einfaktorielle ANCOVA für den 1. *Nachtest* zeigt einen signifikanten Einfluss der Kovariaten *ScorM2* ($F=10,330$; $p=0,002$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=1,305$; $p=0,280$). Das Neutralisieren der Kovariaten führt zu einer geringfügigen Verringerung der Gruppenunterschiede von $d=0,58$ auf $d=0,46$ (vgl. Kapitel 5.1.1.4, Tabelle 50). Die Unterschiede zwischen den Gruppen hinsichtlich der einfachen Aufgaben des 1. *Nachtests* sind somit nicht nur auf die Art des Feedbacks, sondern auch auf den Einfluss der Qualität des 2. Maps ($d=0,91$)

zurückzuführen. Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von 0,043¹⁰² auf 0,353.

Tabelle 77: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 1. *Nachtest für einfache Aufgaben*

Quelle	Quadrat- summe vom Typ III	df	Mittel der Quadrate	F- Wert	p	Partielles Eta- Quadrat	d
<u>Abh. Variable 1. Nachtest (über einfache Aufgaben)**</u>							
ScorM1	11,635	1	11,635	2,453	0,124	0,047	0,44
ScorM2	48,999	1	48,999	10,330	0,002*	0,171	0,91*
ScorM3	5,328	1	5,328	1,123	0,294	0,022	0,30
Gruppe	12,380	2	6,190	1,305	0,280	0,050	0,46

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,277

Die Ergebnisse der einfaktoriellen ANCOVA für die einfachen Aufgaben des 2. Nachtests sind in Tabelle 78 dargestellt. Die einfaktorielle ANCOVA für den 2. Nachtest zeigt keinen signifikanten Einfluss der Kovariaten. Der Gruppenunterschied bleibt dabei nicht signifikant ($F=0,201$; $p=0,818$). Das Neutralisieren der Kovariaten führt zu einer Verringerung der Gruppenunterschiede von $d=0,51$ auf $d=0,21$ (vgl. Kapitel 5.1.1.4, Tabelle 50). Die Aufnahme der Kovariaten führt zu einer Verbesserung der Modellgüte von 0,019¹⁰³ auf 0,053.

Tabelle 78: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 2. *Nachtest für einfache Aufgaben*

Quelle	Quadrat- summe vom Typ III	df	Mittel der Quadrate	F- Wert	p	Partielles Eta- Quadrat	d
<u>Abh. Variable 2. Nachtest (über einfache Aufgaben)**</u>							
ScorM1	1,644	1	1,644	0,267	0,608	0,006	0,16
ScorM2	7,751	1	7,751	1,260	0,268	0,028	0,34
ScorM3	7,499	1	7,499	1,219	0,276	0,028	0,34
Gruppe	2,957	2	1,478	0,240	0,787	0,011	0,21

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat = 0,053

Die Ergebnisse der einfaktoriellen ANCOVA für die komplexen Aufgaben des 1. Nachtests sind in Tabelle 79 dargestellt. Die einfaktorielle ANCOVA für den 1. Nachtest zeigt

¹⁰² vgl. Kapitel 5.1.1.4, Tabelle 50; ANOVA mit der UV Gruppe und der AV 1. Nachtest für einfache Aufgaben

¹⁰³ vgl. Kapitel 5.1.1.4, Tabelle 50; ANOVA mit der UV Gruppe und der AV 2. Nachtest für einfache Aufgaben

einen signifikanten Einfluss der Kovariaten *ScorM2* ($F=29,130$; $p=0,010$) sowie *ScorM3* ($F=41,924$; $p=0,002$). Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=0,077$; $p=0,993$). Das Neutralisieren der Kovariaten führt zu einer deutlichen Verringerung der Gruppenunterschiede von $d=0,60$ auf $d=0,00$ (vgl. Kapitel 5.1.1.4, Tabelle 54). Die Unterschiede zwischen den Gruppen hinsichtlich der komplexen Aufgaben des 1. Nachttests sind somit nicht nur auf die Art des Feedbacks, sondern auch auf den Einfluss der Qualität des 2. und 3. Maps ($d=0,76$ bzw. $d=0,91$) zurückzuführen. Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von $0,048^{104}$ auf $0,323$.

Tabelle 79: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 1. *Nachttest für komplexe Aufgaben*

Quelle	Quadrat- summe vom Typ III	df	Mittel der Quadrate	F- Wert	p	Partielles Eta- Quadrat	d
<u>Abh. Variable 1. Nachttest (über komplexe Aufgaben)**</u>							
ScorM1	0,003	1	0,003	0,001	0,978	0,000	0,00
ScorM2	29,130	1	29,130	7,174	0,010*	0,125	0,76*
ScorM3	41,924	1	41,924	10,325	0,002	0,171	0,91*
Gruppe	0,055	2	0,028	0,077	0,993	0,000	0,00

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,323

Die Ergebnisse der einfaktoriellen ANCOVA für die komplexen Aufgaben des 2. Nachttests sind in Tabelle 80 dargestellt. Die einfaktorielle ANCOVA für den 2. Nachttest zeigt einen signifikanten Einfluss der Kovariaten *ScorM3* ($F=8,507$; $p=0,006$) mit $d=0,89$. Der Gruppenunterschied bleibt jedoch nicht signifikant ($F=0,960$; $p=0,391$). Das Neutralisieren der Kovariaten führt zu einer Erhöhung der Gruppenunterschiede von $d=0,27$ auf $d=0,42$ (vgl. Kapitel 5.1.1.4, Tabelle 54). Die Aufnahme der Kovariaten führt zu einer deutlichen Verbesserung der Modellgüte von $-0,025^{105}$ auf $0,153$.

¹⁰⁴ vgl. Kapitel 5.1.1.4, Tabelle 54; ANOVA mit der UV Gruppe und der AV 1. Nachttest für komplexe Aufgaben

¹⁰⁵ vgl. Kapitel 5.1.1.4, Tabelle 54; ANOVA mit der UV Gruppe und der AV 2. Nachttest für komplexe Aufgaben

Tabelle 80: Ergebnisse der einfaktoriellen ANCOVA mit der UV *Gruppe*, den Kovariaten *ScorM1*, *ScorM2* und *ScorM3* sowie der AV 2. *Nachtest für komplexe Aufgaben*

Quelle	Quadrat- summe vom Typ III	df	Mittel der Quadrate	F- Wert	p	Partielles Eta- Quadrat	d
<u>Abh. Variable 2. Nachtest (über komplexe Aufgaben)**</u>							
ScorM1	1,521	1	1,521	0,187	0,667	0,004	0,13
ScorM2	0,050	1	0,050	0,006	0,938	0,000	0,00
ScorM3	69,027	1	69,027	8,507	0,006	0,165	0,89*
Gruppe	15,588	2	7,794	0,960	0,391	0,043	0,42

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat = 0,153

Um den Einfluss der signifikanten Kovariaten näher zu untersuchen, werden zweifaktorielle ANOVAs bzw. dreifaktorielle ANOVAs mit Messwiederholung mit den dichotomisierten Variablen *ScorM2_nh* und *ScorM3_nh* durchgeführt (vgl. Anhang 5.1.3.11).

Dabei ist es von besonderem Interesse, inwieweit sich die Gruppen hinsichtlich der Ergebnisse des 1. und 2. Nachtests sowie des Wissenszuwachses für einfache und komplexe Aufgaben in Abhängigkeit vom Niveau für *ScorM2* und *ScorM3* unterscheiden. Aus diesem Grund liegt der Fokus insbesondere auf der Analyse der Interaktionseffekte *Gruppe* und *ScorM2_nh* bzw. *ScorM3_nh* der ANOVAs und *Zeit*, *Gruppe* und *ScorM2_nh* bzw. *ScorM3_nh* für die ANOVAs mit Messwiederholung. Hierbei zeigen sich keine signifikanten Interaktionseffekte (vgl. Anhang 5.1.3.11). Nachfolgend werden anhand der graphischen Darstellung der Gruppenmittelwerte des Wissenstests sowie der Wissensentwicklungen für einfache und komplexe Aufgaben über drei Messzeitpunkte in Abhängigkeit vom Niveau für *ScorM2* und *ScorM3* Auffälligkeiten bzw. Unterschiede zur Analyse über alle Aufgaben dargestellt (vgl. Anhänge 5.1.3.12 und 5.1.3.13).

Für *ScorM2 niedrig* ist bei den *einfachen Aufgaben* kurzfristig der höchste Wissenszuwachs für *kein FB* erkennbar (4,17 Punkte; *EiFB*: 4,06 Punkte; *ElFB*: 3,95 Punkte), für *ScorM2 hoch* sind die FB-Gruppen *keinem FB* (4,10 Punkte) hinsichtlich des Wissenszuwachses überlegen (*EiFB*: 6,07 Punkte; *ElFB*: 5,06 Punkte). Über drei Zeitpunkte zeigt sich für *ScorM2 niedrig* ein Vorteil des *EiFB* (-4,31 Punkte) gegenüber *keinem FB* (-5,00 Punkte) und *ElFB* (-5,55 Punkte). Eine zusätzliche ANOVA mit Messwiederholung über drei Zeitpunkte zeigt hierbei einen Effekt von $d=0,26$ des *EiFB* gegenüber *keinem FB*. (vgl. Anhang 5.1.3.14). Für *ScorM2 hoch* ist *kein FB* (-3,35 Punkte) beiden FB-Gruppen überlegen (-6,57 bzw. -6,88 Punkte). Bei den *komplexen Aufgaben* hat das *ElFB* kurzfristig den höchsten Wissenszuwachs (4,00 bzw. 5,38 Punkte; *kein FB*: 3,83 bzw. 3,60

Punkte; *EiFB*: 3,38 bzw. 3,71 Punkte), sowohl bei *ScorM2 niedrig* als auch bei *ScorM2 hoch*. Langfristig zeigt sich ein Vorteil von *keinem FB* gegenüber den beiden *FB*-Gruppen für *ScorM2 niedrig* und *ScorM2 hoch* (-1,75 bzw. -4,20 Punkte; *EiFB*: -2,38 bzw. -5,07 Punkte; *ElFB*: -4,70 bzw. -4,88 Punkte).

Insgesamt zeigt sich bei den *FB*-Gruppen eine ähnliche Tendenz, wie bei der Analyse über alle Aufgaben, kurzfristig sind die Probanden mit *ScorM2 hoch* den Probanden mit *ScorM2 niedrig* überlegen, über drei Zeitpunkte zeigt sich ein gegensätzlicher Effekt.

Für *ScorM3* zeigt sich bei den *einfachen Aufgaben* lediglich kurzfristig für *ScorM3 niedrig* ein Vorteil des *EiFB* (5,45 Punkte; *kein FB*: 3,92 Punkte; *ElFB*: 2,25 Punkte) gegenüber den anderen Gruppen hinsichtlich des Wissenszuwachses. Für *ScorM3 hoch* vom Vor- zum 1. Nachtest (*kein FB*: 5,00 Punkte; *EiFB*: 4,10 Punkte; *ElFB*: 4,72 Punkte) sowie über drei Zeitpunkte (*kein FB*: -4,17 Punkte; *EiFB*: -5,20 Punkte; *ElFB*: -6,31 Punkte) und für *ScorM3 niedrig* über drei Zeitpunkte (*kein FB*: -3,92 Punkte; *EiFB*: -5,45 Punkte; *ElFB*: -4,75 Punkte) ist *kein FB* sowohl dem *EiFB* also auch dem *ElFB* hinsichtlich des Wissenszuwachses überlegen. Bei den *komplexen Aufgaben* unterscheiden sich für *ScorM3 niedrig* die Wissenszuwächse von *EiFB* und *keinem FB* kurzfristig nicht (3,35 Punkte). Bei *ScorM3 hoch* ist *kein FB* den *FB*-Gruppen überlegen (5,17 Punkte vs. 3,90 bzw. 4,81 Punkte). Langfristig zeigt sich ein Vorteil von *keinem FB* gegenüber den beiden *FB*-Gruppen für *ScorM3 niedrig* (-3,54 Punkte vs. -4,10 und -4,25 Punkte) und *ScorM3 hoch* (-2,17 Punkte vs. -2,70 und -4,84 Punkte).

Insgesamt zeigt sich beim *ElFB* eine ähnliche Tendenz wie bei der Analyse über alle Aufgaben. Probanden mit *ScorM3 hoch* sind den Probanden mit *ScorM3 niedrig* kurzfristig überlegen, über drei Zeitpunkte zeigt sich ein gegensätzlicher Effekt. Beim *EiFB* zeigt sich kurzfristig bei *einfachen Aufgaben* ein Vorteil der Probanden mit *ScorM3 niedrig* gegenüber denen mit *ScorM3 hoch*. Bei den *komplexen Aufgaben* ist kurzfristig, sowie bei den *einfachen* und *komplexen Aufgaben* langfristig, ein Vorteil von *ScorM3 hoch* gegenüber *ScorM3 niedrig* für Probanden der Gruppe *EiFB* erkennbar.

5.1.3.5 Zusammenfassung der Ergebnisse und Bewertung der Hypothesen zum Einfluss der Qualität der Maps auf den Wissenserwerb

Nachfolgend werden die Hypothesen (vgl. Kapitel 4.1.) anhand der in Kapitel 5.1.3 durchgeführten Analysen hinsichtlich des Einflusses der Mapqualität auf den Wissenserwerb bewertet. Hierbei werden die Scoringwerte *ScorM1*, *ScorM2* und *ScorM3* sowie der

Wissenszuwachs vom Vor- zum 1. Nachtest und über drei Zeitpunkte (Vortest-1. Nachtest-2. Nachtest) herangezogen.

Zunächst wurde überprüft, inwieweit die Qualität der Maps einen Einfluss auf den 1. und 2. Nachtest hat. Hierbei zeigten sich signifikante Einflüsse von *ScorM2* auf den 1. Nachtest mit $d=0,95$ und signifikante Einflüsse von *ScorM3* auf den 1. und 2. Nachtest mit $d=0,67$ bzw. $d=0,70$. Der Einfluss von *ScorM1* war dabei nicht signifikant.

Um genauere Aussagen zu den Hypothesen treffen zu können, wurden die Einflüsse von *ScorM2* und *ScorM3* mittels Dichotomisierung der Variablen am Median (Einteilung in niedrig und hoch) genauer untersucht.

Bei *ScorM2 niedrig* ist *FB* nicht signifikant besser als *kein FB*, wobei sich sowohl das *EiFB* als auch das *ElFB* nicht signifikant von *keinem FB* unterscheiden. Das *ElFB* ist für den 1. Nachtest signifikant besser als *EiFB*. Im 2. Nachtest sowie hinsichtlich des Wissenszuwachses vom Vor zum 1. Nachtest bzw. über drei Zeitpunkte unterscheiden sich *ElFB* und *EiFB* nicht signifikant voneinander. Demnach müssen die Hypothesen H33, H33a und H33b abgelehnt werden. Die Hypothese H34 kann für den 1. Nachtest angenommen werden.

Für *ScorM2 hoch* ist *FB* hinsichtlich des Wissenszuwachses von Vor- zum 1. Nachtest signifikant besser als *kein FB*, in Bezug auf die Ergebnisse des 1. und 2. Nachtests sowie des Wissenszuwachses über drei Zeitpunkte unterscheiden sich *FB* und *kein FB* nicht signifikant. Sowohl das *EiFB* als auch das *ElFB* unterscheiden sich dabei nicht signifikant von *keinem FB*. Das *ElFB* ist nicht signifikant besser als *EiFB*. Demnach muss die Hypothese H35 für den Wissenszuwachs vom Vor- zum 1. Nachtest abgelehnt, für den 1. und 2. Nachtest sowie den Wissenszuwachs über drei Zeitpunkte kann diese angenommen werden. Die Hypothesen H35a, H35b und H36 können angenommen werden.

Bei *ScorM3 niedrig* zeigen alle Paarvergleiche nicht signifikante Gruppenunterschiede. Demnach müssen die Hypothesen H33, H33a, H33b und 34 abgelehnt, die Hypothesen H35, H35a, H35b und H36 können angenommen werden.

Hinsichtlich des Vergleichs von *ScorM2 niedrig* und *ScorM2 hoch* jeweils separat für die Gruppen *kein FB*, *EiFB* und *ElFB* zeigt sich lediglich ein signifikanter Unterschied beim *EiFB*, wobei die Probanden mit *ScorM2 hoch* denen mit *ScorM2 niedrig* im 1. Nachtest überlegen sind. Die Hypothesen H37, H38 und H39 müssen demnach abgelehnt werden.

Bei den Gruppen *EiFB* und *ElFB* sind die Probanden mit *ScorM3 niedrig* denen mit *ScorM3 hoch* nicht signifikant überlegen. Die Hypothesen H37 und H38 müssen demnach abgelehnt werden. Für *ScorM3 niedrig* verglichen mit *ScorM3 hoch* ist ein signifikanter Vorteil bei den Probanden der Gruppe *kein FB* mit *ScorM3 hoch* gegenüber denen mit *ScorM3 niedrig* im 2. Nachtest erkennbar. Demnach kann Hypothese H39 für den 2. Nachtest angenommen, muss für den 1. Nachtest sowie für die Wissenszuwächse vom Vor- zum 1. Nachtest und über drei Zeitpunkte abgelehnt werden.

In Bezug auf den 1. Nachtest ist *FB* bei geringer Mapqualität des 2. Mappings (*ScorM2 niedrig*) geringfügig besser als *kein FB* ($d=0,16$), wobei lediglich das *ElFB* mit einem mittleren Effekt ($d=0,63$) *keinem FB* überlegen ist, *EiFB* ist *keinem FB* mit einem geringen Effekt ($d=-0,26$) unterlegen. Das *ElFB* zeigt gegenüber dem *EiFB* einen hohen signifikanten Effekt ($d=1,03$). Im 2. Nachtest ist das *FB* der Gruppe *kein FB* mit einem geringen Effekt überlegen, wobei das *EiFB* einen geringen Effekt ($d=0,24$) und das *ElFB* einen mittleren Effekt ($d=0,71$) gegenüber *keinem FB* zeigen. Das *ElFB* ist dem *EiFB* mit einem mittleren Effekt ($d=0,57$) überlegen. Hinsichtlich des Wissenszuwachses zeigen sich vom Vor- zum 1. Nachtest negative Effekte des *FB* gegenüber *keinem FB* ($d=-0,23$), wobei dies auf einen mittleren Effekt von *keinem FB* gegenüber dem *EiFB* zurückzuführen ist ($d=-0,47$). Das *ElFB* zeigt gegenüber *keinem FB* keinen Effekt ($d=-0,06$) und ist dem *EiFB* mit einem mittleren Effekt überlegen ($d=0,45$). Über drei Zeitpunkte zeigt lediglich das *EiFB* einen geringen Vorteil gegenüber *keinem FB* ($d=0,23$). Das *ElFB* unterscheidet sich nicht von *keinem FB* ($d=0,09$) und ist *EiFB* geringfügig unterlegen ($d=-0,16$).

Hohe Mapqualität im 2. Mapping (*ScorM2 hoch*) führt kurzfristig zu einem Vorteil des *FB* gegenüber *keinem FB* (1. NT: $d=0,77$; VT-1. NT: sign.; $d=0,84$). Das *EiFB* sowie das *ElFB* zeigen hierbei mittlere bzw. hohe Effekte gegenüber *keinem FB* (1. NT: $d=0,72$; $d=0,85$; VT-1. NT: $d=0,76$; $d=0,94$). Das *ElFB* zeigt gegenüber *EiFB* kleine Effekte. ($d=0,24$; $d=0,26$). Im 2. Nachtest unterscheiden sich *FB* und *kein FB* nicht ($d=-0,06$). Lediglich das *ElFB* ist *keinem FB* sowie dem *EiFB* mit kleinen Effekten überlegen ($d=0,18$; $d=0,26$). Hinsichtlich des Wissenszuwachses über drei Zeitpunkte ist das *FB* insgesamt ($d=-0,61$) sowie das *EiFB* ($d=-0,56$) und das *ElFB* ($d=-0,63$) jeweils *keinem FB* mit mittleren Effekten unterlegen. Das *EiFB* zeigt dabei einen geringen Vorteil gegenüber dem *ElFB* ($d=-0,20$).

Für *ScorM3 niedrig* zeigt sich sowohl für den 1. als auch für den 2. Nachtest ein Vorteil des *FB* gegenüber *keinem FB* ($d=0,21$; $d=0,44$), wobei sowohl das *EiFB* als auch das *ElFB* mit geringen bis hohen Effekten *keinem FB* überlegen sind (1.NT: $d=0,17$; $d=0,29$; 2.NT: $d=0,33$; $d=0,84$). Das *ElFB* ist dabei lediglich im 2. Nachtest dem *EiFB* überlegen (1. NT: $d=0,11$; 2. NT: $d=0,40$). Im Hinblick auf die Entwicklung des Wissens zeigt sich kurzfristig ein geringer Vorteil des *FB* gegenüber *keinem FB*, wobei dies auf einen kleinen Effekt des *EiFB* gegenüber *keinem FB* zurückzuführen ist ($d=0,18$). Langfristig ist das *FB* insgesamt ($d=-0,19$) sowie das *EiFB* ($d=-0,31$) und das *ElFB* ($d=-0,39$) jeweils *keinem FB* unterlegen. Das *ElFB* zeigt gegenüber dem *EiFB* einen mittleren Effekt ($d=0,57$).

Für *ScorM3 hoch* zeigt sich im 1. Nachtest ein geringer Vorteil des *ElFB* gegenüber *keinem FB* ($d=0,17$). Das *EiFB* ist *keinem FB* ($d=-0,74$) sowie dem *ElFB* ($d=0,61$) unterlegen. Das *FB* zeigt insgesamt keinen Effekt gegenüber *keinem FB* ($d=0,00$). Im 2. Nachtest ($d=-0,56$; $d=-1,71$; $d=-0,50$) sowie hinsichtlich des Wissenszuwachses über zwei ($d=-0,09$; $d=-0,22$; $d=-0,18$) bzw. drei Zeitpunkte ($d=-0,61$; $d=-1,00$; $d=-0,64$) ist das *FB* insgesamt sowie das *EiFB* und das *ElFB* jeweils *keinem FB* unterlegen, wobei sich gerade langfristig die größten Effekte zeigen. Das *ElFB* ist dem *EiFB* im 2. Nachtest sowie hinsichtlich des Wissenszuwachses geringfügig überlegen ($d=0,40$; $d=0,38$), im Hinblick auf den Wissenszuwachs über drei Zeitpunkte geringfügig unterlegen ($d=-0,25$).

Es zeigt sich also bei *niedriger* Mapqualität für das 3. Mapping kurzfristig in Bezug auf den 1. Nachtest ein Vorteil für die Probanden der *FB*-Gruppen gegenüber *keinem FB*, für das 2. Mapping ist lediglich das *ElFB* überlegen. Vom Vor- zum 1. Nachtest zeigt lediglich das *EiFB* für Probanden mit *niedriger Mapqualität* für das 3. Mapping einen Vorteil gegenüber *keinem FB*. Hinsichtlich des 2. Nachtests sind die Probanden der *FB*-Gruppen mit *niedriger Mapqualität* im 2. und 3. Mapping *keinem FB* überlegen. Im Hinblick auf die langfristige Entwicklung wirkt lediglich das *EiFB* bei *niedriger* Qualität im 2. Mapping tendenziell gegenüber *keinem FB*.

Bei *hoher* Mappingqualität wirkt das *FB* kurzfristig (1. NT und VT-1.NT) entgegen der Vermutungen für das 2. Mapping, langfristig wirkt das *FB* bei hoher Qualität im 2. und 3. Mapping kontraproduktiv.

Beim separaten Vergleich der Gruppen im Hinblick auf *ScorM2_nh* zeigt sich für den 1. Nachtest sowie die Entwicklung von Vor- zum 1. Nachtest, dass die Probanden mit

ScorM2 hoch den Probanden mit *ScorM2 niedrig* überlegen sind (1. NT *EiFB*: sign.; $d=1,25$; *ElFB*: $d=0,49$; VT-1. NT *EiFB*: $d=0,96$; *ElFB*: $d=0,76$). Im 2. Nachtest hat das Niveau von *ScorM2* keinen Einfluss ($d=0,06$; $d=-0,09$). Hinsichtlich des Wissenszuwachses über drei Zeitpunkte sind die Probanden der FB-Gruppen mit *ScorM2 niedrig* denen mit *ScorM2 hoch* überlegen ($d=-0,70$; $d=-0,50$) und profitieren somit eher vom Feedback.

Hinsichtlich *ScorM3_nh* zeigen die Probanden mit *ScorM3 niedrig* der Gruppe *EiFB* gegenüber denen mit *ScorM3 hoch* im 1. Nachtest ($d=0,14$) sowie der Entwicklung von Vor- zum 1. Nachtest ($d=-0,06$) keine Effekte, im 2. Nachtest ($d=-0,17$) kleine Effekte. Hinsichtlich der Entwicklung über drei Zeitpunkte ist ein kleiner Effekt von *ScorM3 hoch* gegenüber *ScorM3 niedrig* erkennbar ($d=0,17$). Beim *ElFB* sind die Probanden mit *ScorM3 hoch* denen mit *ScorM3 niedrig* im 1. Nachtest und der Entwicklung vom Vor- zum 1. Nachtest mit mittleren Effekten ($d=0,55$; $d=0,46$) überlegen. Über drei Zeitpunkte führt beim *ElFB* eine niedrigere Qualität zu einem höheren Wissenszuwachs ($d=-0,54$).

Insgesamt kann folglich festgestellt werden, dass kurzfristig in Bezug auf den 1. Nachtest sowie die Entwicklung vom Vor- zum 1. Nachtest entgegen der Vermutungen die Probanden der FB-Gruppen mit *hoher Mapqualität* gegenüber denen mit *niedriger Mapqualität* profitieren (beide FB-Gruppen hinsichtlich Map 2 und *ElFB* für Map3). Langfristig ist das Feedback in Bezug auf die Entwicklung des Wissens über drei Zeitpunkte eher bei *niedriger Mapqualität* im 2. Mapping lernwirksamer. Beim 3. Mapping profitieren lediglich die Probanden der Gruppe *ElFB* mit *ScorM3 niedrig* gegenüber denen mit *ScorM3 hoch*.

Bei der Gruppe *kein FB* zeigen sich zwischen dem 2. und 3. Mapping unterschiedliche Ergebnisse. Für den 1. und 2. Nachtest sind die Probanden mit *ScorM2 hoch* denen mit *ScorM2 niedrig* geringfügig überlegen ($d=0,35$; $d=0,19$), über zwei Zeitpunkte geringfügig unterlegen ($d=-0,28$) und über drei Zeitpunkte unterscheiden sich die Gruppen nicht voneinander ($d=-0,09$). Für das 3. Mapping sind die Probanden mit *ScorM3 hoch* denen mit *ScorM3 niedrig* deutlich überlegen (1. NT: $d=0,75$; 2. NT sign; $d=1,53$; VT-NT: $d=0,41$; VT-1.NT-2.NT: $d=0,96$).

Die Ergebnisse in Bezug auf die Hypothesen werden in nachfolgender Tabelle 81 noch einmal zusammenfassend veranschaulicht.

Hierbei werden die Hypothesen anhand der in Kapitel 5.1.3 durchgeführten Analysen bewertet. Des Weiteren werden die Ergebnisse der paarweisen Vergleiche der Gruppen

für den 1. Nachtest, 2. Nachtest, den Wissenszuwachs über zwei und drei Zeitpunkte anhand der Effektstärken dargestellt, um Tendenzen hinsichtlich der Hypothesen ableiten zu können.

Tabelle 81: Ergebnisse des Einflusses der Mapqualität auf den Wissenserwerb - Prüfung der Hypothesen

Hypothese	Inhalt	Prüfung der Hypothesen	Effektstärken			
			1.NT	2.NT	WZ Vor-1.NT	WZ Vor-1.NT-2.NT
Einfluss der Mapqualität						
ScorM2 niedrig						
H33	FB>kein FB	X	0,16	0,44	- 0,23	0,06
H33a	EiFB>kein FB	X	- 0,26	0,24	- 0,47	0,23
H33b	EIFB>kein FB	X	0,63	0,71	- 0,06	0,09
H34	EIFB>EiFB	✓ (1.NT) bzw. X	1,03*	0,57	0,45	- 0,16
ScorM2 hoch						
H35	FB=kein FB	✓ bzw. X (VT-NT)	0,77	0,06	0,84*	- 0,61
H35a	EiFB=kein FB	✓	0,72	- 0,11	0,76	- 0,56
H35b	EIFB=kein FB	✓	0,85	0,18	0,94	- 0,63
H36	EIFB=EiFB	✓	0,24	0,26	0,20	- 0,20
ScorM3 niedrig						
H33	FB>kein FB	X	0,21	0,44	0,16	- 0,19
H33a	EiFB>kein FB	X	0,17	0,33	0,18	- 0,31
H33b	EIFB>kein FB	X	0,29	0,84	- 0,06	- 0,39
H34	EIFB>EiFB	X	0,11	0,40	- 0,09	0,57
ScorM3 hoch						
H35	FB=kein FB	✓	0,00	- 0,56	- 0,09	- 0,61
H35a	EiFB=kein FB	✓	- 0,74	- 1,71	- 0,22	- 1,00
H35b	EIFB=kein FB	✓	0,17	- 0,50	- 0,18	- 0,64
H36	EIFB=EiFB	✓	0,61	0,40	0,38	- 0,25
ScorM2 hoch vs. niedrig						
H37	EiFB: hoch<niedrig	X	1,25*	0,06	0,96	- 0,70
H38	EIFB: hoch<niedrig	X	0,49	- 0,09	0,76	- 0,50
H39	kein FB: hoch>niedrig	X	0,35	0,19	- 0,28	- 0,09
ScorM3 hoch vs. niedrig						
H37	EiFB: hoch<niedrig	X	0,14	- 0,17	- 0,06	0,17
H38	EIFB: hoch<niedrig	X	0,55	- 0,11	0,46	- 0,54
H39	kein FB: hoch>niedrig	✓ (2.NT) bzw. X	0,75	1,53*	0,41	0,96

* signifikant auf dem Niveau $\alpha = 0,05$

✓ Hypothese angenommen

X Hypothese abgelehnt

kein Effekt, kleiner Effekt, *mittlerer Effekt*, **großer Effekt**

5.2 Analysen zum Umgang mit dem Feedback sowie Analyse der Fehler in den Concept Maps

5.2.1 Analysen zum Umgang mit dem Feedback

Im folgenden Kapitel 5.2.1 werden die Ergebnisse aus dem Fragebogen zur eingeschätzten Nützlichkeit des Feedbacks (EiFBN, EIFBN1, EIFBN2) sowie zum Wunsch nach zusätzlichem Feedback (FBgew, EIFBgew) analysiert, um Aussagen zu den Hypothesen H27 bis H32 treffen zu können.

Zunächst werden die Mittelwerte der Items *FBgew* und *EIFBgew* analysiert, um nähere Aussagen darüber treffen zu können, inwieweit sich die Probanden weiteres Feedback gewünscht haben (vgl. Kapitel 4.6.2). Der Mittelwert für das Item *FBgew* liegt bei 2,83 (trifft eher zu), für das Item *EIFBgew* bei 2,12 (trifft eher nicht zu). Demnach wünschen sich die Probanden der Gruppe *kein FB* eher eine Rückmeldung zu ihren Concept Maps, die Gruppe *EiFB* hält dagegen ein weiteres Feedback zur Besprechung der Fehler in ihren Concept Maps für eher überflüssig. Um zu überprüfen, inwieweit es hinsichtlich des Wunsches nach einem zusätzlichen Feedback einen signifikanten Unterschied zwischen den Gruppen gibt, wird eine univariate ANOVA mit der UV *Gruppe* und der AV *gewünschtes FB allgemein* durchgeführt.¹⁰⁶ Hierzu werden die Variablen *FBgew* und *EIFBgew* zu einer Variablen *FBgew_allgemein* zusammengefasst. Die Ergebnisse sind in nachfolgender Tabelle 82 dargestellt.

Tabelle 82: Ergebnisse der einfaktoriellen ANOVA mit der UV *Gruppe* und der AV *gewünschtes FB allgemein*

Quelle	Quadrat-summe	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<i>Abh. Variable gewünschtes FB allgemein</i>							
Gruppe	6,250**	1	6,250	8,673	0,006*	0,203	1,01

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,180

Tabelle 82 zeigt einen signifikanten Haupteffekt für den Faktor *Gruppe* ($F=8,6773$; $p=0,006$). Die Probanden der Gruppe *kein FB* wünschen sich also signifikant eher eine

¹⁰⁶ Zunächst gilt es, die Voraussetzungen zu prüfen. Die Messwerte in allen Teilstichproben sollten normalverteilt sein. Diese Voraussetzung ist nicht erfüllt. Dennoch kann das Verfahren durchgeführt werden, da es relativ robust bei annähernd gleich großen Stichproben ist (Bühner & Ziegler, 2009, S. 518). Als weitere Voraussetzung gilt es, die Varianzhomogenität zu prüfen. Der Levene-Test auf Gleichheit der Varianzen zeigt eine Varianzhomogenität für *FBgew_allgemein*.

Rückmeldung zu ihren Concept Maps als die Probanden der Gruppe *EiFB* zusätzlich elaboriertes Feedback wünschen, mit einem hohen Effekt von 1,01.

Im nächsten Schritt wird die Variable *EiFBN* genauer untersucht. Hierfür wird eine univariate ANOVA mit der UV *Gruppe* und der AV *EiFBN* durchgeführt um zu analysieren, inwieweit sich die Gruppen *EiFB* und *ElFB* hinsichtlich der Einschätzung zur Nützlichkeit des *EiFB* unterscheiden.¹⁰⁷

Tabelle 83: Ergebnisse der einfaktoriellen ANOVA mit der UV *Gruppe* und der AV *Mittelwert EiFBN*

Quelle	Quadratsumme	df	Mittel der Quadrate	F-Wert	p	Partielles Eta-Quadrat	d
<u>Abh. Variable Mittelwert EiFBN</u>							
Gruppe	5,896**	1	5,896	12,901	0,001*	0,264	1,20

* signifikant auf dem Niveau $\alpha = 0,05$

** korrigiertes R-Quadrat: 0,243

Tabelle 83 zeigt einen signifikanten Haupteffekt für den Faktor *Gruppe* ($F=12,901$; $p=0,001$). Die Probanden der Gruppe *ElFB* schätzen das einfache Feedback mit einem hohen Effekt von $d=1,20$ signifikant nützlicher ein als die Probanden der Gruppe *EiFB*.

Hinsichtlich der Analyse von *ElFBN_1* sowie *ElFBN_3* zeigen sich Mittelwerte von 3,10 bzw. 2,94 (trifft eher zu). Dabei unterscheiden sich diese Werte nur geringfügig vom Mittelwert für *EiFBN* mit 3,15 (trifft eher zu) für die Gruppe *ElFB* (vgl. Kapitel 4.6.2). Demnach schätzen die Probanden der Gruppe *ElFB* das einfache Feedback tendenziell sogar etwas nützlicher als das elaborierte Feedback ein. Für eine statistische Prüfung werden Korrelationsanalysen zwischen *ElFBN_1* und *EiFBN* sowie *ElFBN_3* und *EiFBN* durchgeführt. Es zeigt sich hierbei eine hohe signifikante Korrelation für *EiFBN* mit *ElFBN_3* ($r_s=0,654$ ¹⁰⁸; $p=0,006$). Die Schüler, welche das *EiFB* für das Erkennen von Wissenslücken sowie für das bessere Verstehen von Zusammenhängen nützlich fanden, schätzten ebenso die zusätzliche Fehlerbesprechung sowie die Erklärungen als nützlich ein und umgekehrt. Zwischen *EiFBN* und *ElFBN_1* zeigt sich kein signifikanter Zusammenhang ($r_s=-0,016$; $p=0,954$).

¹⁰⁷ Die Voraussetzung der Normalverteilung ist lediglich bei der Gruppe *EiFB* erfüllt. Die Analyse kann dennoch durchgeführt werden (Bühner & Ziegler, 2009, S. 518). Die Voraussetzung der Varianzhomogenität ist erfüllt.

¹⁰⁸ Nach Cohen (1988, S. 79 f.) wird eine Korrelation von $r=0,1$ als gering, eine Korrelation von $r=0,3$ als mittel und eine Korrelation von $r=0,5$ als hoch eingestuft.

Des Weiteren werden Korrelationsanalysen für die Variablen *FBgew*, *EIFBgew*, *EiFBN*, *EIFBN1* bzw. *EIFBN2*, *1. NT* und *2. NT* sowie für die Variablen *FBgew*, *EIFBgew*, *EiFBN*, *EIFBN_1* bzw. *EIFBN_3*, *ScorM2* und *ScorM3* durchgeführt. Hierbei kann zunächst überprüft werden, inwieweit der Wunsch nach zusätzlichem Feedback sowie die eingeschätzte Nützlichkeit des Feedbacks mit der Lernwirksamkeit bzw. der Qualität der Maps zusammenhängen. Des Weiteren kann überprüft werden, inwieweit es zwischen dem Wunsch nach zusätzlichem elaborierten Feedback und der eingeschätzten Nützlichkeit des einfachen Feedbacks einen Zusammenhang gibt. Insgesamt existieren keine signifikanten Zusammenhänge zwischen den Variablen *FBgew*, *EIFBgew*, *EiFBN*, *EIFBN_1* bzw. *EIFBN_3* mit dem *1. NT* bzw. *2. NT* und mit der Qualität der Maps (*ScorM2* und *ScorM3*). Des Weiteren ist kein Zusammenhang zwischen *EIFBgew* und *EiFBN* erkennbar ($r_s=0,187$; $p=0,471$). Die Ergebnisse der entsprechenden Korrelationsanalysen können den Anhängen 5.2.1.-1 und 5.2.1.-2 entnommen werden.

Hinsichtlich des Umgangs mit dem Feedback werden in einem nächsten Schritt die *Qualität der Feedback-Netze* sowie die *Strategien bezüglich der Verarbeitung des Feedbacks* näher analysiert.

Zunächst wird die Entwicklung der Scoringwerte über die drei Mappings hinweg für die selbsterstellten Maps sowie für die Feedback-Netze zwischen den Gruppen *EIFB* und *EiFB* verglichen. Hierzu wird eine ANOVA mit Messwiederholung mit den UVs *Zeit* und *Gruppe* sowie der AV *Scoringwerte* durchgeführt. Hierbei lassen sich sechs Messzeitpunkte definieren: *ScorM1*, *ScorM1_FB*, *ScorM2*, *ScorM2_FB*, *ScorM3* und *ScorM3_FB*.¹⁰⁹

Die ANOVA mit Messwiederholung über zwei Gruppen zeigt einen signifikanten Haupteffekt für den Faktor *Zeit* ($F=35,985$; $p=0,000$), jedoch keinen signifikanten Interaktionseffekt für die Faktoren *Zeit* und *Gruppe* ($F=2,001$; $p=0,138$). Die Scoring-Werte unterscheiden sich folglich zwischen den sechs Messzeitpunkten signifikant voneinander. Die Veränderung über die *Zeit* hängt jedoch nicht signifikant von der Art des Feedbacks ab.

¹⁰⁹ Für die Durchführung gilt es zunächst die Voraussetzungen zu prüfen. Zur Prüfung der Normalverteilung kann auf die entsprechenden Ausführungen im Kapitel 5.1.2 verwiesen werden. Als weitere Voraussetzung gilt es, die Varianzhomogenität zu prüfen. Der Levene-Test auf Gleichheit der Varianzen zeigt Varianzhomogenität für alle sechs Zeitpunkte. Die Sphärizitätsannahme ist jedoch verletzt. Aus diesem Grund müssen die Freiheitsgrade bei den Tests des Messwiederholungseffekts und der Wechselwirkung korrigiert werden.

In nachfolgender Tabelle 84 sind die Ergebnisse der ANOVA mit Messwiederholung unter Berücksichtigung der Korrekturverfahren nach Greenhouse-Geisser dargestellt.

Tabelle 84: Ergebnisse der zweifaktoriellen ANOVA mit Messwiederholung mit den UVs Zeit und Gruppe und der AV Scoringwerte

Quelle	Quadrat-summe	df	Mittel der Qua-drate	F-Wert	p	Partielles Eta-Quad-rat	d
Zeit _{ScorM1-ScorM1_FB-ScorM2-ScorM2_FB-ScorM3-ScorM3_FB}	2,852	2,193	1,301	35,985	0,000*	0,500	2,00
Zeit _{ScorM1-ScorM1_FB-ScorM2-ScorM2_FB-ScorM3-ScorM3_FB} X Gruppe	0,159	2,193	0,072	2,001	0,138	0,053	0,47

* signifikant auf dem Niveau $\alpha = 0,05$

Nachfolgende Abbildung 36 zeigt die Entwicklung der Qualität der Feedback-Netze im Vergleich zur Qualität der selbsterstellten Maps über die sechs Messzeitpunkte.

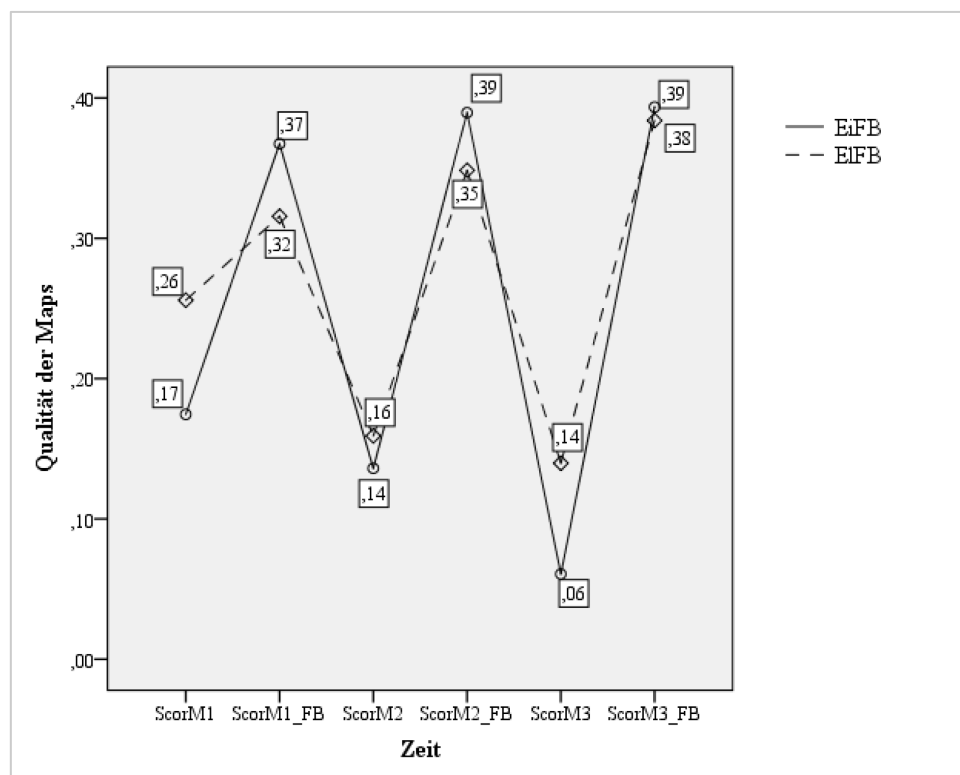


Abbildung 36: Graphische Darstellung der Scoringwerte für die Gruppen *EiFB* und *EIFB* über sechs Messzeitpunkte

Insgesamt ist erkennbar, dass die Mapqualität der Feedback-Netze deutlich über der Qualität der selbsterstellten Maps liegt. Die Scoring-Werte für die Feedback-Netze der Gruppe *EIFB* liegen tendenziell unter den Werten der Gruppe *EiFB*, obwohl die Schüler der Gruppe *EIFB* tendenziell eine höhere Mapqualität bei den selbsterstellten Maps

(ScorM1, ScorM2 und ScorM3) erzielten. Die Unterschiede sind dabei jedoch nicht signifikant (vgl. Kapitel 5.1.2 sowie Anhang 5.2.1.-3).

In Kapitel 4.5.1 wurden die Strategien, welche die Probanden zur Verarbeitung des einfachen Feedbacks nutzten, ausgewertet. Hierbei ergaben sich vier Strategien: Das Streichen bzw. als falsch Kennzeichnen von Propositionen, das Korrigieren falscher Propositionen und das Hinzufügen von falschen, unvollständigen bzw. „halb“ richtigen sowie richtigen Propositionen. Interessant ist es zu analysieren, inwieweit sich die Verarbeitungsstrategien zwischen den drei Mappings sowie zwischen der Gruppe *EiFB* und *EIFB* unterscheiden. Hierzu werden nachfolgend für jedes Mapping die Strategien beider Gruppen dargestellt (vgl. Abbildung 37 bis Abbildung 39).

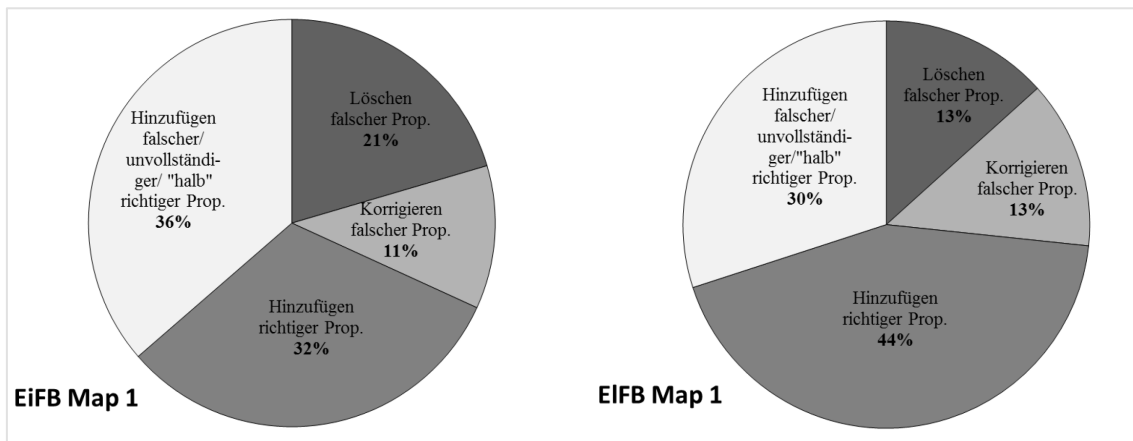


Abbildung 37: Relative Häufigkeiten der Verarbeitungsstrategien für Map 1 für das einfache und elaborierte Feedback

Abbildung 37 zeigt die Verarbeitungsstrategien des 1. Mappings. Hierbei ist erkennbar, dass insgesamt 68 Prozent der von der Gruppe *EiFB* verwendeten Strategien auf das Hinzufügen von Propositionen entfallen, bei der Gruppe *EIFB* sind es 74 Prozent. Das Hinzufügen falscher oder nur teilweise richtiger Propositionen wird dabei von der Gruppe *EiFB* 6 Prozent häufiger verwendet als von der Gruppe *EIFB*.

Insgesamt setzten sich bei der Gruppe *EiFB* mehr Schüler tiefgründiger mit ihren Maps auseinander als bei der Gruppe *EIFB*. Das Korrigieren und Löschen von Propositionen wird zu 31 Prozent von der Gruppe *EiFB* und zu 26 Prozent von der Gruppe *EIFB* verwendet. Aufgrund dessen, dass die Intervention für alle Gruppen bis zu diesem Zeitpunkt identisch erfolgte, lassen sich Gruppenunterschiede nicht auf die Art des Feedbacks zurückführen und werden daher für das 1. Mapping nicht weiter analysiert.

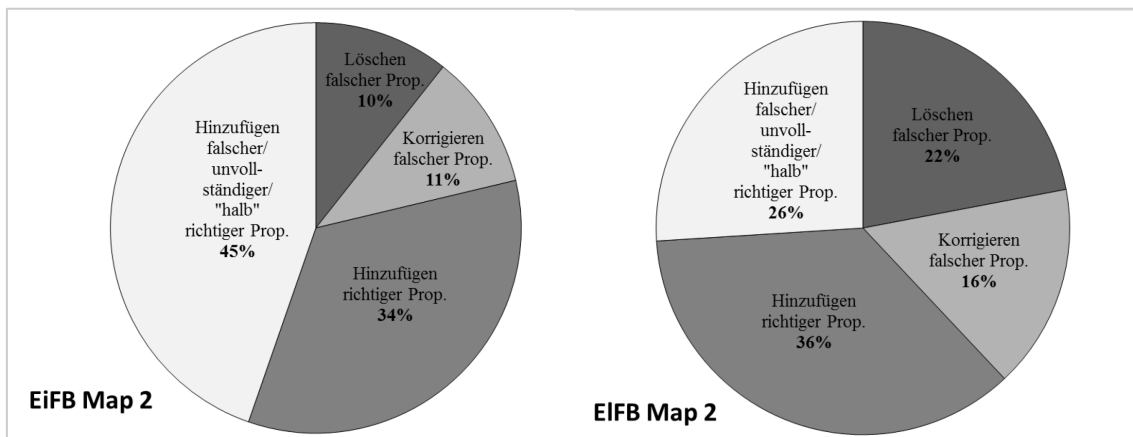


Abbildung 38: Relative Häufigkeiten der Verarbeitungsstrategien für Map 2 für das einfache und elaborierte Feedback

In Abbildung 38 ist erkennbar, dass die Gruppe *EIFB* im 2. Mapping die Strategie des Hinzufügens von Propositionen deutlich weniger anwendet (26 % und 36 %) als die Gruppe *EiFB* (45 % und 34 %). Dabei entfällt bei der Gruppe *EiFB* (45 %) auch ein höherer Anteil auf das Hinzufügen falscher Propositionen als bei der Gruppe *EIFB* (26 %). Des Weiteren setzen sich die Schüler der Gruppe *EIFB* intensiver mit ihren Maps auseinander als die Schüler der Gruppe *EiFB*. Der Anteil des Löschens bzw. Korrigierens falscher Propositionen ist bei der Gruppe *EIFB* (22 % und 16 %) höher als bei der Gruppe *EiFB* (10 % und 11 %).

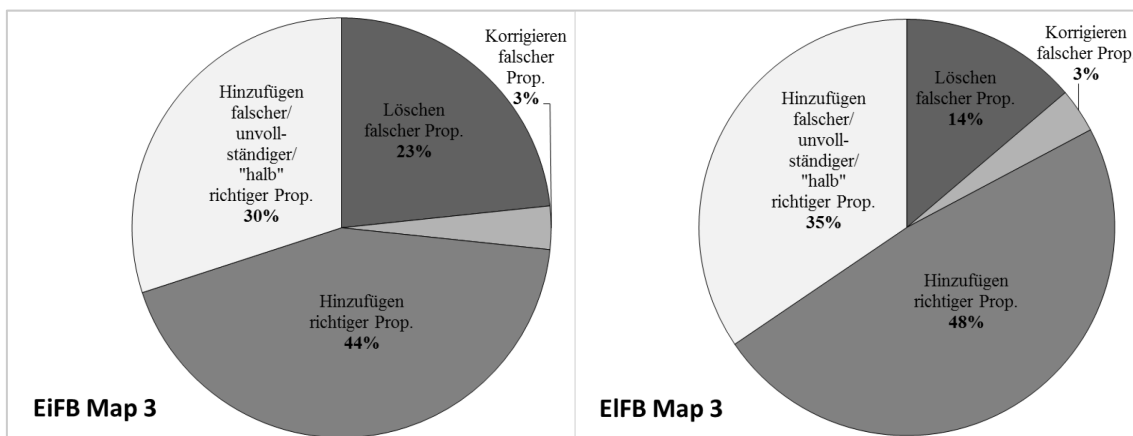


Abbildung 39: Relative Häufigkeiten der Verarbeitungsstrategien für Map 3 für das einfache und elaborierte Feedback

Aus der Abbildung 39 geht hervor, dass im 3. Mapping die Gruppe *EIFB* die Verarbeitungsstrategie des Hinzufügens von Propositionen häufiger anwendet (35 % und 48 %) als die Gruppe *EiFB* (30 % und 44 %). Das Löschen und Korrigieren falscher Propositionen nutzen die Schüler der Gruppe *EiFB* häufiger als (23 %) bzw. genauso viel (3 %) wie die Schüler der Gruppe *EIFB* (14 % und 3 %).

5.2.2 Analysen der Fehler in den Concept Maps

Im folgenden Kapitel werden die Propositionen der selbsterstellten Maps in Bezug auf ihre inhaltliche Richtigkeit näher untersucht. Hierbei wird einerseits analysiert, welche Propositionen häufig genannt wurden, aber nicht im Referenznetz enthalten sind oder andererseits kaum bzw. nicht genannt wurden und im Referenznetz vorkommen. Infolgedessen können Fehlkonzepte im Sinne von falschen Propositionen aufgedeckt werden. Darüber hinaus ist es möglich, fehlende Propositionen zu identifizieren. Infolgedessen können gezielte Hinweise zur inhaltlichen Gestaltung des elaborierten Feedbacks in zukünftigen Studien gewonnen werden. Hierbei werden die Propositionslisten für die drei Mappings separat analysiert.

Zur Identifikation der „häufig genannten“ Propositionen kann auf die im Kapitel 4.5.1 erstellten Modalnetze zurückgegriffen werden. Die Modalnetze berücksichtigen alle erstellten Maps und enthalten die am häufigsten übereinstimmend genannten Propositionen. Auf diese Weise ist ein vereinfachtes Vorgehen möglich, ohne eine Analyse aller Individualnetze vornehmen zu müssen. Als falsche Propositionen werden somit diejenigen angesehen, welche im Modalnetz, aber nicht im Referenznetz enthalten sind. Als fehlende Propositionen gelten diejenigen, welche dreimal oder seltener genannt werden und Teil des Referenznetzes sind. Das Vorgehen, nicht nur die komplett fehlenden Propositionen als fehlend zu betrachten, wird dabei bewusst gewählt. Der Lehrperson lässt sich auf diese Weise ein Hinweis geben, welche Inhalte besonders zu thematisieren sind, da sie von den Lernenden noch nicht beherrscht werden (vgl. Anhänge 5.2.2-1 bis 5.2.2.-3).

Nachfolgende Tabelle 85 zeigt die falschen und fehlenden Propositionen des 1. Mapping. Hierbei zeigt sich ein falsches Verständnis der Schüler vom Begriff Forderungen. Die Schüler verstehen darunter die Produkte, welche sie vom Unternehmen fordern und nicht den bilanziellen Begriff. Die Schüler formulieren des Weiteren eher umgangssprachlich und verwenden die Proposition „Kunden begleichen Rechnungen“, obwohl der korrekte Weg über die Geldzahlung stattfindet (Kunden tätigen Geldzahlungen, Geldzahlungen begleichen Rechnungen). Die Komplexität der Zusammenhänge fällt den Schülern noch schwer zu erkennen, weshalb dies durch das differenzierte Bewertungssystem berücksichtigt wurde. Die Schüler erhalten daher für diese Proposition $\frac{1}{3}$ Punkt (vgl. Kapitel 4.5.2). Dies zeigt sich auch in der häufig genannten Proposition „Materialien sind Voraussetzung für Produkte“. Hierbei wurde ein Schritt übersprungen, die Proposition daher

lediglich mit $\frac{1}{2}$ Punkt bewertet (Materialien gehen an Fertigung, Fertigung erzeugt Produkte). Die Schwierigkeiten im Umgang mit finanzwirtschaftlichen Begriffen zeigen sich in der Proposition „Unternehmen hat Kasse und Bankguthaben“. Aus dem Alltagsverständnis des Schülers ist das richtig. Im Referenznetz werden jedoch lediglich Unternehmensbereiche wie die Fertigung berücksichtigt (z. B. Unternehmen hat Fertigung). „Kasse und Bankguthaben“ stellt eine Bilanzposition dar und ist kein Unternehmensbereich.

Bei den fehlenden Propositionen ist die bereits beschriebene Problematik im Umgang mit dem Begriff Geldzahlungen erkennbar. Die Proposition „Geldzahlungen begleichen Rechnungen“ des Referenznetzes wird dreimal verwendet. Dass die Geldzahlungen in diesem Zusammenhang die Kasse und das Bankguthaben erhöhen oder vermindern können (Geldzahlungen erhöhen/vermindern Kasse und Bankguthaben) und in Folge dessen die Geldzahlungen die Forderungen vermindern bzw. die Verbindlichkeiten erhöhen (Geldzahlungen vermindern Forderungen; Geldzahlungen erhöhen Verbindlichkeiten), thematisieren die Schüler nicht. Darüber hinaus erkennen sie nicht, dass Geldzahlungen in die Fertigung gehen, um Produktionskosten zu decken (Geldzahlungen gehen an Fertigung). Des Weiteren fällt den Schülern das Denken aus Unternehmersicht schwer, da auch Unternehmen Rechnungen erhalten (Rechnungen gehen an Unternehmen). Dies ist auch an der nur zweimal verwendeten Proposition „Mitarbeiter sollen haben Treue“ erkennbar, wobei bei dieser Proposition zu betonen ist, dass diese nachträglich in das Referenznetz aufgenommen wurde und die Bedeutung daher nicht überinterpretiert werden darf. Die Schüler haben weiterhin Schwierigkeiten im Umgang mit dem Begriff Lieferant. Für sie ist ein Lieferant wahrscheinlich eher ein Spediteur, welcher die Produkte liefert. Dies zeigt sich auch in der sieben Mal genannten Proposition „Lieferanten beliefern Kunden“ (vgl. Anhang 5.2.2.-1). Dass die Lieferanten Rechnungen stellen (Lieferanten stellen Rechnungen), vom Unternehmen Treue erwartet (Lieferanten erwarten Treue) und Geldzahlungen an den Lieferanten gehen (Geldzahlungen gehen an Lieferanten), entzieht sich damit dem Verständnis der Schüler.

Tabelle 85: Falsche und fehlende Propositionen des 1. Mapping

Proposition	Häufigkeit der Nennung
<i>falsche Propositionen</i>	
Kunden haben Forderungen	31
Kunden begleichen Rechnungen	23
Materialien sind Voraussetzung für Produkte	17
Unternehmen hat Kasse und Bankguthaben	15
<i>fehlende Propositionen</i>	
Geldzahlungen begleichen Rechnungen	3
Lieferanten stellen Rechnungen	2
Lieferanten erwarten Treue	2
Mitarbeiter sollen haben Treue	2
Rechnungen gehen an Unternehmen	2
Geldzahlungen gehen an Lieferanten	1
Geldzahlungen gehen an Fertigung	0
Geldzahlungen vermindern Kasse und Bankguthaben	0
Geldzahlungen erhöhen Kasse und Bankguthaben	0
Geldzahlungen vermindern Forderungen	0
Geldzahlungen vermindern Verbindlichkeiten	0
Lieferanten sollen haben guter Service	0

Die Tabelle 86 stellt die falschen und fehlenden Propositionen des 2. Mapping dar.

Es zeigt sich, dass die Schüler die Zusammenhänge zwischen Bilanz und GuV nicht verstanden haben. Die Propositionen „Herstellungskosten sind Teil von Aktivseite“, „Gemeinkosten sind Teil von Aktivseite“, „Abschreibungen sind Teil von Aktivseite“, „Nutzungsdauer ist Teil von Aktivseite“, „Gewinn- und Verlustrechnung ist Teil von Passivseite“ und „Aufwand ist Teil von Aktivseite“ sind falsch und lassen eventuell darauf schließen, dass die Aktivseite mit dem Aufwand verwechselt wurde. Des Weiteren verbinden einige Schüler den Ertrag direkt mit dem Gewinn (Ertrag verursacht Gewinn), erkennen den komplexeren Weg über das Periodenergebnis nicht und erhalten daher lediglich $\frac{1}{2}$ Punkt. Dies zeigt sich auch in den nur einmal oder gar nicht genannten Propositionen „Periodenergebnis ist Verlust“, „Periodenergebnis ist Gewinn“, „Ertrag ist Teil von Periodenergebnis“ und „Aufwand ist Teil von Periodenergebnis“ des Referenznetzes. Darüber hinaus verbinden die Schüler die Anschaffungskosten direkt mit der Aktivseite (Anschaffungskosten sind Teil von Aktivseite) und erhalten dafür $\frac{1}{2}$ Punkt. Der im Referenznetz thematisierte Zwischenschritt über das Anlagevermögen wird nicht erkannt, die Proposition „Anlagevermögen verursacht Anschaffungskosten“ des Referenznetzes nicht

verwendet. Die Schüler erkennen ebenso nicht, dass die Abschreibung von den Anschaffungskosten (Abschreibungen sind abhängig von Anschaffungskosten) abhängig ist. Die Verbindungen zwischen dem Eigenkapital und dem Gewinn (Gewinn erhöht Eigenkapital) bzw. Verlust (Verlust vermindert Eigenkapital) im Referenznetz und somit zum Ertrag und Aufwand führen die Bilanz mit der GuV-Rechnung zusammen. Diese Propositionen werden nicht verwendet, da die Schüler, wie bereits oben angedeutet, eventuell annehmen, dass die Aktivseite mit den Aufwendungen gleichzusetzen ist.

Tabelle 86: Falsche und fehlende Propositionen des 2. Mapping

Proposition			Häufigkeit der Nennung
<i>falsche Propositionen</i>			
Herstellungskosten	sind Teil von	Aktivseite	14
Anschaffungskosten	sind Teil von	Aktivseite	13
Gemeinkosten	sind Teil von	Aktivseite	10
Ertrag	verursacht	Gewinn	7
Abschreibungen	sind Teil von	Aktivseite	7
Nutzungsdauer	ist Teil von	Aktivseite	6
Gewinn- und Verlustrechnung	sind Teil von	Passivseite	6
Aufwand	ist Teil von	Aktivseite	6
<i>fehlende Propositionen</i>			
Abschreibungen	ist abhängig von	Anschaffungskosten	3
Anlagevermögen	verursacht	Anschaffungskosten	1
Aufwand	ist Teil von	Periodenergebnis	1
Periodenergebnis	ist	Verlust	1
Ertrag	ist Teil von	Periodenergebnis	1
Gewinn	erhöht	Eigenkapital	0
Verlust	vermindert	Eigenkapital	0
Periodenergebnis	ist	Gewinn	0

In der Tabelle 87 sind die falschen und fehlenden Propositionen des 3. Mapping dargestellt. Wie bereits erläutert, erkannten die Schüler im 2. Mapping nicht, dass die Anschaffungskosten mit dem Anlagevermögen in Verbindung stehen. Daher verbinden sie nun die Maschinen und Anlagen direkt mit den Anschaffungskosten (Maschinen und Anlagen verursachen Anschaffungskosten) und der Zwischenschritt über das Anlagevermögen wird nicht dargestellt (Maschinen und Anlagen sind Teil von Anlagevermögen, Anlagevermögen verursacht Anschaffungskosten), weshalb diese Proposition lediglich mit ½ Punkt bewertet wird. Des Weiteren zeigt die Proposition „Produkte verursachen Herstellungskosten“, dass das erlangte Wissen der Schüler durch das Planspiel eher oberflächlich

ist und daher lediglich mit $\frac{1}{2}$ Punkt bewertet wird. Die Produkte werden gefertigt (Fertigung erzeugt Produkte) und dadurch verursachen sie auch indirekt Herstellungskosten (Fertigung verursacht Herstellungskosten). Die Proposition „Maschinen und Anlagen haben Nutzungsdauer“ ist nicht falsch. Jedoch ist die Nutzungsdauer im Zusammenhang mit den Abschreibungen von Bedeutung, der Zusammenhang wird im Referenznetz daher über die Propositionen „Maschinen und Anlagen führen zu Abschreibungen“ und „Abschreibungen ist abhängig von Nutzungsdauer“ dargestellt, die Schüler erhalten hierfür daher lediglich $\frac{1}{2}$ Punkt. Die Schüler verwenden darüber hinaus die Proposition „Materialien verursachen Anschaffungskosten“. Aus ihrem Verständnis heraus ist es richtig, dass das Material angeschafft wird, jedoch sind die Anschaffungskosten ein Bilanzierungsbegriff, lediglich das Anlagevermögen verursacht Anschaffungskosten (Anlagevermögen verursacht Anschaffungskosten). Die Proposition ist daher falsch. Die Propositionen „große Abnahmemengen führen zu Gewinn“ und „Produkte führen zu Gewinn“ zeigen das einfache Denken der Schüler, das Erkennen der komplexen Zusammenhänge fällt ihnen vermutlich schwer („Kunden sollen haben große Abnahmemengen“/ „Produkte gehen an Kunden“, „Rechnungen gehen an Kunden“, „Rechnungen führen zu Umsatzerlöse“, „Umsatzerlöse sind Teil von Ertrag“, „Ertrag ist Teil von Periodenergebnis“, „Periodenergebnis ist Gewinn“). Beide Propositionen werden daher lediglich mit $\frac{1}{4}$ Punkt bewertet. Das zeigt sich auch in der fehlenden Proposition „Rechnungen führen zu Umsatzerlösen“ des Referenznetzes. Bei den fehlenden Propositionen ist erkennbar, dass der Begriff Gemeinkosten selten verwendet wurde. Die Propositionen „Materiallager verursacht Gemeinkosten“, „Fertigwarenlager verursacht Gemeinkosten“ und „Fertigung verursacht Gemeinkosten“ des Referenznetzes werden lediglich zweimal bzw. einmal genannt. Des Weiteren wird die Proposition „Forderungen sind Teil von Umlaufvermögen“ des Referenznetzes nicht verwendet. Der Begriff Forderungen wurde schon im 1. Mapping falsch verstanden, das Verständnis dieses Konzepts wurde also im Laufe des Spiels nicht korrigiert. Ebenso verhält es sich mit den Verbindlichkeiten, welche schon im 1. Mapping nicht thematisiert wurden. Daher ist es nun auch schwierig zu erkennen, dass die Verbindlichkeiten ein Teil des Fremdkapitals sind (Verbindlichkeiten sind Teil von Fremdkapital). Die Proposition „Produkte sind Teil von Umlaufvermögen“ ist eine weitere fehlende Proposition. Hierbei erkennen die Schüler nicht, dass fertige Erzeugnisse mit Produkten synonym verwendet werden können und somit ein Zusammenhang zum Umlaufvermögen möglich ist.

Tabelle 87: Falsche und fehlende Propositionen des 3. Mapping

Proposition			Häufigkeit der Nennung
<i>falsche Propositionen</i>			
Maschinen und Anlagen	verursachen	Anschaffungskosten	14
Produkte	verursachen	Herstellungskosten	11
Maschinen und Anlagen	hat	Nutzungsdauer	9
Materialien	verursachen	Anschaffungskosten	8
große Abnahmemengen	führen zu	Gewinn	8
Produkte	führen zu	Gewinn	7
<i>fehlende Propositionen</i>			
Materiallager	verursacht	Gemeinkosten	2
Fertigwarenlager	verursacht	Gemeinkosten	2
Fertigung	verursacht	Gemeinkosten	1
Forderungen	sind Teil von	Umlaufvermögen	0
Produkte	sind Teil von	Umlaufvermögen	0
Rechnungen	führen zu	Umsatzerlöse	0
Verbindlichkeiten	sind Teil von	Fremdkapital	0

5.2.3 Zusammenfassung der Ergebnisse zum Umgang mit dem Feedback und der Fehleranalyse

Nachfolgend werden die Ergebnisse zum Umgang mit dem Feedback und zur Fehleranalyse zusammengefasst.

Die Ergebnisse hinsichtlich des Umgangs mit dem Feedback können der Tabelle 88 entnommen werden.

Tabelle 88: Ergebnisse der Analyse zum Umgang mit dem Feedback - Prüfung der Hypothesen

Hypothese	Inhalt	Prüfung der Hypothesen Effektstärken	
H27	FBgew allgemein: kein FB>FB	✓ d=1,01	
H28	EiFBN: EIFB>EiFB	✓ d=1,20	
H29	EiFBN und EIFBN: r>0	✓ r _s =0,654 (EIFBN_3) X r _s =-0,016 (EIFBN_1)	
H30	EIFBgew und EiFBN: r>0	X r _s =0,187	
		1. NT	2. NT
H31a	FBgew und Wissenserwerb: r<0	X r _s =0,213	X r _s =0,331
H31b	EIFBgew und Wissenserwerb: r<0	X r _s =-0,222	X r _s =0,093
H32a	EiFBN und Wissenserwerb: r>0	X r _s =0,273	X r _s =0,084
H32b	EIFBN und Wissenserwerb: r>0	X r _s =-0,120 (EIFBN_1) r _s =0,428 (EIFBN_3)	X r _s =0,093 (EIFBN_1) r _s =0,085 (EIFBN_3)
		ScorM2	ScorM3
H31a	FBgew und Qualität der Maps: r<0	X r _s =0,351	X r _s =-0,066
H31b	EIFBgew und Qualität der Maps: r<0	X r _s =-0,469	X r _s =0,214
H32a	EiFBN und Qualität der Maps: r>0	X r _s =0,136	X r _s =0,258
H32b	EIFBN und Qualität der Maps: r>0	X r _s =0,007 (EIFBN_1) r _s =0,217 (EIFBN_3)	X r _s =-0,263 (EIFBN_1) r _s =0,025 (EIFBN_3)

Es zeigt sich, dass sich die Gruppe *kein FB* signifikant eher eine Rückmeldung zu ihren Concept Maps wünscht als die Probanden der Gruppe *EiFB* zusätzliches Feedback wünschen (d=1,01). Die Hypothese H27 kann demnach angenommen werden.

Die Probanden der Gruppe *EIFB* schätzen das einfache Feedback signifikant nützlicher ein als die Probanden der Gruppe *EiFB* (d=1,20). Die Hypothese H28 kann folglich angenommen werden.

Weiterhin zeigt sich ein signifikanter Zusammenhang von $r_s=0,654$ zwischen *EiFBN* und *EIFBN_3*, d. h. die Probanden der Gruppe *EIFB*, welche das einfache Feedback für das Erkennen von Wissenslücken nützlich fanden, schätzten ebenso das zusätzliche Besprechen der Fehler sowie die Erklärungen als nützlich ein und umgekehrt. Die Hypothese H29 kann folglich für *EIFBN_3* angenommen werden. Für *EIFBN_1* muss diese abgelehnt werden, da sich hier kein signifikanter Zusammenhang zeigt. Probanden, welche das elaborierte Feedback für das Erkennen von Fehlerursachen als nützlich erachten, schätzen nicht gleichzeitig das einfache Feedback als nützlich ein ($r_s=-0,016$).

Es gibt keinen signifikanten Zusammenhang zwischen dem Wunsch nach zusätzlichem elaborierten Feedback und der eingeschätzten Nützlichkeit des einfachen Feedbacks. Die Hypothese H30 muss demnach abgelehnt werden. Es zeigt sich lediglich eine geringe Korrelation von $r_s=0,187$.

Weiterhin existiert kein Zusammenhang zwischen den Variablen zur Einschätzung der Nützlichkeit des Feedbacks (FBgew, *EIFB*gew, *EiFBN*, *EIFBN_1* bzw. *EIFBN_3*) und dem Wissenserwerb (1. NT, 2. NT) bzw. der Qualität der Maps (ScorM2 und ScorM3). Es kann folglich nicht geschlussfolgert werden, dass Probanden, welche das Feedback als nützlich erachten, bessere Ergebnisse in den Wissenstests erzielen bzw. qualitativ bessere Concept Maps erstellen. Des Weiteren schneiden Probanden, welche sich weiteres Feedback gewünscht hätten, in den Wissenstests nicht schlechter ab bzw. erstellen qualitativ schlechtere Concept Maps als die Probanden, welche sich kein zusätzliches Feedback gewünscht hätten. Die Hypothesen H31a, H31b, H32a und H32b müssen abgelehnt werden.

Die Qualität der Feedback-Netze der Gruppe *EIFB* liegt tendenziell unter der der Gruppe *EiFB*. Es zeigen sich jedoch keine signifikanten Unterschiede. Darüber wendeten die Probanden der Gruppe *EIFB* beim 2. Mapping die Strategie des Löschens und Korrigierens falscher Propositionen häufiger an (38 %) als die Probanden der Gruppe *EiFB* (21 %). Auf der anderen Seite wurde das Hinzufügen von Propositionen von den Probanden der Gruppe *EiFB* (79 %) häufiger verwendet als von den Probanden der Gruppe *EIFB* (62 %). Hierbei kann von vergleichbaren Voraussetzungen ausgegangen werden, da sich die Ergebnisse für ScorM2 beider Gruppen nicht signifikant unterscheiden, mit $d=0,00$ (vgl. Kapitel 5.1.2). Es kann geschlussfolgert werden, dass die niedrigere Qualität der Feedback-Netze womöglich auf ein intensiveres Auseinandersetzen der Gruppe *EIFB* mit ihren selbsterstellten Concept Maps zurückzuführen ist. Die Ergebnisse hinsichtlich der

Verarbeitungsstrategien des einfachen Feedbacks für das 3. Mapping der Gruppe *EiFB* und *EIFB* sind schwerer vergleichbar, da die Gruppe *EIFB* mit $d=0,66$ tendenziell eine höhere Qualität im 3. Mapping erzielte (vgl. Kapitel 5.1.2). Das erklärt auch, warum die Schüler der Gruppe *EiFB* das Löschen und Korrigieren falscher Propositionen häufiger anwendeten (26 %) als die Probanden der Gruppe *EIFB* (17 %). Insgesamt ist auffällig, dass bei den Schülern beider Gruppen für die drei Mappings insgesamt 34 Prozent der angewendeten Verarbeitungsstrategien auf das Hinzufügen falscher oder nur teilweise richtiger Propositionen entfallen.

Bei der Analyse der falschen bzw. fehlenden Propositionen in den selbsterstellten Concept Maps ist erkennbar, dass die Schüler ein falsches Verständnis der Begriffe Forderungen und Lieferanten haben. Unter den Forderungen verstehen die Schüler das, was die Kunden vom Unternehmen fordern und nicht den bilanziellen Begriff. Der Lieferant wird bei den Schülern als Spediteur angesehen. Dies zeigt, dass überwiegend einfache und oberflächliche Denken der Schüler. Es fällt ihnen auch schwer, die Komplexität der Zusammenhänge zu erkennen. Oftmals werden Konzepte übersprungen und somit die Komplexität reduziert. Einige Begriffe, wie Geldzahlungen, werden dabei gar nicht thematisiert. Des Weiteren werden die Zusammenhänge zwischen Bilanz und GuV nicht richtig erkannt. Die Schüler setzen die Aktivseite mit den Aufwendungen gleich und ordnen somit z. B. die Herstellungskosten oder die Gemeinkosten der Aktivseite zu. Sachverhalte werden oft vereinfacht, dabei fehlt es an Trennschärfe sowohl bei den Begriffen als auch den Prozessen, z. B. bei den Konzepten Anschaffungs- und Herstellungskosten.

5.3 Interpretation der Ergebnisse

Nachfolgend werden die Ergebnisse der Analysen zum Einfluss der Art des Feedbacks (vgl. Kapitel 5.1) und die Ergebnisse der Analysen zum Umgang mit dem Feedback sowie der Fehleranalyse der Concept Maps (vgl. Kapitel 5.2) interpretiert. Unter Rückbezug auf die im Kapitel 4.1.1 aufgestellten Forschungsfragen bzw. Hypothesen sowie die theoretischen Grundlagen und empirischen Befunde der Kapitel 2 und 3 sollen in diesem Zusammenhang Begründungen für Übereinstimmungen der Ergebnisse mit den bzw. ihrer Abweichungen von den Hypothesen gefunden werden.

Als erste Forschungsfrage sollte beantwortet werden, inwieweit ergänzendes Feedback zu Concept Maps lernwirksamer ist als kein ergänzendes Feedback (H1). Als zweite Forschungsfrage galt es dabei zu klären, ob elaboriertes Feedback besser ist als einfaches Feedback (H2). Die Befunde auf dem Gebiet der Feedbackforschung sowie der Forschung zu Concept Maps (vgl. Kapitel 2.2, 3.2 sowie 4.1.1) haben gezeigt, dass die erste und zweite Forschungsfrage nicht pauschal beantwortet werden können, sondern die Lernwirksamkeit des Feedbacks von verschiedenen Faktoren beeinflusst wird. Daher sollte mit der dritten Forschungsfrage beantwortet werden, inwieweit Vorwissen (H3-H6), Sprachfähigkeit (H7-H10), die Einschätzung des Umgangs mit Fehlern (H11-H14), die Einschätzung des Umgangs mit Concept Mapping (H15-H22), die Aufgabenanforderungen der Lernaufgabe (H23-H26) sowie die Einschätzung der Nützlichkeit des Feedbacks (H27-H32) die Lernwirksamkeit des Feedbacks beeinflussen. Zusätzlich wurde in den Hypothesen H33-H39 der Einfluss der Mapqualität auf den Wissenserwerb untersucht. Die Lernwirksamkeit des Feedbacks wurde zum einen durch den Wissenserwerb (Ergebnisse des 1. und 2. Nachtests; Wissenszuwachs Vortest-1. Nachtest und Vortest-1. Nachtest-2. Nachtest) und zum anderen durch die Qualität der Maps (ScorM1, ScorM2 und ScorM3) operationalisiert. Ergänzend wurden qualitative Analysen durchgeführt, um den Umgang mit dem Feedback und die Fehler in den erstellten Concept Maps näher zu analysieren.

Nachfolgend werden die Ergebnisse hinsichtlich des Wissenserwerbs sowie der Qualität der Maps interpretiert. Danach gilt es, die Ergebnisse des Einflusses der Mapqualität auf den Wissenserwerb zu erläutern. Abschließend werden die Ergebnisse der Analysen zum Umgang mit dem Feedback und zu den Fehlern in den erstellten Concept Maps genauer betrachtet.

Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs

Zunächst werden die Ergebnisse aus Kapitel 5.1.1 zur Analyse des Einflusses der Art des Feedbacks auf den Wissenserwerb näher untersucht.

In den Hypothesen H1, H1a, H1b und H2 wurde behauptet, dass *FB* lernwirksamer ist als *kein FB*, wobei *EIFB* gegenüber *EiFB* einen Vorteil zeigen sollte. Die Hypothesen H1, H1a, und H2 müssen abgelehnt werden. Das *FB* ist insgesamt nicht signifikant besser als *kein FB*, wobei sich *EiFB* und *kein FB* sowie *EIFB* und *EiFB* nicht signifikant unterscheiden. Die Hypothese 1b kann für den 1. Nachtest angenommen werden, muss jedoch für den 2. Nachtest sowie für die Wissenszuwächse über zwei und drei Zeitpunkte abgelehnt werden. *EIFB* und *kein FB* unterscheiden sich in Bezug auf den 1. Nachtest signifikant voneinander.

Beim 1. und 2. Nachtest lassen sich Tendenzen ableiten, die für einen kleinen Vorteil des *FB* gegenüber *keinem FB* sprechen, wobei die Überlegenheit auf die Gruppe *EIFB* ($d=0,69$ bzw. $d=0,37$) zurückzuführen ist.

Die Einflussfaktoren Sprachfähigkeit, Vorwissen, der eingeschätzte Umgang mit Fehlern, die eingeschätzte Interessantheit-Nützlichkeit sowie die Handhabbarkeit des Concept Mapping wurden anschließend als Kovariaten in die Analysen einbezogen. Dabei zeigt sich lediglich ein signifikanter Einfluss des Vorwissens und der Sprachfähigkeit auf die Ergebnisse des 1. und 2. Nachtests. Die anderen Einflussfaktoren zeigen keine signifikanten Effekte. Durch das Herauspartialisieren der Einflussfaktoren sind tendenziell nur noch kurzfristige Effekte zugunsten des Feedbacks erkennbar (*FB* vs. *kein FB* $d=0,39$; *EiFB* vs. *kein FB* $d=0,37$), wobei der Unterschied zwischen *EIFB* und *keinem FB* nicht länger signifikant ist ($d=0,52$). Demnach muss die Hypothese 1b unter Herauspartialisierung der Einflussfaktoren in Bezug auf den 1. Nachtest nun auch abgelehnt werden.

Langfristig, d. h. im 2. Nachtest, verschwinden die kleinen Effekte zugunsten des Feedbacks unter Herauspartialisierung der Einflussfaktoren. Somit resultieren die kleinen Effekte im 2. Nachtest zugunsten des Feedbacks aus dem Einfluss der Kovariaten Sprachfähigkeit und Vorwissen, d. h. nicht die Intervention führte zu den Gruppenunterschieden.

Wird der Wissenszuwachs über zwei (VT-1. NT) bzw. drei (VT-1. NT-2. NT) Zeitpunkte betrachtet, so zeigen sich kurzfristig kleine Effekte zugunsten des Feedbacks. Dies kommt auch hier erneut durch einen mittleren Effekt des *EIFB* gegenüber *keinem FB* zum

Ausdruck ($d=0,46$). Langfristig kehren sich die Effekte sogar um, so dass *kein FB* mit einem kleinen Effekt dem *FB* überlegen ist ($d=-0,29$).

Nun stellt sich die Frage, warum sich der Einfluss des Feedbacks nicht so deutlich zeigt. Zunächst wird hierzu das Vorwissensniveau betrachtet.

In den Hypothesen H3, H3a und H3b wurde behauptet, dass *FB* lediglich bei Probanden mit niedrigem Vorwissen lernwirksamer ist als *kein FB*, wobei sich bei Probanden mit hohem Vorwissen kein Unterschied zeigen sollte (H5, H5a und H5b). In Hypothese H4 wurde postuliert, dass *ElFB* bei Probanden mit niedrigem Vorwissen lernwirksamer sein sollte als *EiFB*. Bei Probanden mit hohem Vorwissen sollten sich *ElFB* und *EiFB* nicht unterscheiden (Hypothese H6).

Die Hypothesen H3, H3a, H3b und 4 müssen abgelehnt werden. Es zeigt sich kein signifikanter Vorteil des Feedbacks bei niedrigem Vorwissen. Dabei lassen sich jedoch Tendenzen ableiten, die kurzfristig (1. NT; VT-1. NT) für einen kleinen bzw. mittleren Vorteil des *ElFB* gegenüber *keinem FB* ($d=0,38$; $d=0,31$) und *EiFB* ($d=0,57$ bzw. $d=0,38$) sprechen. Langfristig kehren sich die Effekte sogar um, so dass *kein FB* mit kleinen bzw. mittleren Effekten dem *FB* überlegen ist ($d=-0,36$; $d=-0,54$). Kurzfristig zeigt sich somit, dass entsprechend der Argumentation in Kapitel 4.1.1 ergänzendes Feedback bei Probanden mit niedrigem Vorwissen die kognitive Belastung beim Erstellen der Concept Maps im Anschluss an die nächste Planspielphase möglicherweise reduzieren kann. Langfristig scheint Feedback für Lernende mit niedrigem Vorwissen hingegen sogar kontraproduktiv zu wirken. Das Feedback scheint demnach für eine langfristige Behaltensleistung zu komplex für die Lernenden zu sein. Im Rahmen des Feedbacks wird das Durcharbeiten von Concept Maps angewendet. Die Referenznetze, welche für das einfache Feedback verwendet werden, stellen vorstrukturierte Maps dar, die anschließend selbstständig durcharbeitet werden sollen. Das Besprechen der Fehler- und Referenznetze im Rahmen des elaborierten Feedbacks stellt ein gemeinsames Durcharbeiten der Maps mit der Lehrperson dar. Hierbei kann auf die Argumentation von O'Donnell und Dansereau (2000) zurückgegriffen werden, welche die Neuheit und Schwierigkeit des Lehr- und Lernmaterials beim Durcharbeiten von Maps als Begründung für eine geringe Lernwirksamkeit bei Lernenden mit niedrigem Vorwissen heranziehen. Des Weiteren fanden Schroeder u. a. (2017) in ihrer Metaanalyse heraus, dass sich größere Effekte für das Concept Mapping zeigen, wenn es über einen längeren Zeitraum angewendet wird, da sich die Lernenden

mit der Zeit besser auf die Verarbeitung der Lerninhalte konzentrieren können (siehe auch Kapitel 3.2.2 und 4.1.1). Das dreimalige Anwenden der Lernstrategie des Concept Mapping in Verbindung mit Feedback als Durcharbeiten von Maps innerhalb von zwei Tagen war möglicherweise nicht ausreichend, um eine tiefgründige Verarbeitung der Lerninhalte über einen längeren Zeitraum zu erwirken. Dabei ist jedoch zu betonen, dass eine häufigere Anwendung der Methode des Concept Mapping allein nicht genügt. Hierbei sollte auch die Komplexität der Lerninhalte berücksichtigt werden. Die im Rahmen des Planspielunterrichts vermittelten Inhalte sind nicht Teil des Lehrplans, sondern werden im Rahmen der Sekundarstufe 2 in berufsbildenden Schulen des Bereichs Wirtschaft behandelt¹¹⁰. Aus diesem Grund würde die Methode des Concept Mapping in Verbindung mit Feedback höchstwahrscheinlich erst in Kombination mit einer kleinschrittigeren Vermittlung der Inhalte über einen längeren Zeitraum hinweg auch eine langfristige lernförderliche Wirkung erzielen.

Die Hypothesen H5, H5b und H6 können angenommen werden. Es zeigt sich kein signifikanter Vorteil des Feedbacks bei hohem Vorwissen. Ausnahme bilden die Ergebnisse des 2. Nachtests. Hier zeigt sich ein signifikanter Vorteil des *EiFB* gegenüber der Gruppe *kein FB* mit einem sehr großen Effekt ($d=3,23$). Die Hypothese H5a kann somit für den 1. Nachtest sowie die Wissenszuwächse über zwei und drei Zeitpunkte angenommen werden, muss jedoch für den 2. Nachtest abgelehnt werden.

Es sind insgesamt Tendenzen erkennbar, die für einen kurz- und langfristigen Vorteil des Feedbacks für Probanden mit hohem Vorwissen sprechen. Das *FB* ist *keinem FB* mit mittleren bzw. hohen Effekten überlegen (zwischen $d=0,48$ und $d=0,68$), wobei sich langfristig das *EiFB* durch hohe Effekte gegenüber *keinem FB* abhebt (2. NT: $d=3,23$; VT-1. NT-2. NT: $d=1,06$). Es ist langfristig auch dem *ElFB* mit kleinen Effekten überlegen (2. NT: $d=-0,38$; VT-1. NT-2. NT: $d=-0,39$). Entgegen der in Kapitel 4.1.1 geäußerten Vermutungen können die Probanden mit hohem Vorwissen die hohe kognitive Belastung, welche das Erstellen von Maps vermutlich verursacht, langfristig nicht kompensieren und benötigen ergänzendes Feedback. Dabei profitieren sie langfristig insbesondere vom *EiFB*. Das *EiFB* ist dem *ElFB* dabei geringfügig überlegen ($d=-0,38$ und $d=-0,39$). Das Präsentieren des Referenzmaps als KCR scheint demnach zu genügen. An dieser Stelle

¹¹⁰ Beispielsweise lernen die angehenden Kaufleute für Büromanagement im 2. Ausbildungsjahr im Lernfeld 6 wie ein Jahresabschluss zu erstellen ist (Sächsisches Staatsministerium für Kultus, 2015).

kann die in Kapitel 4.1.1, unter Bezug auf Krause (2007), dargelegte Begründung bestätigt werden: Vorwissensstarke scheinen die durch das elaborierte Feedback präsentierten zusätzlichen Informationen zumindest langfristig nicht zu benötigen.

Nachfolgend wird der Einfluss der Sprachfähigkeit näher untersucht. In den Hypothesen H7 und H9 wurde behauptet, dass *FB* lediglich bei Probanden mit niedriger Sprachfähigkeit lernwirksamer ist als *kein FB*, wobei sich bei Probanden mit hoher Sprachfähigkeit kein Unterschied zeigen sollte. In Hypothese H8 wurde postuliert, dass *EIFB* bei Probanden mit niedriger Sprachfähigkeit lernwirksamer sein sollte als *EiFB*. Bei Probanden mit hoher Sprachfähigkeit sollten sich *EIFB* und *EiFB* nicht unterscheiden (Hypothese H10).

Die Hypothesen H7 und H7b können für den 1. Nachtest angenommen werden, müssen jedoch für den 2. Nachtest sowie die Wissenszuwächse über zwei und drei Zeitpunkte abgelehnt werden. Bei niedriger Sprachfähigkeit ist das *FB* im 1. Nachtest insgesamt *keinem FB* mit einem großen Effekt signifikant überlegen ($d=0,97$), wobei das *EIFB* gegenüber *keinem FB* einen großen signifikanten Effekt zeigt ($d=1,91$). Die Hypothesen H7a und H8 müssen abgelehnt werden. Es lassen sich Tendenzen ableiten, die kurzfristig auch in Bezug auf den Wissenszuwachs für einen mittleren Effekt des *FB* gegenüber *keinem FB* sprechen ($d=0,42$), wobei der Effekt durch einen hohen Effekt des *EIFB* gegenüber *keinem FB* zum Tragen kommt ($d=0,94$). Langfristig in Bezug auf den 2. Nachtest sind keine Effekte für das Feedback erkennbar, hinsichtlich des Wissenszuwachses über drei Zeitpunkte kehren sich die Effekte sogar um. Das *FB* insgesamt ist *keinem FB* signifikant mit einer hohen Effektstärke unterlegen ($d=-1,11$). Dabei zeigt sich *kein FB* sowohl dem *EiFB* ($d=-1,08$) als auch dem *EIFB* ($d=-1,49$) signifikant überlegen. Kurzfristig ist somit erkennbar, dass entsprechend der Argumentation in Kapitel 4.1.1 ergänzendes Feedback bei Probanden mit niedriger Sprachfähigkeit die kognitive Verarbeitung unterstützt, wobei sich insbesondere ein Vorteil des *EIFB* zeigt. Langfristig, insbesondere hinsichtlich des Wissenszuwachses über drei Zeitpunkte kehren sich die Effekte um. Für Probanden mit niedriger Sprachfähigkeit wirkt das Feedback langfristig scheinbar kontraproduktiv. Wie auch beim Vorwissen könnte die Komplexität des Feedbacks in Kombination mit der zu kurzen Anwendung des Concept Mapping einen Erklärungsansatz bieten. Ein häufigerer und kontinuierlicherer Einsatz der Methode und des Feedbacks würde auch hier

höchstwahrscheinlich erst durch eine ergänzende kleinschrittige Vermittlung der Lerninhalte über einen längeren Zeitraum zu einer Förderung der langfristigen Behaltensleistung beitragen.

Die Hypothesen H9, H9a, H9b und H10 können angenommen werden. Das zusätzliche Feedback zeigt demnach bei Probanden mit hoher Sprachfähigkeit keine signifikante Wirkung. Tendenziell ist dabei das *FB* der Gruppe *kein FB* mit kleinen Effekten sowohl kurzfristig als auch langfristig überlegen (zwischen $d=0,17$ und $d=0,39$). Es zeigt sich hierbei kein eindeutiges Ergebnis für einen Vorteil des *EiFB* bzw. des *ElFB*.

Insgesamt kann sich an die Argumentation in Kapitel 4.1.1 unter Bezug auf Stensvold und Wilson (1990) angelehnt werden. Entsprechend der „Aptitude Treatment Interaction“ sollte Lernenden mit niedrigen verbalen Fähigkeiten eine zusätzliche Unterstützung, z. B. als Feedback, angeboten und Lernenden mit hohen verbalen Fähigkeiten Freiheitsgrade bei der Steuerung ihres Lernens zur Verfügung gestellt werden. Um auch langfristige Effekte des Feedbacks für Lernende mit niedriger Sprachfähigkeit zu erzielen, sollte dabei das Concept Mapping in Verbindung mit ergänzendem Feedback kontinuierlich in den Lernprozess eingebunden werden.

Nachfolgend wird der untersucht, inwieweit die Lernwirksamkeit des Feedbacks in Abhängigkeit von den Aufgabenanforderungen variiert. In den Hypothesen H23, H23a, H23b und H24 wurde behauptet, dass Feedback bei einfachen Aufgaben keine Wirkung zeigen sollte, unabhängig davon, ob die Lernenden *EiFB* oder *ElFB* erhalten. Hypothesen H25, H25a, H25b und H26 postulierten einen Vorteil des *FB* gegenüber *keinem FB*, insbesondere einen Vorteil des *ElFB* bei komplexen Aufgaben. Die Hypothesen H23, H23a, H23b und H24 können angenommen, die Hypothesen H25, H25a, H25b und H26 müssen abgelehnt werden. Es zeigen sich für einfache und komplexe Aufgaben relativ ähnliche Ergebnisse. Tendenziell ist das *FB* kurzfristig der Gruppe *kein FB* mit kleinen Effekten überlegen (zwischen $d=0,22$ und $d=0,38$), wobei dies auf kleine bzw. mittlere Effekte des *ElFB* gegenüber *keinem FB* zurückzuführen ist (zwischen $d=0,26$ und $d=0,67$). Langfristig ist keine eindeutige Richtung erkennbar. Die Analysen in Bezug auf den 2. Nachtest sowie den Wissenszuwachs über drei Zeitpunkte zeigen sowohl für einfache als auch für komplexe Aufgaben gegensätzliche Effekte, die einerseits für einen geringen Vorteil des *FB* gegenüber *keinem FB* und andererseits keine Überlegenheit des Feedbacks zeigen.

Insgesamt stellt sich die Frage, inwieweit der Einfluss der Aufgabenanforderungen auf die Lernwirksamkeit des Feedbacks von weiteren Faktoren abhängt. Zu diesem Zweck wurde der Einfluss des Vorwissens, der Sprachfähigkeit, des eingeschätzten Umgangs mit Fehlern, der eingeschätzten Interessantheit-Nützlichkeit sowie der eingeschätzten Handhabbarkeit analysiert. Es zeigten sich signifikante Einflüsse der Sprachfähigkeit, des Vorwissens sowie des Faktors Interessantheit-Nützlichkeit. Weitere Analysen mit den dichotomisierten Variablen zeigten lediglich einen signifikanten Effekt der Sprachfähigkeit auf den Wissenszuwachs komplexer Aufgaben über drei Zeitpunkte. Hierbei vergrößerten sich die Effekte für das *FB* gegenüber *keinem FB* (von $d=0,30$ auf $d=0,69$) bei hoher Sprachfähigkeit. Der Effekt des *EiFB* gegenüber *keinem FB* wird sogar signifikant (von $d=0,45$ auf $d=1,05$). Insgesamt scheinen Probanden mit hoher Sprachfähigkeit bei komplexen Aufgabenanforderungen langfristig vor allem vom *EiFB* zu profitieren. Demnach sollten Freiheitsgrade in Abhängigkeit von den Aufgabenanforderungen angeboten werden. Um einen langfristigen Behaltenseffekt bei komplexen Anforderungen zu gewährleisten, empfiehlt es sich, Lernende mit hoher Sprachfähigkeit durch einfaches Feedback zu unterstützen. Bei einfachen Anforderungen ist das Erstellen von Concept Maps als Lernstrategie auch ohne ergänzendes Feedback ausreichend.

Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps

Nachfolgend werden die Ergebnisse aus Kapitel 5.1.2 zur Analyse des Einflusses der Art des Feedbacks auf die Qualität der Maps näher untersucht.

Insgesamt wurden drei Concept Maps erstellt. Dabei zeigt sich eine signifikant sinkende Qualität für die Gruppen *kein FB*, *EiFB* und *EIFB* über die drei Mappings hinweg. Eine mögliche Ursache hierfür kann mit Blick auf die Themenbereiche der erstellten Maps gefunden werden. Im ersten zu erstellenden Map sollten die Abläufe in einem Produktionsunternehmen dargestellt werden. Hierbei konnte zusätzlich zum im Planspielunterricht erworbenen Wissen auf erlerntes Wissen aus dem Fach Wirtschaft-Technik-Haushalt, aber auch auf Alltagswissen zurückgegriffen werden. Für die 1. Mapping-Aktivität konnten somit die höchsten Scoringwerte erzielt werden ($\text{ScorM1}_{\text{Median}}=0,22$; vgl. auch Kapitel 4.5.2). Das 2. Mapping zu den Zusammenhängen zwischen linearer Abschreibung, GuV und Bilanz stellte ein vollkommen neues Inhaltsgebiet dar, welches nicht Teil des Lehrplans war und womit die Schüler auch in ihrem Alltag wenig konfrontiert wer-

den. Infolgedessen wurden auch Maps mit einer niedrigeren Qualität erstellt ($ScorM2_{Median}=0,15$). Das 3. Mapping verlangte den Probanden am meisten ab, da sie die Zusammenhänge zwischen 1. und 2. Map herstellen sollten. Die Scoringwerte waren folglich für das 3. Mapping am niedrigsten ($ScorM3_{Median}=0,07$).

Die Hypothesen H1, H1a und H1b können in Bezug auf die Qualität des 3. Maps sowie die Entwicklung der Mapqualität über die drei Mapping-Aktivitäten hinweg angenommen werden. Das *FB* ist signifikant besser als *kein FB*, wobei sowohl *EiFB* als auch *ElFB* der Gruppe *kein FB* signifikant überlegen sind. Dabei zeigen sich mittlere bis große Effekte zwischen $d=0,69$ und $d=1,40$. Für das 2. Mapping müssen die Hypothesen H1, H1a und H1b abgelehnt werden. Die Hypothese H2 muss ebenfalls abgelehnt werden, wobei das *ElFB* mit mittleren Effekten dem *EiFB* hinsichtlich der Qualität des 3. Maps ($d=0,51$) sowie der Entwicklung der Map-Qualität ($d=0,66$) tendenziell überlegen ist.

Beim 2. Mapping zeigen sich eher gegensätzliche Ergebnisse. Hier scheint Feedback tendenziell sogar kontraproduktiv zu wirken. Das *FB* insgesamt ist *keinem FB* ($d=-0,47$) und dabei sowohl dem *EiFB* ($d=-0,55$) als auch dem *ElFB* ($d=-0,42$) mit kleinen bis mittleren Effekten tendenziell überlegen, wobei sich keine Unterschiede zwischen *EiFB* und *ElFB* zeigen.

Die Scoringwerte für das 1. Mapping wurden als Kovariate mit aufgenommen, da bis zum Erstellen des 1. Maps noch keine Intervention stattgefunden hatte und die Gruppenunterschiede daher auf andere Ursachen zurückgeführt werden müssen. Unter Einbezug des Vorwissens, der Sprachfähigkeit, des eingeschätzten Umgangs mit Fehlern, der eingeschätzten Interessantheit-Nützlichkeit und der eingeschätzten Handhabbarkeit des Concept Mapping als weitere Kovariaten vergrößern sich die Unterschiede insbesondere für das 2. Mapping. Dabei zeigt sich lediglich ein signifikanter Einfluss von Interessantheit-Nützlichkeit und Handhabbarkeit. Das *FB* insgesamt sowie das *EiFB* und das *ElFB* jeweils separat sind *keinem FB* nun mit mittleren bzw. hohen Effekten tendenziell unterlegen (zwischen $d=-0,61$ und $d=-0,84$). Der Vorteil von *keinem FB* gegenüber *ElFB* wird dabei signifikant.

Der signifikante Einfluss der eingeschätzten Interessantheit-Nützlichkeit und Handhabbarkeit des Concept Mapping auf die Qualität des 2. Mappings wurde genauer analysiert. In den Hypothesen H15 bis H22 wurde postuliert, dass bei Probanden, welche die Inte-

ressantheit-Nützlichkeit bzw. Handhabbarkeit des Concept Mapping als niedrig einschätzen, das *FB* weniger wirksam ist als *kein FB*, wobei *kein FB* sowohl *EiFB* als auch *ElFB* überlegen sein sollte. Das *ElFB* sollte dabei weniger lernwirksam sein als *EiFB*. Bei Probanden, welche die Interessantheit-Nützlichkeit bzw. Handhabbarkeit des Concept Mapping als hoch einschätzen, sollte *FB* der Gruppe *kein FB* überlegen sein, wobei sich der Vorteil sowohl für *EiFB* als auch für *ElFB* zeigen sollte. *ElFB* sollte dem *EiFB* überlegen sein.

Insgesamt kann lediglich die Hypothese H15b angenommen werden. Probanden welche die Interessantheit-Nützlichkeit des Concept Mapping als niedrig einschätzen und *kein FB* erhalten, sind der Gruppe *ElFB* überlegen ($d=-1,29$). Wie vermutet, scheint das elaborierte Feedback das Lernen für Probanden, die die Interessantheit-Nützlichkeit als niedrig einschätzen, zu behindern. Diese Probanden haben vermutlich noch Schwierigkeiten im Umgang mit der Methode Concept Mapping oder aber sind der Meinung, dass ein ergänzendes Concept Mapping zum Wissenserwerb nicht notwendig ist. Aufgrund dessen, dass das Feedback ebenfalls die Anwendung von Concept Maps beinhaltet, ist die kognitive Belastung infolgedessen zu hoch oder unnötig. Tendenziell zeigt sich ein Nachteil des *FB* insgesamt ($d=-0,68$) sowie des *EiFB* ($d=-0,39$) gegenüber *keinem FB*. Entgegen der Vermutung, dass bei einer hohen Einschätzung das Feedback überlegen ist, zeigt sich auch hier ein tendenzieller Vorteil der Gruppe *kein FB* gegenüber dem *FB* insgesamt ($d=-0,38$) und gegenüber *EiFB* ($d=-0,85$). Lediglich das *ElFB* zeigt gegenüber *keinem FB* keinen Nachteil ($d=-0,11$). Obwohl die Probanden also das Concept Mapping als interessant und nützlich einschätzen, können Sie vom Feedback nicht profitieren. Für eine detailliertere Analyse wurden die Gruppen zusätzlich separat untersucht. Hierbei zeigt sich bei der Gruppe *ElFB*, dass Probanden mit einer hohen Einschätzung für die Interessantheit –Nützlichkeit im Vergleich zu Probanden mit einer niedrigen Einschätzung auch Concept Maps mit einer höheren Qualität erstellen ($d=1,02$). Hier scheint demzufolge ein hoher Einfluss der eingeschätzten Interessantheit-Nützlichkeit vorzuliegen. (vgl. auch Abbildung 29, Kapitel 5.1.2.2). Dieses Ergebnis ist nachvollziehbar, wenn man die Gestaltung des elaborierten Feedbacks berücksichtigt. Um vom elaborierten Feedback profitieren zu können, stellt die Einschätzung des Concept Mapping als interessant und nützlich eine Grundvoraussetzung dar. Im Anschluss an das Vergleichen des erstellten Maps mit dem Referenzmap werden Fehler anhand eines Beispielmaps sowie das Referenzmap

in einem Lehrer-Schüler-Gespräch-besprochen. Sobald Probanden nun das Concept Mapping als nicht nützlich und interessant erachten, sind sie vermutlich auch nicht motiviert bzw. nicht in der Lage den Ausführungen des Lehrers zu folgen, was das Lernen an dieser Stelle deutlich hemmen kann.

Beim Einfluss der eingeschätzten Handhabbarkeit zeigen sich ähnliche Ergebnisse. Unabhängig davon, ob die Handhabbarkeit hoch oder niedrig eingeschätzt wurde, ist das *FB* mit kleinen bis mittleren Effekten *keinem FB* unterlegen (zwischen $d=-0,43$ und $d=-0,64$). Innerhalb der Gruppen *kein FB*, *EiFB* und *EIFB* erstellen Probanden, welche die Handhabbarkeit als niedrig einschätzen tendenziell bessere Maps als Probanden, welche die Handhabbarkeit als hoch einschätzen (vgl. auch Abbildung 30, Kapitel 5.1.2.2). Dieses etwas überraschende Ergebnis könnte mit der Kompetenzillusion (Fürstenau, 2011, S. 47) zusammenhängen. Da die Probanden im Anschluss an das Erstellen der Maps keine individuelle Rückmeldung, sondern kein Feedback bzw. ein Gruppenfeedback erhalten, könnte es sein, dass ein Teil der Probanden die eigenen Fähigkeiten in Bezug auf die Anwendung der Methode des Concept Mapping überschätzt. Infolgedessen wird die Handhabbarkeit als hoch eingeschätzt, obwohl die erstellten Maps eine niedrige Qualität haben. Der andere Teil hingegen ist eher in der Lage, die eigenen Fähigkeiten hinsichtlich der Methode des Concept Mapping zu bewerten. Diese Probanden gehen reflektierter vor und erkennen Schwächen im Umgang mit der Methode. Infolgedessen bewerten sie die Handhabbarkeit als niedrig, obwohl sie Maps mit einer hohen Qualität erstellen.

Insgesamt bleibt zu hinterfragen, warum das Feedback beim 2. Mapping nicht lernwirksam ist und sogar kontraproduktiv wirkt. Für eine mögliche Begründung können die Themenbereiche der drei Maps betrachtet werden. Im ersten Map werden die Abläufe in einem Produktionsunternehmen von der Zulieferung über die Fertigung hin zum Verkauf dargestellt. Im 2. Map gilt es, die Zusammenhänge zwischen linearer Abschreibung, GuV und Bilanz zu finden. Hierzu kann somit kaum auf die durch das Feedback zum ersten Map erworbenen Kenntnisse zurückgegriffen werden. Das Feedback führt möglicherweise sogar zu einer kognitiven Überlastung der Probanden der Gruppen *EiFB* und *EIFB*. Sie erhalten zu viel Input, ohne das Wissen direkt im 2. Mapping anwenden zu können. Für Probanden, die die Interessantheit-Nützlichkeitswert als niedrig einschätzen und elaboriertes Feedback erhalten, könnte darüber hinaus die fehlende Motivation die Wirkung des Feedbacks beeinträchtigen. Hier zeigen sich besonders hohe negative Effekte. Eine hohe Einschätzung der Interessantheit-Nützlichkeitswert geht zumindest damit einher, dass *EIFB*

gegenüber *keinem FB* keinen Nachteil hat. Das *EiFB* können selbst Probanden, welche die Interessantheit-Nützlichkeit als hoch einschätzen, nicht nutzen. In der 3. Mapping-Aktivität sollen Verbindungen zwischen 1. und 2. Map gefunden werden. Die Wirkung des Feedbacks zum 1. und 2. Map kann somit erst hier richtig zum Tragen kommen.

Einfluss der Qualität der Maps auf den Wissenserwerb

Des Weiteren werden die Ergebnisse zu den Analysen des Einflusses der Mapqualität auf den Wissenserwerb interpretiert (Hypothesen H33-H39).

Insgesamt zeigt sich kein signifikanter Vorteil des *FB* gegenüber *keinem FB* bei niedriger Mapqualität. Daher müssen die Hypothesen H33, H33a und H33b abgelehnt werden. Eine niedrige Mapqualität im 2. Mapping führt bei Probanden der Gruppe *EiFB* gegenüber dem *EiFB* im 1. Nachtest zu signifikant besseren Ergebnisse ($d=1,03$). Die Hypothese H34 kann demzufolge dahingehend angenommen werden. In Bezug auf den 2. Nachtest, die Wissensentwicklung über zwei bzw. drei Zeitpunkte sowie hinsichtlich des 3. Map-pings muss die Hypothese allerdings abgelehnt werden.

Die Hypothesen H35a, H35b und H36 können angenommen werden. Bei einer hohen Mapqualität zeigt sich demnach kein signifikanter Vorteil des *EiFB* bzw. *EiFB* gegenüber *keinem FB*. Hierbei spielt die Art des Feedbacks auch keine Rolle. Entgegen der Vermutung nimmt jedoch bei Probanden mit einer hohen Mapqualität im 2. Mapping, welche Feedback erhalten, das Wissen vom Vor- zum 1. Nachtest gegenüber Probanden, welche kein Feedback erhalten signifikant mehr zu ($d=0,84$). Die Hypothese H35 muss somit dahingehend abgelehnt werden. Sie kann jedoch in Bezug auf den 1. Nachtest, den 2. Nachtest sowie die Wissensentwicklung über drei Zeitpunkte angenommen werden.

Tendenziell ist kurzfristig keine eindeutige Richtung der Ergebnisse erkennbar. Die Ergebnisse unterscheiden sich insbesondere zwischen dem 2. und 3. Mapping. Kurzfristig in Bezug auf das 2. Mapping zeigt *FB* gegenüber *keinem FB* bei einer niedrigen Mapqualität einen kleinen Effekt ($d=0,16$; $d=0,21$). Bei einer hohen Mapqualität ist *FB* kurzfristig mit hohen Effekten gegenüber *keinem FB* überlegen ($d=0,77$; $d=0,84$). In Bezug auf das 3. Mapping bestätigt sich dieser Vorteil jedoch nicht.

Langfristig in Bezug auf die Ergebnisse des 2. Nachtests profitieren lediglich Probanden mit einer niedrigen Map-Qualität vom Feedback. Das *FB* allgemein ist dabei mit einem mittleren Effekt (2. Mapping: $d=0,44$; 3. Mapping: $d=0,44$) *keinem FB* überlegen, wobei

der Vorteil auf das *EIFB* zurückzuführen ist ($d=0,71$ bzw. $0,84$). Das *EIFB* ist dabei dem *EiFB* tendenziell überlegen ($d=0,57$; $d=0,40$). Bei hoher Map-Qualität wirkt das Feedback langfristig in Bezug auf den 2. Nachtest sowie die Wissensentwicklung über drei Zeitpunkte kontraproduktiv. Es zeigen sich insbesondere in Bezug auf das 3. Mapping mittlere bis hohe negative Effekte (zwischen $d=-0,50$ und $d=-1,71$).

Entsprechend der in Kapitel 4.1 geäußerten Vermutungen scheinen somit Probanden, die qualitativ schlechtere Maps erstellen, langfristig tendenziell insbesondere vom elaborierten Feedback zu profitieren. Das Feedback lenkt die Aufmerksamkeit auf die Fehler in den Maps. Die Lernenden können folglich direkt aus ihren Fehlern lernen. Es werden Reflexionsprozesse angeregt und somit vermutlich die langfristige Behaltensleistung verbessert. Kurzfristig zeigt sich dieser Effekt aber noch nicht.

Bei einer hohen Qualität wirkt das Feedback langfristig tendenziell sogar hinderlich. Dieses Ergebnis könnte auf den sog. Expertise Reversal Effect¹¹¹ zurückgeführt werden. Möglicherweise benötigen Probanden, welche Maps mit einer hohen Qualität erstellen, keine oder nur wenig instruktionale Unterstützung. Der Wissenserwerb wird durch die redundanten Informationen, welches das Feedback liefert, somit vermutlich behindert.

Zusätzlich wurde der Einfluss der Mapqualität auf den Wissenserwerb jeder der drei Versuchsgruppen separat untersucht. Entgegen der Vermutung zeigt sich bei den beiden Feedbackgruppen keine signifikante Überlegenheit von Probanden, welche Maps mit einer niedrigeren Qualität erstellten gegenüber Probanden, welche qualitativ bessere Maps erstellten. Daher müssen die Hypothesen H37 und H38 abgelehnt werden. Bei Probanden, die kein Feedback erhalten, sollte eine hohe Qualität der Maps gegenüber einer schlechteren Qualität überlegen sein (H39). In Bezug auf das 3. Mapping kann die Hypothese H39 für den 2. Nachtest angenommen werden, muss jedoch für den 1. Nachtest sowie die Wissensentwicklung über zwei und drei Zeitpunkte abgelehnt werden. Hinsichtlich des 2. Mappings muss H39 ebenfalls abgelehnt werden.

¹¹¹ Expertise Reversal Effect bezieht sich auf die Wirkung instruktionaler Unterstützung auf den Lernerfolg in Abhängigkeit von der Expertise. Während Lernende mit geringem Vorwissen eine lernunterstützende didaktische Gestaltung benötigen, kann der Wissenserwerb von Personen mit hohem Vorwissen dadurch sogar behindert werden. Auf Grundlage der Cognitive Load Theory wird dies damit begründet, dass das Arbeitsgedächtnis mit den redundanten Informationen belastet wird (Kalyuga, Ayres, Chandler, & Sweller, 2003). Möglicherweise ist der Effekt auf Probanden mit einer hohen Mapqualität übertragbar.

Tendenziell sind kurzfristig (1.NT; VT-NT), entgegen der Vermutungen, die Probanden der FB-Gruppen mit hoher Mapqualität denen mit niedriger Mapqualität überlegen (zwischen $d=0,49$ und $d=1,25$), wobei sich dieser Effekt beim 3. Mapping lediglich für *EIFB* ($d=0,55$ und $d=0,46$) zeigt. Hinsichtlich des langfristigen Wissenserwerbs ist es gemäß der Vermutungen umgekehrt, so dass sich das Wissensniveau im 2. Nachtest wieder angleicht und es kaum Unterschiede zwischen hoher und niedriger Mapqualität mehr gibt (zwischen $d=-0,17$ und $d=0,06$).

Hinsichtlich der Wissensentwicklung über drei Zeitpunkte muss zwischen dem 2. und 3. Mapping differenziert werden. In Bezug auf das 2. Mapping nimmt das Wissen beider FB-Gruppen bei niedriger Map-Qualität geringer ab als bei hoher Map-Qualität ($d=-0,70$ und $d=-0,50$). Hinsichtlich des 3. Mapping nimmt nur das Wissen der Gruppe *EIFB* bei niedriger Qualität der Maps geringer ab als bei hoher Qualität ($d=-0,54$). Es zeigen sich mittlere bis hohe Effekte.

Kurzfristig profitieren eher Probanden mit hoher Mapqualität als die mit niedriger Mapqualität vom Feedback. Ausnahme bildet das *EiFB* für das 3. Mapping. Langfristig regt niedrige Qualität beim 2. Mapping bei beiden FB-Gruppen eventuell Reflexionsprozesse an, hilft dabei das Feedback für das 2. Mapping besser zu verarbeiten und erhöht somit den langfristigen Behaltenseffekt. In Bezug auf das 3. Mapping wirkt beim *EiFB* die niedrige Qualität gegenüber hoher Qualität eher kontraproduktiv für die Entwicklung des Wissens über drei Zeitpunkte. Das Feedback wird hier ziemlich am Ende der Intervention gegeben, so dass für die Probanden der Gruppe *EiFB* mit niedriger Qualität aufgrund der Komplexität des 3. Maps das Präsentieren der richtigen Lösung für eine angemessene Wissensentwicklung nicht ausreicht. Hinsichtlich der Entwicklung des Wissens über drei Zeitpunkte können die Probanden mit niedriger Qualität für das 3. Mapping der Gruppe *EIFB* gegenüber denen mit hoher Map-Qualität profitieren. Die Probanden der Gruppe *EIFB* besprechen das Map ausführlich in der Gruppe, so dass das Feedback nachhaltiger verarbeitet werden kann und eher Reflexionsprozesse angeregt werden. Kurzfristig hat das noch keine Wirkung bzw. profitieren die Probanden eher von hoher Mapqualität. Aber langfristig wirkt das *EIFB* für Probanden mit niedriger Mapqualität nachhaltiger, weil möglicherweise das Lernen aus den eigenen Fehlern den langfristigen Behaltenseffekt erhöht. Bei den Probanden mit hoher Mapqualität ist das Lernpotential nicht so hoch, sie erhalten eher eine Bestätigung ihrer Leistung durch das elaborierte Feedback

und werden somit weniger zum Lernen aus Fehlern bzw. zum Reflektieren allgemein angeregt.

Bei der Gruppe *kein FB* zeigt sich langfristig beim 3. Mapping ein deutlicher Vorteil der Probanden mit hoher Qualität gegenüber denen mit niedriger Qualität ($d=1,53$ bzw. $d=0,96$). Die Probanden mit einer hohen Qualität für ScorM3 sind in der Lage ihr erlerntes Wissen in einem Concept Map angemessen darzustellen, ohne weitere Unterstützung durch ein ergänzendes Feedback zu erhalten. Für diese Gruppe ist es daher deutlich schwieriger eine hohe Qualität zu erzielen, als für die FB-Gruppen. Sie sind in der Lage neues Wissen in bestehende kognitive Strukturen so zu integrieren, so dass es langfristig gespeichert wird. Verglichen mit allen anderen Versuchsgruppen erzielen sie den höchsten Wissenszuwachs vom Vor- zum 1. Nachtest, wobei sie in der Lage sind, das Wissensniveau vom 1. Nachtest zum 2. Nachtest fast konstant zu halten (vgl. Abbildung 22, Kapitel 5.1.3.3).

Umgang mit dem Feedback und Fehleranalyse der erstellten Concept Maps

Abschließend werden die Ergebnisse zum Umgang mit dem Feedback (H27-H32) sowie die Ergebnisse der Fehleranalyse der Concept Maps näher analysiert.

Die Analysen zur Einschätzung der Nützlichkeit des Feedbacks haben gezeigt, dass sich die Probanden der Gruppe *kein FB* eine Rückmeldung zu ihren Concept Maps gewünscht hätten (vgl. Kapitel 5.2.1). Dieser Wunsch ist dabei signifikant höher ausgeprägt als der Wunsch der Gruppe *EiFB* nach einer weiteren Fehlerbesprechung, also nach elaboriertem Feedback ($d=1,01$). Entsprechend der Vermutung kann daher Hypothese H27 bestätigt werden. Es zeigt sich ein mittlerer signifikanter Zusammenhang zwischen der Einschätzung zur Nützlichkeit des einfachen und des elaborierten Feedbacks (für EIFBN_3; $r_s=0,654$). Hypothese H29 kann daher für das Item EIFBN_3 angenommen werden. Probanden der Gruppe *EIFB*, welche das einfache Feedback als nützlich erachten, schätzen auch das elaborierte Feedback als nützlich ein und umgekehrt. Beide Feedback-Arten stehen also in einem engen Zusammenhang und sollten aufeinander abgestimmt werden. Zwischen den Variablen zur Einschätzung des Feedbacks und dem Wissenserwerb sowie der Qualität der Maps gibt es keinen Zusammenhang. Die Hypothesen H31a, H31b, H32a und H32b müssen demnach abgelehnt werden.

Die Schüler der Gruppe *ElFB* schätzen das einfache Feedback signifikant nützlicher ein, als die Schüler der Gruppe *EiFB*. Hypothese H28 kann daher bestätigt werden ($d=1,20$). Hierbei liegt die Qualität der Feedback-Netze (der korrigierten Maps) der Gruppe *ElFB* jedoch deutlich unter der Qualität der Gruppe *EiFB*. Die Schüler der Gruppe *ElFB* wenden im 2. Mapping die Strategie des Streichens und Korrigierens von falschen Propositionen häufiger an als die Gruppe *EiFB* und setzen sich somit intensiver mit den erstellten Maps auseinander. Auf der anderen Seite wird das Hinzufügen von Propositionen von der Gruppe *EiFB* häufiger angewandt als von der Gruppe *ElFB*, d. h. sie übernehmen mehr Propositionen aus dem Referenznetz des 2. Mapping. An dieser Stelle könnten Motivationsaspekte eine Rolle spielen. Bei der Gruppe *ElFB* folgte die Besprechung mit der Lehrperson, wohingegen die Schüler der Gruppe *EiFB* ihre Maps lediglich mit den Referenzmaps verglichen, ohne weitere Informationen zu erhalten. Daher sind die Schüler der Gruppe *EiFB* weniger motiviert, ihre eigenen Maps zu durchdenken und Fehler zu erkennen und zu berichtigen. Dies führt zu einer niedrigeren Einschätzung der Nützlichkeit des Feedbacks. Das einfache Feedback allein scheint aus ihrer Sicht nicht auszureichen. Das einfache Übernehmen der Propositionen aus dem Referenznetz führt auf der anderen Seite zu einer höheren Qualität der Feedback-Netze. Insgesamt ist auffällig, dass circa ein Drittel der verwendeten Verarbeitungsstrategien des einfachen Feedbacks bei beiden Gruppen auf das Hinzufügen falscher oder nur teilweise richtiger Propositionen entfallen. Dies deutet darauf hin, dass die Schüler unkonzentriert waren oder die Zusammenhänge nicht ausreichend verstanden haben.

Die Fehleranalyse ergab, dass die Schüler z. T. ein falsches Verständnis der Begriffe und Zusammenhänge haben, welches sich im Zeitverlauf des Erstellens der drei Mappings auch nicht korrigierte (z. B. Forderungen, Lieferant). Des Weiteren wurde insbesondere das Zusammenspiel zwischen GuV und Bilanz nicht richtig verstanden. Die Schüler verwenden die Begriffe Aufwendungen und Aktivseite synonym und ordnen somit z. B. die Herstellungskosten oder die Gemeinkosten der Aktivseite zu. Sachverhalte werden oft vereinfacht, dabei fehlt es an Trennschärfe sowohl bei den Begriffen als auch den Prozessen, z. B. bei den Konzepten Anschaffungs- und Herstellungskosten. Die Denkweise der Schüler ist insgesamt sehr oberflächlich. Das Darstellen der im Planspiel abgebildeten komplexen Zusammenhänge bereitet ihnen noch große Schwierigkeiten.

6 Schlussbetrachtung

6.1 Zusammenfassung

Im Rahmen der Untersuchung war es Ziel zu analysieren, inwieweit das Lernen im Planspielunterricht durch ergänzendes Feedback zu den in Concept Maps dargestellten betriebswirtschaftlichen Zusammenhängen unterstützt werden kann und welche Art von Feedback lernwirksamer ist. Die lernförderliche Wirkung sowohl von Concept Maps als auch von Feedback wurde dabei in zahlreichen Studien nachgewiesen (vgl. Kapitel 2 und 3). Inwieweit das Erstellen von Concept Maps als Lernstrategie durch ein Feedback als ergänzende instruktionale Maßnahme unterstützt werden sollte, wurde bisher jedoch kaum untersucht. Diesem Defizit wurde im Rahmen der Untersuchung begegnet. Hieraus ergaben sich drei Forschungsfragen.

- F1) Inwieweit ist ergänzendes Feedback zu Concept Maps lernwirksamer als kein ergänzendes Feedback?
- F2) Ist elaboriertes Feedback lernwirksamer als einfaches Feedback?
- F3) Inwieweit beeinflussen das Vorwissen, die Sprachfähigkeit, die Einschätzung des Umgangs mit Concept Mapping, die Einschätzung der Nützlichkeit des Feedbacks, die Einschätzung des Umgangs mit Fehlern sowie die Aufgabenanforderungen der Lernaufgabe die Lernwirksamkeit des Feedbacks?

Die Lernwirksamkeit des Feedbacks wurde zum einen durch den Wissenserwerb (kurzfristig: 1. Nachtest, Wissenszuwachs Vortest-1. Nachtest; langfristig: 2. Nachtest, Wissensentwicklung Vortest-1. Nachtest-2. Nachtest) und zum anderen durch die Qualität der drei erstellten Maps (ScorM1, ScorM2, ScorM3) sowie die Entwicklung der Qualität über die drei Mapping-Aktivitäten hinweg operationalisiert. Zur Analyse der Qualität der Maps wurden zunächst zwei Bewertungsverfahren durchgeführt. Zunächst wurde der Übereinstimmungsgrad mit dem jeweiligen Referenznetz auf Basis der Galanter-Metrik berechnet. Zusätzlich wurde ein differenziertes Scoringverfahren ebenfalls auf Basis der Galanter-Metrik entwickelt, welches die Besonderheiten der Zielgruppe berücksichtigte. Aufgrund der Komplexität der dargebotenen Inhalte waren die Probanden nicht in der Lage, die Zusammenhänge in der gebotenen Tiefe in den Concept Maps darzustellen. Sie „übersprangen“ oftmals mehrere Konzepte, verwendeten die falsche Relation und/ oder gaben eine andere Richtung an, so dass in der Folge Propositionen dargestellt wurden, die es zwar so nicht im Referenznetz gab, die aber weder falsch noch richtig waren. Diese

„halbrichtigen“ Verbindungen wurden bei dem differenzierten Scoringschema mit berücksichtigt. Für eine weitere Analyse der Concept Maps wurden beide Bewertungsverfahren gegenübergestellt. Es zeigte sich eine hohe Korrelation zwischen den Werten beider Verfahren. Für die Studie wurden die Werte des differenzierten Verfahrens herangezogen, da diese das tatsächliche Wissen der Probanden besser widerspiegeln.

Zunächst werden die Ergebnisse hinsichtlich des Wissenserwerbs zusammengefasst. Es zeigt sich ein kurzfristiger signifikanter Vorteil des *EIFB* gegenüber *keinem FB* ($d=0,69$) in Bezug auf die Ergebnisse des 1. Nachttests. Die Hypothese H1b konnte somit zunächst angenommen werden. Um zu analysieren, inwieweit Vorwissen, Sprachfähigkeit, der eingeschätzte Umgang mit dem Concept Mapping (Interessantheit-Nützlichkeit sowie Handhabbarkeit) sowie der Umgang mit Fehlern einen Einfluss auf die Lernwirksamkeit des Feedbacks haben, wurden diese Faktoren als Kovariaten einbezogen. Unter Herauspartialisieren der Kovariaten verschwinden die signifikanten Vorteile des *EIFB* gegenüber *keinem FB* im 1. Nachttest, es zeigt sich nur noch ein mittlerer tendenzieller Effekt ($d=0,52$), die Hypothese H1b musste nun abgelehnt werden. Des Weiteren zeigt das *EiFB* nun einen kleinen Effekt gegenüber *keinem FB* ($d=0,37$), *EiFB* und *EIFB* unterscheiden sich nicht. In Bezug auf den Wissenszuwachs vom Vor- zum 1. Nachttest zeigen sich ähnliche Ergebnisse (*EIFB*>*kein FB*, $d=0,46$; *EiFB*=*kein FB*, $d=0,09$). Das *EIFB* ist dabei dem *EiFB* mit einem kleinen Effekt hinsichtlich des Wissenszuwachses vom Vor- zum 1. Nachttest ($d=0,35$) tendenziell überlegen.

Kurzfristig zeigt sich in Bezug auf die Ergebnisse des 1. Nachttests und des Wissenszuwachses vom Vor- zum 1. Nachttest somit eine tendenzielle Lernwirksamkeit insbesondere des elaborierten Feedbacks.

Langfristig in Bezug auf die Ergebnisse im 2. Nachttest verschwinden die kleinen Effekte zugunsten des Feedbacks unter Herauspartialisieren der Einflussfaktoren, es zeigen sich keine Effekte zugunsten des *EiFB* bzw. *EIFB*. Hinsichtlich des Wissenszuwachses über drei Zeitpunkte kehren sich die Effekte sogar um, so dass *kein FB* sowohl dem *EiFB* ($d=-0,29$) als auch dem *EIFB* ($d=-0,33$) mit einem kleinen Effekt tendenziell überlegen ist.

Anschließend wurde der Einfluss weiterer personaler Faktoren untersucht. Hierzu wurde der Einfluss des Vorwissens und der Sprachfähigkeit genauer analysiert, da bei diesen Kovariaten ein signifikanter Einfluss auf die Wissenstestergebnisse festgestellt wurde.

Hierzu wurden die Variablen Vorwissen und Sprachfähigkeit am Median dichotomisiert in hoch und niedrig eingeteilt.

Kurzfristig in Bezug auf den 1. Nachtest zeigen sich signifikante Effekte zugunsten des Feedbacks bei Probanden mit niedriger Sprachfähigkeit. Das *FB* insgesamt ist dabei *keinem FB* signifikant überlegen ($d=0,97$), wobei der Einfluss aus der Überlegenheit des *ElFB* gegenüber *keinem FB* resultiert ($d=1,91$). Die Hypothesen H7 und H7b konnten demnach angenommen werden. Das *EiFB* ist *keinem FB* tendenziell mit einem mittleren Effekt ($d=0,79$) überlegen. In Bezug auf den Wissenszuwachs zeigt sich ein kleiner Effekt des *FB* gegenüber *keinem FB* ($d=0,42$), wobei der Einfluss hier ebenfalls aus dem Vorteil des *ElFB* gegenüber *keinem FB* resultiert ($d=0,94$). Das *ElFB* ist dem *EiFB* mit mittleren Effekten kurzfristig tendenziell überlegen ($d=0,64$ bzw. $d=0,57$).

Hinsichtlich des Vorwissens zeigen sich ähnliche Effekte, wobei diese nicht signifikant sind. Bei Probanden mit niedrigem Vorwissen ist das *ElFB* kurzfristig in Bezug auf den 1. Nachtest ($d=0,38$) sowie den Wissenszuwachs von Vor- zu 1. Nachtest ($d=0,31$) *keinem FB* tendenziell überlegen. Das *EiFB* zeigt gegenüber *keinem FB* keinen Effekt, so dass *ElFB* dem *EiFB* mit einem mittleren Effekt tendenziell überlegen ist ($d=0,57$).

Kurzfristig wirkt folglich insbesondere elaboriertes Feedback bei Probanden mit niedriger Sprachfähigkeit und niedrigem Vorwissen.

Langfristig verschwindet dieser Effekt bzw. kehrt sich um. Bei niedriger Sprachfähigkeit zeigt *FB* hinsichtlich der Ergebnisse des 2. Nachtests gegenüber *keinem FB* keinen Effekt, in Bezug auf die Wissensentwicklung über drei Zeitpunkte ist *FB* *keinem FB* sogar signifikant mit hohen Effekten unterlegen. Bei niedrigem Vorwissen wirkt das Feedback ebenfalls kontraproduktiv, so dass sich kleine bis mittlere negative Effekte zeigen.

Das Feedback (einfaches und elaboriertes Feedback) wirkt somit langfristig bei niedriger Sprachfähigkeit und niedrigem Vorwissen kontraproduktiv. Nun stellt sich die Frage, warum ergänzendes Feedback bei geringen verbalen Fähigkeiten und einem niedrigen Vorwissen die kognitive Belastung beim Erstellen der Maps scheinbar nur kurzfristig reduzieren kann. Das Feedback ist möglicherweise zu komplex, um eine langfristige Behaltensleistung zu erwirken. Letztlich stellt das einfache Feedback im Rahmen der Studie ein selbstständiges Durcharbeiten vorstrukturierter Maps dar, das elaborierte Feedback entspricht einem gemeinsamen Durcharbeiten der Fehler- und Referenzmaps mit der

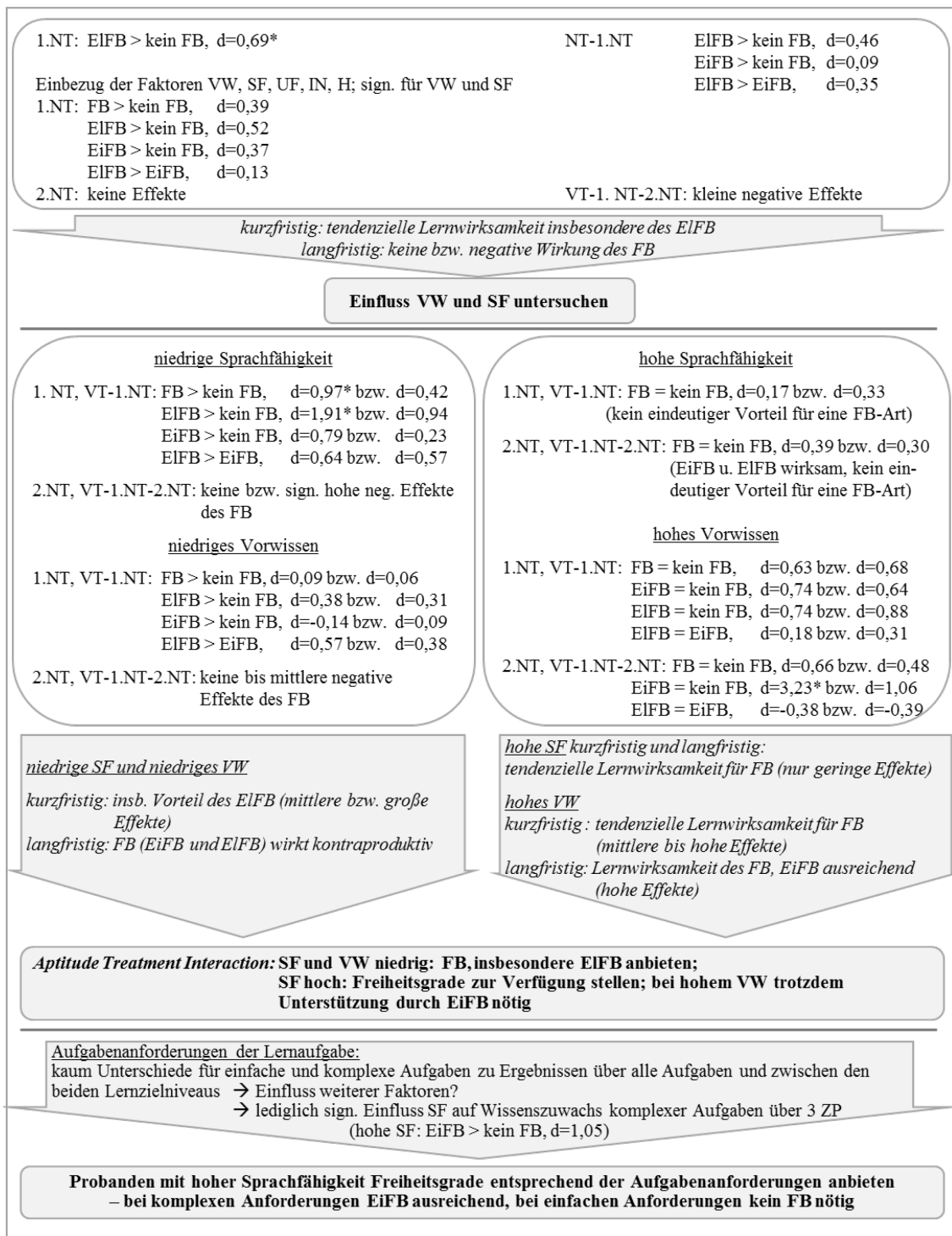
Lehrperson. Die Neuheit und Schwierigkeit der Methode verhindert möglicherweise einen langfristigen Behaltenseffekt. Schroeder u. a. (2017) fanden heraus, dass sich größere Effekte für das Concept Mapping erst dann zeigen, wenn es über einen längeren Zeitraum angewendet wird, da sich die Lernenden mit der Zeit besser auf die Verarbeitung der Lerninhalte konzentrieren können. Möglicherweise war das Concept Mapping in Verbindung mit dem Feedback in Form eines Durcharbeitens von Maps nicht ausreichend, um eine tiefgründige und somit langfristige Verarbeitung der Lerninhalte zu erwirken. Eine weitere Ursache für die nur kurzfristige Wirkung könnte auf die Komplexität der Lerninhalte zurückgeführt werden. Die im Rahmen des Planspiels vermittelten Inhalte sind eigentlich Teil mehrerer Lerneinheiten der Sekundarstufe II im berufsbildenden Bereich für Wirtschaft und wurden im Rahmen der Studie an zwei Tagen vermittelt. Infolgedessen würde die Methode des Concept Mapping in Verbindung mit Feedback höchstwahrscheinlich erst in Kombination mit einer kleinschrittigeren Vermittlung der Inhalte über einen längeren Zeitraum hinweg auch eine langfristige lernförderliche Wirkung insbesondere für Lernende mit einer niedrigen Sprachfähigkeit und niedrigem Vorwissen erzielen.

Die Hypothesen H5, H5b und H6 können angenommen werden. Bei hohem Vorwissen zeigen das *FB* insgesamt sowie das *EIFB* gegenüber *keinem FB* keine signifikanten Unterschiede, wobei sich *EIFB* und *EiFB* nicht signifikant unterscheiden. Entgegen der Vermutungen können die Probanden mit hohem Vorwissen jedoch die hohe kognitive Belastung, welche das Erstellen von Maps verursacht, langfristig nicht kompensieren und benötigen ergänzendes Feedback. Dabei profitieren sie langfristig insbesondere vom *EiFB*. Hypothese H5a kann demnach nicht angenommen werden. In Bezug auf den 2. Nachtest zeigt sich dabei ein großer signifikanter Effekt des *EiFB* gegenüber *keinem FB* von $d=3,23$ und in Bezug auf die Wissensentwicklung über drei Zeitpunkte zeigt sich ein großer Effekt von $d=1,06$. Kurzfristig sind mittlere bzw. hohe Effekte sowohl für das *EiFB* als auch für das *EIFB* erkennbar (zwischen $d=0,64$ und $d=0,88$). Langfristig ist das *EiFB* dem *EIFB* tendenziell überlegen ($d=-0,38$; $d=-0,39$). Das Präsentieren des Referenzmaps als KCR scheint demnach zu genügen. An dieser Stelle kann auf die Ergebnisse von Krause (2007) zurückgegriffen werden. Vorwissensstarke scheinen die durch das elaborierte Feedback präsentierten zusätzlichen Informationen zumindest langfristig nicht zu benötigen.

Bei hoher Sprachfähigkeit zeigt sich kein signifikanter Vorteil für das Feedback, weshalb die Hypothesen H9, H9a, H9b und H10 angenommen werden können. Tendenziell sind lediglich kleine Effekte für das Feedback erkennbar. Es zeigt sich hierbei kein eindeutiges Ergebnis für einen Vorteil des *EiFB* bzw. des *EIFB*.

Entsprechend der „Aptitude Treatment Interaction“ sollte Lernenden mit niedriger Sprachfähigkeit bzw. niedrigem Vorwissen eine zusätzliche Unterstützung in Form von elaboriertem Feedback angeboten und Lernenden mit hoher Sprachfähigkeit Freiheitsgrade bei der Steuerung ihres Lernens zur Verfügung gestellt werden. Um auch langfristige Effekte des Feedbacks für Lernende mit niedriger Sprachfähigkeit und niedrigem Vorwissen zu erzielen, sollte dabei das Concept Mapping in Verbindung mit ergänzendem Feedback kontinuierlich in den Lernprozess eingebunden werden. Probanden mit hohem Vorwissen benötigen die Unterstützung von Feedback, wobei einfaches Feedback zu genügen scheint.

Um den Einfluss der Aufgabenanforderungen der Lernaufgabe zu untersuchen, wurden Analysen jeweils für einfache und komplexe Aufgaben durchgeführt. Für die Entwicklung der Aufgaben wurde die Lernzieltaxonomie nach Anderson und Krathwohl (2001) genutzt, welche aus der Wissensdimension und der Dimension der kognitiven Prozesse besteht. Aufgaben, welche sich auf das „Faktenwissen erinnern“ sowie das „Konzeptuelle Wissen erinnern“ beziehen, stellen einfache Aufgaben dar. Das „Konzeptuelle Wissen verstehen“ sowie das „Prozedurale Wissen anwenden“ repräsentieren komplexe Aufgaben. Dabei zeigten sich kaum Unterschiede zu den Ergebnissen über alle Aufgaben und zwischen beiden Lernzielniveaus. Zusätzlich wurde überprüft, ob der Einfluss der Aufgabenanforderungen von weiteren Faktoren moderiert wird. Es zeigte sich lediglich ein signifikanter Einfluss der Sprachfähigkeit auf die Wissensentwicklung komplexer Aufgaben über drei Zeitpunkte. Weitere paarweise Vergleiche mit der dichotomisierten Variable für die Sprachfähigkeit (niedrig/hoch) zeigten dabei lediglich einen signifikanten Vorteil des *EiFB* gegenüber *keinem FB* ($d=1,05$) bei hoher Sprachfähigkeit. Probanden mit hoher Sprachfähigkeit sollten demnach Freiheitsgrade entsprechend der Aufgabenanforderungen angeboten werden. Bei komplexen Anforderungen empfiehlt es sich somit, einfaches Feedback anzubieten. Bei einfachen Anforderungen ist das Erstellen von Concept Maps als Lernstrategie auch ohne ergänzendes Feedback ausreichend. Nachfolgende Abbildung 40 zeigt die Ergebnisse zur Analyse der Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs.



* Signifikant auf dem Niveau $\alpha = 0,05$

Abbildung 40: Ergebnisse der Analysen zur Lernwirksamkeit des Feedbacks hinsichtlich des Wissenserwerbs

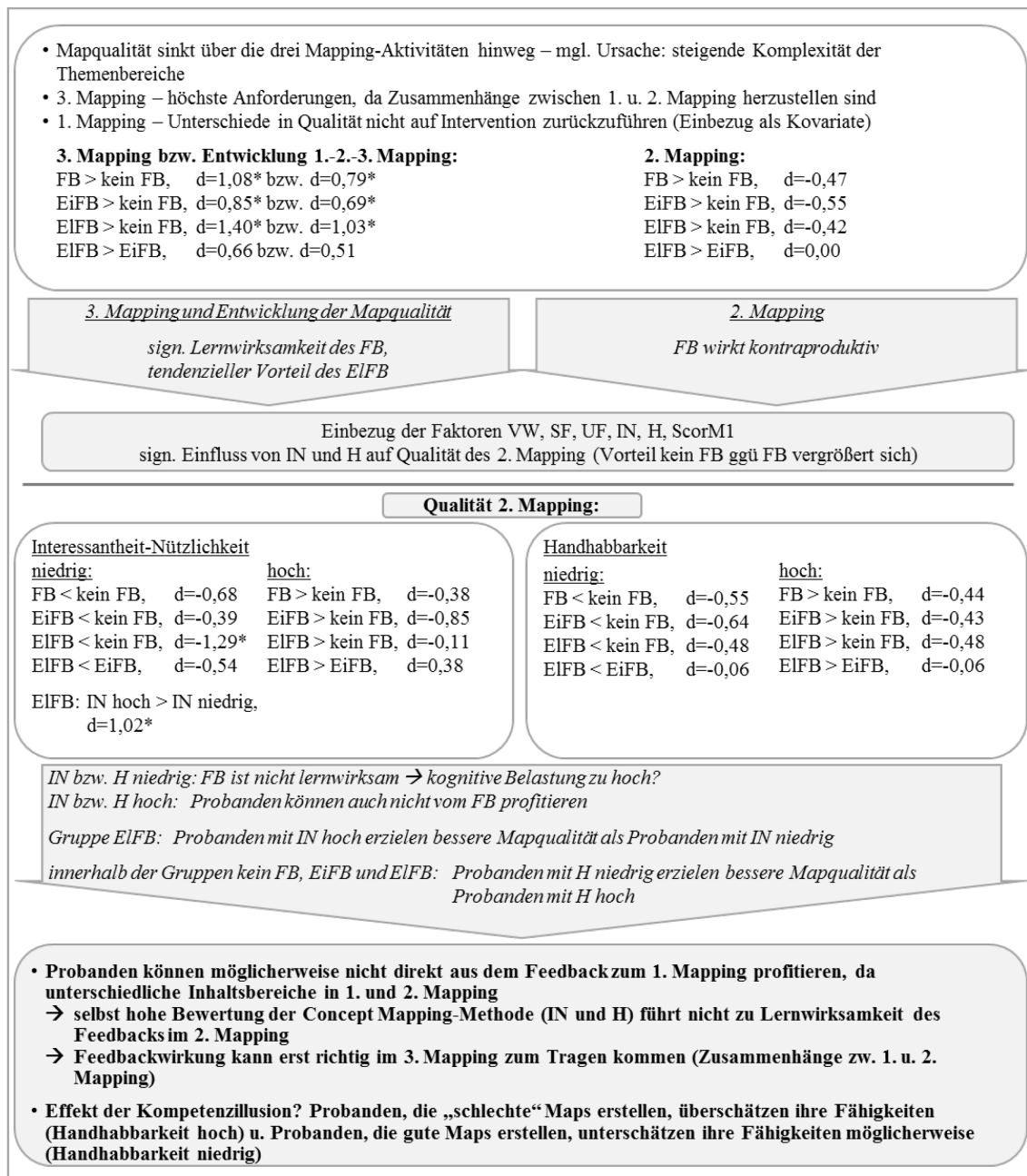
In der Abbildung 40 werden die Richtungen der Zusammenhänge zwischen den Gruppen, wie im Rahmen der Hypothesen vermutet, angegeben und die Ergebnisse anhand der Effekstärken dargestellt. Des Weiteren werden lediglich eindeutige Ergebnisse im Hinblick auf die Hypothesenprüfung berücksichtigt.

Die Qualität der Maps sinkt über die drei Mappingaktivitäten hinweg, wobei als mögliche Ursache die Komplexität der Themenbereiche genannt werden kann. Die Unterschiede in der Mapqualität des 1. Mapping können nicht auf die Intervention zurückgeführt werden, weshalb diese Variable als Kovariate in die Analysen einbezogen wurde. In Bezug auf die Ergebnisse zeigen sich Unterschiede zwischen dem 2. und 3. Mapping. Ein signifikanter Vorteil für das Feedback (sowohl für das einfache als auch für das elaborierte Feedback) ist für das 3. Mapping sowie für die Entwicklung der Mapqualität über die drei Mapping-Aktivitäten hinweg erkennbar. Die Hypothesen H1, H1a und H1b können daher in Bezug auf das 3. Mapping und die Entwicklung der Mapqualität angenommen werden. Hierbei zeigen sich hohe Effekte des *FB* gegenüber *keinem FB* ($d=1,08$ bzw. $d=0,79$), des *EiFB* gegenüber *keinem FB* ($d=0,85$ bzw. $d=0,69$) und des *ElFB* gegenüber *keinem FB* ($d=1,40$ bzw. $d=1,03$), wobei das *ElFB* dem *EiFB* tendenziell überlegen ist ($d=0,66$ bzw. $d=0,51$). Hinsichtlich des 2. Mappings wirkt das Feedback eher kontraproduktiv, so dass das *FB* einen mittleren negativen Effekt gegenüber *keinem FB* zeigt, dabei unterscheiden sich *EiFB* und *ElFB* nicht voneinander. Um Ursachen der unterschiedlichen Feedbackwirkung beim 2. und 3. Mapping zu untersuchen, wurden Vorwissen, Sprachfähigkeit, der eingeschätzte Umgang mit dem Concept Mapping (Interessantheit-Nützlichkeit sowie Handhabbarkeit) sowie der Umgang mit Fehlern als Kovariaten in die Analysen einbezogen. Infolgedessen vergrößerte sich der Vorteil von *keinem FB* gegenüber dem *FB* für das 2. Mapping. Die Ergebnisse in Bezug auf das 3. Mapping veränderten sich kaum. Der Einfluss der Variablen Interessantheit-Nützlichkeit und Handhabbarkeit auf die Ergebnisse des 2. Mapping wurde genauer analysiert, da sich für diese beiden Kovariaten signifikante Einflüsse zeigten. Hierzu wurden beide Variablen am Median dichotomisiert und in hoch und niedrig eingeteilt. Entsprechend der Vermutungen konnten Probanden welche die Interessantheit-Nützlichkeit bzw. Handhabbarkeit als niedrig einschätzten, nicht vom Feedback profitieren, weil die kognitive Belastung durch das Feedback möglicherweise zu hoch war. Dabei zeigte sich jedoch lediglich beim Vergleich von *keinem FB* und *ElFB* für IN niedrig ein signifikanter Nachteil für das *ElFB* ($d=1,29$). Die Hypothese H15b konnte demnach angenommen werden. Alle anderen Paarvergleiche zeigten lediglich eine tendenzielle Unterlegenheit des Feedbacks. Entgegen der Vermutungen zeigte sich aber auch kein Vorteil des Feedbacks bei Probanden, welche die Interessantheit-Nützlichkeit bzw. Handhabbarkeit als hoch einschätzten. Die Feedback-Gruppen sind der Gruppe *kein FB* tendenziell unterlegen. Eine zusätzliche Analyse zeigte, dass Proban-

den der Gruppe *EIFB* mit *IN hoch* signifikant bessere Concept Maps im 2. Mapping erstellen als Probanden mit *IN niedrig* ($d=1,02$). Der hohe Einfluss des Interesses und der Nützlichkeit an der Methode Concept Mapping ist schlüssig, da das elaborierte Feedback das Besprechen der Fehler- und Referenzmaps beinhaltet. Die Einschätzung der Methode als interessant und nützlich stellt eine Grundvoraussetzung dar, um vom Feedback profitieren zu können. Zur Untersuchung des Einflusses der Handhabbarkeit wurden zusätzlich innerhalb der Gruppen *kein FB*, *EiFB* und *EIFB* die Probanden mit *H hoch* und *H niedrig* verglichen. Dabei erstellten Probanden, welche die Handhabbarkeit als niedrig einschätzten, tendenziell bessere Maps als die Probanden, welche die Handhabbarkeit als hoch einschätzten. Diese Ergebnisse könnten auf eine mögliche Kompetenzillusion zurückgeführt werden. Ein Teil der Probanden überschätzt seine Fähigkeiten in Bezug auf die Anwendung der Methode des Concept Mapping und bewertet die Handhabbarkeit als hoch, obwohl die erstellten Maps eine niedrige Qualität haben. Der andere Teil ist in der Lage, seine Fähigkeiten im Umgang mit der Methode zu reflektieren und Schwächen zu erkennen. Infolgedessen wird die eigene Kompetenz unterschätzt, obwohl die Maps eine hohe Qualität aufweisen.

Insgesamt bleibt zu hinterfragen, warum das Feedback auf die Qualität des 2. Maps kontraproduktiv wirkt. Möglicherweise können die Probanden beim Erstellen der Maps nicht direkt vom vorherigen Feedback zum 1. Mapping profitieren, da es sich auf unterschiedliche Inhaltsbereiche bezieht. Es scheint sogar zu einer kognitiven Überlastung zu führen. Des Weiteren hat die Bewertung der Concept Mapping Methode (Interessantheit-Nützlichkeit, Handhabbarkeit) einen signifikanten Einfluss auf die Qualität des 2. Maps. Insbesondere eine niedrige Bewertung führt zu einer deutlichen Unterlegenheit der Feedback-Gruppen gegenüber den Probanden, die kein Feedback erhalten. Dennoch können Probanden mit einer hohen Einschätzung der Methode auch nicht vom Feedback profitieren, was möglicherweise auf die nicht direkt anwendbaren Informationen aus dem Feedback zum 1. Mapping zurückgeführt werden kann.

Die Wirkung des Feedbacks kann erst beim 3. Mapping richtig zum Tragen kommen, da hier die Zusammenhänge zwischen dem 1. und dem 2. Mapping dargestellt werden sollen. Nachfolgende Abbildung 41 zeigt die Ergebnisse zur Analyse der Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps.



* signifikant auf dem Niveau $\alpha = 0,05$

Abbildung 41: Ergebnisse der Analysen zur Lernwirksamkeit des Feedbacks hinsichtlich der Qualität der Maps

Die Unterschiede hinsichtlich der Wirkung zwischen dem Wissenserwerb und der Qualität der Maps können möglicherweise auf die verschiedenen Testformate zurückgeführt werden. Das Erstellen der Concept Maps erfüllt eine Doppelfunktion. Zum einen soll es den Planspieleinsatz als Lernstrategie unterstützen. Zum anderen dient es der Wissensdiagnose, um die Qualität der Maps im Rahmen der Analyse der Lernwirksamkeit des Feedbacks zu analysieren. Dabei wird die Maperstellung somit direkt in den Lernprozess integriert. Das Feedback beinhaltet das Durcharbeiten der Referenzmaps. Unabhängig vom

Vorwissens- und Sprachniveau profitieren die Probanden somit vom Feedback und erstellen Maps mit einer signifikant besseren Qualität gegenüber Probanden, die kein Feedback erhalten. Möglicherweise fällt den Probanden das Erstellen der Maps auch leichter, da von einer netzwerkartigen Struktur des Gedächtnisses ausgegangen werden kann. Aufwändige Umkodierungsprozesse sind nicht notwendig (vgl. Kapitel 3.1.3). Die im Rahmen des Feedbacks erhaltenen Informationen sind somit auch direkt mit der kognitiven Struktur verknüpfbar.

Der Wissenstest beinhaltet lediglich zwei Aufgaben, die Ähnlichkeiten mit dem Aufgabenformat Concept Mapping haben (Aufgabe 4 und 10; vgl. Kapitel 4.3.1). Bei den anderen Aufgaben sind daher aufwändigere Umkodierungsprozesse notwendig. Die Übertragung des im Planspielunterrichts erworbenen Wissens auf die im Wissenstest geforderten Aufgaben gelingt den Probanden, welche elaboriertes Feedback erhalten gegenüber Probanden, welche kein Feedback erhalten, kurzfristig, wobei der Effekt nicht signifikant ist. Signifikante kurzfristige Effekte des Feedbacks gegenüber keinem Feedback zeigen sich bei Probanden mit niedriger Sprachfähigkeit. Eine langfristige Übertragung gelingt lediglich Probanden mit hohem Vorwissen, welche einfaches Feedback erhalten. Inwieweit sich die Lernwirksamkeit des Feedbacks operationalisiert durch die Qualität der Maps auch langfristig gezeigt hätte, kann nicht beurteilt werden, da im Rahmen des 2. Nachttests lediglich der Wissenstest eingesetzt wurde.

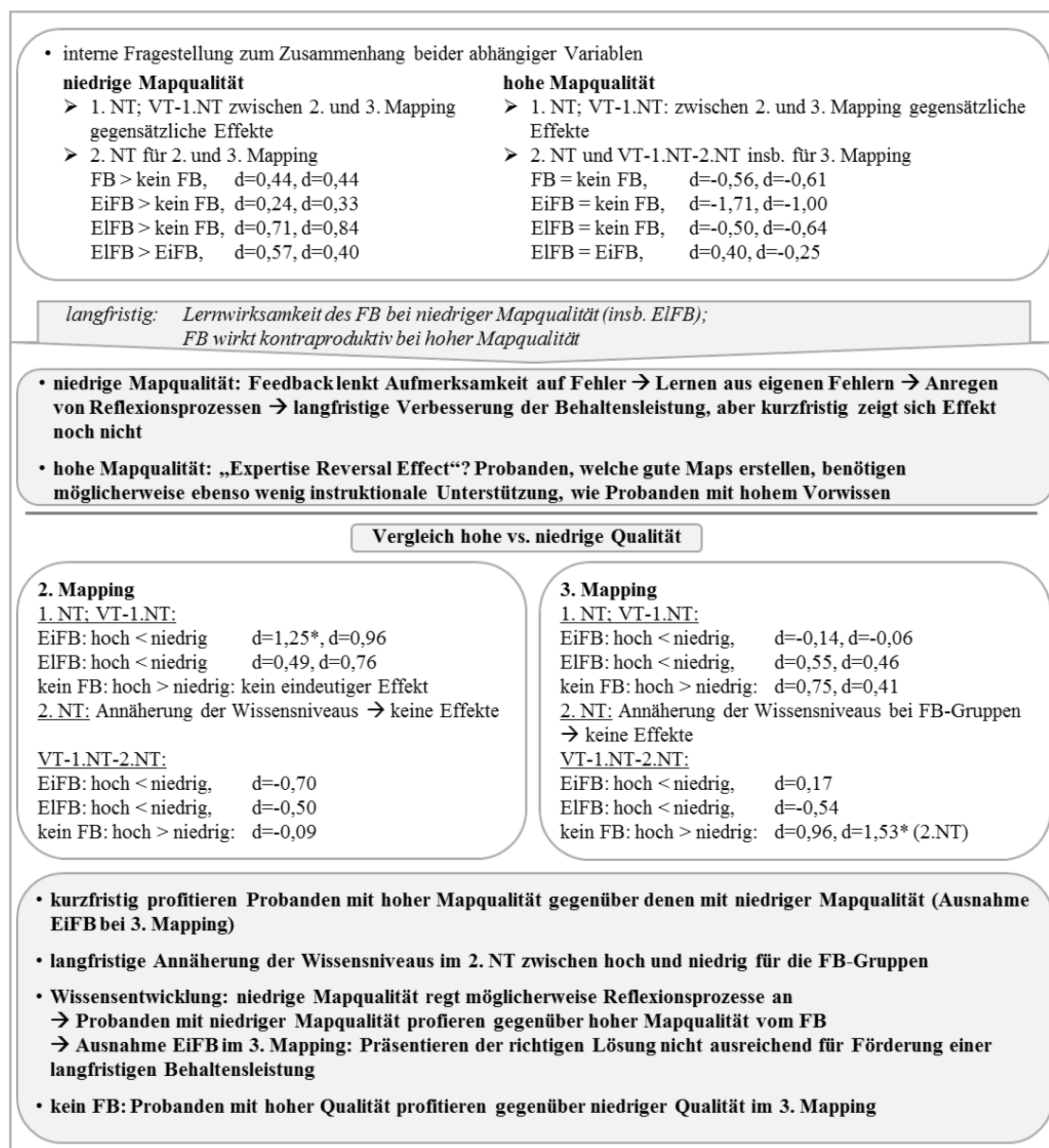
Hinsichtlich der Einflussfaktoren können die Ergebnisse für den Wissenserwerb sowie für die Qualität der Maps nicht zusammengefasst werden, da verschiedene Faktoren einen signifikanten Einfluss haben. Beim Wissenserwerb beeinflussen das Vorwissen sowie die Sprachfähigkeit die Lernwirksamkeit des Feedbacks, bei der Qualität der Maps sind es Interessantheit-Nützlichkeit sowie die Handhabbarkeit des Concept Mapping.

Des Weiteren wurde der Einfluss der Mapqualität auf den Wissenserwerb analysiert. Es wurde dabei angenommen, dass insbesondere Probanden, welche qualitativ schlechtere Maps erstellen, vom Feedback profitieren. Hinsichtlich des 1. Nachttests sowie des Wissenszuwachses vom Vor- zum 1. Nachttest zeigen sich zwischen 2. und 3. Mapping gegensätzliche Ergebnisse. Probanden, welche im 2. Mapping qualitativ schlechtere Maps erstellen und elaboriertes Feedback erhalten, erzielen im 1. Nachttest signifikant bessere Ergebnisse als Probanden, die einfaches Feedback erhalten ($d=1,03$), Hypothese H34 konnte daher dahingehend angenommen werden. Insgesamt lässt sich kurzfristig jedoch

kein Vorteil des Feedbacks bei niedriger Mapqualität erkennen. In Bezug auf den 2. Nachtest zeigt sich ein tendenzieller Vorteil des *FB* gegenüber *keinem FB* bei niedriger Mapqualität sowohl im 2. als auch im 3. Mapping ($d=0,44$; $d=0,44$). Hierbei können die Probanden insbesondere vom *EiFB* profitieren ($d=0,71$; $d=0,84$). Bei hoher Mapqualität zeigen sich hinsichtlich des 1. Nachtests sowie des Wissenszuwachses vom Vor- zum 1. Nachtest ebenfalls gegensätzliche Ergebnisse zwischen 2. und 3. Mapping. Insgesamt lässt sich kurzfristig entsprechend der Vermutungen kein Vorteil des Feedbacks bei hoher Mapqualität erkennen. Sowohl im 2. Nachtest als auch in Bezug auf die Wissensentwicklung über drei Zeitpunkte zeigen sich tendenzielle negative Effekte für das *FB* gegenüber *keinem FB* ($d=-0,56$; $d=-0,61$). Hierbei sind die negativen Effekte bei Probanden, die *EiFB* erhalten, gegenüber *keinem FB* besonders groß ($d=-1,71$; $d=-1,00$). Die Effekte sind jedoch nicht signifikant, so dass die Hypothesen H35, H35a, H35b und H36 angenommen werden können. Entsprechend der Vermutungen zeigt sich bei hoher Mapqualität kein Vorteil für das Feedback. Tendenziell wirkt Feedback langfristig sogar kontraproduktiv. Bei niedriger Mapqualität lenkt insbesondere das elaborierte Feedback möglicherweise die Aufmerksamkeit auf die eigenen Fehler. Infolgedessen werden Reflexionsprozesse angeregt. Die Probanden lernen aus ihren Fehlern und die Behaltensleistung verbessert sich somit langfristig. Kurzfristig hat das noch keine Auswirkungen. Bei hoher Mapqualität kommt möglicherweise der „Expertise Reversal Effect“ zum Tragen. Probanden, welche gute Maps erstellen, benötigen eventuell keine instruktionale Unterstützung durch ein ergänzendes Feedback.

Zusätzlich wurde innerhalb der Gruppen hohe und niedrige Qualität gegenübergestellt. Entgegen der Vermutungen profitieren im 1. Nachtest sowie hinsichtlich des Wissenszuwachses vom Vor- zum 1. Nachtest Probanden mit hoher Mapqualität tendenziell gegenüber denen mit niedriger Mapqualität vom Feedback (zwischen $d=0,46$ und $d=1,25$). Im 2. Nachtest nähert sich das Wissensniveau zwischen hoher und niedriger Qualität für die Feedbackgruppen an (keine Effekte). Hinsichtlich der Entwicklung über drei Zeitpunkte profitieren Probanden mit niedriger Mapqualität gegenüber Probanden mit hoher Mapqualität vom Feedback (zwischen $d=-0,50$ und $d=-0,70$). Ausnahme bildet das einfache Feedback im 3. Mapping. Hier zeigt sich ein kleiner Effekt ($d=0,17$) für hohe gegenüber niedriger Mapqualität. Wie bereits oben erläutert, regt niedrige Qualität möglicherweise Reflexionsprozesse an, wobei das Präsentieren der richtigen Lösung als einfaches

Feedback dabei nicht zu genügen scheint, um eine langfristige Behaltensleistung zu fördern. Bei Probanden, welche kein Feedback erhalten, ist beim 3. Mapping langfristig ein deutlicher Vorteil von hoher gegenüber niedriger Qualität erkennbar ($d=1,53$; $d=0,96$). Für diese Probanden ist es besonders schwierig, das erlernte Wissen in einem Concept Map angemessen darzustellen, ohne ergänzendes Feedback zu erhalten. Das 3. Mapping stellt die Zusammenhänge zwischen 1. und 2. Map dar und hat somit die höchste Komplexität. Eine hohe Qualität zeugt damit von einem hohen Tiefenverständnis der erlernten Inhalte. Die Abbildung 42 zeigt die Ergebnisse der Analysen zum Einfluss der Mapqualität auf den Wissenserwerb.



* signifikant auf dem Niveau $\alpha = 0,05$

Abbildung 42: Ergebnisse der Analysen zum Einfluss der Mapqualität auf den Wissenserwerb

Im Anschluss wurde analysiert, wie die Probanden den Umgang mit dem Feedback einschätzen. Hierbei zeigte sich, dass sich Probanden, welche kein Feedback erhalten haben, eine Rückmeldung zu ihren Concept Maps gewünscht hätten. Dabei ist dieser Wunsch signifikant höher ausgeprägt, als der Wunsch der Gruppe *EiFB* nach einer weiteren Fehlerbesprechung ($d=1,01$). Hypothese H27 konnte daher angenommen werden. Zwischen der eingeschätzten Nützlichkeit des einfachen und des elaborierten Feedbacks zeigt sich ein signifikanter Zusammenhang ($r_s=0,654$). Hypothese H29 konnte angenommen werden. Beide Feedback-Arten sollten daher aufeinander abgestimmt werden. Zwischen den Variablen zur Einschätzung des Feedbacks und dem Wissenserwerb bzw. der Qualität der Maps konnte entgegen der Vermutungen kein Zusammenhang festgestellt werden. Das elaborierte Feedback wurde von den Probanden der Gruppe *ElFB* signifikant nützlicher erachtet als das einfache Feedback ($d=1,20$). Hypothese H28 konnte daher bestätigt werden.

Des Weiteren wurde die Qualität der Feedback-Netze analysiert. Die Probanden beider FB-Gruppen erhielten zunächst einfaches Feedback in Form des jeweiligen Referenznetzes, welches sie mit dem selbsterstellten Map vergleichen sollten. Zusätzlich sollten sie die Maps korrigieren, indem sie Änderungen in einer Kopie der zuvor angefertigten Concept Maps vornahmen. Die überarbeiteten Maps wurden als Feedback-Netze bezeichnet. Zur Analyse der Qualität der Feedback-Netze wurden analog zur Analyse der Qualität der selbsterstellten Maps die Scoringwerte herangezogen. Tendenziell lag die Qualität der Feedback-Netze beider FB-Gruppen über der Qualität der selbsterstellten Maps. Hierbei erzielte die Gruppe *EiFB* ($M1=0,37$, $M2=0,39$, $M3=0,39$) tendenziell eine höhere Qualität für die Feedback-Netze als die Gruppe *ElFB* ($M1=0,32$, $M2=0,35$, $M3=0,38$). Gleichzeitig schätzte jedoch die Gruppe *EiFB* das einfache Feedback signifikant schlechter ein als die Gruppe *ElFB*. Um Ursachen hierfür zu finden, wurden die Verarbeitungsstrategien des einfachen Feedbacks analysiert. Insgesamt konnten vier Verarbeitungsstrategien identifiziert werden: das Löschen falscher Propositionen, das Korrigieren falscher Propositionen, das Hinzufügen richtiger Propositionen oder das Hinzufügen falscher/ unvollständiger/ „halb“ richtiger Propositionen. Hierbei konnten die Strategien lediglich für das 2. und 3. Mapping analysiert werden, da bis zum 1. Mapping keine Unterschiede hinsichtlich der Intervention stattfanden. Die Probanden beider FB-Gruppen erzielten im 2. Mapping ähnliche Scoringwerte ($d=0,00$), so dass hier von vergleichbaren Voraussetzun-

gen ausgegangen werden kann. Probanden der Gruppe *EIFB* (38 %) löschten und korrigierten öfter Propositionen als die Probanden der Gruppe *EiFB* (21 %) und setzten sich so intensiver mit den erstellten Maps auseinander. Auf der anderen Seite fügten Probanden der Gruppe *EiFB* öfter Propositionen hinzu (79 %) als die Gruppe *EIFB* (62 %), sie übernahmen somit häufiger Propositionen aus dem Referenznetz. Möglicherweise führte eine mangelnde Motivation der Gruppe *EiFB* dazu, dass sie einfach nur Propositionen des Referenzmaps hinzufügten, ohne im Gegenzug die erstellten Propositionen zu löschen oder zu korrigieren, da sich bei ihnen kein weiteres Feedback anschloss. Darauf kann auch die signifikant schlechtere Einschätzung der Nützlichkeit des einfachen Feedbacks zurückgeführt werden. Das einfache Übernehmen der Propositionen aus dem Referenznetz führt auf der anderen Seite zu einer höheren Qualität der Feedback-Netze für die Gruppe *EiFB* im Vergleich zur Gruppe *EIFB*. Im 3. Mapping lag die Qualität für die Gruppe *EIFB* tendenziell über der Qualität der Gruppe *EiFB* ($d=0,66$), so dass die Verarbeitungsstrategien hier schwerer vergleichbar sind. Dies erklärt auch, warum die Probanden der Gruppe *EiFB* öfter falsche Propositionen löschten und korrigierten (26 %) als die Probanden der Gruppe *EIFB* (17 %). Insgesamt ist auffällig, dass bei den Probanden beider FB-Gruppen für die drei Mapping-Aktivitäten insgesamt 34 Prozent der angewendeten Verarbeitungsstrategien auf das Hinzufügen falscher oder nur teilweise richtiger Propositionen entfallen. Dies deutet darauf hin, dass die Schüler unkonzentriert waren oder die Zusammenhänge nicht ausreichend verstanden haben.

Abschließend wurden die Propositionen der selbsterstellten Maps in Bezug auf ihre inhaltliche Richtigkeit näher untersucht, um Fehlkonzepte im Sinne von falschen Propositionen und fehlende Propositionen zu identifizieren. Falsche Propositionen sind die, welche häufig genannt wurden, jedoch nicht im Referenznetz enthalten sind. Hierzu wurden die Modalnetze für die drei Mapping-Aktivitäten erstellt, da sie alle erstellten Probanden-netze berücksichtigen und die am häufigsten übereinstimmend genannten Propositionen enthalten. Auf diese Weise ist ein vereinfachtes Vorgehen möglich, ohne eine Analyse aller Individualnetze vornehmen zu müssen. Als falsche Propositionen wurden somit solche angesehen, die im Modalnetz, aber nicht im Referenznetz enthalten sind. Als fehlende Propositionen gelten die, welche dreimal oder seltener genannt werden und Teil des Referenznetzes sind. Das Vorgehen, nicht nur die komplett fehlenden Propositionen als fehlend zu betrachten, wird dabei bewusst gewählt. Der Lehrperson lässt sich auf diese Weise

ein Hinweis geben, welche Inhalte besonders zu thematisieren sind, da sie von den Lernenden noch nicht beherrscht werden. Die Fehleranalyse ergab, dass die Schüler z. T. ein falsches Verständnis der Begriffe und Zusammenhänge haben, welches sich im Zeitverlauf des Erstellens der drei Mappings auch nicht korrigierte (z. B. Forderungen, Lieferant, Zusammenspiel zwischen GuV und Bilanz, Abgrenzung der Begriffe Aufwendungen und Aktivseite). Darüber hinaus werden Sachverhalte oft vereinfacht, dabei fehlt es an Trennschärfe sowohl bei den Begriffen als auch bei den Prozessen. Die Denkweise der Schüler ist insgesamt sehr oberflächlich. Sie sind nicht in der Lage die komplexen Zusammenhänge angemessen darzustellen.

In Bezug auf die eingangs gestellten Forschungsfragen lässt sich somit keine eindeutige Antwort finden. Feedback (insbesondere elaboriertes Feedback) zu Concept Maps im betriebswirtschaftlichen Planspielunterricht scheint insbesondere für Probanden mit niedrigem Vorwissen und niedriger Sprachfähigkeit lernwirksam zu sein, wobei sich diese Effekte nur kurzfristig zeigen. Probanden, welche Feedback erhalten, sind in der Lage Concept Maps mit einer höheren Qualität zu erzielen, als Probanden, welche kein Feedback erhalten. Dabei zeigt sich ein tendenzieller Vorteil des elaborierten gegenüber dem einfachen Feedback.

6.2 Kritische Würdigung, Optimierungsvorschläge und Ausblick

Nachfolgend werden die Gütekriterien Objektivität, Reliabilität und Validität der Untersuchung überprüft (Döring & Bortz, 2016, S. 440 ff.).

Objektivität

Für eine optimale Objektivität hätte der Planspielunterricht in allen Versuchsgruppen von denselben Personen durchgeführt werden müssen. Dies war aus forschungsökonomischen Gesichtspunkten nicht möglich. Um dennoch eine Durchführungsobjektivität zu gewährleisten, wurde mit den Lehrpersonen vorab eine intensive Schulung zum Planspiel Easy BusinessTM, der Methode des Concept MappingTM sowie zum Feedback durchgeführt. Zur Durchführung des Planspielunterrichts wurden detaillierte Verlaufspläne zur Verfügung gestellt. Um mit den Probanden die Methode des Concept Mapping zu trainieren, wurde ebenfalls ein detaillierter Leitfaden bereitgestellt. Zur Gestaltung der Mapping-Aktivitäten sowie der Feedbackmaßnahmen wurden den Versuchsleitern schriftliche Hinweise gegeben. Zunächst wurden Fehler anhand eines Schülermaps besprochen. Die Auswahl

des Netzes erfolgte dabei eher spontan anhand von Kriterien wie Lesbarkeit und Nachvollziehbarkeit der dargestellten Propositionen. Die Rahmenbedingungen ließen es nicht zu, eine systematische Auswahl zu treffen. Bei einer längeren Studiendauer wäre es darüber hinaus möglich gewesen, die Maps einer detaillierten Fehleranalyse zu unterziehen, um die anschließende Fehlerbesprechung strukturierter zu gestalten. Hierbei kann erneut auf die Problematik Bezug genommen werden, dass papierbasiertes Mapping kein sofortiges Feedback zulässt. Für den zweiten Teil des elaborierten Feedbacks wurden die Versuchsleiter instruiert, die Referenzmaps in einem Schüler-Lehrer-Gespräch zu erläutern. Hierzu wurden die Referenzmaps in Themenblöcken erarbeitet. Dabei gingen die Versuchsleiter eher intuitiv vor. Möglicherweise hätte ein standardisiertes Vorgehen die Durchführungsobjektivität erhöht. Darauf wurde jedoch bewusst verzichtet, um die Natürlichkeit der Unterrichtssituation an dieser Stelle zu gewährleisten und den Schülern somit die Angst zu nehmen, bei Verständnisproblemen nachzufragen.

Vortest, 1. Nachtest und 2. Nachtest sowie der Fragebogen beinhalteten standardisierte Anweisungen an die Probanden, so dass hier von einer Durchführungsobjektivität ausgegangen werden kann. Die verbalen Erläuterungen der Versuchsleiter beschränkten sich auf das Besprechen des einzutragenden Codes sowie den Hinweis auf einen anonymen Umgang mit den Daten.

Um eine objektive Auswertung der Wissenstests zu gewährleisten, wurde das Codiervorgehen in einem Codierhandbuch dokumentiert. Dieses enthält zahlreiche Ankerbeispiele, die das Auswertungsvorgehen verdeutlichen, weshalb insgesamt eine hohe Auswertungsobjektivität konstituiert wird. Des Weiteren ist die Interpretationsobjektivität zu betrachten, die als hoch eingeschätzt werden kann, da für die Wissenstests numerische Ergebnisse als erreichte Punktzahlen festgehalten wurden.

Die Auswertungsobjektivität der Concept Maps kann als hoch betrachtet werden, da das Vorgehen bei der kategorialen und strukturellen Inhaltsanalyse ausführlich beschrieben und durch detaillierte Codierregeln mit Beispielen standardisiert wurde. Das Vorgehen bei der Bewertung der Concept Maps anhand des differenzierten Scoringverfahrens wurde mithilfe von vorgegebenen Regeln ebenfalls standardisiert. Die Interpretationsobjektivität wird ebenso als hoch eingeschätzt, da ein numerisches Ergebnis in Form von Scoringwerten als differenzierte Übereinstimmungsgrade zwischen Schüler- bzw. Feedback-Netz und Referenznetz vorliegt.

Der Fragebogen umfasst geschlossene und lediglich eine teil-offene Frage, so dass eine standardisierte Auswertung ohne größere Schwierigkeiten möglich war. Somit können die Auswertungs- sowie die Interpretationsobjektivität gewährleistet werden.

Der WIT-2 hat sich als objektives Instrument bewährt. Die Objektivität in Bezug auf die Durchführung wird aufgrund eines separaten Instruktionshefts zur fehlerfreien und komfortablen Instruktion gewährleistet. Zur Auswertung des Tests stehen ebenfalls ausführliche Instruktionsmaterialien mit optionaler PC-Unterstützung zur Verfügung. Die Interpretationsobjektivität ist durch die Vorgabe Gesamt- sowie alters- und bildungsspezifischer Normen gegeben (Kersting u. a., 2008, S. 56 ff.).

Reliabilität

Hinsichtlich der Reliabilität sind für die Wissenstests insbesondere die Paralleltestreliabilität und die Intercoderreliabilität relevant. Im Vorfeld der Untersuchung wurde mit einer anderen Stichprobe mittels der Formel von Cureton (1971, zitiert nach Lienert & Raatz, 1998, S. 182) eine Paralleltestreliabilität von $r_{tt} = 0,7902$ ermittelt. Döring und Bortz (2016, S. 83) geben an, dass in der Forschung ein Wert von 0,8 oft als Mindestmaß für Reliabilitätskoeffizienten angesehen wird. Sie betonen, dass die Beurteilung jedoch unter Berücksichtigung der Besonderheiten der jeweiligen Studie erfolgen muss. Lienert und Raatz (1998, S. 14) stellen heraus, dass ein Reliabilitätskoeffizient zwischen 0,5 und 0,7 bereits genügt, insofern der Test als Forschungsinstrument zum Vergleich von Gruppen dienen soll. Infolgedessen kann von vergleichbaren Testformen ausgegangen werden. Die Bewertung der Wissenstests wurde von zwei fachkundigen Personen unabhängig voneinander vorgenommen. Mit Hilfe der Korrelation nach Spearman wurde für die Intercoderreliabilität ein Wert von $r_s = 0,968$ (Spearman's Rho) ermittelt, was bedeutet, dass zwischen beiden Codierern bei unabhängiger Codierung der Tests eine Übereinstimmung von 96,8 Prozent erzielt wurde. Mit beiden Werten konnte nachgewiesen werden, dass die verwendeten Wissenstests zuverlässig messen und Sorgfalt bei der Anwendung gegeben war.

Um das Codiervorgehen der Concept Maps zu überprüfen, wurde bei 11 Prozent der Probandennetze die Inter- und Intracoderreliabilität berechnet. Für die Intercoderreliabilität konnte ein Koeffizient von $C = 0,83$, für die Intracoderreliabilität ein Koeffizient von $C = 0,81$ ermittelt werden. Döring und Bortz (2016, S. 83) geben an, dass in der Forschung ein Wert von 0,8 oft als Mindestmaß für Reliabilitätskoeffizienten angesehen wird. Die

inhaltsanalytische Auswertung der Concept Maps erforderte ein aufwendiges Codiervorgehen, weshalb ein Wert größer als 0,8 als zufriedenstellend angesehen werden kann. Das Codiervorgehen sowie dessen Anwendung kann daher als zuverlässig eingeschätzt werden. Um das differenzierte Bewertungsverfahren zu überprüfen, wurde bei 16 Prozent der Probandennetze die Intercoderreliabilität berechnet. Die Netze wurden von einem zweiten Codierer anhand der vorliegenden Bewertungsregeln erneut bewertet. Mit Hilfe der Korrelation nach Spearman wurde für die Intercoderreliabilität ein Wert von $r_s=0,926$ (Spearman's Rho) ermittelt, was bedeutet, dass zwischen beiden Codierern bei unabhängiger Bewertung der Maps eine Übereinstimmung von 92,6 Prozent ($p=0,000$) erzielt wurde. Abschließend fand bei strittigen Bewertungen eine Konsensbildung zwischen Erst- und Zweitcodierer statt, sodass im Ergebnis eine von beiden Codierern akzeptierte Bewertung der Maps für die weitere Auswertung vorlag.

Für den Teil-Fragebogen zur Einschätzung der Methode des Concept Mapping wurde die interne Konsistenz für die Bereiche Interessantheit, Handhabbarkeit und Nützlichkeit berechnet. Für den Bereich Interessantheit liegt dieser bei 0,776, für die Handhabbarkeit bei 0,693 und für die Nützlichkeit bei 0,287. Diese Werte zeigen, dass die interne Konsistenz für den Bereich Interessantheit und Handhabbarkeit akzeptabel ist, wohingegen die interne Konsistenz für den Bereich Nützlichkeit darauf hindeutet, dass die Items diesen Bereich unzureichend repräsentieren. Cronbachs Alpha konnte auch durch den Ausschluss von Variablen nicht erhöht werden. Aus diesem Grund wurde anschließend eine Faktorenanalyse durchgeführt. Diese sollte Aufschluss darüber geben, inwieweit die Variablen wirklich auf die drei Dimensionen Interessantheit, Handhabbarkeit und Nützlichkeit reduziert werden können. Die Faktorenanalyse führte zu einer Zwei-Faktoren-Lösung. Auf dem ersten Faktor laden die Items des Bereiches Interessantheit und Nützlichkeit. Der Faktor wurde für folgende Analysen als *Interessantheit-Nützlichkeit* bezeichnet. Die theoretische Zuordnung konnte daher nicht beibehalten werden. Es zeigt sich, dass die Interessantheit mit der Nützlichkeit in Bezug auf den Umgang mit Concept Mapping zusammenhängt. Dem zweiten Faktor sind alle Variablen des Bereiches *Handhabbarkeit* zugewiesen. Somit konnte die theoretische Zuordnung bestätigt werden.

Des Weiteren wurde die interne Konsistenz für die drei Items zur Einschätzung der Nützlichkeit des Einfachen Feedbacks berechnet. Mit einem Wert von 0,324 ist diese inakzeptabel, konnte jedoch durch Ausschluss eines Items auf einen akzeptablen Wert von 0,767

erhöht werden. Mittels einer anschließend durchgeführten Faktorenanalyse konnten die beiden Items auf einen Faktor verdichtet werden.

In einem nächsten Schritt wurde die interne Konsistenz für die Items zur Einschätzung der Nützlichkeit des elaborierten Feedbacks berechnet. Mit einem Wert von 0,157 ist diese inakzeptabel. Cronbachs Alpha konnte auch durch den Ausschluss von Variablen nicht erhöht werden. Die Items *EIFBN_1* und *EIFBN_3* wurden jeweils einzeln für weitere Analysen einbezogen und das Item *EIFBN_2* wurde ausgeschlossen, da die Frage möglicherweise nicht richtig verstanden wurde.

In einem nächsten Schritt wurde die interne Konsistenz für alle neun Items der Skala Umgang mit Fehlern berechnet. Mit einem Wert von 0,741 kann diese als akzeptabel angesehen werden. Die Skala wurde durch Spychiger, Kuster und Oser (2006) bereits statistisch validiert. Des Weiteren ist kein Ausschluss einer Variablen vonnöten, um eine akzeptable interne Konsistenz zu erreichen. Alle neun Items repräsentieren die Skala gut. Daher ist es nicht notwendig, mithilfe einer Faktorenanalyse die Subskalen zu überprüfen bzw. neue Faktoren zu identifizieren.

Der WIT-2 hat sich als zuverlässiges Instrument bewährt. Die Module des Tests zeichnen sich durch eine hohe interne Konsistenz von $>0,80$ aus (Kersting u. a., 2008, S. 56 ff.).

Validität

Hinsichtlich der Inhaltsvalidität sollten die einzelnen Testitems das Zielkonstrukt in seinen inhaltlichen Bedeutungsaspekten vollständig und sinngemäß widerspiegeln. Die Wissenstests sowie die Referenzmaps wurden auf die Lernziele des Planspiels abgestimmt, so dass von einer Inhaltsvalidität ausgegangen werden kann. Hinsichtlich der Konstruktvalidität sollten die Tests mit bestehenden Konstruktdefinitionen und Theorien übereinstimmen. Bei der Testkonstruktion sowie der Erstellung der Referenzmaps wurde darauf geachtet, dass die Testfragen bzw. Propositionen die betriebswirtschaftlichen Grundlagen in Bezug auf Wertschöpfungsprozess, lineare Abschreibungen, GuV und Bilanz widerspiegeln. Dabei dienten die Planspielinhalte als Orientierungshilfe. Diese spiegeln zum einen die Bereiche Beschaffung, Produktion und Verkauf der Wertschöpfungskette nach Porter (1985) und zum anderen die im HGB gesetzlich geregelten Inhalte zur linearen Abschreibung, GuV und Bilanz stark vereinfacht wider. Insgesamt erzielten die Probanden eine relativ niedrige Mapqualität in allen drei Mappings (MW zwischen 0,0815 und

0,2195). Daher sollte für weitere Studien über eine mögliche Anpassung der in den Referenznetzen verwendeten Propositionen an das Schülerniveau nachgedacht werden. Einige Zusammenhänge wurden von den Probanden nicht erkannt. Es empfiehlt sich dabei die Nachvollziehbarkeit der Referenzmaps vorab zu überprüfen.

Die Kriteriumsvalidität gibt an, in welchem Grad die mit einem Messinstrument erzielten Resultate mit anderen relevanten Außenkriterien übereinstimmen. Hierbei hätten die Ergebnisse der Tests bzw. der erstellten Maps z. B. über die Noten des Faches WTH validiert werden können. Da die Ergebnisse hierzu nicht vorliegen, kann zur Kriteriumsvalidität keine Aussage getätigt werden.

Probanden, welche Feedback erhielten, waren in der Lage, signifikant bessere Maps zu erstellen als Probanden, welche kein Feedback erhielten. Hinsichtlich der Wissenstestergebnisse zeigte sich dieser Vorteil nicht. Um Ursachen zu ergründen, wird der Zusammenhang zwischen den Wissenstestergebnissen sowie der Qualität der erstellten Concept Maps als Instrumente zur Messung der Lernwirksamkeit ermittelt. Eine entsprechende Korrelationsanalyse zeigt geringe Korrelationen für die Scoringwerte des 3. Mappings (ScorM3) mit den Testergebnissen des Vortests ($r_s=0,313$, $p=0,019$), 1. Nachtests ($r_s=0,367$, $p=0,005$) und 2. Nachtests ($r_s=0,320$, $p=0,025$). Des Weiteren zeigt sich eine geringe Korrelation zwischen ScorM1 und den Ergebnissen des Vortests ($r_s=0,280$, $p=0,037$) sowie zwischen ScorM2 und den Ergebnissen des 1. Nachtests ($r_s=0,345$, $p=0,009$). Für einen Zusammenhang zwischen dem 1. Nachtest bzw. 2. Nachtest mit der Qualität des 3. Mappings spricht auch der Ausreißer El07, welcher bei beiden Tests sowie für ScorM3 identifiziert werden konnte (vgl. Kapitel 4.4). Der Zusammenhang zwischen den Tests und dem 3. Mapping ist nachvollziehbar, da im 3. Mapping die Zusammenhänge zwischen 1. und 2. Mapping abgebildet und somit die gesamten Inhalte des Planspielunterrichts dargestellt werden. Das 3. Referenzmap stellt einige Zusammenhänge im Vergleich zu den Tests noch tiefgründiger dar. Beispielsweise werden die Bestandteile des Umlaufvermögens abgebildet, im Test wird lediglich eine Zuordnung des Umlaufvermögens zur Aktiv- bzw. Passivseite verlangt. Möglicherweise hätten weitere Aufgaben hierzu zu einer Verbesserung der Korrelation beigetragen. Um genauere Aussagen zu den Zusammenhängen der einzelnen Themenbereiche zwischen Tests und Mapping zu analysieren, könnten die Testergebnisse entsprechend der Themenbereiche der drei Mapping-Aktivitäten geclustert und den Scoringwerten für die drei Mappings gegenübergestellt werden. Aus forschungsökonomischen Gründen wird darauf jedoch verzichtet.

Neben den inhaltlichen Aspekten könnte auch die unterschiedliche Struktur eine Rolle spielen. Möglicherweise fällt es den Probanden leichter, Maps mit einer hohen Qualität zu erstellen, als gute Ergebnisse im Wissenstest zu erzielen, da aufwändige Umkodierungsprozesse aufgrund der Annahme einer netzwerkartigen Struktur des Gedächtnisses beim Erstellen der Maps nicht nötig sind. Aufgrund dessen, dass sich das Feedback ebenfalls der Methode Concept Mapping bedient, sind die hierbei erhaltenen Informationen ebenso direkt mit der kognitiven Struktur verknüpfbar. Infolgedessen hätte eine Anpassung der Aufgaben im Wissenstest an die Struktur der Concept Maps möglicherweise auch zu einem eindeutigeren Vorteil des Feedbacks geführt.

Um den langfristigen Wissenserwerb zu überprüfen, hätten die Probanden ergänzend zum 2. Nachtest aufgefordert werden können, die Inhalte in einem Concept Map darzustellen. Auf diese Weise wäre eine Analyse der langfristigen Entwicklung der Mapqualität möglich gewesen.

Beim Teilfragebogen zur Einschätzung der Methode des Concept Mapping kann von einer Inhaltsvalidität ausgegangen werden. Der Fragebogen enthält die wichtigsten Dimensionen, die zur Bewertung der Methode des Concept Mapping vonnöten sind, um insbesondere Hinweise zur Wirksamkeit des Concept Mapping-Trainings zu erhalten. Der Fragebogen orientiert sich an einem von Ryssel, Fürstenau und Thieme (2018) auf Basis einer Akzeptanzbefragung von Stracke (2004) und Eckert (1998) entwickelten Fragebogen. Eckert (1998) verwendete einen Fragebogen auf der Grundlage des LIST-Inventars, welches bereits an verschiedenen Stichproben validiert wurde (Wild, 2000). Dem Fragebogen zur Bewertung des Concept Mapping liegt jedoch kein theoretisches Modell zugrunde, so dass die Konstruktvalidität nur bedingt gegeben ist. Zur Prüfung der Kriteriumsvalidität hätten retrospektive Interviews durchgeführt werden können. Um insbesondere die Validität des Bereichs Handhabbarkeit zu prüfen, hätte ein weiterer Test durchgeführt werden können, der das Wissen über die Methode des Concept Mapping misst. Die Ergebnisse dieses Tests sollten mit der Selbsteinschätzung zur Handhabbarkeit der Methode hoch korrelieren.

Bei der Formulierung der Items zur Einschätzung der Nützlichkeit des einfachen und des elaborierten Feedbacks bzw. zum Wunsch nach zusätzlichem Feedback im Rahmen des Teil-Fragebogens zur Einschätzung der Nützlichkeit des Feedbacks dienten die Items zur Einschätzung der Nützlichkeit des Concept Mapping als Orientierungshilfe. Zur Ein-

schätzung der Nützlichkeit des elaborierten Feedbacks wurde ein zusätzliches Item aufgenommen, welches im Rahmen der Reliabilitätsanalyse jedoch wieder ausgeschlossen werden musste. Insgesamt kann daher von einer Inhaltsvalidität ausgegangen werden. Zur Konstruktvalidität kann auch hier keine Aussage getroffen werden, da dem Teil-Fragebogen kein theoretisches Modell zugrunde liegt. Zur Prüfung der Kriteriumsvalidität hätten hier ebenfalls retrospektive Interviews durchgeführt werden können.

Die Teil-Fragebögen zur Einschätzung des Concept Mapping sowie zur Einschätzung der Nützlichkeit des Feedbacks hätten nach den drei Mapping-Aktivitäten differenziert werden können. Infolgedessen hätte festgestellt werden können, inwieweit das 3. Mapping bzw. das Feedback zum 3. Mapping als nützlicher wahrgenommen wurde, als das 1. und 2. Mapping. Somit wären detailliertere Aussagen möglich gewesen, nach welcher Mapping-Aktivität ein ergänzendes Feedback überhaupt gewünscht war und welches Feedback (einfach oder elaboriert) wann als nützlich angesehen wird.

Der Teil-Fragebogen zur Einschätzung des Umgangs mit Fehlern wurde in Anlehnung an Spychiger, Oser, Mahler und Hascher (1998) sowie Spychiger, Kuster und Oser (2006) für alle vier Versuchsgruppen eingesetzt. Dazu wurde die Skala „Lernorientierung“ von Spychiger u. a. (1998, 2006) übernommen. Die Skala wurde von Spychiger u. a. mehrfach eingesetzt und weiterentwickelt. Insgesamt kann daher davon ausgegangen werden, dass das Zielkonstrukt vollständig erfasst wurde und die Inhaltsvalidität somit gegeben ist. Die Skala als Teil des Schülerfragebogens von Spychiger u. a. basiert auf einer Pilotstudie von Kaufmann (1996), der sich theoretisch sehr intensiv mit dem Fehlerbegriff auseinandersetzte und ein erstes Konzept zur Fehlerkultur in der Schule entwickelte. Daher kann von einer Konstruktvalidität ausgegangen werden. Zur Überprüfung der Kriteriumsvalidität wurden in den Studien keine Hinweise gegeben, so dass hierzu keine Aussage getroffen werden kann.

Der WIT-2 hat sich als gültiges Instrument bewährt. Die Kriteriums- sowie Konstruktvalidität konnte durch umfassende Validitätsstudien bestätigt werden. Zusätzlich wurden Akzeptanzwerte empirisch erhoben (Kersting u. a., 2008, S. 56 ff.).

Weitere kritische Aspekte und Optimierungsvorschläge und Ausblick

Insgesamt gilt es zu hinterfragen, inwieweit die Unterschiede hinsichtlich der Intervention zwischen den Versuchsgruppen zu gering waren. Aufgrund der Komplexität der Lerninhalte und des Aufgabenformats wurde entschieden, beiden Feedbackgruppen das

Referenzmap als richtige Lösung zu präsentieren (einfaches Feedback) sowie den Probanden das jeweilige Referenzmap mit dem erstellten Map vergleichen und korrigieren zu lassen. Erst im Anschluss erhielt eine der beiden Gruppen elaboriertes Feedback. Das Auseinandersetzen mit den erstellten Maps und das anschließende Korrigieren auf Basis der Referenzmaps regte möglicherweise schon Reflexionsprozesse an, bevor überhaupt elaboriertes Feedback gegeben wurde. Insofern den Probanden lediglich die Referenzmaps präsentiert worden wären, ohne diese zu korrigieren, hätten ggf. größere Unterschiede zwischen beiden Feedback-Gruppen hervorgerufen werden können. Die Gruppe *EIFB* hätte dann erst im Anschluss an das elaborierte Feedback zur Korrektur ihrer Maps aufgefordert werden können. An dieser Stelle ist jedoch kritisch anzumerken, dass sich die beiden Feedback-Gruppen dann auch hinsichtlich des Korrigierens unterschieden hätten.

Letztlich regt das Erstellen der Maps selbst schon Denkprozesse an, welche durch kognitive und metakognitive Prompts unterstützt wurden. Insbesondere metakognitive Prompts als Hinweise auf eine abschließende Kontrolle der erstellten Concept Maps unterstützten die Probanden dabei zu hinterfragen, inwieweit die Maps das erworbene Wissen widerspiegeln. Infolgedessen könnte also auch der Unterschied zwischen der Gruppe „kein Feedback“ und den Feedbackgruppen zu gering ausgefallen sein. Letztendlich stellt sich die Frage, ob nun ergänzendes Feedback zu Concept Maps überhaupt gegeben werden sollte. Kurzfristig hat sich gezeigt, dass elaboriertes Feedback eine lernwirksame Unterstützung darstellt. Darüber hinaus war der Wunsch nach ergänzendem Feedback eher hoch ausgeprägt ($MW=2,833$; $SD=0,786$).

Die Ergebnisse haben weiterhin gezeigt, dass es scheinbar gar nicht darum geht, ob Concept Maps um Feedback ergänzt werden sollten oder nicht, sondern dass dies von Einflussfaktoren wie Vorwissen oder Sprachfähigkeit abhängt. Probanden mit niedrigem Vorwissen und niedriger Sprachfähigkeit scheinen insbesondere vom elaborierten Feedback zu profitieren, wohingegen Probanden mit hohem Vorwissen und hoher Sprachfähigkeit maximal einfaches Feedback benötigen. Vor dem Einsatz der Methode Concept Mapping in Verbindung mit Feedback sollte daher das Vorwissen und die Sprachfähigkeit überprüft werden und entsprechend der Fähigkeiten Feedback angeboten werden. Bei hoher Sprachfähigkeit sollten darüber hinaus die Aufgabenanforderungen beachtet werden. Bei hohen Anforderungen benötigen die Probanden einfaches Feedback, bei niedrigen Anforderungen ist kein Feedback notwendig.

Diese Effekte zeigten sich lediglich kurzfristig. Um auch langfristige Effekte zu erzielen, ist es ratsam, die Methode Concept Mapping mehr zu üben und in den Lernprozess zu integrieren. Studien (Schroeder u. a., 2017; Pankrätius, 1990) haben bestätigt, dass der Effekt von Concept Maps erst bei längerfristigem Einsatz zum Tragen kommt. Schroeder u. a. (2017) begründen ihre Ergebnisse damit, dass die kognitive Belastung, welche insbesondere mit dem Format der Selbstkonstruktion verbunden ist, mit der Zeit durch eine erhöhte Vertrautheit verringert wird und sich der Lernende besser auf die kognitive Verarbeitung der Lerninhalte fokussieren kann. Hier sind jedoch Langzeitstudien vonnöten.

Insbesondere im Unterrichtskontext könnten Lehrpersonen das Concept Mapping als kontinuierliche Methode integrieren und in Abhängigkeit von individuellen Voraussetzungen und Aufgabenanforderungen um Feedback ergänzen. Die Anwendungsgebiete des Concept Mapping als Lernstrategie und als Diagnoseinstrument könnten dabei verbunden werden. Dem Argument des zeitlichen Aufwands könnte der ebenfalls anfallende Aufwand beim Korrigieren von Klassenarbeiten etc. entgegengebracht werden. Die Maps müssen inhaltsanalytisch ausgewertet und bewertet werden. Im Rahmen der Studie wurde ein differenziertes Verfahren vorgestellt. Hierbei zeigte sich jedoch eine hohe Korrelation mit der Berechnung eines Übereinstimmungsgrades nach Galanter, dessen Aufwand überschaubar ist. Der Einsatz von computergestützten Concept Mapping Tools könnte eine weitere zeitersparende und objektive Auswertung ermöglichen, die darüber hinaus der zunehmenden Digitalisierung gerecht wird. Das Training müsste dementsprechend angepasst werden.

Im Rahmen der Studie wurde festgestellt, dass sich das Wissensniveau zwischen den Probanden, welche Maps mit einer niedrigen und hohen Qualität erstellen, langfristig angleicht. Mit dem Einsatz der Selbstkonstruktion von Concept Maps als Lernstrategie und ergänzendem Diagnoseinstrument wäre es somit ggf. möglich, das Leistungsniveau innerhalb der Klasse bzw. Lerngruppe anzugleichen. Ergänzende Fehleranalysen der Maps könnten der Lehrperson dabei helfen, falsche und fehlende Zusammenhänge zu identifizieren. Insbesondere resistente Alltagsvorstellungen könnten somit thematisiert sowie die entsprechenden Begrifflichkeiten und Zusammenhänge im Rahmen des elaborierten Feedbacks zielgruppengerecht erläutert werden.

Für eine zukünftige Forschung zur Lernwirksamkeit von Feedback zu Concept Maps wäre es somit zu empfehlen, Langzeitstudien durchzuführen, die die Selbstkonstruktion von Concept Maps als kontinuierliche Methode in den Lernprozess einbinden. Hierbei

stellt ein detailliertes Training vorab eine Grundvoraussetzung dar. Um zu erkennen, inwieweit die Methode angemessen beherrscht wird, könnten entsprechende Tests Aufschluss geben. Darüber hinaus sollten die verbalen Fähigkeiten sowie das Vorwissen vorab überprüft werden, um den Probanden im Rahmen der Studie entsprechendes Feedback anbieten zu können. Die eingesetzten Wissenstests sollten an die Struktur von Concept Maps angepasst und eine weitere Concept Mapping-Aktivität zur Überprüfung des langfristigen Wissenserwerbs eingesetzt werden. Um das elaborierte Feedback optimal zu gestalten, könnten die Erkenntnisse aus der Fehleranalyse der Concept Maps genutzt und den Probanden gezielte Hinweise für eine tiefgründige Verarbeitung der erlernten Inhalte gegeben werden.

Die Methode Concept Mapping als Lernstrategie unterstützt dabei, sehr komplexe Inhalte strukturell darzustellen und fördert darüber hinaus das Verstehen und Behalten neuen Wissens, indem neue Informationen mit bestehendem Wissen vernetzt werden können. Das Feedback kann dabei bei sehr komplexen Themenbereichen eine tiefergehende Verarbeitung der Lerninhalte unterstützen. Indem das Feedback entsprechend der Bedürfnisse der Lernenden mit der Methode des Concept Mapping verbunden wird, ist es somit möglich den Planspielunterricht so zu unterstützen, dass eine Tiefenverarbeitung der erlernten Inhalte erreicht wird und die Lernenden auf die komplexen Anforderungen des beruflichen Alltags vorbereitet werden.

Literaturverzeichnis

- Achtenhagen, F. (1995). Komplexe Lehr-Lern-Arrangements. In R. Dubs & R. Dörig (Hrsg.), *Dialog Wissenschaft und Praxis. Berufsbildungstage Sankt Gallen 23. bis 25. Februar 1995*. (S. 374–428). Sankt Gallen: Institut für Wirtschaftspädagogik.
- Achtenhagen, F., Tramm, T., & Preiss, P. (1992). *Lernhandeln in komplexen Situationen. Neue Konzepte der betriebswirtschaftlichen Ausbildung*. Wiesbaden: Gabler.
- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*(1), 3–27. <http://dx.doi.org/10.1037/0033-2909.102.1.3>
- Aebli, H. (1983). *Zwoelf Grundformen des Lehrens. Eine Allgemeine Didaktik auf psycholog. Grundlage*. Stuttgart: Klett.
- Álvarez-Montero, F. J., Jacobo-García, H., & Rocha-Ruiz, E. (2015). Feedback in Computer-Based Concept Mapping Tools: A Short Review. In P. Zaphiris & A. Ioannou (Hrsg.), *Learning and Collaboration Technologies* (Bd. 9192, S. 187–198). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20609-7_18
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of educational outcomes: Complete edition*. New York: Longman.
- Arbinger, R. (1997). *Psychologie des Problemlösens : eine anwendungsorientierte Einführung*. Darmstadt : Wiss. Buchges.
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360° feedback. *Journal of Vocational Behavior*, *66*(3), 532–548.
- Ausubel, D. P. (1974). *Psychologie des Unterrichts 1* (2., völlig überarb. Aufl.). Weinheim: Beltz.
- Ausubel, D. P. (1978). Die Förderung bedeutungsvollen verbalen Lernens. *Unterrichtswissenschaft*, *6*(1), 58–66.
- Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2015). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (14. Aufl.). Berlin Heidelberg: Springer Gabler.
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Theories of Cognitive Self-Regulation*, *50*(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Bandura, A. (1997). *Self Efficacy: The Exercise of Control*. New York: Worth.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238. <http://dx.doi.org/10.2307/1170535>
- Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*(04), 689–725. <https://doi.org/10.1017/S0272263113000429>
- Barenholz, H., & Tamir, P. (1992). A Comprehensive Use of Concept Mapping in Design Instruction and Assessment. *Research in Science & Technological Education*, *10*(1), 37–52. <https://doi.org/10.1080/0263514920100104?needAccess=true>

- Becker, D., Oldenbürger, H. A., & Piehl, J. (1987). Motivation und Emotion. In G. Lüer (Hrsg.), *Allgemeine experimentelle Psychologie* (S. 431–470). Stuttgart: Gustav Fischer.
- Behrendt, H., & Reiska, P. (2001). Abwechslung im Naturwissenschaftsunterricht mit Concept Mapping. *Plus Lucis*, 1, 9–12.
- Bell, S. T., & Arthur, W. (2008). Feedback acceptance in developmental assessment centers: the role of feedback message, participant personality, and affective response to the feedback session. *Journal of Organizational Behavior*, 29(5), 681–703.
- Bernd, H., & Jüngst, K. L. (1999). Lernen mit Concept Maps: Lerneffektivität von Selbstkonstruktion und Durcharbeiten. In W. K. Schulz (Hrsg.), *Aspekte und Probleme der didaktischen Wissensstrukturierung*. (S. 113–129). Frankfurt, Main: Lang.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Blair, A., & McGinty, S. (2013). Feedback-dialogues: exploring the student perspective. *Assessment & Evaluation in Higher Education*, 38(4), 466–476.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain* (Bd. 19). New York: David McKay Co Inc.
- Bonato, M. (1990). *Wissensstrukturierung mittels Struktur-lege-Techniken : Eine graphentheoretische Analyse von Wissensnetzen*. Frankfurt: Lang.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Mit 156 Abbildungen und 87 Tabellen*. (4., überarb. Aufl.). Berlin u.a.: Springer.
- Bortz, J., & Schuster, C. (2016). *Statistik für Human- und Sozialwissenschaftler: Limitierte Sonderausgabe* (7. Aufl.). Berlin Heidelberg: Springer.
- Brookhart, S. M. (2010). *Wie sag ich 's meinem Schüler? So kommt Ihr Feedback wirklich an [für alle Schulformen, Fächer und Klassenstufen]*. Weinheim: Beltz.
- Brosius, F. (2011). *SPSS 19* (1. Aufl.). Heidelberg: Verlagsgruppe Hüthig-Jehle-Rehm.
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97–111. <https://doi.org/10.1016/j.learninstruc.2008.02.003>
- Bühl, A. (2012). *SPSS 20: Einführung in die moderne Datenanalyse* (13. Aufl.). München: Pearson Studium.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler* (1. Aufl.). München: Pearson Studium.

- Bürgermeister, A., Klieme, E., Rakoczy, K., Harks, B., & Blum, W. (2014). Formative Leistungsbeurteilung im Unterricht: Konzepte, Praxisberichte und ein neues Diagnoseinstrument für das Fach Mathematik. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 41–60). Göttingen: Hogrefe.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281. <https://doi.org/10.3102/00346543065003245>
- Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(6), 1491–1494. <https://doi.org/10.1037/0278-7393.27.6.1491>
- Cañas, A. J., Coffey, J. W., Carnot, M.-J., Feltovich, P., Hoffman, R. R., Feltovich, J., & Novak, J. D. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. *Report to the Chief of Naval Education and Training, Pensacola, Florida, IHMC*. Abgerufen von <https://ihmc.us/users/acanas/publications/conceptmaplitreview/ihmc%20literature%20review%20on%20concept%20mapping.pdf>
- Chandler, P., & Sweller, J. (1992). The Split-attention Effect as a Factor in the Design of Instruction. *British Journal of Educational Psychology; Edinburgh*, 62(2). <https://doi.org/10.1111/j.2044-8279.1992.tb01017.x>
- Chang, K.-E., Sung, Y.-T., & Chen, I.-D. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, 17(1), 21–33. <https://doi.org/10.1111/j.1365-2729.2001.00156.x>
- Chang, K.-E., Sung, Y.-T., & Chen, I.-D. (2002). The Effect of Concept Mapping to Enhance Text Comprehension and Summarization. *The Journal of Experimental Education*, 71(1), 5–23. <https://doi.org/10.1080/00220970209602054>
- Chase, J. A., & Houmanfar, R. (2009). The Differential Effects of Elaborate Feedback and Basic Feedback on Student Performance in a Modified, Personalized System of Instruction Course. *Journal of Behavioral Education*, 18(3), 245–265. <https://doi.org/10.1007/s10864-009-9089-2>
- Cheng, W., & Warren, M. (1997). Having second thoughts: student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233–239.
- Chiu, M. M. (2008). Flowing Toward Correct Contributions During Group Problem Solving: A Statistical Discourse Analysis. *Journal of the Learning Sciences*, 17(3), 415–463. <https://doi.org/10.1080/10508400802224830>
- Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16(1), 64–73. <https://doi.org/10.2307/3150876>
- Cianci, A. M., Schaubroeck, J. M., & McGill, G. A. (2010). Achievement Goals, Feedback, and Task Performance. *Human Performance*, 23(2), 131–154.
- Clariana, R. B. (1993). A review of multiple-try feedback in traditional and computer-based instruction. *Journal of Computer-Based Instruction*, 20(3), 67–74.

- Clariana, R. B. (2000). Feedback in computer-assisted learning. In *NETg University of Limerick Lecture Series*. Abgerufen von https://www.researchgate.net/profile/Roy_Clariana/publication/246177453_Feedback_in_computer-assisted_learning/links/597762a70f7e9b277721c66b/Feedback-in-computer-assisted-learning.pdf
- Clariana, R. B., & Koul, R. (2005). Multiple-Try Feedback and Higher-Order Learning Outcomes. *International Journal of Instructional Media*, 32(3), 239.
- Clariana, R. B., & Koul, R. (2006). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. *British Journal of Educational Psychology*, 76(2), 259–270.
- Clariana, R. B., Wagner, D., & Murphy, L. C. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development*, 48(3), 5–22. <https://doi.org/10.1007/BF02319855>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale (2. Aufl.). Hillsdale, N.J: Lawrence Erlbaum Associates Publishers.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240–247. [https://doi.org/10.1016/S0022-5371\(69\)80069-1](https://doi.org/10.1016/S0022-5371(69)80069-1)
- Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (S. 245–252). Seattle, Washington, USA: ACM.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa: Psychological Assessment Resources.
- Dansereau, D. F. (1978). The Development of a Learning Strategies Curriculum I. In H. F. O’Neil (Hrsg.), *Learning Strategies* (S. 1–29). New York: Academic Press. <https://doi.org/10.1016/B978-0-12-526650-5.50006-X>
- De Bleser, R. (2012). Aufbau und Funktionen der Sprache. In H.-O. Karnath & P. Thier (Hrsg.), *Kognitive Neurowissenschaften* (S. 423–428). Berlin: Springer.
- de Kleijn, R. A. M., Mainhard, M. T., Meijer, P. C., Brekelmans, M., & Pilot, A. (2013). Master’s thesis projects: student perceptions of supervisor feedback. *Assessment & Evaluation in Higher Education*, 38(8), 1012–1026.
- de Kleijn, R. A. M., Meijer, P. C., Pilot, A., & Brekelmans, M. (2014). The relation between feedback perceptions and the supervisor–student relationship in master’s thesis projects. *Teaching in Higher Education*, 19(4), 336–349.
- Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In *Nebraska Symposium on Motivation, 1990: Perspectives on motivation*. (S. 237–288). Lincoln, NE, US: University of Nebraska Press.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, 22(4), 499–533.

- Dempsey, J. V., Driscoll, M. P., & Swindell, L. K. (1993). Text-based feedback. In J. V. Dempsey & G. C. Sales (Hrsg.), *Interactive instruction and feedback* (S. 21–54). Englewood Cliffs, New Jersey: Educational Technology Publications.
- Dempsey, J. V., Litchfield, B. C., & Driscoll, M. P. (1993). Feedback, Retention, Discrimination Error, and Feedback Study Time. *Journal of Research on Computing in Education*, 25(3), 303–326. <https://doi.org/10.1080/08886504.1993.10782053>
- DePasque, S., & Tricomi, E. (2015). Effects of intrinsic motivation on feedback processing during learning. *NeuroImage*, 119, 175–186. <https://doi.org/10.1016/j.neuroimage.2015.06.046>
- Doerr, K. H., Mitchell, T. R., Klastorin, T. D., & Brown, K. A. (1996). Impact of material flow policies and goals on job outcomes. *Journal of Applied Psychology*, 81(2), 142–152. <http://dx.doi.org/10.1037/0021-9010.81.2.142>
- Donovan, J., & Hafsteinsson, L. (2006). The Impact of Goal-Performance Discrepancies, Self-Efficacy, and Goal Orientation on Upward Goal Revision 1. *Journal of Applied Social Psychology*, 36(4), 1046–1069.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer. Abgerufen von <http://dx.doi.org/10.1007/978-3-642-41089-5>
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* (3. Aufl.). Stuttgart: Kohlhammer.
- Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (2016). *Serious games : foundations, concepts and practice*. Switzerland: Springer. Abgerufen von <http://link.springer.com/10.1007/978-3-319-40612-1>
- Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Hrsg.), *Focus on form in classroom second language acquisition* (S. 114–138). Cambridge: Cambridge University Press.
- Duden online. (2017). *Stichwort: Feedback*. Abgerufen 22. Februar 2017, von <http://www.duden.de/rechtschreibung/Feedback>
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040–1048. <https://doi.org/10.1037/0003-066X.41.10.1040>
- Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computerunterstützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs-Technik (NET)*. Lengerich: Pabst.
- Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *Modern Language Journal*, 94(1), 1–21. <http://doi:10.1111/j.1540-4781.2009.00980.x>
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of personality and social psychology*, 54(1), 5.
- Fengler, J. (2009). *Feedback geben: Strategien und Übungen* (4. Aufl.). Weinheim: Beltz.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <http://dx.doi.org/10.1037/0003-066X.34.10.906>

- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8410–8415.
- Friedrich, H. F. (1995). *Training und Transfer reduktiv-organisierender Strategien für das Lernen mit Texten*. Münster: Aschendorff.
- Friedrich, H. F., & Mandl, H. (1992). Lern- und Denkstrategien - ein Problemaufriß. In H. Mandl & H. F. Friedrich (Hrsg.), *Lern- und Denkstrategien. Analyse und Intervention* (S. 3–54). Göttingen: Hogrefe.
- Friedrich, H. F., & Mandl, H. (1997). Analyse und Förderung selbstgesteuerten Lernens. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich D, Praxisgebiete. Serie 1, Pädagogische Psychologie. Bd. 4, Psychologie der Erwachsenenbildung*. (S. 237–293). Göttingen u.a.: Hogrefe.
- Friedrich, H. F., & Mandl, H. (2006). Lernstrategien: Zur Strukturierung des Forschungsfeldes. In H. Mandl & H. F. Friedrich (Hrsg.), *Handbuch Lernstrategien*. (S. 7–23). Göttingen: Hogrefe.
- Fürstenau, B. (2001). Empirische Prüfung zweier Lehr-Lern-Arrangements für Industriekaufleute. Evaluation von Wissen mit Hilfe von Netzwerken. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 97(2), 247–270.
- Fürstenau, B. (2009). Planspiel und Simulation. In K.-H. Arnold, U. Sandfuchs, & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 240–243). Stuttgart: UTB GmbH.
- Fürstenau, B. (2011). Concept Maps im Lehr-Lern-Kontext. *DIE-Zeitschrift für Erwachsenenbildung*, (1), 46–48.
- Fürstenau, B. (2016). Theorie des bedeutungsvollen verbalen Lernens. In B. Fürstenau (Hrsg.), *Lehr-Lern-Theorien / Behaviorismus, Kognitivismus, Konstruktivismus: Lernen und Expertise verstehen und fördern* (Bd. 6, S. 27–41). Hohengehren: Schneider.
- Fürstenau, B. (2018). Begriff, Einsatzmöglichkeiten und Training von Concept Mapping. In B. Fürstenau (Hrsg.), *Concept Mapping - Trainings- und Instruktionsbeispiele; Dresdner Beiträge zur Wirtschaftspädagogik (1/2018)* (S. 1–11). Dresden: TU Dresden, Professur für Wirtschaftspädagogik.
- Fürstenau, B., Ryssel, J., & Kunath, J. (2010). Concept Mapping versus Summary Writing as Instructional Devices for Understanding Complex Business Problems Symposium: Fostering the Acquisition and Application of Domain-Specific Knowledge through Concept Mapping. In K. Gomez, L. Lyons, & J. Radinsky (Hrsg.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010)* (S. 14–16). Chicago IL: International Society of the Learning Sciences.
- Fürstenau, B., & Trojahnner, I. (2005). Prototypische Netzwerke als Ergebnis struktureller Inhaltsanalysen. In P. Gonon, F. Klauser, R. Nickolaus, & R. Huisinga (Hrsg.), *Kompetenz, Kognition und Neue Konzepte der beruflichen Bildung* (S. 191–202). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-86895-4_13

- Gabelica, C., van den Bossche, P., Segers, M., & Gijsselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, 7(2), 123–144. <https://doi.org/10.1016/j.edurev.2011.11.003>
- Gagné, R. M. (1980). *Die Bedingungen des menschlichen Lernens*. (B. Meyer & H. Skowronek, Übers.) (5., neubearb. Aufl. nach d. 3., amerikan. Aufl.). Hannover: Schroedel.
- Gärtner, H. (2013). Wirksamkeit von Schülerfeedback als Instrument der Selbstevaluation von Unterricht. In J. R. Hense (Hrsg.), *Forschung über Evaluation. Bedingungen, Prozesse und Wirkungen*. (S. 107–124). Münster u.a.: Waxmann.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Unravelling Peer Assessment*, 20(4), 304–315. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Han, Z. (2002). A Study of the Impact of Recasts on Tense Consistency in L2 Output. *TESOL Quarterly*, 36(4), 543–572.
- Hank, B. (2015). Conceptual Change - relevant für die Hochschullehre? In B. V. Berendt (Hrsg.), *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten. [Teil] A. Lehren und Lernen. Psychologische Aspekte*. (S. A 2.8, S. 7-24). Berlin: Raabe.
- Hardy, I., & Stadelhofer, B. (2006). Concept Maps wirkungsvoll als Strukturierungshilfen einsetzen. Welche Rolle spielt die Selbstkonstruktion? *Zeitschrift für pädagogische Psychologie*, 20(3), 175–187.
- Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, 34(3), 269–290. <https://doi.org/10.1080/01443410.2013.785384>
- Hascher, T., & Hagenauer, G. (2010). Lernen aus Fehlern. In C. Spiel, B. Schober, P. Wagner, & R. Reimann (Hrsg.), *Bildungspsychologie* (S. 377–381). Göttingen u.a.: Hogrefe.
- Hattie, J. (2014). *Lernen sichtbar machen für Lehrpersonen: Überarbeitete deutschsprachige Ausgabe von „Visible Learning for Teachers“: Überarbeitete ... Ausgabe von „Visible Learning for Teachers“*. (K. Zierer & W. Beywl, Hrsg.) (1. korr. Aufl.). Baltmannsweiler: Schneider.
- Hattie, J. (2015). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von „Visible Learning“*. (W. Beywl & K. Zierer, Hrsg.) (3. erw. Aufl.). Baltmannsweiler: Schneider Hohengehren.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hattie, J., & Wollenschläger, M. (2014). A conceptualization of feedback. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 135–149). Münster: Waxmann.
- Haugwitz, M., Nesbit, J. C., & Sandmann, A. (2010). Cognitive ability and the instructional efficacy of collaborative concept mapping. *Learning and Individual Differences*, 20(5), 536–543. <https://doi.org/10.1016/j.lindif.2010.04.004>

- Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16(2), 416–431. <https://doi.org/10.1017/S0958344004001120>
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)*. (M. Hasselhorn, H. Marx, & W. Schneider, Hrsg.). Göttingen: Beltz Test.
- Heslin, P. A., & Latham, G. P. (2004). The Effect of Upward Feedback on Managerial Behavior. *Applied Psychology: An International Review*, 53(1), 23–37.
- Hilbert, T. S., Nückles, M., & Matzel, S. (2008). Concept mapping for learning from text: evidence for a worked-out-map-effect. In *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 1* (S. 358–365). International Society of the Learning Sciences. Abgerufen von <http://dl.acm.org/citation.cfm?id=1599855>
- Hilbert, T. S., Nückles, M., Renkl, A., Minarik, C., Reich, A., & Ruhe, K. (2008). Concept Mapping zum Lernen aus Texten: Können Prompts den Wissens- und Strategieerwerb fördern? *Zeitschrift für Pädagogische Psychologie*, 22(2), 119–125. <https://doi.org/10.1024/1010-0652.22.2.119>
- Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: what characterizes good and poor mappers? *Instructional Science*, 36(1), 53–73. <https://doi.org/10.1007/s11251-007-9022-9>
- Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Including the Special Issue: State of the Art Research into Cognitive Load Theory*, 25(2), 267–274. <https://doi.org/10.1016/j.chb.2008.12.006>
- Holsti, O. R. ; (1969). *Content analysis for the social sciences and humanities*. Reading, Mass.: Addison-Wesley.
- Homburg, C., & Giering, A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte. Ein Leitfaden für die Marketingforschung. *Marketing : ZFP; journal of research and management*, 18(1), 3–24.
- Horton, P. B., McConney, A. A., Gallo, M., Woods, A. L., Senn, G. J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education*, 77(1), 95–111. <https://doi.org/10.1002/sce.3730770107>
- Hurrelmann, K., Grundmann, M., & Walper, S. (2008). *Handbuch Sozialisationsforschung* (7. Aufl.). Weinheim: Beltz.
- Huth, K. (2004). *Entwicklung und Evaluation von fehlerspezifischem informativem tutoriellen Feedback (ITF) für die schriftliche Subtraktion*. Technische Universität, Dresden. Abgerufen von <http://nbn-resolving.de/urn:nbn:de:swb:14-1105354057406-47158>
- Hwang, G.-J., Wu, P.-H., & Ke, H.-R. (2011). An interactive concept map approach to supporting mobile learning activities for natural science courses. *Computers & Education*, 57(4), 2272–2280. <https://doi.org/10.1016/j.compedu.2011.06.011>
- Hyland, F., & Hyland, K. (2001). Sugaring the Pill: Praise and Criticism in Written Feedback. *Journal of Second Language Writing*, 10(3), 185–212.

- Ifenthaler, D. (2010). Bridging the gap between expert-novice differences: The model-based feedback approach. *Journal of Research on Technology in Education*, 43(2), 103–117.
- Ilgen, D. R., & Davis, C. A. (2000). Bearing Bad News: Reactions to Negative Performance Feedback. *Applied Psychology: An International Review*, 49(3), 550.
- Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Janssen, J., & Laatz, W. (2017). *Statistische Datenanalyse mit SPSS: Eine anwendungsorientierte Einführung in das Basissystem und das Modul Exakte Tests* (9. Aufl.). Berlin: Springer Gabler.
- Jehn, K. A., & Shah, P. P. (1997). Interpersonal Relationships and Task Performance: An Examination of Mediating Processes in Friendship and Acquaintance Groups. *Journal of Personality & Social Psychology*, 72(4), 775–790.
- Jüngst, K. L. (1992). *Lehren und Lernen mit Begriffsnetzdarstellungen. Zur Nutzung von concept-maps bei der Vermittlung fachspezifischer Begriffe in Schule, Hochschule, Aus- und Weiterbildung*. Frankfurt, Main: Afra.
- Jüngst, K. L., & Strittmatter, P. (1995). Wissensstrukturdarstellung: Theoretische Ansätze und praktische Relevanz. *Unterrichtswissenschaft*, 23(3), 194–207.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Kaufmann, S. (1996). *Mal den Fehler an die Wand - Konzept der Fehlerkultur in der Schule* (Lizentiatsarbeit). Pädagogisches Institut der Universität Freiburg.
- Kersting, M., Althoff, K., & Jäger, A. O. A. (2008). *Wilde-Intelligenz-Test 2 / WIT-2*. (Deutsche Gesellschaft für Personalwesen, Hrsg.). Göttingen: Hogrefe. Abgerufen von <http://d-nb.info/990045765/04>
- King, P. E., Schrodtt, P., & Weisel, J. J. (2009). The Instructional Feedback Orientation Scale: Conceptualizing and Validating a New Measure for Assessing Perceptions of Instructional Feedback. *Communication Education*, 58(2), 235–261.
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues & Practice*, 30(4), 28–37.
- Klauer, Karl Josef; (1987). *Kriteriumsorientierte Tests / Lehrbuch d. Theorie u. Praxis lehrzielorientierten Messens*. Göttingen: Hogrefe.
- Klauer, Karl Josef. (1993). *Denktraining für Jugendliche. Ein Programm zur intellektuellen Förderung. Handanweisung*. Göttingen: Hogrefe.
- Kline, F. M., Schumaker, J. B., & Deshler, D. D. (1991). Development and validation of feedback routines for instructing students with learning disabilities. *Learning Disability Quarterly*, 14(3), 191–207.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>

- Kopp, B., & Mandl, H. (2014). Aspekte der Feedbacknachricht. In H. Ditton & A. Müller (Hrsg.), *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder* (S. 151–162). Münster: Waxmann.
- Krampen, G. (1985). Differentielle Effekte von Lehrerkommentaren zu Noten bei Schülern. [Different effects of teachers' comments on the work of students.]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *17*(2), 99–123.
- Krampen, G. (1987). Differential effects of teacher comments. *Journal of Educational Psychology*, *79*(2), 137–146.
- Krause, U.-M. (2007). *Feedback und kooperatives Lernen* (Bd. 60). Münster: Waxmann.
- Kuckartz, U. (2014). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (2., durchgesehene Aufl.). Weinheim; Basel: Beltz Juventa.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of educational research*, *47*(2), 211–232.
- Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, *63*(5), 505–512. <https://doi.org/10.1037/h0033243>
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.
- Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. V. Dempsey & G. C. Sales (Hrsg.), *Interactive instruction and feedback* (S. 3–20). Englewood Cliffs, New Jersey: Educational Technology Publications.
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*, *10*(3), 285–291. [https://doi.org/10.1016/0361-476X\(85\)90025-6](https://doi.org/10.1016/0361-476X(85)90025-6)
- Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79–97. <https://doi.org/10.2307/1170349>
- Kung, M. C. (2008). *Why and how do people seek success and failure feedback: A closer Look at motives, methods and cultural differences* (Unpublished doctoral dissertation). Florida Institute of Technology, Florida.
- Lambiotte, J., & Dansereau, D. (1992). Effects of Knowledge Maps and Prior Knowledge on Recall of Science Lecture Content. *Journal of Experimental Education*, *60*(3), 189–201.
- Lambiotte, J. G., Dansereau, D. F., Cross, D. R., & Reynolds, S. B. (1989). Multirelational semantic maps. *Educational Psychology Review*, *1*(4), 331–367. <https://doi.org/10.1007/BF01320098>
- Lambiotte, J. G., Skaggs, L. P., & Dansereau, D. F. (1993). Learning from lectures: Effects of knowledge maps and cooperative review strategies. *Applied Cognitive Psychology*, *7*(6), 483–497.
- Langer, I., Schulz von Thun, F., & Tausch, R. (2011). *Sich verständlich ausdrücken* (9., neu gestaltete Aufl.). München: Reinhardt.

- Lau, D. (2011). *Algebra und Diskrete Mathematik 1: Grundbegriffe der Mathematik, Algebraische Strukturen 1, Lineare Algebra und Analytische Geometrie, Numerische Algebra und Kombinatorik* (3. Aufl.). Berlin Heidelberg: Springer-Verlag. Abgerufen von //www.springer.com/de/book/9783642194429
- Lee, D., Smith, P., & Savenye, W. (1991). The effects of feedback and second try in computer-assisted instruction for rule-learning task. In *Proceedings of selected research papers, Association for Educational Communications and Technology, Research and Theory Division* (S. 432–441). Orlando.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, 63(1), 13–22. <https://doi.org/10.1093/elt/ccn010>
- Lee, O. M. (1985). *The effect of type of feedback on rule learning in computer based instruction (mathematics, numbers, base conversion, micro-computers, high school, superpilot, cbi)* (Unpublished doctoral dissertation). The Florida State University.
- Lehmenkühler, A., Roscher, H., & Theis, W. (1976). Feedback: Anmerkungen zu Funktion und Form. In M. Sader, W. Schäuble, & W. Theis (Hrsg.), *Verbesserung von Interaktion durch Gruppendynamik: Psycholog. Grundlagen u. pragmat. Strukturierungen*. Münster: Aschendorff.
- Leopold, C. (2009). *Lernstrategien und Textverstehen: Spontaner Einsatz und Förderung von Lernstrategien*. Münster: Waxmann Verlag.
- Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory: The Interactions Between the Effects of Implicit and Explicit Feedback. *The Modern Language Journal*, 97(3), 634–654. <https://doi.org/10.1111/j.1540-4781.2013.12030.x>
- Lienert, G., & Raatz, U. (1998). *Testaufbau und Testanalyse. 6. Aufl.* Weinheim: Psychologie Verlags Union.
- Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 739.
- Litfin, T., Teichmann, M.-H., & Clement, M. (2000). Beurteilung der Güte von explorativen Faktorenanalysen im Marketing. *Wirtschaftswissenschaftliches Studium: WiSt: Zeitschrift für Studium und Forschung*, 29(5), 283–286.
- Luque, M. F. S. D., & Sommer, S. M. (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review*, 25(4), 829–849.
- Lyden, J. A., Chaney, L. H., Danehower, V. C., & Houston, D. A. (2002). Anchoring, Attributions, and Self-Efficacy: An Examination of Interactions. *Contemporary Educational Psychology*, 27(1), 99.
- Lyster, R. (2004). Differential Effects of Prompts and Recasts in form-Focused Instruction. *Studies in Second Language Acquisition*, 26(3), 399–432. <http://dx.doi.org/10.1017/S0272263104263021>
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in second language acquisition*, 22(4), 471–497.

- Maier, U. (2010). Formative assessment—A promising concept for improving instruction and classroom assessment. *Zeitschrift für Erziehungswissenschaft*, 13(2), 293–308. <https://doi.org/10.1007/s11618-010-0124-9>
- Maier, U. (2014). Formative Leistungsdiagnostik in der Sekundarstufe. Grundlegende Fragen, domänenspezifische Verfahren und empirische Befunde. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik*. (S. 19–39). Göttingen: Hogrefe.
- Mandl, H., & Fischer, F. (2000). Mapping-Techniken und Begriffsnetze in Lern- und Kooperationsprozessen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen - Wissensmanagement mit Mapping-Techniken* (S. 3–12). Göttingen: Hogrefe. Abgerufen von https://www.wiso-net.de/document/ECON__534523064
- Mandl, H., & Friedrich, H. F. (2006). *Handbuch Lernstrategien*. Göttingen: Hogrefe. Abgerufen von <http://www.ciando.com/ebook/bid-5634>
- Mandl, H., Friedrich, H. F., & Hron, A. (1988). Theoretische Ansätze zum Wissenserwerb. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie* (S. 123–160). München: Psychologie Verlags Union.
- Mason, B. J., & Bruning, R. (2001, Januar 1). *Providing Feedback in Computer-based Instruction: What the Research Tells Us* (CLASS Research Report No. 9). Center for Instructional Innovation, University of Nebraska-Lincoln.
- Matsui, T., Kakuyama, T., & Onglatco, M. U. (1987). Effects of goals and feedback on performance in groups. *Journal of Applied Psychology*, 72(3), 407–415. <https://doi.org/10.1037/0021-9010.72.3.407>
- Mayer, R. E. (1979). Can Advance Organizers Influence Meaningful Learning? *Review of Educational Research*, 49(2), 371–383.
- Mayer, R. E. (2005). Introduction to Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 1–16). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.002>
- Mayer, R. E. (2009). *Multimedia Learning* (2. Aufl.). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>
- Mayer, R. E., & Anderson, R. B. (1991). Animation Need Narrations: An Experimental Test of a Dual-Coding Hypothesis. *Journal of Educational Psychology*, 83(4), 484.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4), 444–452. <http://dx.doi.org/10.1037/0022-0663.84.4.444>
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312–320. <http://dx.doi.org/10.1037/0022-0663.90.2.312>
- Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, 12(1), 107–119.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

- Mayring, P. (2015). *Qualitative Inhaltsanalyse / Grundlagen und Techniken*. Weinheim: Beltz.
- McCagg, E. C., & Dansereau, D. F. (1991). A convergent paradigm for examining knowledge mapping as a learning strategy. *The Journal of Educational Research*, 84(6), 317–324.
- McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475–492. [https://doi.org/10.1002/\(SICI\)1098-2736\(199904\)36:4<475::AID-TEA5>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2736(199904)36:4<475::AID-TEA5>3.0.CO;2-O)
- McKendree, J. (1990). Effective Feedback Content for Tutoring Complex Skills. *Human-Computer Interaction*, 5(4), 381–413. https://doi.org/10.1207/s15327051hci0504_2
- Moore, D. W., & Readence, J. E. (1984). A Quantitative and Qualitative Review of Graphic Organizer Research. *Journal of Educational Research*, 78(1), 11.
- Moreno, R. (2004). Decreasing Cognitive Load for Novice Students: Effects of Explanatory versus Corrective Feedback in Discovery-Based Multimedia. *Instructional Science*, 32(1), 99–113. <https://doi.org/10.1023/B:TRUC.0000021811.66966.1d>
- Mory, E. H. (2004). Feedback research revisited. In *Handbook of research on educational communications and technology*, 2nd ed. (S. 745–783). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Miller, J., Newton, D. P., & Gregson, M. (2005). *Frameworks for Thinking: A Handbook for Teaching and Learning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511489914>
- Müller, A., & Ditton, H. (2014). Feedback: Begriff, Formen und Funktionen. In *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. (S. 11–28). Münster: Waxmann.
- Mummendy, H. D., & Grau, I. (2008). *Die Fragebogen-Methode / (5., überarb. und erw. Aufl.)*. Göttingen: Hogrefe.
- Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior & Human Performance*, 23(3), 309–338. [https://doi.org/10.1016/0030-5073\(79\)90001-1](https://doi.org/10.1016/0030-5073(79)90001-1)
- Narciss, S. (2006). *Informatives tutorielles Feedback*. Münster: Waxmann.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. van Merriënboer, & M. P. Driscoll (Hrsg.), *Handbook of research on educational communications and technology* (S. 125–143). New York: Erlbaum.
- Narciss, S. (2014). Modelle zu den Bedingungen und Wirkungen von Feedback in Lehr-Lernsituationen. In *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. (S. 43–82). Münster.
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, R. Brünken, & D. Leutner (Hrsg.), *Instructional design for multimedia learning* (S. 181–195). Münster: Waxmann.

- Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*(4), 310–322. <http://dx.doi.org/10.1016/j.learninstruc.2006.07.003>
- Nesbit, J. C., & Adesope, O. O. (2006). Learning With Concept and Knowledge Maps: A Meta-Analysis. *Review of Educational Research, 76*(3), 413–448.
- Nesbit, J. C., & Adesope, O. O. (2013). Concept maps for learning. *Learning through visual displays. Charlotte, NC: Information Age Publishing*, 303–328.
- Niegemann, H. M., Domagk, S., Hessel, S., Hein, A., Hupfer, M., & Zobel, A. (2008). *Kompendium multimediales Lernen*. Berlin u.a.: Springer. Abgerufen von <http://dx.doi.org/10.1007/978-3-540-37226-4>
- Novak, J. D., & Cañas, A. J. (2008). The theory underlying concept maps and how to construct and use them. Abgerufen von <http://eprint.ihmc.us/5/2/TheoryUnderlyingConceptMaps.pdf>
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.
- Nückles, M., Gurlitt, J., Pabst, T., & Renkl, A. (2004). *Mind Maps und Concept Maps: Visualisieren–Organisieren–Kommunizieren*. Beck-Wirtschaftsberater im dtv, München. München: Deutscher Taschenbuchverlag.
- Nunnally, J. C. ; (1978). *Psychometric theory* (2. Aufl.). New York: McGraw-Hill.
- Oberauer, K., Mayr, U., & Kluwe, R. H. (2006). Gedächtnis und Wissen. In H. Spada (Hrsg.), *Lehrbuch Allgemeine Psychologie* (S. 115–188). Bern: Huber.
- O'Donnell, A. M. (1994). Learning from knowledge maps: The effects of map orientation. *Contemporary Educational Psychology, 19*(1), 33.
- O'Donnell, A. M., & Dansereau, D. F. (2000). Interactive Effects of Prior Knowledge and Material Format on Cooperative Teaching. *Journal of Experimental Education, 68*(2), 101.
- O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge Maps as Scaffolds for Cognitive Processing. *Educational Psychology Review, 14*(1), 71–86. <https://doi.org/10.1023/A:1013132527007>
- Oldenbürger, H.-A. (1986). *Zur Erhebung und Repräsentation kognitiver Strukturen: methodenheurist. Überlegungen u. Untersuchungen*. Braunschweig: Techn. Univ., Fb 9 - Abt. Sozialarbeitswiss. d. Seminars für Soziologie u. Sozialarbeitswiss.
- Oldenbürger, H.-A. (1992). *Netz-Werk-Zeug 1: Zählwerk für beliebige Variablenkombinationen mit beliebigen Filtern. Version a. z. (Programm und Dokumentation)*. Göttingen: Seminar für Wirtschaftspädagogik.
- Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens. In W. Althof (Hrsg.), *Fehlerwelten: Vom Fehlermachen Und Lernen Aus Fehlern. Beiträge Und Nachträge Zu Einem Interdisziplinären Symposium Aus Anlaß Des 60. Geburtstags Von Fritz Oser* (S. 11–42). Opladen: Vs Verlag für Sozialwissenschaften.
- Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors, 35*(4), 737–743. <https://doi.org/10.1177/001872089303500412>

- Paivio, A. (1986). *Mental representations / a dual coding approach* (Bd. 9). Oxford: Oxford Univ. Press.
- Pankratius, W. J. (1990). Building an organized knowledge base: Concept mapping and achievement in secondary school physics. *Journal of Research in Science Teaching*, 27(4), 315–333. <https://doi.org/10.1002/tea.3660270404>
- Phye, G. D. (1979). The processing of informative feedback about multiple-choice test performance. *Contemporary Educational Psychology*, 4(4), 381–394. [https://doi.org/10.1016/0361-476X\(79\)90057-2](https://doi.org/10.1016/0361-476X(79)90057-2)
- Phye, G. D., & Andre, T. (1989). Delayed retention effect: Attention, perseveration, or both? *Contemporary Educational Psychology*, 14(2), 173–185. [https://doi.org/10.1016/0361-476X\(89\)90035-0](https://doi.org/10.1016/0361-476X(89)90035-0)
- Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology*, 14(2), 97–110. [https://doi.org/10.1016/0361-476X\(89\)90028-3](https://doi.org/10.1016/0361-476X(89)90028-3)
- Popitz, H. (1980). *Die normative Konstruktion von Gesellschaft* (1. Aufl.). Tübingen: Mohr Siebeck.
- Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance* (4 5 6 7 8 9 10). New York : London: The Free Press.
- Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation*. München: Ernst Reinhardt.
- Preiß, P. (1995). Methodenfreiheit oder Handlungsorientierung? Vortrag auf der Bezirksversammlung des VLWN-Bezirksverbandes Göttingen am 9. 5. 1995 in Northeim.
- Raab, A. E., Poost, A., & Eichhorn, S. (2008). *Marketingforschung: ein praxisorientierter Leitfaden*. Stuttgart: W. Kohlhammer Verlag.
- Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63–73. <https://doi.org/10.1016/j.learninstruc.2013.03.002>
- Rakoczy, K., Klieme, E., Bürgermeister, A., & Harks, B. (2008). The Interplay Between Student Evaluation and Instruction: Grading and Feedback in Mathematics Classrooms. *Zeitschrift Für Psychologie / Journal of Psychology*, 216(2), 111–124. <https://doi.org/10.1027/0044-3409.216.2.111>
- Rebich, S., & Gautier, C. (2005). Concept Mapping to Reveal Prior Knowledge and Conceptual Change in a Mock Summit Course on Global Climate Change. *Journal of Geoscience Education*, 53(4), 355–365.
- Rebmann, K. (2001). *Planspiel und Planspieleinsatz / theoretische und empirische Explorationen zu einer konstruktivistischen Planspieldidaktik* (Bd. 4). Hamburg: Kovač.
- Reigeluth, C. M. (Hrsg.). (1983). *Instructional-design theories and models*. Hillsdale, N.J: Lawrence Erlbaum Associates.
- Reinhold, P., Lind, G., & Friege, G. (1999). Wissenszentriertes Problemlösen in Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 5(1), 41–62.

- Renkl, A., Hilbert, T. S., & Schworm, S. (2007). Cognitive skill acquisition from complex examples. A taxonomy of examples and tentative instructional guidelines. In M. Prenzel (Hrsg.), *Studies on the educational quality of schools. The final report on the DFG Priority Programme*. (S. 239–249). Münster: Waxmann.
- Restle, F. (1959). A metric and an ordering on sets. *Psychometrika*, 24(3), 207–220. <https://doi.org/10.1007/BF02289843>
- Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (Bd. 10, S. 178–186). Göttingen: Hogrefe.
- Rheinberg, F., & Fries, S. (2010). Bezugsnormorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 61–68). Weinheim: Beltz.
- Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching*, 35(10), 1103–1127. [https://doi.org/10.1002/\(SICI\)1098-2736\(199812\)35:10<1103::AID-TEA4>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1098-2736(199812)35:10<1103::AID-TEA4>3.0.CO;2-P)
- Roberts, M. A. (1995). Awareness and the efficacy of error correction. In R. W. Schmidt (Hrsg.), *Attention and Awareness in Foreign Language Learning* (S. 163–182). University of Hawai'i, Second Language Teaching & Curriculum Center.
- Röhle, J. (2007). *Wissensentwicklung von Experten und Novizen. Eine Untersuchung zum Thema Unternehmensgründung*. Dresden: TU, Lehrstuhl Wirtschaftspädagogik.
- Roth, W.-M. (1994). Student views of collaborative concept mapping: An emancipatory research project. *Science Education*, 78(1), 1–34. <https://doi.org/10.1002/sce.3730780102>
- Ruiz-Primo, M. A. (2000). On the use of concept maps as an assessment tool in science: What we have learned so far. *REDIE: Revista Electrónica de Investigación Educativa*, 2(1), 3.
- Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, 38(2), 260–278.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569–600. [https://doi.org/10.1002/\(SICI\)1098-2736\(199608\)33:6<569::AID-TEA1>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1098-2736(199608)33:6<569::AID-TEA1>3.0.CO;2-M)
- Rutke, U. (2007). *Schülervorstellungen und wissenschaftliche Vorstellungen zur Entstehung und Entwicklung des menschlichen Lebens* (PhD Thesis). LMU München, Fakultät für Biologie. Abgerufen von https://edoc.ub.uni-muenchen.de/8625/1/Rutke_Ulrike.pdf
- Ryssel, J., & Fürstenau, B. (2011). Unterstützung des Lernens betriebswirtschaftlicher Inhalte durch Concept Maps oder Textzusammenfassungen - eine vergleichende Untersuchung im Rahmen des Planspielunterrichts. In U. A. Faßhauer (Hrsg.), *Lehr-Lernforschung und Professionalisierung. Perspektiven der Berufsbildungsforschung*. (S. 111–121). Opladen; Farmington Hills, Mich.: Budrich. Abgerufen von <http://www.pedocs.de/volltexte/2013/7035>

- Ryssel, J., Fürstenau, B., Förster-Kuschel, J., Thieme, R., & Lützner, S. (2018). Anwendung von Concept Maps als ergänzende Lernstrategie im betriebswirtschaftlichen Planspielunterricht: Trainings- und Instruktionsbeispiele. In B. Fürstenau (Hrsg.), *Concept Mapping - Trainings- und Instruktionsbeispiele; Dresdner Beiträge zur Wirtschaftspädagogik (1/2018)* (S. 55–91). Dresden: TU Dresden, Professur für Wirtschaftspädagogik.
- Ryssel, J., Fürstenau, B., & Thieme, R. (2018). Entwicklung eines Fragebogens zur Evaluation von Concept Mapping durch die Teilnehmenden. In B. Fürstenau (Hrsg.), *Concept Mapping - Trainings- und Instruktionsbeispiele; Dresdner Beiträge zur Wirtschaftspädagogik (1/2018)* (S. 92–104). Dresden: TU Dresden, Professur für Wirtschaftspädagogik.
- Sächsisches Staatsministerium für Kultus. (2009). *Lehrplan Mittelschule. Wirtschaft-Technik-Haushalt-Soziales*. Abgerufen von http://www.schule.sachsen.de/lpdb/web/downloads/lp_ms_wirtschaft_technik_haushalt_soziales_2009.pdf?v2
- Sächsisches Staatsministerium für Kultus (Hrsg.). (2015). *Arbeitsmaterial für die Berufsschule - Kaufmann für Büromanagement, Kauffrau für Büromanagement*. Radebeul: Sächsisches Bildungsinstitut. Abgerufen von https://www.schule.sachsen.de/lpdb/web/downloads/1729_lp_bs_kfm_bueromanagement_gesamt_2015.pdf?v2
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Safayeni, F., Derbentseva, N., & Cañas, A. J. (2005). A theoretical note on concepts and the need for Cyclic Concept Maps. *Journal of Research in Science Teaching*, 42(7), 741–766. <https://doi.org/10.1002/tea.20074>
- Schau, C., Mattern, N., Zeilik, M., Teague, K. W., & Weber, R. J. (2001). Select-and-Fill-in Concept Map Scores as a Measure of Students' Connected Understanding of Science. *Educational and Psychological Measurement*, 61(1), 136–158.
- Schiefele, U. (2009). Motivation. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 151–177). Berlin: Springer. https://doi.org/10.1007/978-3-540-88573-3_7
- Schimmel, B. J. (1983). A Meta-Analysis of Feedback to Learners in Computerized and Programmed Instruction. Gehalten auf der Annual Meeting of the American Educational Research Association, Montreal.
- Schimmel, B. J. (1988). Providing meaningful feedback in courseware. In *Instructional designs for microcomputer courseware*. (S. 183–195). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Schmid, R. F., & Telaro, G. (1990). Concept Mapping as an Instructional Strategy for High School Biology. *The Journal of Educational Research*, 84(2), 78–85. <https://doi.org/10.1080/00220671.1990.10885996>
- Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 49–70). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.005>

- Schnotz, W. (2011). *Pädagogische Psychologie kompakt / mit Online-Materialien* (2., überarb. und erw. Aufl.). Weinheim: Beltz. Abgerufen von <https://katalog-beta.slub-dresden.de/id/0000553727/>
- Schnotz, W., Vosniadou, S., & Carretero, M. (1999). *New perspectives on conceptual change*. Oxford: Pergamon Press.
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2017). Studying and Constructing Concept Maps: a Meta-Analysis. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-017-9403-9>
- Schulz von Thun, F. (2015). *Miteinander reden 1: Störungen und Klärungen: Allgemeine Psychologie der Kommunikation* (52. Aufl.). Reinbek bei Hamburg: Rowohlt Taschenbuch.
- Semmer, N. K., & Jacobshagen, N. (2010). Feedback im Arbeitsleben - eine Selbstwert-Perspektive. *Gruppendynamik und Organisationsberatung*, 41(1), 39–55. <https://doi.org/DOI.10.1007/s11612-010-0104-9>
- Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Hrsg.), *Conversational Interaction in Second Language Acquisition* (S. 301–322). Oxford: Oxford University Press.
- Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32(2), 203–234. <http://dx.doi.org/10.1017/S0272263109990507>
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Sievers, H.-P. (1984). *Lernen - Wissen - Handeln: Untersuchungen zum Problem der didaktischen Sequenzierung. Dargest. am Wirtschaftslehre-Curriculum in d. Sekundarstufe II*. Frankfurt, Main: Fischer.
- Smith, T. A., & Kimball, D. R. (2010). Learning From Feedback: Spacing and the Delay—Retention Effect. *Journal of Experimental Psychology. Learning, Memory & Cognition*, 36(1), 80–95.
- Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments*, 16(2), 183–193.
- Snizek, J. A., May, D. R., & Sawyer, J. E. (1990). Social uncertainty and interdependence: A study of resource allocation decisions in groups. *Organizational Behavior and Human Decision Processes*, 46(2), 155–180. [https://doi.org/10.1016/0749-5978\(90\)90027-7](https://doi.org/10.1016/0749-5978(90)90027-7)
- Snow, R. E. (1989). Aptitude-treatment interaction as a framework of research in individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Hrsg.), *Learning and individual differences / advances in theory and research* (S. 13–59). New York: Freeman.
- Sommer, S., Fürstenau, B., Ryssel, J., & Kunath, J. (2009). Einsatz verschieden stark vorstrukturierter Concept Maps zur Unterstützung des Lernens betriebswirtschaftlicher Inhalte. Eine Untersuchung im Zusammenhang mit dem Planspielunter-

- richt. In D. D. Münk (Hrsg.), *Forschungserträge aus der Berufs- und Wirtschaftspädagogik. Probleme, Perspektiven, Handlungsfelder und Desiderata der beruflichen Bildung in der Bundesrepublik Deutschland, in Europa und im internationalen Raum*. (S. 20–29). Opladen: Budrich.
- Spychiger, M., Kuster, R., & Oser, F. (2006). Dimensionen von Fehlerkultur in der Schule und deren Messung. Der Schülerfragebogen zur Fehlerkultur im Unterricht für Mittel- und Oberstufe. *Schweizerische Zeitschrift für Bildungswissenschaften*, 28(1), 87–110.
- Spychiger, M., Oser, F., Hascher, T., & Mahler, F. (1999). Entwicklung einer Fehlerkultur in der Schule. In W. Althof (Hrsg.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern. Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlaß des 60. Geburtstags von Fritz Oser* (S. 43–70). Opladen: Vs Verlag für Sozialwissenschaften.
- Spychiger, M., Oser, F., Mahler, F., & Hascher, T. (1998). *Fehlerkultur aus der Sicht von Schülerinnen und Schülern. Der Fragebogen S-UFS: Entwicklung und erste Ergebnisse*. Freiburg, Schweiz: Pädag. Inst. der Univ.
- Stajkovic, A. D., & Sommer, S. M. (2000). Self-Efficacy and Causal Attributions: Direct and Reciprocal Links. *Journal of Applied Social Psychology*, 30(4), 707–737.
- Stark, R., Kopp, V., & Fischer, M. R. (2009). Förderung der Diagnosekompetenz bei Studierenden der Medizin durch situiertes, fallbasiertes Lernen mit Lösungsbeispielen: der Einfluss von Fehlern und Feedback. *Psychologie in Erziehung und Unterricht*, 56(2), 137–149.
- Steiner, G. (2006). Wiederholungsstrategien. In H. Mandl & H. F. Friedrich (Hrsg.), *Handbuch Lernstrategien*. (S. 101–116). Göttingen: Hogrefe.
- Stensvold, M. S., & Wilson, J. T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. *Science Education*, 74(4), 473–480. <https://doi.org/10.1002/sce.3730740407>
- Stone, T. H. (1971). Effects of mode of organization and feedback level on creative task groups. *Journal of Applied Psychology*, 55(4), 324–330. <https://doi.org/10.1037/h0031534>
- Stracke, I. (2004). *Einsatz computerbasierter concept maps zur Wissensdiagnose in der Chemie. Empirische Untersuchungen am Beispiel des chemischen Gleichgewichts*. Münster; New York; München; Berlin: Waxmann.
- Straka, G. A., & Macke, G. (1979). *Lehren und Lernen in der Schule: eine Einführung in Lehr-Lern-Theorien*. Stuttgart: Kohlhammer.
- Strijbos, J.-W., & Müller, A. (2014). Personale Faktoren im Feedbackprozess. In *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. (S. 83–134). Münster: Waxmann.
- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291–303. <https://doi.org/10.1016/j.learninstruc.2009.08.008>

- Strijbos, J.-W., Ochoa, T. A., Sluijsmans, D. M., Segers, M. S., & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in (web-based) collaborative learning environments. In C. Mourlas, N. Tsianos, & P. Germanakos (Hrsg.), *Cognitive and emotional processes in web-based education: Integrating human factors and personalization* (S. 375–395). Hershey, PA: IGI Global.
- Sweller, J. (1999). *Instructional Design in Technical Areas*. Camberwell: Australian Council Educational Research.
- Sweller, J., & Chandler, P. (1994). Why Some Material Is Difficult to Learn. *Cognition and Instruction*, 12(3), 185–233.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296.
- Swindell, L. K. (1992). Certitude and the constrained processing of feedback. *Contemporary Educational Psychology*, 17(1), 30–37. [https://doi.org/10.1016/0361-476X\(92\)90043-X](https://doi.org/10.1016/0361-476X(92)90043-X)
- Tenorth, H.-E., & Tippelt, R. (2012). *Beltz Lexikon Pädagogik* (1. Aufl.). Weinheim: Beltz.
- Tergan, S.-O. (2006). Individuelles Wissens- und Informationsmanagement mit Concept Maps beim ressourcenbasierten Lernen. In H. Mandl & H. F. Friedrich (Hrsg.), *Handbuch Lernstrategien* (S. 307–324). Göttingen: Hogrefe.
- Tergan, S.-O., Gräber, W., & Neumann, A. (2006). Mapping and managing knowledge and information in resource-based learning. *Innovations in Education and Teaching International*, 43(4), 327–336.
- Tindale, R. S. (1989). Group vs individual information processing: The effects of outcome feedback on decision making. *Organizational Behavior and Human Decision Processes*, 44(3), 454–473. [https://doi.org/10.1016/0749-5978\(89\)90019-8](https://doi.org/10.1016/0749-5978(89)90019-8)
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208. <http://dx.doi.org/10.1037/h0061626>
- Tramm, T., & Rebmann, K. (1997). Handlungsorientiertes Lernen in und an komplexen, dynamischen Modellen. Die Modellierungsperspektive als notwendige Ergänzung des handlungsorientierten Ansatzes in der Wirtschaftsdidaktik. In G. Lübke & B. Riesebieter (Hrsg.), *Zur Theorie und Praxis des SIMBA-Einsatzes in der kaufmännischen Aus- und Weiterbildung. Werkstattberichte - Unternehmensbeschreibung Designermöbel GmbH*. Markhausen: Lübke.
- Tramm, T., & Rebmann, K. (1999). Veränderungen im Tätigkeitsprofil von Handelslehrern unter dem Signum handlungsorientierter Curricula. In T. Tramm, D. Sembill, & E. G. John (Hrsg.), *Professionalisierung kaufmännischer Berufsbildung / Beiträge zur Öffnung der Wirtschaftspädagogik für die Anforderungen des 21. Jahrhunderts ; Festschrift zum 60. Geburtstag von Frank Achtenhagen* (S. 231–259). Frankfurt am Main: Lang.
- Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Hrsg.), *Conversational Interaction in Second Language Acquisition* (S. 144–171). Oxford: Oxford University Press.

- Trojahner, I., & Fürstenau, B. (2016). Die Kognitive Theorie Multimedialen Lernens. In B. Fürstenau (Hrsg.), *Lehr-Lern-Theorien / Behaviorismus, Kognitivismus, Konstruktivismus: Lernen und Expertise verstehen und fördern* (Bd. 6, S. 61–76). Hohengehren: Schneider.
- Überla, K. (2013). *Faktorenanalyse: Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozial- Wissenschaftler* (2. Aufl.). Berlin: Springer.
- van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education, 58*(1), 263–272.
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning & Instruction, 20*(4), 280–290.
- Vollmeyer, R., & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System. *Zeitschrift für Pädagogische Psychologie, 12*(1), 11–23.
- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction, 15*(6), 589–602. <https://doi.org/10.1016/j.learninstruc.2005.08.001>
- Wadouh, J. (2008). *Vernetzung und kumulatives Lernen im Biologieunterricht der Gymnasialklasse 9* (Dissertation). Universität Duisburg-Essen, Fakultät für Biologie. Abgerufen von http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet/Derivate-20492/Dissertation_komplett_5_bib_end.pdf
- Wager, S. U. (1983). *The Effect of Immediacy and Type of Informative Feedback on Retention in a Computer-Assisted Task*. (Unpublished doctoral dissertation). Florida State University.
- Walberg, H. J. (1986). Synthesis of research on teaching. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (3. Aufl., S. 214–229). New York: Macmillan.
- Weaver, M. R. (2006). Do Students Value Feedback? Student Perceptions of Tutors' Written Responses. *Assessment & Evaluation in Higher Education, 31*(3), 379–394.
- Weber, S. (1994). *Vorwissen in der betriebswirtschaftlichen Ausbildung / eine struktur- und inhaltsanalytische Studie*. Wiesbaden: Dt. Univ.-Verl.
- Weber, S., & Schumann, M. (2000). Das Concept Mapping Software Tool (COMASOTO) zur Diagnose strukturellen Wissens. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 158–179). Göttingen: Hogrefe.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573. <http://dx.doi.org/10.1037/0033-295X.92.4.548>
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Hrsg.), *Handbook of Research on Teaching: A Project of the American Educational Research Association, a Department of the National Education Association* (3. Aufl., S. 315–327). New York: Rand McNally.

- Wender, K. F. (1988). Semantische Netzwerke als Bestandteil gedächtnispsychologischer Theorien. In H. Mandl & H. Spada (Hrsg.), *Wissenspsychologie* (S. 55–73). München: Psychologie Verlags Union.
- Wiegmann, D. A., Dansereau, D. F., McCagg, E. C., Rewey, K. L., & Pitre, U. (1992). Effects of knowledge map characteristics on information processing. *Contemporary Educational Psychology*, *17*(2), 136–155. [https://doi.org/10.1016/0361-476X\(92\)90055-4](https://doi.org/10.1016/0361-476X(92)90055-4)
- Wild, K.-P. (2000). *Lernstrategien im Studium. Strukturen und Bedingungen*. Münster: Waxmann.
- Wollenschläger, M., Möller, J., & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift für Pädagogische Psychologie*. <https://doi.org/10.1024/1010-0652/a000040>
- Wollenschläger, M., Möller, J., & Harms, U. (2012). Ist kompetenzielles Fremdfeedback überlegen, weil es als effektiver wahrgenommen wird? *Unterrichtswissenschaft*, *40*(3), 197–212.
- Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology*, *72*(3), 416–425. <http://dx.doi.org/10.1037/0021-9010.72.3.416>
- Woolfolk, A. (2008). *Pädagogische Psychologie* (10. Aufl.). München: Pearson.
- Wu, P.-H., Hwang, G.-J., Milrad, M., Ke, H.-R., & Huang, Y.-M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, *43*(2), 217–232. <https://doi.org/10.1111/j.1467-8535.2010.01167.x>
- Yang, Y.-F. (2015). Automatic Scaffolding and Measurement of Concept Mapping for EFL Students to Write Summaries. *Educational Technology & Society*, *18*(4), 273–286.
- Yilmaz, Y., & Granena, G. (2016). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*, *19*(01), 147–161. <https://doi.org/10.1017/S136672891400090X>