

Wortspeicher im Sachunterricht

Eine empirische Untersuchung zum Nutzen von sprachlichen Hilfen
im naturwissenschaftlichen Lernen

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Anne Mayele Otte

aus Wipperfürth

Köln

2023

Berichterstatter:innen:
(Gutachter:innen)

Prof.' Dr.' Daniela Schmeinck

Prof. Dr. Jörg Großschedl

Tag der mündlichen Prüfung:

28.09.2023

DANKSAGUNG

Mein besonderer Dank gilt zunächst meiner Betreuerin Prof.' Dr.' Daniela Schmeinck für ihre Beratung und die Unterstützung meines Forschungsvorhabens, sowie meinem Zweitbetreuer Prof. Dr. Jörg Großschedl für seine fortwährenden Orientierungshilfen und Ratschläge. Des Weiteren gilt mein Dank der Graduiertenschule für die zahlreichen Fortbildungs- und Unterstützungsangebote und insbesondere Dr. Matthias Krepf für seine umfassende methodische Beratung. Die Unterstützung und der kollegiale Austausch mit meinen Kolleg:innen im Institut für Didaktik des Sachunterrichts trug neben der angenehmen Arbeitsatmosphäre ebenfalls maßgeblich zum Gelingen der Forschungsarbeit bei.

Ganz besonders möchte ich mich an dieser Stelle auch für das mir entgegengebrachte Vertrauen und das große Engagement der Schulleitungen, Lehrkräfte und natürlich der Schüler:innen der beteiligten Grundschulen bedanken, ohne die diese Arbeit nicht möglich gewesen wäre. Darüber hinaus möchte ich mich bei allen bedanken, die mir Bildmaterial für die Unterrichtssequenz zur Verfügung gestellt haben.

Ein tief empfundenes Dankeschön möchte ich schließlich meiner Familie und meinen Freund:innen und Kolleg:innen aussprechen, vor allem Christale, Nici, Saskia, Vinzent, Vivi, Ulla & Berndt und Michel & Sami, deren Rückhalt und Unterstützung unverzichtbar waren.

ZUSAMMENFASSUNG

Im Zentrum des Interesses dieses Forschungsprojekts steht der Nutzen von Wortspeichern für das Verfassen von naturwissenschaftlichen Versuchsprotokollen im Sachunterricht. Wortspeicher sind Wort- und Formulierungshilfen, die Schüler:innen im Unterricht auf verschiedene Weise zur Verfügung gestellt werden. Es wird untersucht, wie häufig Kinder die Sprachmittel aus visuell oder auditiv dargebotenen Wortspeichern zu einem späteren Zeitpunkt beim Schreiben eines Versuchsprotokolls nutzen. Ebenso von Interesse ist die Frage, ob Kinder nach der visuellen oder auditiven Vermittlung von Wortspeichern mehr bildungssprachliche Wörter und Formulierungen in ihren Versuchsprotokollen integrieren. Eine dritte Teiluntersuchung prüft, ob sich der Inhalt durch die Vermittlung visueller oder auditiver Wortspeicher verbessern lässt. In einer vierten Teiluntersuchung wird der Nutzen von Wortspeichern auf Basis der ersten drei Teiluntersuchungen übergreifend eruiert. Zur Untersuchung aller Forschungsfragen wurden Versuchsprotokolle von Kindern aus drei Klassen des dritten Jahrgangs – eine Kontrollgruppe, eine Gruppe mit auditiv und eine Gruppe mit visuell dargebotenen Wortspeichern – zu drei verschiedenen Messzeitpunkten analysiert. Das Treatment bestand aus einer achtstündigen Unterrichtssequenz. Die Auswertung erfolgte im Mixed-Methods-Design maßgeblich mit qualitativer Erhebung und quantitativer Testung von Häufigkeitsunterschieden. Es wurden 138 Versuchsprotokolle von insgesamt 46 Kindern ausgewertet. Vor allem visuelle Wortspeicher, d.h. Wortspeicher in Plakatform, weisen in dieser Studie einen Nutzen auf. Die Befunde der Studie deuten darauf hin, dass das visuelle Wortspeicher-Treatment kurzfristig mit einer höheren Verwendung der Wortspeicher und einer höheren Verwendung von Bildungssprache einhergeht. Weiterhin weisen die Versuchsprotokolle der Kinder mit visuellem Wortspeicher-Treatment ein erhöhtes Vorkommen erklärender Sprachhandlungen im Zusammenhang mit korrekt beurteilten Erklärungen auf. Ein Erklärungsansatz hierfür ist möglicherweise eine gesteigerte Sprachbewusstheit der Kinder. Durch das Bewusstsein, dass eine angemessene Sprache eine wichtige Rolle in Versuchsprotokollen spielt und durch die fehlende Erinnerung an die exakten Formulierungen aus dem Wortspeicher, verwenden die Kinder in kreativer Weise ähnliche Formulierungen. Diese wiederum unterstützen ihre Erklärungsversuche. Eine Verbesserung oder eine Verschlechterung in den inhaltlichen Beurteilungen der Versuchsprotokolle ist in dieser Studie zu keinem Zeitpunkt festzustellen. Als Schlussfolgerung wird für den vermehrten und langfristigen Einsatz von Wortspeichern in Plakatform im unterrichtlichen Kontext plädiert.

INHALTSVERZEICHNIS

ABBILDUNGS-, TABELLEN- UND ABKÜRZUNGSVERZEICHNIS	VIII
1 EINLEITUNG.....	1
2 THEORETISCHER RAHMEN.....	5
2.1 Bildungssprache	5
2.1.1 Kennzeichen von Sprache.....	5
2.1.2 Sprachregister - eine begriffliche Klärung	10
2.1.3 Bedeutung von Bildungssprache im schulischen Kontext	16
2.1.4 Bildungssprache in der Grundschule	17
2.2 Vermittlung der Naturwissenschaften in der Grundschule.....	21
2.2.1 Inhalte einer naturwissenschaftlichen Grundbildung	21
2.2.2 Der Forschungskreislauf.....	23
2.2.3 Sprache im naturwissenschaftlichen Sachunterricht	25
2.3 Wortspeicher	29
2.3.1 Wortspeicher als Methode der Sprachbildung.....	29
2.3.2 Psycholinguistische Erkenntnisse	34
2.3.3 Grundzüge der Wortschatzarbeit	38
2.3.4 <i>Scaffolding</i> - Lernen mit Gerüst	42
2.3.5 Forschungsstand zur Wirksamkeit von <i>Scaffolding</i> und Wortspeichern im naturwissenschaftlichen Lernen	47
2.4 Zusammenfassung.....	51
3 FORSCHUNGSFRAGEN	52
4 METHODE	56
4.1 Stichprobe	56
4.2 Rahmenbedingungen und Verlauf	58
4.2.1 Vorstudie.....	60
4.2.2 Pilotstudien	61
4.2.3 Verlauf der Hauptstudie.....	64
4.2.4 Thematische und methodische Begründungen zur Unterrichtssequenz ...	65
4.3 Datenerhebung	70

4.3.1	Sprachstandserhebung (C-Test).....	70
4.3.2	Versuchsprotokolle als Erhebungsinstrumente	72
4.4	Datenauswertung.....	73
4.4.1	C-Test.....	73
4.4.2	Mixed-Methods als übergreifendes Forschungsdesign	74
4.4.3	Erste Teilstudie (F1)	76
4.4.4	Zweite Teilstudie (F2)	77
4.4.5	Dritte Teilstudie (F3)	83
4.4.6	Vierte Teilstudie (F4)	87
5	ERGEBNISSE.....	88
5.1	Gütekriterien	88
5.1.1	Sprachstandsvergleich mithilfe des C-Tests (Validität)	88
5.1.2	Intercoder-Übereinstimmung (Objektivität und Reliabilität).....	89
5.2	F 1 – Nutzung der Wortspeicher	93
5.2.1	Deskriptive Ergebnisse	93
5.2.2	Voraussetzungen für Inferenzstatistik	94
5.2.3	Gruppenunterschiede – Wortspeicher	97
5.2.4	Zeitliche Unterschiede – Wortspeicher	97
5.2.5	Zusammenfassung	99
5.3	F 2 – Sprachregister	100
5.3.1	Kodierungen der Bildungssprache (Sprachhandlungen)	100
5.3.2	Kodierungen der Bildungssprache (Lexikon).....	103
5.3.3	Kodierungen der Bildungssprache (Morphosyntax).....	105
5.3.4	Kodierungen der Bildungssprache insgesamt.....	107
5.3.5	Gruppenunterschiede – Bildungssprache	110
5.3.6	Zeitliche Unterschiede – Bildungssprache	111
5.3.7	Zusammenfassung	113
5.4	F 3 – Inhalt.....	114
5.4.1	Kodierungen des Inhalts	114
5.4.2	Inhaltliche Punktzahl insgesamt	119
5.4.3	Gruppenunterschiede – Inhalt.....	120
5.4.4	Zeitliche Unterschiede – Inhalt.....	120
5.4.5	Zusammenfassung	122
5.5	F 4 – Datenintegration	123
5.5.1	Ergebnisse insgesamt.....	123
5.5.2	Vergleich von Vpn mit/ ohne Nutzung der Wortspeicher.....	139

6 DISKUSSION.....	144
6.1 Interpretation der Ergebnisse	144
6.1.1 Interpretation der Ergebnisse zu F1 – Nutzung der Wortspeicher	144
6.1.2 Interpretation der Ergebnisse zu F2 – Sprachregister.....	147
6.1.3 Interpretation der Ergebnisse zu F3 – Inhalt	148
6.1.4 Interpretation der Ergebnisse zu F4 – Datenintegration	150
6.2 Zusammenfassende Schlussfolgerung	155
6.3 Methodische Kritik	156
6.4 Ausblick.....	159
6.4.1 Wissenschaftlicher Ausblick	159
6.4.2 Schulpraktische Implikationen	163
7 LITERATURVERZEICHNIS	167
8 ANHANG	178
8.1 Anhang A: Test-Stunden	179
8.1.1 Anhang A.1: C-Test.....	179
8.1.2 Anhang A.2: Versuchsanweisung im Prä-Test.....	182
8.1.3 Anhang A.3: Versuchsanweisung im Post-Test	183
8.1.4 Anhang A.4: Versuchsanweisung im Follow-Up-Test.....	184
8.1.5 Anhang A.5: Vorlage für ein Versuchsprotokoll im Prä-Test.....	185
8.1.6 Anhang A.6: Impressionen aus der Unterrichtssequenz.....	187
8.2 Anhang B: Kategoriensysteme.....	189
8.2.1 Anhang B.1: Kategoriensystem – Bildungssprache	189
8.2.2 Anhang B.2: Kategoriensystem – Inhalt.....	195
8.3 Anhang C: Ergebnisse der statistischen Berechnungen	202
8.3.1 Anhang C.1: Statistische Ergebnisse zum C-Test	202
8.3.2 Anhang C.2: Statistische Ergebnisse zur ersten Teilstudie (F1)	204
8.3.3 Anhang C.3: Statistische Ergebnisse zur zweiten Teilstudie (F2)	211
8.3.4 Anhang C.4: Statistische Ergebnisse zur dritten Teilstudie (F3)	222
8.4 Anhang D: Weitere Ergebnisse.....	228
8.4.1 Anhang D.1: Häufigkeitsanalyse einfachkodierter Sprachhandlungen..	228
8.4.2 Anhang D.2: Häufigkeitsanalyse einfachkodierte Lexikon	229
8.4.3 Anhang D.3: Häufigkeitsanalyse einfachkodierte Morphosyntax.....	230

ABBILDUNGS-, TABELLEN- UND ABKÜRZUNGSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

Abb. 1: Physikalische Kommunikationskompetenz	6
Abb. 2: Kennzeichen von Sprache	8
Abb. 3: Modell zur Bildungssprache in der Grundschule.....	18
Abb. 4: Auszug aus der Ratingskala (RaBi) des Easi-Science-L-Projektes	19
Abb. 5: Der Forschungskreislauf	24
Abb. 6: Ausgewählte Methoden der Sprachbildung auf Wortebene	30
Abb. 7: Ausgewählte Methoden der Sprachbildung auf Satzebene.....	31
Abb. 8: Ausgewählte Methoden der Sprachbildung auf Textebene	32
Abb. 9: Bereiche des mentalen Lexikons.....	35
Abb. 10: Modell inhaltlicher Wortnetze	35
Abb. 11: Forschungsdesign	58
Abb. 12: Verlauf des Forschungsvorhabens	59
Abb. 13: Gedächtnismodell zum Lauten Denken	60
Abb. 14: Verlauf der ersten Pilotstudie.....	62
Abb. 15: Verlauf der Hauptstudie	64
Abb. 16: Verwendeter Wortspeicher in der Hauptstudie.....	68
Abb. 17: Thematische Unterrichtssequenz der Hauptstudie	69
Abb. 18: Auszug aus dem C-Test	71
Abb. 19: Paralleles Design.....	75
Abb. 20: Ober- und Unterkategorien zur Kodierung von Bildungssprache	80
Abb. 21: Auszug aus dem Kategoriensystem zur Kodierung von Bildungssprache ...	81
Abb. 22: Ober- und Unterkategorien zur Kodierung des Inhalts.....	85
Abb. 23: Gruppenunterschiede bezogen auf die Verwendung der Wortspeicher (AV)	93
Abb. 24: Boxplot zur Analyse von Ausreißern bezogen auf die Verwendung der Wortspeicher (AV).....	96
Abb. 25: Zeitlicher Verlauf bezogen auf die Verwendung der Wortspeicher (AV)....	98

Abb. 26: Gruppenunterschiede bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV).....	109
Abb. 27: Gruppenunterschiede bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV).....	109
Abb. 28: Zeitlicher Verlauf bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV).....	111
Abb. 29: Zeitlicher Verlauf bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV).....	112
Abb. 30: Gruppenunterschiede bezogen auf die Nennung der Materialien (AV)	116
Abb. 31: Gruppenunterschiede bezogen auf die inhaltliche Punktzahl (AV).....	119
Abb. 32: Zeitlicher Verlauf bezogen auf die inhaltliche Punktzahl (AV)	121
Abb. 33: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der KG	131
Abb. 34: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der TG1	132
Abb. 35: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der TG2.....	133
Abb. 36: Prozentualer Anteil aller Kodierungen der Bildungssprache bezogen auf Sprachhandlungen, Lexikon, Morphosyntax und Wortspeicher	137
Abb. 37: Prozentualer Anteil aller inhaltlichen Kodierungen bezogen auf <i>korrekte, unzureichende</i> und <i>fehlerhafte</i> Abschnitte.....	138
Abb. 38: Inhaltliche Beurteilung (AV) in Abhängigkeit der tatsächlichen Wortspeicher-Nutzung (Mehrfachkodierungen).....	142
Abb. 39: Boxplot bezogen auf die Punktzahl im C-Test (AV).....	203
Abb. 40: Boxplot zur Analyse von extremen Ausreißern in SPSS bezogen auf die Verwendung der Wortspeicher (AV).....	207
Abb. 41: Boxplot bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV).....	214
Abb. 42: Boxplot bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV).....	215
Abb. 43: Boxplot bezogen auf die inhaltliche Punktzahl (AV)	224

TABELLENVERZEICHNIS

Tab. 1: Konzeptionelle Mündlich- und Schriftlichkeit	10
Tab. 2: Beschreibende Merkmale von Alltags-, Schul- und Fachsprache.....	13
Tab. 3: Zusammensetzung der Stichprobe	57
Tab. 4: Überblick über die Datenerhebung.....	72
Tab. 5: Inhaltsanalyse zu F2	78
Tab. 6: Inhaltsanalyse zu F3	84
Tab. 7: Intercoder-Übereinstimmung in 14 Versuchsprotokollen	90
Tab. 8: Häufigkeitsanalyse bildungssprachlicher Sprachhandlungen (Mehrfachkodierungen)	102
Tab. 9: Häufigkeitsanalyse bildungssprachlicher Lexika (Mehrfachkodierungen)...	104
Tab. 10: Häufigkeitsanalyse bildungssprachlicher Morphosyntax (Mehrfachkodierungen)	106
Tab. 11: Häufigkeitsanalyse bildungssprachlicher Kodierungen insgesamt (Mehrfachkodierungen)	108
Tab. 12: Häufigkeitsanalyse bildungssprachlicher Kodierungen insgesamt (Einfachkodierungen)	108
Tab. 13: Häufigkeitsanalyse der genannten Materialienanzahl	115
Tab. 14: Häufigkeitsanalyse der inhaltlichen Kodierungen.....	118
Tab. 15: Wortspeicher, Bildungssprache und Inhalt im Post-Test	123
Tab. 16: Wortspeicher, Bildungssprache und Inhalt im Follow-Up-Test.....	124
Tab. 17: Verwendung des Wortspeichers pro Abschnitt	125
Tab. 18: Wortspeicher (qualitativ) pro Abschnitt des Versuchsprotokolls	127
Tab. 19: Verwendung von Bildungssprache pro Abschnitt.....	128
Tab. 20: Bildungssprache (qualitativ) in den Erklärungen der Post-Tests	129
Tab. 21: Überschneidungen von Bildungssprache und korrekter Erklärung (qualitativ)	135
Tab. 22: Anzahl der Kodierungen von Wortspeicher, Bildungssprache und Inhalt (AV) in Abhängigkeit der tatsächlichen Wortspeicher-Nutzung	139
Tab. 23: Gemeinsames Vorkommen von Wortspeichern und bildungssprachlichen Kategorien in Versuchsprotokollen mit & ohne Wortspeicher-Nutzung	140

Tab. 24: Fehlerhafte Vermutungen in den Dokumenten ohne Wortspeicher	141
Tab. 25: Deskriptive Statistik bezogen auf die Punktzahl im C-Test (AV)	202
Tab. 26: Test auf Normalverteilung bezogen auf die Punktzahl im C-Test (AV).....	202
Tab. 27: Test auf Varianzhomogenität bezogen auf die Punktzahl im C-Test (AV)	202
Tab. 28: 3x1 ANOVA bezogen auf die Punktzahl im C-Test (AV).....	203
Tab. 29: Deskriptive Statistik bezogen auf die Verwendung der Wortspeicher (AV)	204
Tab. 30: Test auf Normalverteilung bezogen auf die Verwendung der Wortspeicher (AV)	205
Tab. 31: Histogramme bezogen auf die Verwendung der Wortspeicher (AV)	206
Tab. 32: Test auf Varianzhomogenität bezogen auf die Verwendung der Wortspeicher (AV)	208
Tab. 33: Test auf Sphärizität bezogen auf die Verwendung der Wortspeicher (AV)	208
Tab. 34: Ergebnis der 3x1 ANOVA bezogen auf die Verwendung der Wortspeicher (AV)	209
Tab. 35: 3x1 ANOVA-Effektgröße ^{a,b} bezogen auf die Verwendung der Wortspeicher (AV)	209
Tab. 36: Ergebnis der 1x3 ANOVA bezogen auf die Verwendung der Wortspeicher (AV)	210
Tab. 37: 1x3 ANOVA-Effektgröße bezogen auf die Verwendung der Wortspeicher (AV)	210
Tab. 38: Deskriptive Statistik bezogen auf die Verwendung von Bildungssprache (AV)	211
Tab. 39: Test auf Normalverteilung bezogen auf die Verwendung von Bildungssprache (AV)	213
Tab. 40: Test auf Varianzhomogenität bezogen auf die Verwendung von Bildungssprache (AV)	216
Tab. 41: Ergebnis der 3x1 ANOVA bezogen auf die Verwendung von Bildungssprache (AV)	217
Tab. 42: 3x1 ANOVA-Effektgröße ^{a,b} bezogen auf die Verwendung von Bildungssprache (AV)	218

Tab. 43: Test auf Sphärizität bezogen auf die Verwendung von Bildungssprache (AV)	219
Tab. 44: Ergebnis der 1x3 ANOVA bezogen auf die Verwendung von Bildungssprache (AV)	220
Tab. 45: 1x3 ANOVA-Effektgröße bezogen auf die Verwendung von Bildungssprache (AV)	221
Tab. 46: Deskriptive Statistik bezogen auf die inhaltliche Punktzahl (AV).....	222
Tab. 47: Test auf Normalverteilung bezogen auf die inhaltliche Punktzahl (AV)....	223
Tab. 48: Test auf Varianzhomogenität bezogen auf die inhaltliche Punktzahl (AV)	225
Tab. 49: Ergebnis der 3x1 ANOVA bezogen auf die inhaltliche Punktzahl (AV)....	226
Tab. 50: Test auf Sphärizität bezogen auf die inhaltliche Punktzahl (AV)	227
Tab. 51: Ergebnis der 1x3 ANOVA bezogen auf die inhaltliche Punktzahl (AV)....	227
Tab. 52: 1x3 ANOVA-Effektgröße bezogen auf die inhaltliche Punktzahl (AV)	227

ABKÜRZUNGSVERZEICHNIS

KG	Kontrollgruppe
LRS	Lese-Rechtschreibschwäche
MZP	Messzeitpunkt
ndH	nicht-deutsche Herkunftssprache
TG1	Treatmentgruppe 1 (auditive Wortspeicher)
TG2	Treatmentgruppe 2 (Wortspeicher in Plakatform)
Vpn	Versuchspersonen

1 EINLEITUNG

Sprache ist Macht!

Sprachliche Unterschiede in der Gesellschaft wurden bereits in den 1960er Jahren von dem englischen Bildungssoziologen Bernstein (1964) analysiert. In der damals stark arbeitsteiligen Gesellschaft unterschied Bernstein die sogenannte *working class* und *middle class*, die im Folgenden mit *Arbeiterschicht* und *Mittelschicht* übersetzt werden (Leufer, 2016, S. 77).¹ Er ging der Frage nach, warum die Kinder aus der *Arbeiterschicht* in der Regel weniger Erfolg im Bildungssystem hatten (ebd., S. 82). In seiner soziolinguistischen These geht er davon aus, dass dies auf unterschiedliche Sprechweisen zurückzuführen ist. Er unterscheidet zwischen einem situationsabhängigen **restringierten Code** und einem situationsunabhängigen **elaborierten Code** (Bernstein, 1964, S. 57). Bei einer Szenenbeschreibung einer Verkaufssituation könnte eine Aussage im elaborierten Code *Die Käuferin gibt dem Verkäufer das Geld und geht aus dem Supermarkt* sein. Im restringierten Code bezieht sich die sprechende Person auf die Situation und setzt die Szene als gemeinsames Vorwissen mit dem Hörer voraus: *Sie gibt ihm das Geld und geht heraus* ist somit das Beispiel für einen restringierten Code, bezogen auf die gleiche Szenenbeschreibung. Bernstein (1964, S. 66-67) führt auf, dass Kinder aus manchen Bereichen der *Arbeiterschicht* eher sprachökonomisch und situationsgebunden sprechen und somit ausschließlich den restringierten Code verwenden. Demgegenüber ist für Kinder, die innerhalb der *Mittelschicht* sozialisiert sind, die Verwendung beider Sprachsysteme – des restringierten als auch elaborierten Codes – typisch (ebd., S. 66-67). Bernstein verdeutlicht, dass beide Codes gleichwertig und auf ihre Art und Weise wichtig für die Lebensführung sind (ebd., S. 66). Erst die Gesellschaft führt zu einer ungleichen Betrachtung der Sprechweisen, indem auf institutioneller Ebene der elaborierte Code erwartet und positiv bewertet wird (ebd., S. 66-67). Kinder, die auf den restringierten Code limitiert sind, sind dadurch benachteiligt (ebd., S. 67). Bernstein stellt die positive Bewertung des elaborierten Codes als makrosoziologisches Problem dar (Leufer, 2016, S. 92). Da die Arbeitsteilung in der damaligen Gesellschaft mit

¹ Die Begriffe werden kursiv gesetzt, da sie sich nicht auf das heutige Gesellschaftssystem übertragen lassen (vgl. Leufer, 2016, S. 77).

unterschiedlichen Machtverhältnissen zusammenhing, betrachtet er die Institution Schule als eine Institution der *Mittelschicht*, die Kinder mit restringiertem Sprechcode systematisch benachteiligt und somit die bestehenden ungleichen Machtverhältnisse manifestiert (ebd., S. 76, 88). Letzten Endes wird hiermit der Ausdruck *Sprache ist Macht* deutlich.

Der Zusammenhang von Sprache und Macht lässt sich auf das heutige deutsche Schulsystem übertragen. Anstelle einer dualen Unterscheidung zwischen restringiertem und elaboriertem Code werden im Deutschen verschiedene Sprachregister klassifiziert. Die Verwendung eines passenden Sprachregisters in einer bestimmten Situation ist für eine erfolgreiche Kommunikation zwingend notwendig und somit auch im Unterricht relevant. Sehr deutlich wird dies im **naturwissenschaftlichen Bereich**. Oftmals zeichnen sich naturwissenschaftliche Inhalte durch Abstraktion aus (Schroeter-Brauss, Wecker, & Henrici, 2018, S. 12). Abstrakte Inhalte bedürfen einer präzisen Sprache. Aus Sicht von Schüler:innen ist die Kenntnis und Beherrschung einer präzisen Sprache sowohl essenziell, um abstrakte naturwissenschaftliche Texte und Äußerungen zu verstehen (Rezeption) als auch sich selbst über naturwissenschaftliche Themen verständlich zu äußern (Produktion). Hinzu kommt eine wechselseitige Bedingung von Sprache und Denken, sodass naturwissenschaftliche Inhalte erst durch gewisse sprachliche Kenntnisse verstanden werden (ebd., S. 12). Kempert, Schalk und Saalbach (2018) konkludieren unter Einbezug empirischer Studien: „Im mathematischen oder naturwissenschaftlichen Unterricht kann beispielsweise die Verwendung komprimierter, dekontextualisierter Sprache als eine bestimmte fachspezifische Sprachform gesehen werden, die gleichzeitig Ausdruck und Voraussetzung für komplexes fachliches Lernen ist“ (S. 8).

Die Bedeutung von Sprache im naturwissenschaftlichen Unterricht wird vor allem in der Leistungsbeurteilung deutlich. Eine empirische Studie zur Beurteilung schriftlicher Texte von Schüler:innen einer 7. Klasse liefert hierzu relevante Erkenntnisse (Tajmel, 2010, S. 174-175; 2017b, S. 257-259). In zwei ausgewählten schriftlichen Texten geht es um Begründungen, ob und warum ein Baumstamm bzw. eine Metallplatte schwimmen oder sinken (Tajmel, 2017b, S. 253). Während fünf Physiker:innen die Texte unter Berücksichtigung des schulischen Lernstandes als physikalisch korrekt beurteilen, vergeben 73 naturwissenschaftlich ausgebildete Lehrkräfte jedwedem Spektrum zwischen 0 und 5 Punkten (ebd., S. 255-257). Neben inhaltlichen Begründungen der Leistungsbeurteilung fallen u.a. folgende Phrasen auf: Das Kind kann „[...] die Erklärung nicht in richtige Worte fassen“ (ebd., S. 258), es fällt ihm/ihr schwer „[...] zu argumentieren und sich zu artikulieren“ (ebd., S. 258) oder es werden „[...] Probleme der Artikulationsfähigkeit unterstellt [da die Texte sehr kurz sind]“ (ebd., S. 259). Durch die implizite oder explizite Beurteilung sprachlicher Aspekte erhalten Kinder mit guten sprachlichen Kenntnissen und Fähigkeiten in den Naturwissenschaften einen Selektionsvorteil. Sprache in den Naturwissenschaften ist damit von gesellschaftlicher Relevanz.

Ein **Problem** liegt darin begründet, dass die benötigte Sprache im naturwissenschaftlichen Unterricht nicht immer vermittelt wird. Die Diskrepanz zwischen fehlender Vermittlung und der Beurteilung sprachlicher Aspekte spiegelt sich in dem Begriff des **geheimen Lehrplans** (*hidden curriculum*) wider (Christie, 1985, S. 38). Das heißt: Einerseits erwarten Lehrkräfte von ihren Schüler:innen, dass sie sich angemessen ausdrücken können. Andererseits waren die sprachlichen Fähigkeiten bis vor einiger Zeit oftmals nicht im Lehrplan verankert, sodass den Schüler:innen der Zugang zur Sprache verwehrt blieb. Schleppegrell (2004, S. 41) verdeutlicht, dass aufgrund dieser Diskrepanz eine außerschulisch bedingte sprachliche Heterogenität verstärkt wird. Zur Entschärfung des Problems ist die explizite Vermittlung von Sprache im naturwissenschaftlichen Unterricht notwendig. Die Bedeutung der gezielten Vermittlung von Sprache hat auch die Kultusministerkonferenz (KMK) erkannt, sodass sprachliche Bildung mittlerweile in den Bildungsstandards und in den Lehrplänen der Länder ausgewiesen ist (Kultusministerkonferenz, 2019, S. 4). **Wie** Sprache im naturwissenschaftlichen Unterricht realitätsnah vermittelt werden kann, und welchen jeweiligen Nutzen dies hat, ist Gegenstand des vorliegenden empirischen Forschungsvorhabens. Realitätsnah bedeutet in diesem Fall, dass der naturwissenschaftliche Unterricht weiterhin im Vordergrund steht und der sprachliche Fokus einen geringen Anteil der Unterrichtszeit einnimmt. Um das institutionelle Problem möglichst frühzeitig anzugehen, bezieht sich die Untersuchung auf den naturwissenschaftlichen Sachunterricht in einer dritten Jahrgangsstufe.

Für das geschilderte Forschungsvorhaben ist zunächst ein differenzierterer Einblick in das Thema Sprache erforderlich. Im **ersten Theorie-Kapitel** (Kapitel 2.1) geht es um einen Überblick der verschiedenen Sprachregister. Es wird herausgestellt, aus welchen Gründen die Bildungssprache im Fokus der Untersuchung steht und mit welchen Kriterien die Bildungssprache von Kindern im Grundschulalter annähernd erfasst werden kann. Die Zusammenstellung der Kriterien ist wesentlich für eine sprachliche Auswertung schriftlicher Produkte von Kindern im Grundschulalter. Da der naturwissenschaftliche Sachunterricht im Fokus des vorliegenden Forschungsvorhabens steht, bietet das **zweite Theorie-Kapitel** (2.2) hierzu einen inhaltlichen Einblick. Die Auseinandersetzung mit dem naturwissenschaftlichen Sachunterricht verdeutlicht, dass sich Sprache und Inhalt gegenseitig bedingen. Bestimmte Wörter und Formulierungen erleichtern beispielsweise das Verfassen eines naturwissenschaftlichen Versuchsprotokolls. Die Vermittlung von Wörtern und Formulierungen kann in Form von Wortspeichern erfolgen, die sehr unterschiedliche Ausprägungen aufweisen. Im **dritten Theorie-Kapitel** (2.3) werden daher verschiedene Gestaltungsoptionen für Wortspeicher dargelegt. Weitere didaktische Anmerkungen und Konzepte zur Wortschatzarbeit dienen der didaktischen Fundierung zur Verknüpfung von Sprache und Inhalt. Auf Grundlage der vorgestellten Theorie wird ersichtlich, dass sich verschiedene Ausprägungen von Wortspeichern zur Sprachbildung in den Naturwissenschaften eignen. Für ein detaillierteres

Verständnis zum Nutzen von Wortspeichern im naturwissenschaftlichen Lernen werden vier konkrete Forschungsfragen formuliert (**Kapitel 3**). Darauf basierend werden Forschungsdesign (**Kapitel 4**) und Ergebnisse (**Kapitel 5**) der empirischen Untersuchung vorgestellt. Das letzte Kapitel bietet einen Einblick in mögliche Schlussfolgerungen und Implikationen für Wissenschaft und Schule (**Kapitel 6**).

Die empirische Untersuchung zur realitätsnahen Vermittlung von Sprache im naturwissenschaftlichen Sachunterricht ist ein wichtiges Forschungsdesiderat. Durch die Untersuchung zweier verschiedener Ausprägungen der Sprachvermittlung – in Form von auditiven und visuellen Wortspeichern – werden Lösungsmöglichkeiten zur anfänglich beschriebenen Problematik praxisnah geprüft.

2 THEORETISCHER RAHMEN

Die vorgestellte Theorie stellt die Grundlage für eine Untersuchung zum Nutzen von Wortspeichern² im naturwissenschaftlichen Sachunterricht dar. Neben den theoretischen Betrachtungen von Sprache (Kapitel 2.1), der Vermittlung der Naturwissenschaften im Sachunterricht (Kapitel 2.2) und didaktischer Möglichkeiten der Spracharbeit (Kapitel 2.3) werden in den einzelnen Kapiteln empirische Befunde vorgestellt, die die theoretischen Argumentationen komplementieren. Für ein Verständnis von Sprache im unterrichtlichen Kontext bietet sich zunächst die Auseinandersetzung mit dem Begriff *Bildungssprache* an.

2.1 Bildungssprache

Im zweiten Kapitel *Sprachregister – eine begriffliche Klärung* wird eine Definition von Bildungssprache vorgestellt. Im Kapitel *Bedeutung von Bildungssprache im schulischen Kontext* geht es um grundlegende Funktionen, die der Bildungssprache zugeschrieben werden. Da Bildungssprache ein umfassendes Konstrukt darstellt, wird dieses Konstrukt im dritten Kapitel *Bildungssprache in der Grundschule* weiter eingegrenzt. Als Basis für eine Definition von Bildungssprache ist ein allgemeines Verständnis von Kommunikation und Sprache erforderlich. Dieses allgemeine Verständnis liefert vorab das erste Kapitel, in dem es um *Kennzeichen von Sprache* geht.

2.1.1 Kennzeichen von Sprache

Sprache ist ein wichtiges **Werkzeug der Kommunikation** (Smirnova & Mortelmans, 2010, S. 51). Welche Komponenten bei der Kommunikation von Sprecher:innen und Empfänger:innen eine Rolle spielen, wird in verschiedenen Kommunikationsmodellen deutlich. Im Folgenden werden wichtige Aspekte der Kommunikation anhand des

² Für eine Definition von *Wortspeichern* s. S. 29

Kommunikationsmodells von Kulgemeyer (2010a, S. 9) bzw., daran anknüpfend, von Kulgemeyer und Schecker (2013, S. 227) vorgestellt. Da es sich bei diesem Modell – dargestellt in Abb. 1 – um ein Kommunikationsmodell aus der Physikdidaktik handelt und sich die Autoren vor allem mit dem Erklären physikalischer Phänomene auseinandersetzen (ebd., S. 226), ist das Kommunikationsmodell nicht 1:1 auf andere Kommunikationssituationen übertragbar. Dennoch eignet sich das Modell für einen Einblick in die Kommunikation zwischen Kommunikatoren (z.B. ein/e Schüler:in) und Adressaten (z.B. eine fiktive fachfremde Person) und den diesbezüglich notwendigen Kommunikationskompetenzen in einem naturwissenschaftlichen Fach (ebd., S. 226).

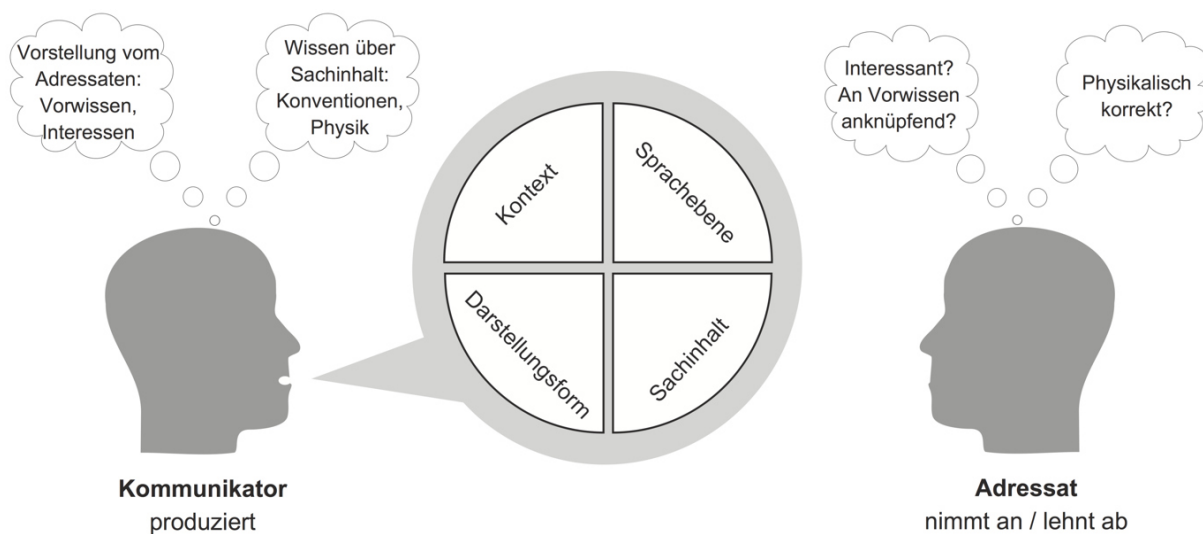


Abb. 1: Physikalische Kommunikationskompetenz (Kulgemeyer, 2010a, S. 9; Kulgemeyer & Schecker, 2013, S. 227)

In ihrem ursprünglichen Modell unterscheiden die Autoren zwischen zwei Dimensionen: Perspektive und Aspekte (ebd., S. 229). Zur Perspektive gehört die Fähigkeit, einen physikalischen Sachverhalt **adressatengerecht und sachgerecht** zu formulieren – siehe Gedankenblasen des Kommunikators in Abb. 1 (ebd., S. 227). Nur wenn ein Kommunikator auf die Bedürfnisse, das Interesse und das Vorwissen des Adressaten eingeht (adressatengemäße Perspektive), nimmt der Adressat die Kommunikation an (ebd., S. 227). Gleichzeitig muss der Sachverhalt korrekt und gemäß physikalischer Darstellungskonventionen dargestellt werden (sachgerechte Perspektive) (ebd., S. 227). Sowohl die Adressatengemäßheit als auch die Sachgemäßheit werden erreicht, indem der Kommunikator die Aspekte verändert (ebd., S. 227). Zu den Aspekten zählen der Kontext, die Darstellungsform, die Sprachebene und der Sachinhalt (ebd., S. 227). Diese sind in Abb. 1 im Kreis illustriert. So könnte der Kommunikator bei einem Gespräch mit einer fachfremden Person beispielsweise auf Abbildungen (Darstellungsform) oder

einem alltagstauglichen Beispiel (Kontext) zurückgreifen, um in der Alltagssprache (Sprachebene) ein vereinfachtes Thema (Sachinhalt) darzustellen. Alles in allem wird durch die hier beispielhaft aufgeführte Variation der Aspekte eine sach- und adressatengemäße Kommunikation ermöglicht.

Die Autoren konnten mithilfe eines erweiterten Modells und einer diesbezüglichen empirischen Studie feststellen, dass hohes Fachwissen in Physik nicht unbedingt mit hoher Kommunikationskompetenz zusammenhängt (ebd., S. 235). Im Gegenteil: Bezogen auf die adressatengerechte Kommunikation ist das „*Fachwissen [...] [anscheinend] nur notwendig, um eine bestimmte Schwelle an Kommunikationskompetenz zu überschreiten, danach hilft ein Zuwachs an Fachwissen nicht weiter oder ist der höheren Ausprägung von Kommunikationskompetenz sogar hinderlich* [Hervorhebung im Original]“ (Kulgemeyer, 2010b, S. 310). Zwar ist das Kommunikationsmodell von Kulgemeyer und Schecker samt der empirischen Studie domänenspezifisch für das Fach Physik angelegt. Dennoch liegt die Vermutung nahe, dass Fachwissen und Kommunikation auch in anderen Schulfächern als getrennte Bereiche gesehen werden müssen, die nicht zwangsläufig positiv miteinander korrelieren. Umso wichtiger scheint es, neben der sachgemäßen Kommunikation auch die adressatengemäße Kommunikation in den Fokus zu rücken.

Die adressatengemäße Kommunikation steht im Mittelpunkt der **Theorie der systemisch-funktionalen Grammatik** (Halliday & Matthiessen, 2014). In dieser Theorie wird Sprache als ein großes Netzwerk von Systemen beschrieben (ebd., S. 23). Die Grammatik ist ein solches Systemnetzwerk, da es sich hierbei nicht einfach um ein Inventar an einzelnen grammatischen Strukturen handelt, sondern die Strukturen miteinander in Verbindung stehen und gemeinsam das Gesamtbild von Sprache beeinflussen (ebd., S. 20, 23). Die Kommunikatoren treffen andauernd eine Wahl der sprachlichen Mittel aus dem großen Systemnetzwerk der Sprache beziehungsweise dem Systemnetzwerk der Grammatik. Sprache ist funktional, da durch die paradigmatische Wahl der Sprachmöglichkeiten bestimmte Bedeutungen kreiert werden (ebd., S. 22-23). Dies wird wie folgt zusammengefasst: „*A language is a resource for making meaning, and meaning resides in systemic patterns of choice* [Hervorhebung d. Verf.]“ (Halliday & Matthiessen, 2014, S. 23).

Abb. 2 zeigt, wie das System genau beschrieben ist, und welche Wahlmöglichkeiten den Kommunikatoren laut der systemisch-funktionalen Grammatik zur Verfügung stehen.

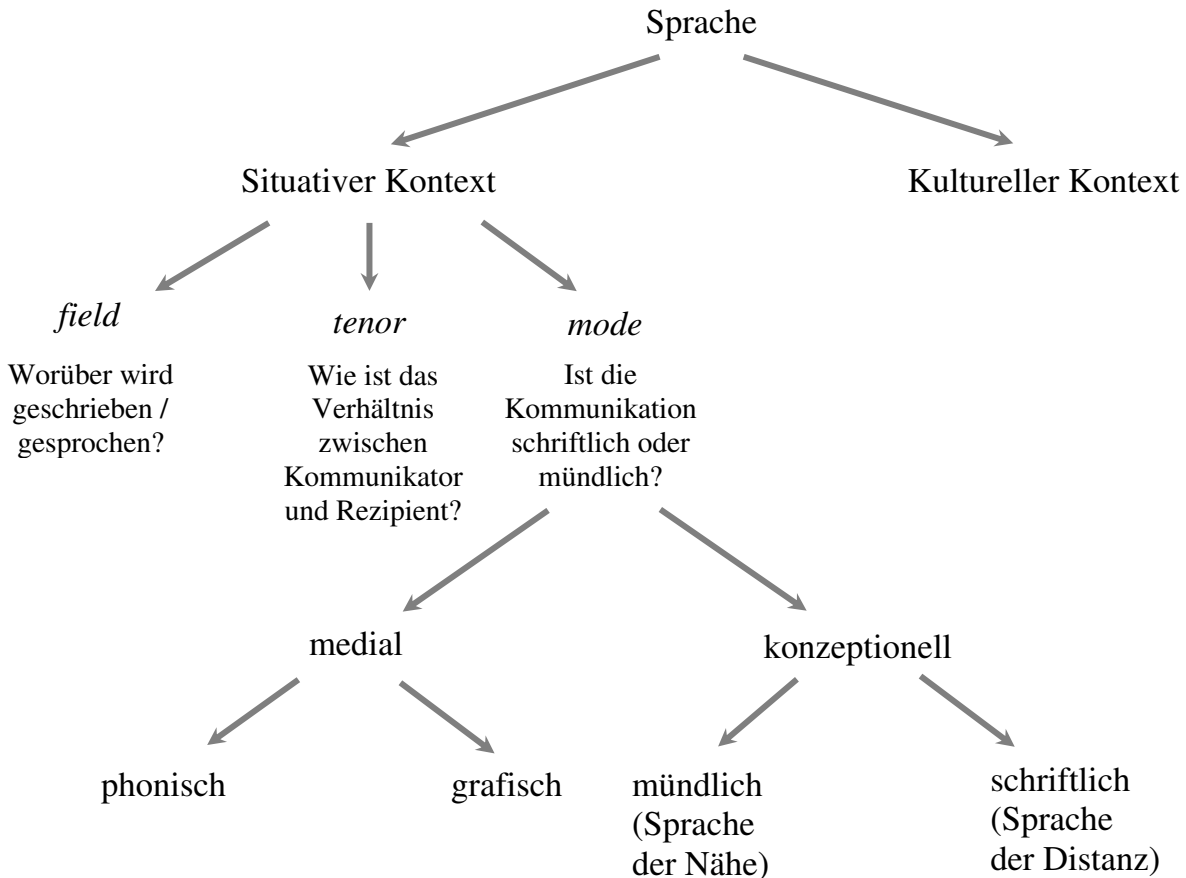


Abb. 2: Kennzeichen von Sprache

Eigene Darstellung mit Informationen, entnommen aus Halliday und Hasan (1985), Gibbons (2014), Söll und Hausmann (1985) sowie Koch und Österreicher (1986)

Wie in Abb. 2 illustriert, wird Sprache immer im Kontext verwendet (ebd., S. 32-33). Der kulturelle Kontext ist für das vorliegende Untersuchungsvorhaben zum Nutzen von Wortspeichern unerheblich. Der situative Kontext umfasst drei Kategorien: *field*, *tenor* und *mode* (ebd., S. 12). Bei der Kategorie *field* geht es um das Themenfeld und somit um die Frage: Worüber wird geschrieben oder gesprochen (Gibbons, 2014, S. 5)? Es macht einen Unterschied, ob sich Personen inhaltlich über den neusten Modetrend oder über die Genomsequenzierung einer Gorillapopulation austauschen. Je nach Themenfeld variiert die Sprache. Die Kategorie *tenor* spezifiziert die soziale Beziehung zwischen Sprecher und Adressat. Die passende Frage zu dieser Kategorie lautet: Wie ist das Verhältnis zwischen Kommunikator und Rezipient (ebd., S. 5)? Je nachdem, ob ein Kommunikator zum Beispiel mit einer gleichaltrigen befreundeten Person oder einer unbekannteren älteren Person im Rahmen eines Bewerbungsgesprächs kommuniziert, wird eine andere Sprache gewählt. Bei der Kategorie *mode* geht es unter anderem um den Kommunikationskanal und somit um die Frage: Ist die Kommunikation schriftlich oder mündlich (ebd., S. 5)? Es macht beispielsweise einen Unterschied, ob sich ein

Kommunikator mündlich am Postschalter über einen nicht zugestellten Brief beschwert oder diese Beschwerde schriftlich verfasst. Die Sprache variiert somit nicht nur je nach Thema (*field*) und sozialer Beziehung (*tenor*), sondern auch je nach Kommunikationskanal (*mode*). Alle drei Kategorien beeinflussen die Wahl des Kommunikators aus dem Systemnetzwerk der Grammatik und verhelfen diesem zu einer erfolgreichen Kommunikation.

Der Kommunikationskanal muss wiederum differenzierter betrachtet werden, denn an dieser Stelle kann zwischen **Mündlichkeit und Schriftlichkeit bezüglich des Mediums** oder **bezüglich der Konzeption** unterschieden werden (Söll & Hausmann, 1985, S. 19). Während die mediale Mündlichkeit bei der Übertragung von Schallwellen gegeben ist, zeichnet sich die mediale Schriftlichkeit durch Schriftzeichen aus (Kniffka & Siebert-Ott, 2012, S. 18). Ersteres wird als phonischer Kode und letzteres als grafischer Kode bezeichnet (Söll & Hausmann, 1985, S. 17). Es handelt sich hierbei um eine dichotome Aufteilung: Ein Text kann entweder gesprochen oder geschrieben sein (Koch & Österreicher, 1986, S. 17). Allerdings macht es einen großen Unterschied, ob es sich bei dem gesprochenen Text zum Beispiel um ein Gespräch mit einer befreundeten Person oder um einen Vortrag handelt. Auch im medial schriftlichen Text kann es sich einerseits um eine informelle Chat-Nachricht oder andererseits um eine Verwaltungsvorschrift handeln. Diese Unterschiede werden mit der konzeptionellen Mündlichkeit und Schriftlichkeit näher beschrieben. Im Gegensatz zur dichotomen medialen Aufteilung wird die konzeptionelle Ebene als ein Kontinuum gesehen (ebd., S. 17). Koch und Österreicher (1985) gehen genauer auf die konzeptionelle Ebene ein. Anstelle einer konzeptionellen Mündlich- oder Schriftlichkeit bezeichnen sie die Pole des Kontinuums mit **Sprache der Nähe** und **Sprache der Distanz** (ebd., S. 21). Die Einordnung von Texten innerhalb des Kontinuums ist relativ und wird nicht genau bestimmt (ebd., S. 18). In den genannten Beispielen ist das Gespräch mit einer befreundeten Person dementsprechend eine phonisch realisierte Kommunikation, die konzeptionell in Richtung *Sprache der Nähe* einzuordnen ist. Der Vortrag ist ebenfalls eine phonisch realisierte Kommunikation, die jedoch mehr Elemente der *Sprache der Distanz* aufweist. Gleichmaßen sind Chat-Nachricht und Verwaltungsvorschrift beide grafisch realisiert, unterscheiden sich jedoch in ihrer Nähe zu den Polen *Sprache der Nähe* (Chat-Nachricht) und *Sprache der Distanz* (Verwaltungsvorschrift) (vgl. Koch & Österreicher, 1986, S. 18; Maak, 2018, S. 11).

Die Eigenschaften von konzeptioneller Mündlichkeit und Schriftlichkeit beziehungsweise *Sprache der Nähe* und *Sprache der Distanz* werden in folgender Tab. 1 aus Archie, Rank und Franz (2014, S. 200) zusammenfassend dargestellt. Ähnlich wie dem elaborierten Code zugeschrieben, zeichnet sich konzeptionelle Schriftlichkeit zum Beispiel durch ihre Abstraktheit und Situationsunabhängigkeit aus. Sowohl der Vortrag als auch die Verwaltungsvorschrift enthalten kompakt viele Informationen und sind

unpersönlich. Außerdem weisen sie komplexe Satzstrukturen und einen vielseitigen Wortschatz mit Nominalisierungen und Komposita auf. Durch diese Merkmale können die konzeptionell schriftlichen Texte auch von Personen verstanden werden, die sich nicht zur gleichen Zeit und am gleichen Ort wie der Kommunikator befinden. Im Gegensatz dazu ist der Satz *Ich meine den Becher da!* unverständlich für Adressaten, die sich nicht zur gleichen Zeit am gleichen Ort befinden. Er wäre somit der grafischen Realisierung der konzeptionellen Mündlichkeit zuzuordnen (ebd., S. 200).

Tab. 1: Konzeptionelle Mündlich- und Schriftlichkeit (Archie et al., 2014, S. 200)

Konzeptionell mündlich bedeutet	Konzeptionell schriftlich bedeutet
konkret und situationsabhängig	abstrakt und situationsunabhängig
geringe Informationsdichte	hohe Informationsdichte
dialogisch und interaktiv	monologisch und und [sic!] persönlich
unvollständige Sätze, keine durchgängige grammatikalische Struktur	vollständige Sätze mit komplexen Satzstrukturen (wie Nebensatz-, Partizipial-, Passivkonstruktionen)
unterstützt durch Gestik und Mimik	differenzierter Wortschatz (wie Nominalisierungen, Komposita)
<i>Beispiel: „Ich meine den Becher da!“</i>	<i>Beispiel: „Ich meine den großen, durchsichtigen Plastikbecher mit 0,5 l Inhalt“</i>

Konzeptionelle Mündlichkeit und konzeptionelle Schriftlichkeit stehen mit den Sprachregistern im Zusammenhang. Die in der Tabelle genannten Merkmale werden oftmals genutzt, um konzeptionelle Mündlichkeit dem Sprachregister Alltagsprache und **konzeptionelle Schriftlichkeit** dem Sprachregister **Bildungssprache** zuzuordnen (Fornol, 2017, S. 178). Die genauen Begrifflichkeiten bezüglich Sprachregister und ihrer Abgrenzung werden auf Grundlage der vorgestellten Kennzeichen von Sprache im folgenden Kapitel deutlich.

2.1.2 Sprachregister - eine begriffliche Klärung

Ein Sprachregister beschreibt laut Eckhardt (2008) „[...] typische Rede- und Schreibweisen innerhalb [...] einer bestimmten Kommunikationssituation [...]“ (S. 64). Beispiele für Sprachregister sind die Alltagsprache, Bildungssprache, Schulsprache und

die Fachsprache. Daneben gibt es weitere Register, die – je nach Autor:in – teils unterschiedlich benannt oder definiert sind (vgl. Ahrenholz, Knoblich, & Reichel, 2018, S. 169). Im vorangehenden Kapitel sind die allgemeinen Kennzeichen von Sprache beschrieben, die zur Definition eines Sprachregisters bedeutsam sind. Hierzu gehören die Begriffe *field*, *tenor* und *mode* und die unter *mode* beschriebene Abgrenzung zwischen konzeptioneller Mündlichkeit und konzeptioneller Schriftlichkeit. Wenngleich die Zuordnung der Alltags- und der Bildungssprache zur konzeptionellen Mündlichkeit (Alltagssprache) und Schriftlichkeit (Bildungssprache) vermuten lässt, dass ein Register nur durch den Kommunikationskanal (*mode*) beschrieben wird, so betont Tajmel (2017a, S. 254), dass alle drei Kategorien *field*, *tenor* und *mode* ein Register genauer definieren. In einigen Veröffentlichungen wird dieser Bereich aus der im vorherigen Kapitel beschriebenen systemisch-funktionalen Grammatiktheorie daher auch als Register-Theorie nach Halliday und Hasan bezeichnet (vgl. Quehl & Scheffler, 2008, S. 75).

Da sich der Kontext in jeder Kommunikationssituation aufgrund der oben genannten Kategorien *field*, *tenor* und *mode* unterscheidet, sind **Register jedoch nicht klar voneinander abzugrenzen** (Bochnik, 2017, S. 28). Auch Rincke (2010, S. 242) hält fest, dass es beispielsweise nicht *die* Alltagssprache gibt. Stattdessen können zwar typische grammatische und lexikalische Merkmale eines Sprachregisters genannt werden, doch letzten Endes entscheidet der Kontext darüber, ob ein entsprechender Sprachstil der Situation angemessen ist und somit Sprache ihre Funktion erfüllt (ebd., S. 239). Trotz der schwierigen Abgrenzung zwischen den verschiedenen Sprachregistern wird im Folgenden eine annähernde Beschreibung der Bildungssprache unternommen, die in Kapitel 2.1.4 auf den Kontext *Grundschule* appliziert wird. Darüber hinaus werden auch die weiteren Sprachregister in einem strukturorientierten Überblick kurz dargestellt und – soweit möglich – von Bildungssprache abgegrenzt.

Die **Bildungssprache** wird im Englischen mit *academic language* oder *academic discourse* bezeichnet (Gogolin & Lange, 2011, S. 110). Nach der Definition von Habermas ist Bildungssprache

[...] die Sprache, die überwiegend in den Massenmedien, in Fernsehen, Rundfunk, Tages- und Wochenzeitungen benutzt wird. Sie unterscheidet sich [...] von Fachsprachen dadurch, daß [sic!] sie grundsätzlich für alle offensteht, die sich mit den Mitteln der allgemeinen Schulbildung ein Orientierungswissen verschaffen können. (Habermas, 1978, S. 330)

Das Themenfeld (*field*) umfasst nach dieser Definition die Themen des öffentlichen Interesses. Allerdings ist Bildungssprache thematisch nicht nur auf Themen der öffentlichen Medien begrenzt, sondern wird im Allgemeinen zur „[...] Vermittlung und Aneignung von Wissen [...]“ (Gogolin & Duarte, 2016, S. 485) genutzt. Deshalb weisen Gogolin und Duarte (2016, S. 485) dem Themenfeld (*field*) der Bildungssprache die

schulische Bildung zu. Die Gesprächspartner sind einander meist relativ fremd, emotional distanziert und geben durch die Bildungssprache eine gewisse Hierarchie zum Ausdruck (*tenor*) (ebd., S. 486). Schlussendlich weist die Bildungssprache viele Elemente der konzeptionellen Schriftlichkeit beziehungsweise der *Sprache der Distanz* auf (*mode*) (ebd., S. 485-486). Dies bedeutet, dass „Zeit, Personen, Orte [und] Sachverhalte [...] mit sprachlichen Mitteln beschrieben werden [müssen]“ (Weis, 2013, S. 15). Dafür ist ein großer Wortschatz (ebd., S. 15) und ein präziser Wortgebrauch notwendig (Leisen, 2015, S. 133). Füllwörter – d.h. Wörter mit geringer Aussage, wie z.B. *gewissermaßen* – werden nicht benutzt (ebd., S. 133). Die Sätze sind komplex und vollständig und weisen keine grammatikalischen Fehler auf (ebd., S. 133). Die Komplexität der Sätze zeichnet sich durch hypotaktische Strukturen aus (Weis, 2013, S. 15). Es gibt wenige Wiederholungen und keine Gedankensprünge (Leisen, 2015, S. 133).

Gegenüber der Beschreibung von Bildungssprache sind in Tab. 2 die weiteren Sprachregister hinsichtlich *field*, *tenor*, *mode* und besonderer Kennzeichen aufgeführt. Da Schulsprache und Bildungssprache schwer trennbar erscheinen, wird im Folgenden eine Abgrenzung unternommen. Als Beispiel für Schulsprache dient die Forderung von Lehrkräften, dass die Schüler:innen in ganzen Sätzen sprechen. Feilke (2012b, S. 149-150) stellt fest, dass dies zwar typisch für den Schulalltag, aber untypisch für die Realität ist. Mit weiteren Beispielen zeigt Feilke (2012a, S. 6), dass die Normen der Schulsprache didaktisch konstruiert und somit artifiziell sind. Die Normen werden von der Institution Schule bestimmt und vorgegeben (Feilke, 2012b, S. 153). Damit unterscheidet sich die Schulsprache gravierend von der Bildungssprache. Denn während die Schulsprache die sprachlichen Mittel beschreibt, die zu didaktischen Zwecken genutzt werden, haben sich die sprachlichen Mittel der Bildungssprache im Laufe der Jahre durch ihren empirischen Gebrauch herausgestellt (Feilke, 2012a, S. 5-6). Als weiteres Beispiel dient die Erörterung. Während das Erörtern eine bildungssprachliche Handlung ist, schreiben die Schüler:innen in der Schule eine Erörterung im didaktischen Sinne (ebd., S. 5-6). Die didaktische Gattung der Erörterung, wie sie im Unterricht gelehrt wird, ist unabhängig von der Realität: „Außerhalb der Schule schreibt niemand Erörterungen“ (ebd., S. 5). **Dennoch dient die didaktische Erörterung mit ihren artifiziellen sprachlichen Erwartungen einem Zweck, denn sie stellt den Übergang zum Erwerb des bildungssprachlichen Erörterns dar.** Anders formuliert: Die Schulsprache, die sich im Schreiben einer didaktischen Erörterung findet, dient als *transitorisches Mittel* zur Erreichung eines bildungssprachlichen Ziels (Feilke, 2012b, S. 170; 2019, S. 2).

Tab. 2: Beschreibende Merkmale von Alltags-, Schul- und Fachsprache

	Alltagssprache	Schulsprache	Fachsprache
<i>field</i>	<ul style="list-style-type: none"> alltägliche Dinge (vgl. L. Hoffmann, 2019, S. 1) Einstellungen und Gefühle (vgl. Biber, 1995, S. 3) 	<ul style="list-style-type: none"> schulische Bildung (Feilke, 2012b, S. 166) 3 Definitionen von Schulsprache: <ol style="list-style-type: none"> 1) die Sprache, wie sie im Deutschunterricht gelehrt wird und die Sprache, die im Fachunterricht ihre Verwendung findet (Vollmer & Thürmann, 2010, S. 108) 2) synonym zur Bildungssprache (vor allem im anglophonen Raum) (Feilke, 2019, S. 1) 3) stärker auf Schule ausgerichtetes Register (Feilke, 2019, S. 1) → Im Folgenden wird die dritte Definition zur Beschreibung der Schulsprache übernommen. 	<ul style="list-style-type: none"> Jeweiliges Fachgebiet <ul style="list-style-type: none"> - beispielsweise die naturwissenschaftlichen, technischen oder geisteswissenschaftlichen Fachsprachen (Steinmüller & Scharnhorst, 1987, S. 4), - die noch einmal für jede einzelne Disziplin wie Mathematik oder Psychologie genauer ausdifferenziert werden können (vgl. M. Hoffmann, 2007, S. 26; Rincke, 2010, S. 238) thematisch auf einer allgemein-gesellschaftlichen Ebene zu lokalisieren (Kniffka & Roelcke S. 57)
<i>tenor</i>	<ul style="list-style-type: none"> Gesprächspartner:innen sind für gewöhnlich miteinander vertraut (L. Hoffmann, 2019, S. 1). 	<ul style="list-style-type: none"> Kommunikation zwischen Schüler:innen untereinander, zwischen Schüler:innen und Lehrkräften oder zwischen Schüler:innen und einem fiktiven Adressaten 	<ul style="list-style-type: none"> Die soziale Beziehung der Gesprächspartner:innen ist distanziert. Fachsprache wird entweder von Forschenden unter sich, zwischen Forschenden und Studierenden oder zwischen Forschenden und Laien genutzt (M. Hoffmann, 2007, S. 23)
<i>mode</i>	<ul style="list-style-type: none"> viele Elemente der konzeptionellen Mündlichkeit (Fornol, 2017, S. 178) 	<ul style="list-style-type: none"> im Bereich der konzeptionellen Schriftlichkeit (Vollmer & Thürmann, 2010, S. 109) 	<ul style="list-style-type: none"> im Bereich der konzeptionellen Schriftlichkeit (Weis, 2013, S. 15)
Kennzeichen	<ul style="list-style-type: none"> viele deiktische Begriffe, wie <i>hier</i>, <i>da</i> und <i>dort</i> aufweist (Weis, 2013, S. 15) geringer Wortschatz (Weis, 2013, S. 15) 	<ul style="list-style-type: none"> „[...] <i>prägnant, präzise, vollständig, komplex, strukturiert, objektiv, distanzant, emotionsfrei, eindeutig</i>, 	<ul style="list-style-type: none"> spezieller Wortschatz & spezielle syntaktische Formulierungen (Steinmüller & Scharnhorst, 1987, S. 5)

	<ul style="list-style-type: none"> • die verwendeten Wörter sind unpräzise (Leisen, 2012, S. 35) • Verwendung vieler Füllwörter, wie <i>ja, nee, doch</i> (Leisen, 2012, S. 35; Weis, 2013, S. 15). • meist unvollständige Sätze (Leisen, 2012, S. 35) • grammatikalische Fehler (Leisen, 2012, S. 35) • Wiederholungen und häufige Gedankensprünge (Leisen, 2012, S. 35) 	<p><i>situationsungebunden</i> und <i>dekontextualisiert</i> [Hervorhebung im Original]“ (Vollmer & Thürmann, 2010, S. 109)</p> <ul style="list-style-type: none"> • viele Fremdwörter & Fachbegriffe (Schmölzer-Eibinger, 2013, S. 26) • komplexe syntaktische Strukturen (Schmölzer-Eibinger, 2013, S. 26) • hoher Anteil an Passivkonstruktionen (Schmölzer-Eibinger, 2013, S. 26) • unpersönliche Ausdrücke (Schmölzer-Eibinger, 2013, S. 26) 	<ul style="list-style-type: none"> • „[...] terminologische Exaktheit und Eindeutigkeit [...]“ (Roelcke, 2019, S. 1) • „[...] Entlehnungen aus fremden Sprachen (insbesondere aus dem Lateinischen, Griechischen und Englischen) [...]“ (Roelcke, 2019, S. 2) • syntaktische Besonderheiten: U.a. <ul style="list-style-type: none"> - Funktionsverbgefüge (z.B. <i>Antrag stellen</i>), - komplexe Attribute (z.B. <i>Das vorbeifahrende große salzstreuende Fahrzeug...</i>) - passivischer oder unpersönlicher Stil (z.B. <i>Man fügt...</i>) (Leisen (1999, S. 6)
Nahestehende Begriffe	<ul style="list-style-type: none"> • Umgangssprache <ul style="list-style-type: none"> - synonym zur Alltagssprache (vgl. Habermas, 1978, S. 328) - ... oder als eigene Sprachvariante definiert (Glück & Rödel, 2016, S. 733) • Handlungssprache (Leisen, 2012, S. 35; 2015, S. 133) 	<ul style="list-style-type: none"> • classroom language (Feilke, 2012a, S. 7) <ul style="list-style-type: none"> - bezeichnet lediglich die organisatorische Sprache in der Institution Schule - Bsp.: „Holt die Hefte raus!“ (Feilke, 2012a, S. 7) • Unterrichtssprache (Leisen, 2012, S. 30) <ul style="list-style-type: none"> - „[...] Sprache, die vom Vokabular und ihren Formulierungen her in mündlicher wie auch schriftlicher Form typischerweise beim Lehren und Lernen im unterrichtlichen Kontext benutzt wird“ (Leisen, 2012, S. 30) 	<ul style="list-style-type: none"> • Wissenschaftssprache Habermas (1978, S. 3) <ul style="list-style-type: none"> - fachgebietsübergreifend (im Gegensatz zu den Fachsprachen) (M. Hoffmann, 2007, S. 26)

In der vorliegenden Arbeit wird der Begriff **Bildungssprache** verwendet. Dies hat mehrere Gründe. Zum einen stellt sich die Frage, welches Sprachregister neben der alltäglich vorkommenden Alltagssprache in der Grundschule erwartet und verwendet wird. Hierzu zunächst ein Blick darauf, warum der Begriff Bildungssprache gegenüber der Fachsprache in diesem Forschungsvorhaben bevorzugt wird. Beyer (2015, S. 34) geht davon aus, dass erst in der Oberstufe von Fachsprache als Register gesprochen werden kann. Dagegen schreibt Luchtenberg (2011, S. 24), dass Kinder bereits ab dem Kindergarten durch das Fernsehen, wie z.B. durch die Betrachtung der *Sendung mit der Maus* oder durch spezifische Hobbys, mit Fachsprache in Kontakt kommen. Allerdings zeigen die folgenden zwei Beispiele die Schwierigkeit, bestimmte lexikalische und grammatische Merkmale im Rahmen des Grundschulunterrichts ausschließlich der Fachsprache und nicht zusätzlich der Bildungssprache zuzuordnen. So kann das Wort *der Stromkreislauf* im Grundschulkontext unterrichtet und aufgrund seiner terminologischen Exaktheit berechtigterweise der Fachsprache zugeordnet werden. Allerdings findet das Wort als Teil eines allgemeinen Orientierungswissens ebenso in den Massenmedien Verwendung und kann somit gleichzeitig als bildungssprachlich gelten. Gleiches gilt für die Verwendung des Passivs. Wenn diese grammatikalisch komplexe Struktur in der Grundschule gefordert wird, kann diese im fachlichen Zusammenhang als Fachsprache und gleichzeitig der allgemein zugänglichen Bildungssprache zugeordnet werden. Die Beispiele zeigen, dass die Bildungssprache das allgemeinere Sprachregister darstellt, das die Fachsprache oder fachsprachliche Elemente teilweise umfasst (Gogolin & Lange, 2011, S. 112). Im Gegensatz zu definitiv fachsprachlichen Wörtern, wie beispielsweise die *DNA-Methylierung* im Kontext der *Epigenetik*, ist in der Grundschule nicht von fachsprachlichen Wörtern auszugehen, die nicht auch in den Massenmedien vorkommen würden. Durch die Verwendung des allgemeineren Begriffs Bildungssprache können insofern fachsprachliche Elemente, die in der Grundschule vorkommen, mit eingeschlossen sein.

Im Vergleich zwischen Bildungssprache und Schulsprache ist der Begriff *Bildungssprache* ebenfalls der allgemeinere Begriff.³ Das heißt: Die Bildungssprache umfasst zum Teil Merkmale der Schulsprache (Gogolin & Lange, 2011, S. 112; Rincke & Markic, 2018, S. 40). Der Hauptunterschied ist laut der ausführlich dargestellten Abgrenzung zwischen Schulsprache und Bildungssprache in der artifiziellen Beschaffenheit der Schulsprache als Übergang zur Bildungs- oder Fachsprache zu sehen. Schreiben die Kinder in einem Versuchsprotokoll im schulischen Kontext *Ich vermute, dass...*, so

³ Beyer (2015, S. 32) argumentiert, dass der Begriff **Standardsprache** aufgrund seiner Wertneutralität bevorzugt anstelle der beiden Sprachregister Schulsprache oder Bildungssprache verwendet werden sollte. Laut Einschätzung der Autorin ist dieser Begriff nicht in allen Forschungsgebieten klar definiert. Dies wird u.a. daran deutlich, dass die Standardsprache bei Juska-Bacher und Beckert (2015, S. 125) in Abgrenzung zur Bildungssprache verwendet wird. Insofern wird auf den Begriff der Standardsprache in diesem Untersuchungsvorhaben nicht näher eingegangen.

handelt es sich um Schulsprache, denn in der Realität ist solch eine Formulierung im Rahmen des wissenschaftlichen Arbeitens unüblich. Andererseits werden im schulischen Versuchsprotokoll auch Wörter, Formulierungen und Strukturen angewandt, die nicht nur im schulischen Kontext von Bedeutung sind. Dazu zählen das Passiv oder die Verwendung von Konnektoren, wie *daraufhin*. Als Schlussfolgerung kann das Schreiben eines Versuchsprotokolls im schulischen Kontext sowohl schulsprachliche als auch bildungssprachliche Elemente aufweisen. Um einer sprachlichen Konfundierung in diesem Forschungsvorhaben vorzubeugen, wird der allgemeine Begriff Bildungssprache gewählt. **Bedeutsam ist jedoch das Verständnis, dass diese Eingrenzung nicht ausschließt, dass einzelne Wörter oder Formulierungen und Strukturen ebenfalls der Schulsprache oder gar einer Fachsprache zuzuordnen sind.**

2.1.3 Bedeutung von Bildungssprache im schulischen Kontext

Dem Register Bildungssprache als eigenständige Kategorie werden darüber hinaus verschiedene Funktionen zugeschrieben, die auch im schulischen Kontext von Bedeutung sein können.

Morek und Heller (2012, S. 70) nennen drei grundlegende Funktionen von Bildungssprache. Sie beschreiben Bildungssprache als *Medium von Wissenstransfer* (kommunikative Funktion), als *Werkzeug des Denkens* (epistemische Funktion) und als *Eintritts- und Visitenkarte* (sozialsymbolische Funktion) (ebd., S. 70).

Sie verdeutlichen, dass erst mithilfe der Bildungssprache komplexe Inhalte kommuniziert werden können (*kommunikative Funktion*) (ebd., S. 74). Tajmel und Hägi-Mead (2017, S. 16) verwenden den Begriff der **Sprachhandlungsfähigkeit**. Kinder im Bildungssystem sollen sprachhandlungsfähig sein, das heißt sie sollen im unterrichtlichen Kontext alles klar und präzise ausdrücken können (ebd., S. 16). Die Betonung der Sprachhandlungsfähigkeit wird auch in einer Leitlinie deutlich, die sich auf die Spracharbeit im Physikunterricht bezieht (Rincke & Leisen, 2015, S. 652). In dieser Leitlinie ist das sprachliche Ziel im Unterricht nicht die fehlerfreie Anwendung von Sprache, sondern die Fähigkeit, einen sprachlich angemessenen Text zum Ausdruck zu bringen (ebd., S. 652). Die gewünschte Sprachhandlungsfähigkeit im Sachunterricht stellt die Basis des vorliegenden Forschungsvorhabens dar.

Hinsichtlich der *epistemischen Funktion* wird die in der Einleitung beschriebene Verknüpfung zwischen Sprache und Denken auch von Morek und Heller (2012, S. 74-76) herausgestellt. Sie formulieren: „Wer Bildungssprache adäquat verwenden kann, der ist auch in der Lage zu den damit in Zusammenhang stehenden komplexen kognitiven Operationen (wie z.B. Abstraktion, Verallgemeinerung, Kausalität)“ (Morek & Heller, 2012, S. 75). Empirisch untermauert wird diese zitierte Aussage von einer

Längsschnittstudie aus der Mathematikdidaktik. Diese zeigt einen signifikanten Zusammenhang zwischen sprachlichen Fähigkeiten und konzeptuellem Wissen bei Grundschulkindern im Fach Mathematik (Ufer, Reiss, & Mehringer, 2013, S. 197-198).

Bezogen auf die *sozialsymbolische Funktion* verdeutlichen Morek und Heller (2012, S. 79), dass die Bildungssprache zu einer sozialen Identität innerhalb einer akademischen Gemeinschaft führt und diese bewusst angenommen oder abgelehnt werden kann, weshalb **die identitätsstiftende Funktion im Unterricht mit berücksichtigt werden sollte** (ebd., S. 79-82). Erreichbar ist dies, indem Lernenden die Möglichkeit angeboten wird in eine Rolle zu schlüpfen und somit die Bildungssprache anzunehmen ohne sie als ihre eigene Sprache ausgeben zu müssen (ebd., S. 81-82).

2.1.4 Bildungssprache in der Grundschule

Das **Vorkommen der Bildungssprache im frühen Kindesalter** ist empirisch belegt. Durch eine Analyse mündlicher Versuchsbeschreibungen zum Thema *Schwimmen und Sinken* konnten Rank, Hartinger, Wildemann und Tietze (2018, S. 124) im Rahmen des Projekts *Easi-Science-L* zeigen, dass Kinder in der Vorschule zwar hauptsächlich die Alltagssprache nutzen. Dennoch finden sich hier bereits bildungssprachliche Elemente (Rank, Hartinger, et al., 2018, S. 124; Rank, Wildemann, Hartinger, & Tietze, 2018, S. 188; Wildemann, Rank, Hartinger, & Sutter, 2016, S. 79). Dabei beeinflussen eine hohe Intelligenz, das männliche Geschlecht sowie ein hohes Alter der Vorschulkinder die mündliche Anwendung der Bildungssprache positiv (Rank, Hartinger, et al., 2018, S. 125-126). So ist es nicht verwunderlich, wenn nach dem Vorschulalter die Grundschulkindern vermehrt bildungssprachliche Elemente anwenden können. Dies wird gestützt von Juska-Bacher und Beckert (2015, S. 53), die im Rahmen des Projekts *Wortschatz und Wortlesen* (WuW) den 6-7 jährigen Grundschulkindern in der Schweiz vielfältige bildungssprachliche „[...] Fähigkeiten bzw. Ansätze [...]“ (S. 124) attestieren.⁴ Dabei wurden die bildungssprachlichen Fähigkeiten auch hier – wie im *Easi-Science-L*-Projekt – in mündlich stattfindenden Interviews analysiert.

Sowohl die Studie zur Bildungssprache in der Vorschule als auch die Studie zur Bildungssprache in der 1. Klasse geben einen Einblick in **mögliche Kriterien zur Beschreibung und Messung der kindlichen Bildungssprache**. Juska-Bacher und Beckert (2015, S. 64) nehmen zur Beschreibung der Bildungssprache im Kindesalter eine allumfassende Perspektive ein. In ihrem *Wortschatz und Wortlesen*-Modell (WuW-Modell) – dargestellt in Abb. 3 – unterscheiden sie eine **sprachliche, eine kognitiv-**

⁴ Die Durchführung der Studie fand in einer Region der Schweiz statt, in der die Kinder größtenteils einen Dialekt und Standarddeutsch sprechen (Juska-Bacher & Beckert, 2015, S. 30). Entsprechend ist davon auszugehen, dass auch in Deutschland bildungssprachliche Fähigkeiten im Grundschulalter vorhanden sind.

reflexive und eine entwicklungspsychologische Dimension. Während die sprachliche Dimension die Bereiche *Syntax*, *Morphologie* und *Lexik* umfasst, gehen die anderen beiden Dimensionen darüber hinaus. So ist nicht nur die sprachliche Kompetenz im Allgemeinen, sondern auch die Fähigkeit zur Sprachreflexion und der Gebrauch von Bildungssprache in unterschiedlichen Situationen (kognitiv-reflexive Dimension) sowie eine positive Einstellung gegenüber konzeptioneller Schriftlichkeit (entwicklungspsychologische Dimension) von Bedeutung (ebd., S. 64, 128).

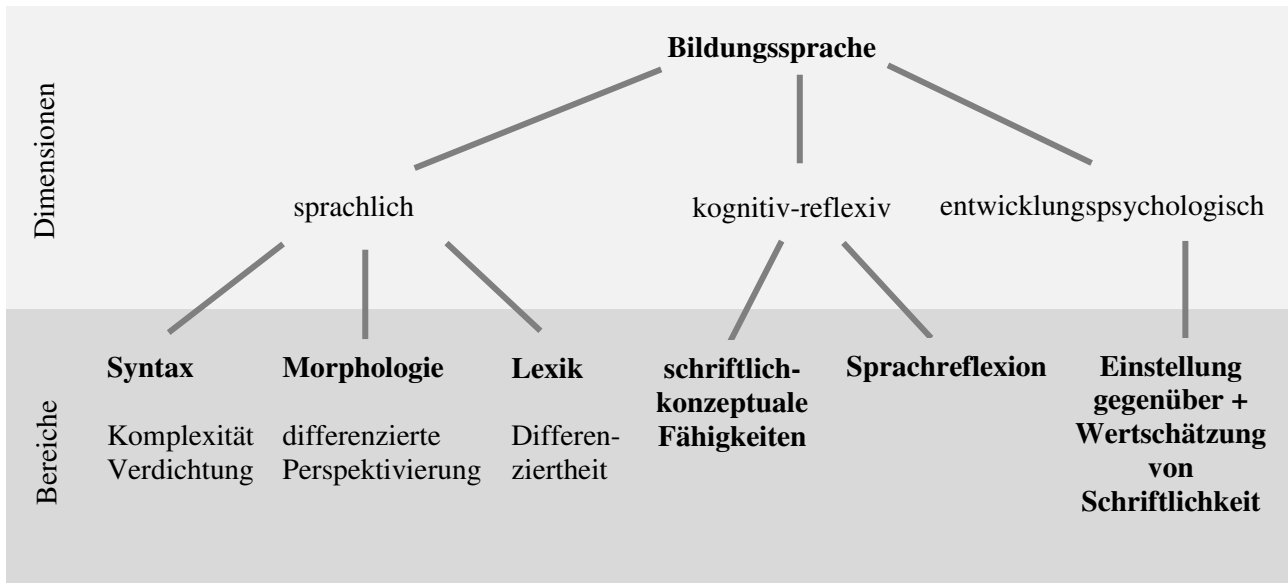


Abb. 3: Modell zur Bildungssprache in der Grundschule (Juska-Bacher & Beckert, 2015, S. 39)

Zur Messung der Bildungssprache in der Vorschule (*Easi-Science-L*) entwickelten Tietze, Rank und Wildemann (2016) auf theoretischer Grundlage und mit empirischer Anpassung eine Ratingskala – auch RaBi genannt. Diese umfasst drei linguistische Felder: **Sprachhandlungen, Lexikon und Morphosyntax** (vgl. ebd). Alle drei linguistischen Felder umfassen Subkategorien, sodass etwa die Lexik durch die Subkategorien Nomen, Verben und Adjektive genauer definiert ist. Jede Subkategorie kann auf einer Skala von vier verschiedenen Niveaustufen (0-3) beurteilt werden, wobei Niveaustufe 0 die Alltagssprache und Niveaustufe 3 die Bildungssprache darlegt. Abb. 4 zeigt einen Ausschnitt aus der Ratingskala. In diesem Fall handelt es sich um einen Ausschnitt zur Kodierung von Nomen.

		Nomen			
		Ausprägung	0	1	2
Lexikon	Definition	Rudimentärer Wortschatz	Alltagssprachliche Verwendung von Nomen ohne konkreten Fachbezug	Alltagssprachliche Verwendung von Nomen mit konkreten Fachbezug	Bildungssprachliche Verwendung von Nomen mit Fachbezug
	Kriterium	<p>Das Kind verwendet keine kontextbezogenen Fachbegriffe.</p> <p>Kein differenzierter Wortschatz (ohne den Zusammenhang zu kennen wird an der Äußerung nicht deutlich, was gemeint ist, weil die verwendeten Ausdrücke unpräzise sind.)</p> <p>Verwendete Nomen beruhen auf vorheriger Instruktion oder werden von päd. Fachkraft ergänzt und unterstützt und noch nicht selbstständig gebraucht.</p>	<p>Das Kind verwendet selbstständig Begriffe, die aus dem Alltag bekannt sind.</p> <p>Verwendung alltagssprachlicher Ausdrücke ohne konkreten Fachbezug.</p>	<p>Das Kind verwendet Fachbegriffe verwendet [sic!], die leicht von Alltagsbegriffen abzuleiten sind, sich aber auf den konkreten Fachkontext beziehen.</p> <p>Komposita, die aus dem Alltagsverständnis heraus bereits als Komposita erlernt werden und bei denen keine Neubildung durch das Kind erforderlich ist.</p> <p>Vereinzelt: Begriffe die fachsprachlich gebraucht werden.</p>	<p>Das Kind verwendet korrekte Fachbegriffe, die einen hohen Abstraktionsgrad erfordern.</p> <p>Differenzierende Nomen, die entweder eine Doppelbedeutung mit einem Alltagsbegriff aufweisen oder die i.d.R. nur fachsprachlich gebraucht werden.</p> <p>Komposita, die in der Regel nicht alltagssprachlich gebraucht werden.</p>
	Ankerbeispiele	<p>„Gib mir des [sic!] <i>Ding</i> da“</p> <p>„Das macht so“</p>	<p>Beim Experiment Schwimmen und Sinken: „Teller“, „Mund“, „Boot“, „Holz“, „Wasser“</p>	<p>„Gewicht“, „Tischbein“, „Holzknopf“,</p> <p>Beim Experiment Schwimmen und Sinken: „Metall“, „Murmel“</p>	<p>„Experiment“, „Druck“, „Dichte“, „Gleichgewicht“, „Fläche“, „Wasserspannung“</p>

Abb. 4: Auszug aus der Ratingskala (RaBi) des *Easi-Science-L*-Projektes (Tietze et al., 2016, Anhang)

Im Gegensatz zum Modell von Bildungssprache des WuW-Projekts (Abb. 3) berücksichtigt das Kategoriensystem der Ratingskala (Abb. 4) nur die sprachliche und nicht die kognitiv-reflexive und entwicklungspsychologische Dimension von Bildungssprache. Auf sprachlicher Ebene werden außerdem die Sprachhandlungen hinzugenommen.

Unter **Sprachhandlungen** verstehen Tietze et al. (2016, S. 6) beispielsweise das Erklären, Begründen, Vermuten, Beschreiben und das Stellen von Informationsfragen. Damit sind bereits die wichtigsten Sprachhandlungen in der Grundschule genannt (ebd., S. 6). Da die genannten Sprachhandlungen unterschiedlich schwierige sprachliche Mittel voraussetzen, kann die Sprachhandlung zur Einschätzung der bildungssprachlichen Kompetenz hinzugezogen werden (ebd., S. 16). Als Beispiel geben die Autorinnen an, dass zur Aufstellung einer Vermutung (Sprachhandlung) die Beherrschung des Konjunktivs (Sprachmittel) notwendig ist (ebd., S. 6). Somit kann – sicherlich mit Ausnahmen – davon ausgegangen werden, dass sich Kinder bildungssprachlich äußern, wenn sie eine Vermutung aufstellen. Neben der Vermutung zählen das Erklären und das Begründen zu den Sprachhandlungen, die mit der Anwendung der Bildungssprache assoziiert werden und somit Niveaustufe 3 der RaBi-Skala repräsentieren (ebd., S. 13).

Zusätzlich zu den bisher beschriebenen möglichen Kriterien zur Beurteilung von Bildungssprache in der Grundschule ist die **Kontextualisierungskompetenz** von Bedeutung. Sie ist eine von drei Kompetenzen innerhalb der Diskurskompetenz (Morek & Heller, 2012, S. 88). Konkret sollen die Lernenden mithilfe der Kontextualisierungskompetenz die sprachlichen Kriterien der Bildungssprache im Kontext korrekt anwenden können (ebd., S. 88-89). Generell lässt sich festhalten, dass Bildungssprache stets im Kontext angewandt, und daher der Kontext bei der Messung von Bildungssprache eine große Rolle einnehmen sollte (ebd., S. 89).

Das vorliegende Kapitel zeigt, auf welche Weise Bildungssprache in Texten von Grundschulkindern identifiziert werden kann. Die **RaBi-Skala** mit den Kategorien *Morphosyntax*, *Lexikon* und *Sprachhandlungen* dient in diesem Forschungsvorhaben als **Grundlage zur Identifikation bildungssprachlicher Wörter und Formulierungen in medial schriftlichen Texten**. Grund hierfür ist der Fokus auf die *sprachliche Dimension* von Bildungssprache. Für eine umfassende Darstellung von Bildungssprache in der Grundschule sei jedoch auf die zusätzliche kognitiv-reflexive und entwicklungspsychologische Dimension von Bildungssprache laut dem WuW-Modell (Abb. 3) hingewiesen. Da der Kontextualisierungskompetenz ebenfalls eine hohe Bedeutung zukommt, **soll Bildungssprache in diesem Forschungsvorhaben qualitativ ermittelt werden**. Im Kontext unangemessen verwendete Bildungssprache wird somit ausgeschlossen.

Im nächsten Kapitel wird herausgestellt, warum sich das naturwissenschaftliche Lernen in besonderem Maße zur Vermittlung von Bildungssprache anbietet.

2.2 Vermittlung der Naturwissenschaften in der Grundschule

Kinder sind neugierig und möchten die Welt verstehen. Den Bezug zur Lebenswelt bietet der Sachunterricht. In diesem Unterrichtsfach der Grundschule sollen Kinder laut der Gesellschaft für Didaktik des Sachunterrichts (GDSU, 2013) wichtige Themen aus ihrer Lebenswelt aus insgesamt fünf Perspektiven kennenlernen: einer sozialwissenschaftlichen, naturwissenschaftlichen, geographischen, historischen und technischen Perspektive. Zu einem Verständnis der Lebenswelt gehört demnach auch ein Verständnis der Naturwissenschaften und damit einhergehend eine naturwissenschaftliche Grundbildung. Diese wird im folgenden Kapitel *Inhalte einer naturwissenschaftlichen Grundbildung* näher beschrieben. Eine wichtige Komponente der naturwissenschaftlichen Grundbildung stellt der naturwissenschaftliche Erkenntnisweg dar. Eine grundschulgerechte Variante des Erkenntnisweges wird im Kapitel *Der Forschungskreislauf* vorgestellt. Das letzte Kapitel *Sprache im naturwissenschaftlichen Sachunterricht* verdeutlicht die Bedeutung der Sprachbildung für das naturwissenschaftliche Lernen, insbesondere für das Erlernen und Durchführen des Forschungskreislaufes, im Grundschulalter.

2.2.1 Inhalte einer naturwissenschaftlichen Grundbildung

Während *Literalität* mit der Beherrschung der konzeptionellen Schriftlichkeit beschrieben werden kann (Kaltenbacher & Klages, 2008, S. 136), bezieht sich der Begriff *Scientific Literacy* auf die Wissenschaft. Er kann mit naturwissenschaftlicher Allgemein- oder naturwissenschaftlicher Grundbildung übersetzt werden (Gräber, 2002, S. 3). Im weiten Sinn schließt der Begriff auch die sozialwissenschaftliche Allgemein- beziehungsweise Grundbildung mit ein (Siarova, Sternadel, & Szönyi, 2019, S. 11). In diesem Kapitel geht es um die naturwissenschaftliche Grundbildung. Die Vorzüge einer naturwissenschaftlichen Grundbildung sind laut Gräber und Nentwig (2002, S. 7, 8) auf individueller und gesellschaftlicher Ebene zu sehen. Beispielhaft beziehen sich Gräber und Nentwig (2002, S. 7, 8) auf Argumente, wie die Notwendigkeit „[...] naturwissenschaftlich und technologisch gebildete[r] Arbeitskräfte [...]“, den Beitrag zur individuellen Entscheidungsfindung in naturwissenschaftlich und technisch geprägten Fragen des Alltags, den kulturellen Beitrag für ein modernes Leben und der Ermöglichung zu naturwissenschaftlich und technisch begründeten Entscheidungsprozessen in einer Demokratie.

Fakt ist, dass die TIMS-Studie von 2019 den Kindern in Deutschland am Ende der Grundschulzeit einen naturwissenschaftlichen Kompetenzwert von durchschnittlich 518 Punkten zuweist (Steffensky, Scholz, Kasper, & Köller, 2020, S. 138). Im internationalen Vergleich liegt Deutschland zwar im Mittelwert-Bereich der EU-Staaten (522 Punkte), doch gerade signifikant unter dem Mittelwert der OECD-Staaten (526 Punkte)

(ebd., S. 138-140). Im Vergleich zu den Ländern mit der höchsten durchschnittlichen Punktzahl – Singapur und der Republik Korea – liegt Deutschland umgerechnet etwa eine Kompetenzstufe zurück (ebd., S. 140). Der internationale Vergleich zeigt die Notwendigkeit zur Verbesserung der naturwissenschaftlichen Kompetenzen, zumal sich die durchschnittliche Punktzahl zwischen der TIMS-Studie im Jahr 2015 (528 Punkte) und der TIMS-Studie im Jahr 2019 (518 Punkte) verringerte (Steffensky, Kleickmann, Kasper, & Köller, 2016, S. 160; Steffensky et al., 2020, S. 138).

Es stellt sich die Frage, was genau die naturwissenschaftliche Grundbildung umfasst und daher gefördert werden sollte. An dieser Stelle gibt es unterschiedliche Modelle. In der TIMS-Studie 2019 werden folgende **Ziele einer *Scientific Literacy*** angegeben, die auch Grundlage zur Erfassung der naturwissenschaftlichen Kompetenzen ist: „*Scientific Literacy* umfasst Wissen über zentrale Begriffe, Konzepte und Theorien (inhaltsbezogenes Wissen), Wissen über naturwissenschaftliche Denk- und Arbeitsweisen (prozessbezogenes Wissen) und epistemisches Wissen sowie motivationale Orientierungen und Einstellungen gegenüber den Naturwissenschaften“ (Steffensky et al., 2020, S. 120-121). Zusammengefasst wird *Scientific Literacy* an dieser Stelle als Zusammenschluss inhaltlicher, prozeduraler und motivationaler Aspekte sowie dem Selbstbild gesehen. Weitere Modelle legen andere Schwerpunkte. So präsentiert Gräber (2002, S. 14) ein Modell zur *Scientific Literacy*, das die Komponenten *Wissen*, *Handeln* und *Werten* umfasst. Während *Wissen* und *Handeln* dem inhalts- und prozessbezogenen Wissen entspricht, wird mit *Werten* eine ethische Kompetenz hinzugefügt. In seinem Punkt *Handeln* nennt Gräber (2002, S. 14) außerdem explizit die Kommunikationskompetenz. In einer Studie für das Europaparlament werden als Hauptpunkte einer wissenschaftlichen Grundbildung folgende fünf Aspekte genannt:

- 1) Grundlegende Literalität
 - 2) Wissenschaftliches Wissen und wissenschaftliche Kompetenzen
 - 3) Kontextuelles wissenschaftliches Verstehen
 - 4) Kritisches Denken und
 - 5) Aktive Teilnahme (an der Gesellschaft)
- (Siarova et al., 2019, S. 15-18).

Im ersten Punkt *Grundlegende Literalität* wird unter anderem die wissenschaftliche Sprache als Ausgangspunkt der wissenschaftlichen Grundbildung beschrieben. Auch das kritische Denken sowie die aktive Teilnahme und das Engagement in der Gesellschaft kommen als neue Schwerpunktsetzung hinzu.

Allen drei Konzeptionen zur *Scientific Literacy* ist gemein, dass es bei einer naturwissenschaftlichen Grundbildung nicht nur um naturwissenschaftliches Wissen geht. Stattdessen wird dem **Handeln** (*prozessbezogenes Wissen, Handeln, wissenschaftliche*

Kompetenzen, aktive Teilnahme) ein ebenso großer Platz eingeräumt. Im naturwissenschaftlichen Kontext hat das Handeln vor allem bei der Planung, Durchführung und Interpretation von eigenen Versuchen einen hohen Stellenwert. Um Versuche schulgerecht durchzuführen und zu dokumentieren, eignet sich der *Forschungskreislauf*, der im folgenden Kapitel vorgestellt wird.

2.2.2 Der Forschungskreislauf

Grundlage von sowohl Experimenten als auch Versuchen⁵ ist der **naturwissenschaftliche Erkenntnisweg**. Es gibt unterschiedliche Heuristiken dieses Erkenntnisweges, die im Bildungssystem vorkommen (Kosler, 2017, S. 2). Zweck dieser Heuristiken ist es, den Kindern die reale naturwissenschaftliche Vorgehensweise didaktisch aufbereitet zu vermitteln (ebd., S. 2). Als Beispiel für eine didaktisch aufbereitete Herangehensweise dient der in Abb. 5 dargestellte **Forschungskreislauf** von Marquardt-Mau (2011).

Die folgenden Erläuterungen zu den acht zentralen Schritten basieren auf einer Beschreibung von Marquardt-Mau (2011, S. 32-36). Voraussetzung für den Forschungskreislauf ist die vorherige Ermittlung der Präkonzepte von Kindern zu einem bestimmten Thema. Daraus lässt sich eine Fragestellung entwickeln (Schritt 1), aus der sich Vermutungen (Schritt 2) und ein mögliches Experiment ergeben. Nach der Durchführung des Experiments, die – wenn möglich – in Teams von vier bis fünf Kindern stattfindet (Schritt 3 und 4), wird das Ergebnis genau beobachtet (Schritt 5). Daraufhin werden alle bisher genannten Schritte, sowie das beobachtete Ergebnis, dokumentiert und letzteres interpretiert (Schritt 6 bis 8). Nach einer daran anschließenden Reflexion können sich neue Fragestellungen ergeben, die gegebenenfalls in einen erneuten Durchlauf des Forschungskreislaufs münden. Zuletzt wird betont, dass der Forschungskreislauf didaktisch aufbereitet ist und somit nicht exakt den realen naturwissenschaftlichen Forschungsprozess widerspiegelt (ebd., S. 32, 36).

⁵ Grygier und Hartinger (2009, S. 13) unterscheiden im Schulkontext zwischen Experimenten, Versuchen, dem Laborieren und Explorieren. Der Versuch grenzt sich von einem Experiment durch die Anweisungen ab. Während bei einem Experiment eine Fragestellung auf kreative Art und Weise beantwortet werden kann, wird bei Versuchen eine Handlungsanweisung vorgegeben (Hartinger, Grygier, Treter, & Ziegler, 2013, S. 5-6). In diesem Forschungsvorhaben werden Versuche durchgeführt und protokolliert.

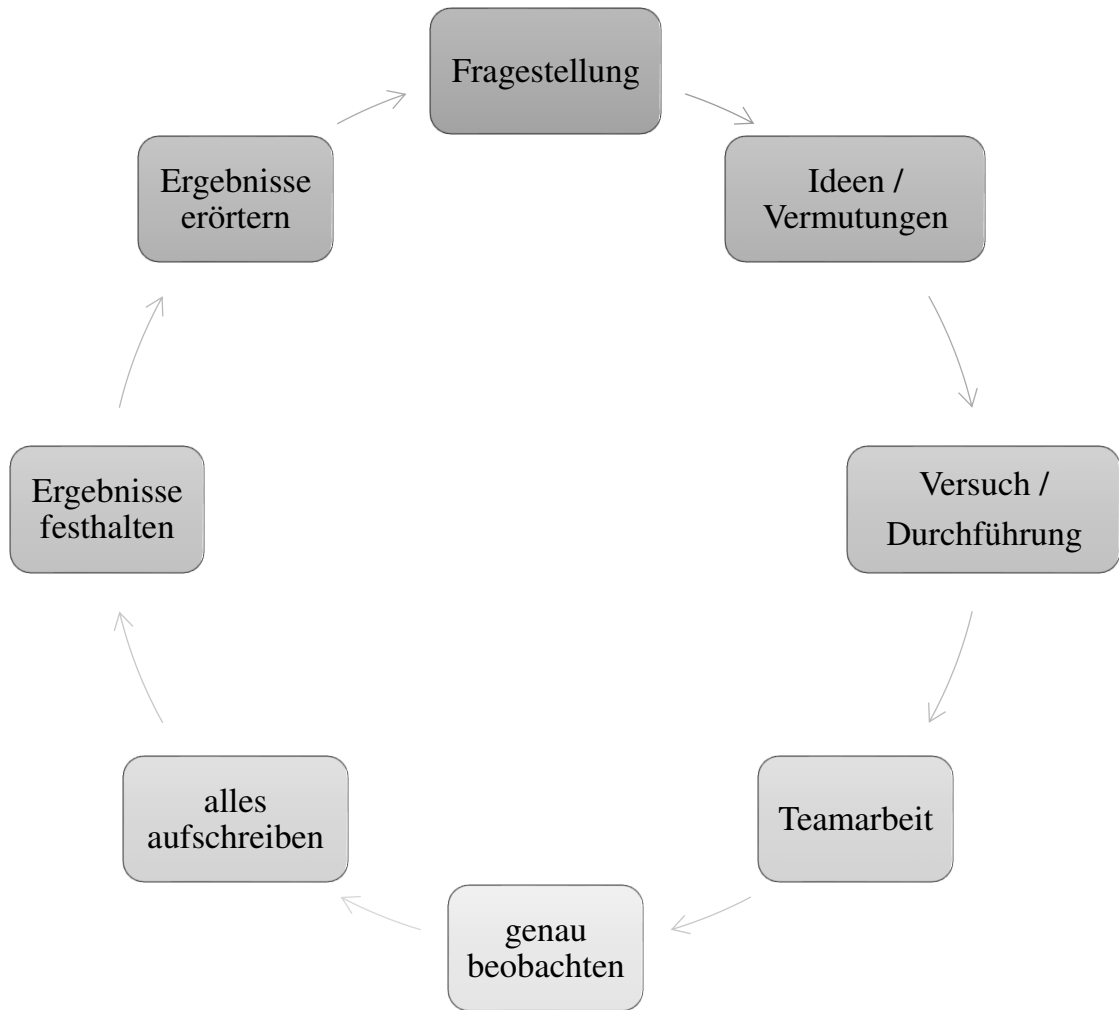


Abb. 5: Der Forschungskreislauf nach Marquardt-Mau (2011, S. 37)

Um Kindern in der Grundschule den Forschungskreislauf nahezulegen, ist eine schriftliche Begleitung eines angeleiteten Versuchs hilfreich. Als **strukturierte Vorlage eines Versuchsprotokolls** werden vereinfacht folgende Überschriften vorgeschlagen, die sich aus dem Forschungskreislauf ableiten lassen:

- 1) Fragestellung
- 2) Vermutung
- 3) Material
- 4) Durchführung
- 5) Beobachtung
- 6) Erklärung

Bei angeleiteten Versuchen ist die Fragestellung (1) vorgegeben. Die Formulierung eigener Vermutungen (2) ist eine vereinfachende Hinführung zur Hypothesenbildung in den Naturwissenschaften. Das Auflisten von Material (3) und die Beschreibung der

Durchführung (4) sind wichtiger Bestandteil der Untersuchungsplanung und können auch bei angeleiteten Versuchen zusammenfassend dargestellt werden. Nach Durchführung des Versuchs folgt die Beschreibung der Beobachtung (5). Schlussendlich werden die Ergebnisse erklärt (6).

Die dargestellten sechs Abschnitte stellen die Grundlage für ein Versuchsprotokoll in diesem Forschungsvorhaben dar. Die getroffene Auswahl der sechs Abschnitte soll keine Allgemeingültigkeit auf die Strukturierung eines Versuchsprotokolls implizieren. Die didaktisch reduzierte Struktur des Versuchsprotokolls erleichtert den Kindern die Dokumentation ihres durchgeführten Versuches und soll sie zur Reflexion animieren, denn: „[...] [E]ntscheidend [ist], dass die Schüler ihren eigenen Erkenntnisprozess reflektieren können. In der Reflexion über ihr eigenes Vorgehen im Forschungsprozess erkennen sie die Zusammenhänge der einzelnen Schritte und die Genese wissenschaftlichen Wissens“ (Grygier, 2005, S. 180). In diesem Fall stellt die versuchsbegleitende Dokumentation der einzelnen Schritte eine Reflexion der wissenschaftlichen Erkenntnisentwicklung dar.

Zusammenfassend wurde im vorherigen Kapitel von den Ergebnissen der TIMS-Studie 2019 (Steffensky et al., 2020) abgeleitet, dass Kinder in deutschen Grundschulen bezogen auf naturwissenschaftliche Kompetenzen stark gefördert werden müssen. Ein großer Teil einer naturwissenschaftlichen Grundbildung stellt das Handeln dar. Das vorliegende Kapitel zeigt eine Möglichkeit, wie das Handeln durch die Dokumentation von Versuchen in Form der sechs zusammengefassten Abschnitte – angelehnt an den Forschungskreislauf von Marquardt-Mau (2011) – zu einer Reflexion und somit zu einem tieferen naturwissenschaftlichen Verständnis des eigenen Handelns führen kann. Im nächsten Kapitel wird der Einsatz von Sprache in diesem Zusammenhang exploriert.

2.2.3 Sprache im naturwissenschaftlichen Sachunterricht

„Sprachbildung umfasst den angemessenen Erwerb der Bildungssprache“ (Leisen, 2023). Das Zusammenspiel von Sprachbildung und naturwissenschaftlichem Lernen ist sowohl im **Lehrplan Sachunterricht** des Ministeriums für Schule und Bildung des Landes Nordrhein-Westfalen (Bildungsministerium NRW, 2021, S. 179) als auch im **Perspektivrahmen Sachunterricht** (GDSU, 2013, S. 11) verankert. Schmitt (2017) stellt die Verknüpfung von Sprachbildung mit dem Kompetenzmodell des Perspektivrahmens wie folgt heraus: „Das große Potenzial des Sachunterrichts liegt in seiner Vielperspektivität, also darin, den Kontext, in dem bildungssprachliche Elemente erlernt werden, möglichst breit zu gestalten, und die Bildungssprache somit in andere Bereiche zu übertragen und flexibel einzusetzen“ (S. 32). Nicht nur die inhaltliche Vielperspektivität ist hier positiv hervorzuheben, sondern auch die prozeduralen Denk-, Arbeits-

und Handlungsweisen (DAH) des Kompetenzmodells. So erwähnt Rau-Patschke (2019, S. 159), dass die DAH geradezu einen sprachlichen Austausch fordern würden.

In einer detaillierten Ausführung zu Wagenscheins Konzeption des **exemplarischen Lehrens** gelangt Loser (2004) zu der Erkenntnis, dass ein exemplarischer Sachunterricht „[...] als Sprachunterricht verstanden werden [kann]“ (S. 14). Im Idealfall eines exemplarischen Lehrens werden Kinder inhaltlich durch *Beispiele* und sprachlich durch die Alltagssprache abgeholt und schrittweise zum wissenschaftlichen Inhalt und zur Bildungssprache geführt (ebd., S. 12). **Inhalt und Sprache führen jeweils vom Konkreten zum Abstrakten, d.h. vom konkreten Inhalt (in der Alltagssprache) zur abstrakten Theorie (in der Bildungssprache)** (Beinbrech, 2013, S. 276).⁶ Die beiden miteinander verwobenen Bereiche (Inhalt und Sprache) bereichern das Allgemeine und führen zu sogenannten *fundamentalen Erfahrungen*, die mit neuen Einsichten über Mensch und Welt einhergehen (Loser, 2004, S. 14).

Die Verknüpfung von Sprache und Inhalt wurde in einer explorativen Studie von Rincke (2010) bezogen auf den Physikunterricht untersucht. In dieser Studie wurden 20 Schüler:innen über 9 Stunden videografiert (ebd., S. 248). In jeder Stunde wurde die Verwendung von Alltags- und Fachsprache analysiert (ebd., S. 248). Im Ergebnis verlief das sprachliche Verhalten der Schüler:innen nicht stringent in Richtung Fachsprache, sondern Alltags- oder Fachsprache wurde je nach Stunde und somit je nach inhaltlichem Lernmaterial mehr oder weniger verwendet (ebd., S. 251-257). Das Ergebnis wurde u.a. mit einem Zurückfallen zur Alltagssprache bei emotional aufgeladenem oder schwierigerem Inhalt interpretiert (ebd., S. 256).⁷

Die beschriebene Verknüpfung von Sprache und Inhalt ist auch auf den naturwissenschaftlichen Erkenntnisweg übertragbar. Ähnlich wie Loser (2004) die Verbindung von Inhalt und Sprache im exemplarischen Sachunterricht illustriert, so beschreiben Quehl und Trapp (2013) diese Verbindung in Bezug auf das Experimentieren im Sachunterricht. Danach gelangen die Kinder schrittweise zu einer inhaltlich und sprachlich nächsthöheren Stufe (ebd., S. 54-55). Während in einem ersten Schritt zunächst neue *Erfahrungen* durch Experimente gemacht und mithilfe neuer Wörter und sprachlicher Strukturen in *Wissen* überführt werden, wird dieses *Wissen* in einem zweiten Schritt in eigenen Worten kontextungebunden wiedergegeben (ebd., S. 54-55). Bei diesen Schritten steigert sich die Komplexität des Wissens und der Sprache (ebd., S. 55). Insgesamt wird beim Experimentieren im naturwissenschaftlichen Sachunterricht in idealer Weise

⁶Ein konkretes Beispiel für eine derartige Verknüpfung im Unterricht bieten Kniffka und Neuer (2008, S. 129) mit ihrem Schaubild zu einer geographischen Unterrichtseinheit zum Umgang mit dem Kompass.

⁷ Rincke (2010, S. 256) bezieht sich bei der Interpretation seiner Studie auf die Theorie der Fossilierung nach Selinker (1972, S. 215, 216).

Sprache, Handeln und Denken verknüpft (vgl. Beinbrech, 2013, S. 277; Kempert et al., 2018, S. 12).

Auch Agel, Beese und Krämer (2012, S. 37) stellen heraus, dass sich besonders die Naturwissenschaften zur Sprachbildung eignen. Dies läge u.a. daran, dass die **Naturwissenschaften wenige Textsorten aufweisen** und diese klar strukturiert seien (ebd., S. 37). **Außerdem würden die Textsorten wiederkehrende sprachliche Muster aufweisen, die – gut gelernt – schnell zu einem beständigen Erfolgserlebnis führten** (ebd., S. 37). Sie resümieren: „Mit dem Erwerb eines begrenzten textsortenspezifischen Sprachrepertoires können die Lerner bereits viele, fachsprachlich angemessene Texte verfassen“ (Agel et al., 2012, S. 37). Eine dieser Textsorten ist das – im vorherigen Kapitel vorgestellte – Versuchsprotokoll, das in wissenschaftlichen Kreisen stark vertreten ist (Sutton & Williamson, 2017, S. 77). Das Verfassen eines Versuchsprotokolls bietet während des Experimentierens die Möglichkeit, die neuen Erfahrungen in Sprache, Handeln und Denken in einer wissenschaftlich angemessenen Form festzuhalten.

In einer vergleichenden Fallstudie zur Verknüpfung von Sprachbildung und Experimentieren im Chemieunterricht ließ sich das besondere Potenzial des Versuchsprotokolls zur Sprachbildung ableiten (Pineker-Fischer, 2017). Es muss berücksichtigt werden, dass es sich in der Studie um die medial mündlichen Äußerungen in einer 10. Jahrgangsstufe handelt (ebd., S. 203), sodass das Ergebnis nicht unreflektiert auf medial schriftliche Lernprodukte in der Grundschule übertragbar ist. Das Ergebnis zeigt: „In Klassen, in denen häufiger anspruchsvolle Sprachhandlungen, wie Erklärungen/ Begründungen, Bewertungen oder Hypothesenbildung verlangt werden, sind die Schülerantworten länger“ (ebd., S. 357). Das Versuchsprotokoll für die Grundschule, das im vorherigen Kapitel beschrieben wurde, beinhaltet bereits zwei der in der Studie als anspruchsvoll beschriebenen Sprachhandlungen: Das *Aufstellen einer Vermutung* (als didaktisch reduzierte Form der Hypothesenbildung) und das *Erklären eines Versuches*. Diese zwei Sprachhandlungen könnten auch in der Grundschule ggf. zu längeren Antworten in Versuchsprotokollen führen.

Zur **konkreten Vermittlung** des Versuchsprotokolls in der Grundschule bietet sich im Unterricht zunächst die Erarbeitung des allgemeinen Aufbaus eines Versuchsprotokolls an (Wildemann & Fornol, 2016, S. 305). Daran anknüpfend ist ein spezifisches Sprachrepertoire notwendig. In Bezug auf Vermutungen sind sprachliche Formen des Möglichen wichtig, wie: „Ich denke, es wird...“ oder „Ich vermute, dass...“ (Kempert et al., 2018, S. 12). Ebenso sind Konditionalsätze, wie „Je..., desto...“ und „Wenn..., dann...“ relevant (ebd., S. 12). Bei der Beschreibung der Materialien kommen neue Adjektive, wie „rau“ oder „porös“ zum Einsatz (ebd., S. 12). Das Beschreiben der Durchführung benötigt Formulierungen, die das Handeln begleiten und somit verinnerlichen (ebd., S. 12). Die abschließende Auswertung fördert Diskurs- und

Argumentationsformen (ebd., S. 12). Als Beispiel könnte hier die Formulierung *Das liegt (vermutlich) daran, dass ...* genannt werden. Die hier genannten Beispiele unterstreichen, dass **das schriftliche Verfassen von Versuchsprotokollen das Sprachregister Bildungssprache voraussetzt und gleichzeitig den Kindern näherbringt**. Werden gewisse Satz- und Textstrukturen vermittelt, hat dies zu Auswirkungen auf das Versuchsprotokoll (bspw. durch die Kenntnis, wie Vermutungen formuliert und Abläufe chronologisch präsentiert werden) und zum anderen auf die allgemeine Befähigung der Kinder, „[...] sich präzise, überzeugend und nachvollziehbar auszudrücken [...]“ (Schroeter-Brauss et al., 2018, S. 13).

Zusammengefasst repräsentiert das Versuchsprotokoll die ideale Textsorte zur Sprachbildung. Erstens: Das Versuchsprotokoll mit den sechs vorgestellten Abschnitten *Fragestellung (1)*, *Vermutung (2)*, *Material (3)*, *Durchführung (4)*, *Beobachtung (5)* und *Erklärung (6)* **stellt verschiedene geforderte Sprachhandlungen (z.B. das Erklären) dar**. Zweitens: Um ein Versuchsprotokoll schreiben zu können, sind **konkrete Sprachstrukturen (z.B. *Wenn ..., dann ...*)** notwendig, die die Kinder auch im Allgemeinen für ihren Ausdruck benötigen.

Wie genau die konkreten Sprachstrukturen didaktisch am besten vermittelt werden, ist Thema des nächsten Kapitels.

2.3 Wortspeicher

Laut Scheuer, Kleffken und Ahlborn-Gockel (2010, S. 248) beinhaltet die Verknüpfung von Sprachbildung⁸ und naturwissenschaftlichem Sachunterricht, dass ein jeweiliges Sachthema bezüglich *Wörter, syntaktischer Mittel* und *Grammatik* analysiert und die möglichen sprachlichen Schwierigkeiten im Unterricht differenziert vermittelt werden. Auf welche Art und Weise die Vermittlung der Wörter und Strukturen geschieht, ist vielfältig. Eine Methode ist die Verwendung von Wortspeichern. **Wortspeicher sind sowohl Wörter als auch Formulierungen, die den Kindern eine sprachliche Hilfestellung bieten (Fuchs, Haberfellner, & Öhlerer, 2014, S. 24).** Das Kapitel *Wortspeicher als Methode der Sprachbildung* verdeutlicht die Vielfältigkeit der sonstigen Vermittlungsmethoden und zeigt auf, warum gerade Wortspeicher im Interesse des Forschungsvorhabens stehen. Für ein Verständnis kognitiver Lernprozesse im Zusammenhang mit der Vermittlung von Wörtern und Strukturen ist das Kapitel *Psycholinguistische Erkenntnisse* dienlich. Darauf basierend lassen sich *Grundzüge der Wortschatzarbeit* ableiten. In dem darauffolgenden Kapitel wird *Scaffolding* als ein konkretes Konzept vorgestellt, das die Arbeit mit Wortspeichern umfasst. Das Kapitel *Forschungsstand zur Wirksamkeit von Scaffolding und Wortspeichern im naturwissenschaftlichen Lernen* rundet die Ausführungen mit bisherigen wissenschaftlichen Ergebnissen ab.

2.3.1 Wortspeicher als Methode der Sprachbildung

Vermutlich könnte kein Schaubild der Vielfältigkeit der Sprachbildung gerecht werden. Um die vielzähligen Methoden ansatzweise zu kategorisieren, können sie in Methoden auf Wort-, Satz- und Textebene unterteilt werden. Abb. 6-8 geben beispielhaft ausgewählte Möglichkeiten zur Sprachbildung an und zeigen eine potenzielle Einordnung in die drei genannten Kategorien. Die eigene Darstellung beruht u.a. auf Methoden, die in Budke und Kuckuck (2017, S. 29) gelistet, sowie in Lehmann, Pilz und Sarich (2013) beschrieben sind. Methoden, die die bildungssprachliche Grammatik fokussieren, sind ausgeblendet. **Es fällt auf, dass die Methode *Wortspeicher / Sprachliche Hilfen* eine übergreifende Möglichkeit ist, um Hilfestellungen sowohl auf Wort-, Satz- als auch auf Textebene anzubieten.** Da ein Versuchsprotokoll Sprache auf Wort-, Satz- und auf Textebene voraussetzt, bietet sich zum Erstellen eines Versuchsprotokolls die Vermittlung von Wörtern und Formulierungen an – folglich: die Vermittlung von Wortspeichern.

⁸ Für eine Definition von *Sprachbildung* siehe S. 25

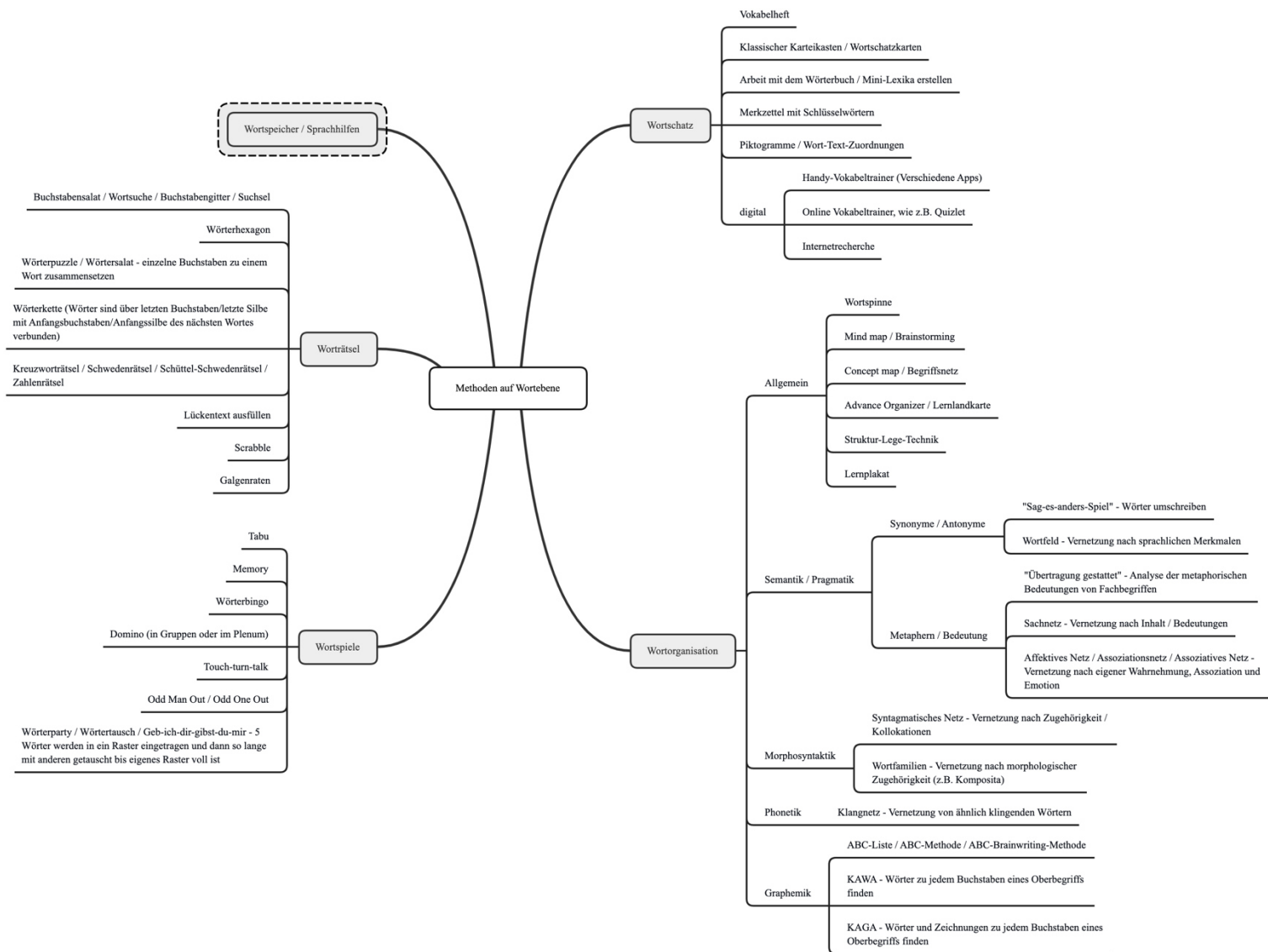


Abb. 6: Ausgewählte Methoden der Sprachbildung auf Wortebene

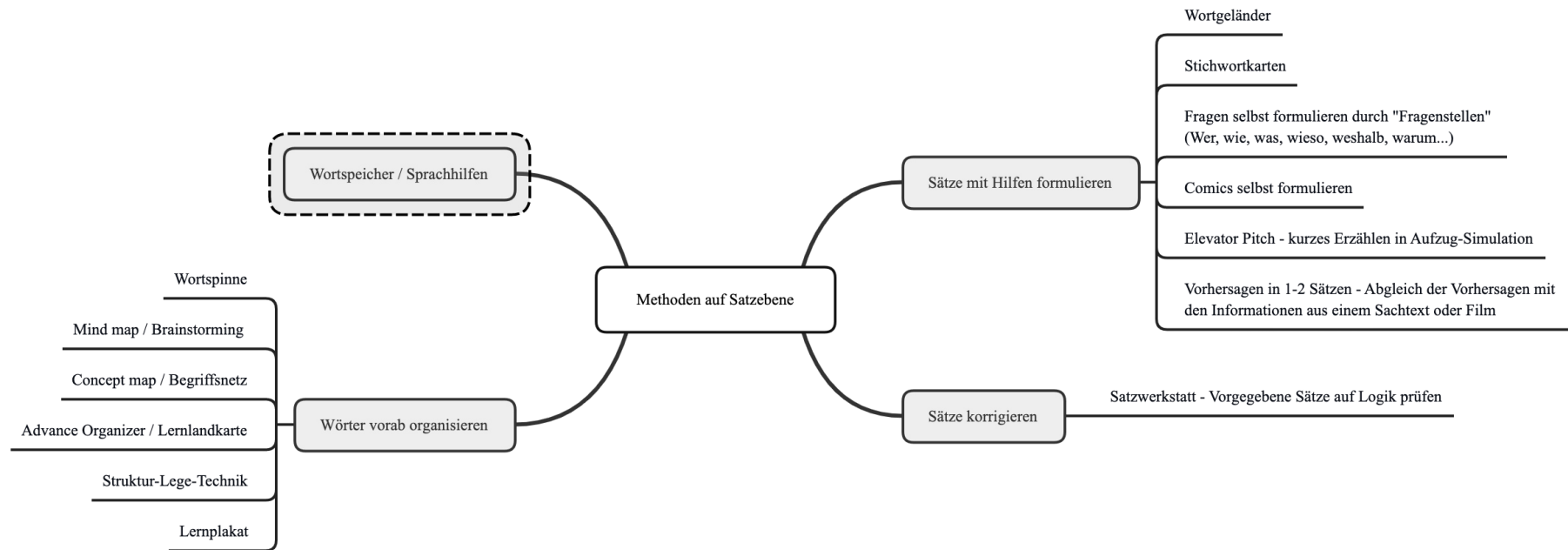


Abb. 7: Ausgewählte Methoden der Sprachbildung auf Satzebene

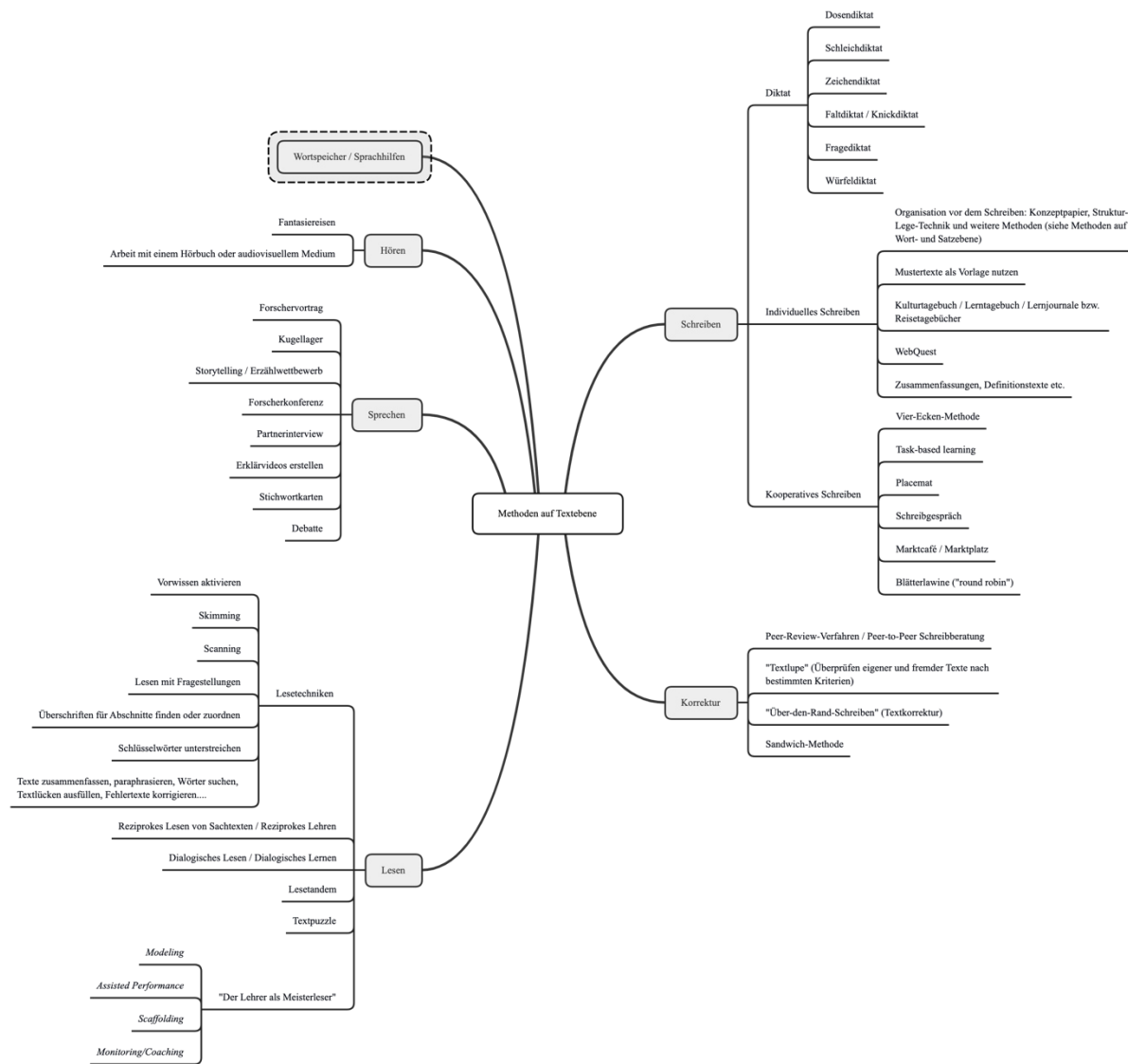


Abb. 8: Ausgewählte Methoden der Sprachbildung auf Textebene

Wortspeicher bzw. sprachliche Hilfen auf den drei Ebenen können wie folgt beschrieben werden:

Auf Wortebene: Einzelne Wörter werden mithilfe eines Wortspeichers vermittelt oder erarbeitet: Zum Beispiel das Substantiv *der Messbecher*.

Auf Satzebene: Einzelne Wörter oder Satzphrasen werden mithilfe eines Wortspeichers vermittelt oder erarbeitet, sodass ein Satz gebildet werden kann: Zum Beispiel das Wort *hinzufügen* oder die Satzphrase *Man kann beobachten, dass ...* .

Auf Textebene: Einzelne satzverbindende Wörter oder Satzphrasen werden mithilfe eines Wortspeichers vermittelt oder erarbeitet, um einen kohärenten Text zu verstehen oder zu verfassen: Zum Beispiel die Wörter *daraufhin, anschließend* oder die Satzphrase *Dies liegt daran, dass ...* .

Wortspeicher werden sehr unterschiedlich ausgelegt. Während einige Lehrkräfte Wortspeicher in Plakatform nutzen und diese im Klassenraum aushängen (Götze, 2015, S. 34; Verboom, 2017, S. 60), nutzen andere Lehrkräfte die Wortspeicher in Form von Wort- und Formulierungs-Listen, die sie den Kindern in Papierform an die Hand geben. Des Weiteren können Wortspeicher von der Lehrkraft vorgegeben, gemeinsam mit Kindern erarbeitet oder ausschließlich von Kindern erstellt werden (ebd., S. 34-35). Allen Auslegungen des Wortspeichers ist gemein, dass sich die Wörter und Formulierungen explizit auf ein bestimmtes Thema beziehen. Dies ist wichtig, da der Wortschatz immer im fachlichen Kontext erarbeitet werden sollte (Wildemann & Fornol, 2016, S. 313). Da Wortspeicher für den Mathematikunterricht in der Grundschule gut erforscht sind, können im Folgenden weitere Kriterien genannt werden, die sich auf Wortspeicher im Sachunterricht übertragen lassen. So sollte ein guter Wortspeicher **die Wörter visualisieren und strukturieren** (anstelle einer einfachen Auflistung), **verschiedene Wortgruppen berücksichtigen** (Substantive von Verben und Satzphrasen trennen), sowie **zusammengehörende Wörter als Satzphrase oder als Satz** angeben (Götze, 2015, S. 35-40, 50). Weiterhin sollte ein Wortspeicher im Laufe einer Unterrichtsreihe **kontinuierlich ergänzt** werden (ebd., S. 35). Insgesamt geben die Wörter und Formulierungen aus den Wortspeichern den Kindern eine sprachliche Hilfestellung und Orientierung. Die Kinder wählen die sprachlichen Mittel eigenständig aus und verändern beziehungsweise ergänzen diese passend zu ihrem Thema (ebd., S. 43). Das Ziel ist die Übernahme der Wörter und Satzphrasen in den eigenen Sprachgebrauch (ebd., S. 43). Wie dies psycholinguistisch betrachtet möglich ist, wird im folgenden Kapitel exploriert.

2.3.2 Psycholinguistische Erkenntnisse

Für dieses Kapitel lassen sich zwei Fragen formulieren, die nacheinander erörtert werden. Beide Fragestellungen sind für das Forschungsvorhaben essenziell, da die didaktischen Grundzüge der Wort- und Sprachstruktur-Vermittlung unter Berücksichtigung der Psycholinguistik leichter nachvollziehbar sind.

- A. Welche Prozesse finden im Gedächtnis statt, wenn Kinder Wörter oder Formulierungen lernen?
- B. Welche Prozesse finden im Gedächtnis statt, wenn Kinder Wörter oder Formulierungen später abrufen?

Zur Erörterung der ersten Frage gibt das Mehrspeichermodell von Atkinson und Shiffrin (1968, S. 93) Aufschluss.⁹ Laut diesem Modell durchlaufen aufgenommene Informationen drei verschiedene Gedächtnisspeicher. Alle Reize werden zunächst durch das **sensorische Register** aufgenommen (Imhof, 2012, S. 51). Werden Wortspeicher genutzt, können Kinder die Wörter und Formulierungen erstmalig als akustische oder visuelle Information aufnehmen. Gleichzeitig mit der Wort-Information nimmt das Kind jegliche weitere Reize auf, die genau in dem Augenblick mit den Sinnesorganen erfasst werden. Aus der Überfülle an Reizen werden relevante Informationen selektiert. Diese gelangen ins **Arbeitsgedächtnis**, das bei Erwachsenen etwa fünf bis neun Einträge zeitgleich aufnehmen kann (ebd., S. 52). Bei Kindern im Alter der Einschulung werden drei bis vier Einträge gespeichert (Schuchardt & Mähler, 2019). Daraus lässt sich für Kinder in der dritten Jahrgangsstufe eine Speicherung von mindestens drei bis maximal neun Einträgen ableiten. Im Fall der Wortspeicher-Vermittlung ist ein Eintrag im Arbeitsgedächtnis beispielsweise ein aufgenommenes Wort. Bezogen auf den Wortspeicher bedeutet dies, dass Kinder – würden weitere Informationen ausgeschlossen, die ein Kind zeitgleich aufnimmt und als relevant einstuft – die genannte Anzahl an Wörtern kurzzeitig, das heißt weniger als 30 Sekunden lang, in seinem Arbeitsgedächtnis aktiv halten kann (Imhof, 2012, S. 52). Da die Kapazität des Arbeitsspeichers unabhängig von der Länge der Einträge ist, lassen sich genauso viele zusammenhängende und somit größere Elemente kurzzeitig speichern (ebd., S. 52). Das Zusammenfassen von Elementen wird als *Chunking* bezeichnet (ebd., S. 52). Formulierungen stellen solche *Chunks* dar. Aktiv verarbeitete Informationen aus dem Arbeitsgedächtnis werden in das **Langzeitgedächtnis** überführt und dort prinzipiell lebenslang gespeichert (ebd., S. 52).

⁹ Das Modell musste seit seiner Veröffentlichung an einigen Stellen aufgrund neuerer Testergebnisse revidiert werden (Hergovich, 2021, S. 65, 68). Da es weiterhin als grundlegendes Modell in der Psychologie Verwendung findet, werden die nicht-angefochtenen Grundzüge in diesem Kapitel beschrieben. Für ein tieferes Verständnis sei an dieser Stelle auf erweiterte Gedächtnismodelle verwiesen.

Im Grunde gelangen die Wörter und Formulierungen in einen genauer definierten Teil des Langzeitgedächtnisses – dem **mentalen Lexikon** (Reber & Schönauer-Schneider, 2016, S. 64). Hier sind Wörter auf vielfältige Weise miteinander vernetzt (ebd., S. 64). Ein einzelner Worteintrag ist mit vielen Informationen bezüglich seines Inhalts (Lemma) und seiner Form (Lexem) verknüpft (Levelt, 1995, S. 18-19). Abb. 9 gibt einen Überblick über die zwei großen Bereiche Lemma und Lexem.

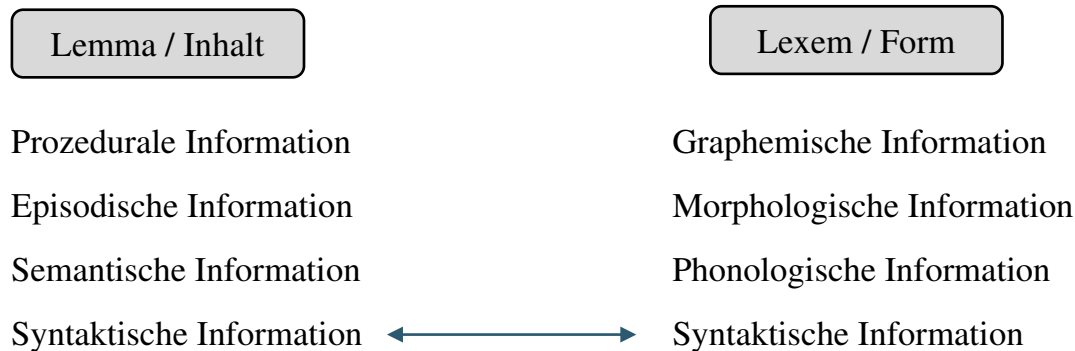


Abb. 9: Bereiche des mentalen Lexikons (verändert nach Reber & Schönauer-Schneider, 2016, S. 65)

Die Bereiche Lemma und Lexem werden im Folgenden anhand des Wortes *Versuchsprotokoll* veranschaulicht.

Abb. 10 zeigt einen möglichen **inhaltlichen Eintrag (Lemma)** zum Begriff *Versuchsprotokoll*.

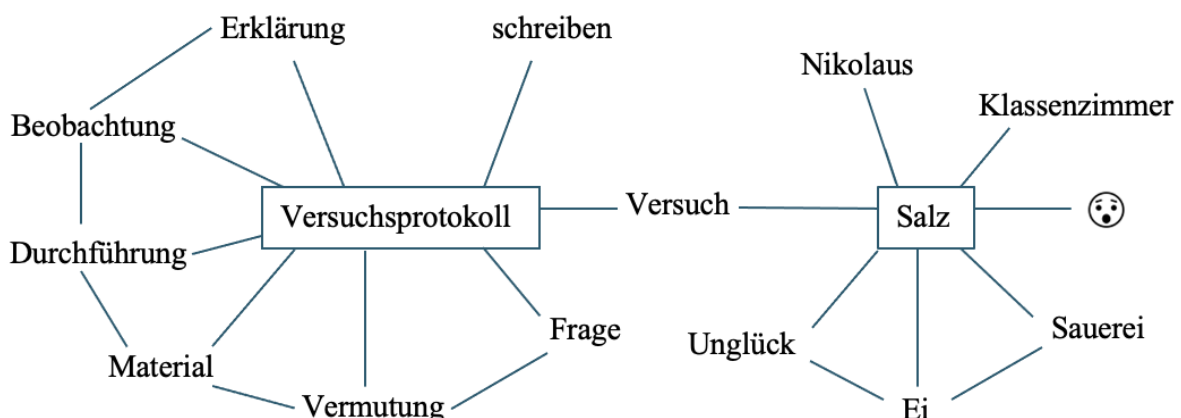


Abb. 10: Modell inhaltlicher Wortnetze (verändert nach Imhof, 2012, S. 53)

Zunächst das Beispiel zur **prozeduralen Information** in Abb. 10: Mit dem Begriff *Versuchsprotokoll* ist in diesem Fall das Verb *schreiben* verknüpft. Ein Kind, das sich das

Wort *Versuchsprotokoll* einprägt, wird sich vermutlich auch die **prozedurale Information** des *Protokollierens* merken und beides – *Versuchsprotokoll* und *schreiben* – gedanklich miteinander verbinden. Das rechts abgebildete Begriffsnetz in Abb. 10 zeigt eine Vernetzung aufgrund von persönlichen und individuellen Assoziationen beziehungsweise Erlebnissen und Erinnerungen. Es handelt sich um eine **episodische Information**. In diesem fiktiven Fall hat ein Kind am *Nikolaustag* im *Klassenzimmer* einen Versuch durchgeführt und musste ein *rohes Ei* vorsichtig in Salzwasser legen. Das Ei ist allerdings geplatzt, und es gab eine riesige *Sauerei*. Das Kind war entsprechend sehr *erschrocken* – hier dargestellt durch den erschrockenen Smiley. Die geschilderte Situation zur Verdeutlichung einer beispielhaften episodischen Information ist in starker Anlehnung an eine ähnlich geschilderte Situation nach Imhof (2012, S. 53-54) beschrieben. Die episodische Information des Erlebten verbindet das Kind fortan mit dem Schreiben eines Versuchsprotokolls. In dem linken Begriffsnetz in Abb. 10 sind weiterhin inhaltlich naheliegende Wörter vernetzt. Da ein Kind eventuell gelernt hat, dass ein Versuchsprotokoll mit den Kategorien *Frage*, *Vermutung*, *Material*, *Durchführung*, *Beobachtung* und *Erklärung* strukturiert wird, tragen diese Begriffe zur Wortbedeutung des Wortes Versuchsprotokoll bei. Es handelt sich um **semantische Informationen**. Die Bedeutung eines Wortes kann Kindern dadurch vermittelt werden, dass ein Wort erklärt oder umschrieben (intensionaler Erwerb) oder mithilfe von Beispielen eingegrenzt wird (extensionaler Erwerb) (vgl. Reber & Schönauer-Schneider, 2016, S. 65). Schlussendlich gibt die **syntaktische Information** Aufschluss darüber, wie das Wort *Versuchsprotokoll* im Satz verwendet wird (ebd., S. 67). Da es sich um ein Nomen handelt, wird das Wort als solches behandelt. Inhaltlich ist dementsprechend folgender Satz möglich: *Ich schreibe ein Versuchsprotokoll zum Thema Salz.*

Durch die Verwendung im Satz gibt die syntaktische Information auch Hinweise auf die **Form des Wortes (Lexem)** (ebd., S. 67). So wird klar, dass das Wort Versuchsprotokoll niemals als Verb benutzt werden kann: *Ich *versuchsprotokolle zum Thema Salz.* Weiterhin zeigt die syntaktische Information, dass es sich um ein Nomen mit sächlichem Artikel handelt (ebd., S. 67). Die syntaktische Information wird folglich als Schnittstelle zwischen Lemma und Lexem gesehen (ebd., S. 67). Weiterhin könnte ein Eintrag zur Form (Lexem) des Wortes *Versuchsprotokoll* durch folgende Informationen komplettiert sein. Das Kind speichert die Information, wie das Wort geschrieben wird (**graphemische Information**; hier: V-e-r-s-u-c-h-s-p-r-o-t-o-k-o-l-l), aus welchen bedeutungstragenden Bausteinen das Wort besteht (**morphologische Information**; hier: *Versuch*, *Protokoll* und gegebenenfalls der Fugenlaut *s*) und es speichert Informationen zur Aussprache des Wortes (**phonologische Information**; hier: /fɛʁˈzuːxspʁotoˈkɔl/) (ebd., S. 64-67).

Die geschilderte Aufnahme von Wörtern in das Gedächtnis hat mehrere Implikationen für die vorliegende Forschungsarbeit. Für das Erlernen neuer oder bekannter Wörter und

Formulierungen kann festgehalten werden, dass Kinder nicht zu viele Wörter in einer Stunde neu erlernen sollten, da diese entsprechend ohne große Verknüpfungen abgelegt werden (Reber & Schönauer-Schneider, 2016, S. 67). Von Vorteil ist es andersherum, **wenige Wörter und Formulierungen** intensiv zu besprechen und somit ein möglichst weitreichendes Verbindungsnetz im mentalen Lexikon aufzubauen. So sind **5-10 neu gelernte Wörter in einer Unterrichtsstunde mit jungen Kindern** bereits angemessen (ebd., S. 70).

Das konkrete Beispiel von Lemma und Lexem anhand des Wortes *Versuchsprotokoll* verdeutlicht, wie vielfältig die Vernetzungen im mentalen Lexikon sind. Wenn wenig Zeit für die intensive Besprechung von Wörtern und Formulierungen in einer Unterrichtsstunde zur Verfügung steht, sollten Kinder zumindest auf die **semantische und phonologische Information** zugreifen können. Denn weiß ein Kind um die Bedeutung und Aussprache eines Wortes oder einer Formulierung, so kann es diese bereits selbstständig aktiv nutzen (ebd., S. 67). **Entsprechend bietet sich als Basis einer Wortschatzarbeit die auditive Vermittlung von Wörtern und Formulierungen an, die noch einmal erklärt oder anhand eines Beispiels erläutert werden.** Die auditive Vermittlung führt dazu, dass Kinder die entsprechende phonologische Information zu dem Wort oder zu der Formulierung im mentalen Lexikon abspeichern, während die Erklärung zum Abspeichern von semantischer Information dient. Steht im Unterricht zusätzlich Zeit zur Verfügung, sind weitere Informationen hilfreich, um die Verknüpfungen im mentalen Lexikon weiter auszubauen. Als Beispiel sei an dieser Stelle die graphemische Information genannt, da sich diese gut vermitteln lässt. **Durch visuelles Aufzeigen von Wörtern und Formulierungen können Kinder die graphemische Information zusätzlich zur phonologischen und semantischen Information im mentalen Lexikon speichern und alle Informationen miteinander verknüpfen.**

Im Folgenden wird die zweite – zu Beginn des Kapitels gelistete – Frage erörtert, d.h. wie ein Kind auf Wörter und Formulierungen im mentalen Lexikon zurückgreifen kann. Hierzu ist ein Modell von Levelt (1995, S. 14) hilfreich. Laut diesem Modell wird davon ausgegangen, dass eine Sprechabsicht zunächst konzeptualisiert, danach formuliert und in einem dritten Schritt artikuliert wird (ebd., S. 14). Für die Formulierung werden aus dem mentalen Lexikon die passenden Wörter für die konzeptuelle Sprechabsicht gefunden (ebd., S. 17). Es werden zuerst Informationen auf Lemma-Ebene und erst im zweiten Schritt Informationen auf Lexem-Ebene abgerufen (ebd., S. 18). Denken Kinder an ein Versuchsprotokoll, so wird das Lemma *Versuchsprotokoll* mit seinen syntaktischen Eigenheiten aktiviert. Es können auch semantisch nahestehende Begriffe, wie hier vielleicht das Wort *Experimentieren*, aktiviert werden. Da das Wort *Versuchsprotokoll* in diesem Beispiel mehr Aktivierungen erhält als das Wort *Experimentieren*, wird letztendlich das Wort *Versuchsprotokoll* ausgewählt (ebd., S. 18). Daraufhin wird das Wort grammatikalisch an die gesamte Sprechabsicht angepasst, sodass die Syntax des Satzes

Sinn ergibt (ebd., S. 17). In einem letzten Schritt werden die Informationen auf Lexem-Ebene abgerufen (ebd., S. 18). In diesem Fall gehören dazu unter anderem die Aussprache /fɛgˈzuːxspʁotoˈkɔl/, deren einzelne phonologische Silben und die dazu gehörenden Artikulationsformen (ebd., S. 18).

Wie das Modell zeigt, optimiert die **vielfältige Vernetzung** von Wörtern auch ihr späteres Abrufen aus dem Langzeitgedächtnis (vgl. Imhof, 2012, S. 53-54; Levelt, 1995, S. 18). Denn je mehr Wege es gibt, die auf die Aktivierung eines Wortes hinweisen, desto eher wird dieses Wort – und kein anderes – ausgewählt. Ein weiterer Vorteil: Ist bereits ein weitreichendes Wortnetz im mentalen Lexikon vorhanden, so können Analogien und Verknüpfungen leichter gebildet und neue Wörter schneller gelernt werden (Kempert et al., 2018, S. 8). Das bedeutet, dass die Arbeitsgedächtnisressourcen für weiterführendes Denken genutzt werden können (ebd., S. 8). Die Vernetzungen umfassen sowohl die beschriebene Lexem- als auch die Lemma-Ebene. Weiterhin ist die Verbindung von alltagssprachlichen als auch bildungssprachlichen oder fachsprachlichen Wörtern im mentalen Lexikon zu empfehlen (Both, Pechstein, & Siehr, 2013, S. 259).

Insgesamt sind die Vernetzungen im mentalen Lexikon als ein **dynamisches System** zu betrachten. Gerade mit Hinblick auf die semantischen Informationen konstatiert Ulrich (2013, S. 318), dass neue Bedeutungen eines Wortes jederzeit aufgenommen, aber auch abgelegt oder jeweils in den Vorder- beziehungsweise in den Hintergrund gerückt werden können. **Dies ist der Grund, warum Wortschatzarbeit selbst dann Sinn ergibt, wenn Kinder die gelernten Wörter und Formulierungen bereits kennen.** In diesem Fall führt die Wortschatzarbeit zwar nicht zwangsläufig zu einer Wortschatzerweiterung. „Sie führt aber [...] zumindest zur Entdeckung und Verstärkung der Verknüpfungen der Lexeme und somit zu ihrer besseren Verfügbarkeit beim Sprachhandeln“ (ebd., S. 316). In dem vorliegenden Forschungsvorhaben stehen diese Verknüpfungen im Fokus. Es geht entsprechend nicht um das Erlernen von neuen Wörtern und Formulierungen, sondern um den Abruf und die Nutzung der Sprachmittel im Kontext von Versuchsprotokollen im naturwissenschaftlichen Lernen durch verstärkte Verknüpfungen im mentalen Lexikon.

Die Erkenntnisse aus diesem Kapitel bilden die Grundlage für wichtige Ansätze der Wortschatzarbeit.

2.3.3 Grundzüge der Wortschatzarbeit

Viele Ansätze der Wortschatzarbeit integrieren die oben beschriebenen Vernetzungen im mentalen Lexikon in ihre normative Betrachtung der Wortinstruktion (vgl. Selimi, 2014, S. 54-56; Vasylyeva & Kurtz, 2015, S. 240). Zunächst werden allgemeine Prinzipien für gute Wortschatzarbeit in den Blick genommen.

Bekannt sind die von Leisen (2010, S. 185) zusammengeführten 9 Leitlinien der Wortschatzarbeit:

Wortschatzarbeit

1. führt neue Begriffe und Sprachstrukturen nicht isoliert ein
2. semantisiert im fachlich relevanten Kontext
3. verwendet neue Begriffe und Sprachstrukturen und grenzt sie in bekannten Wortfeldern ab
4. führt neue Begriffe und Sprachstrukturen über mehrere Stufen sprachlicher Fassungen ein
5. liegt knapp über dem jeweiligen Entwicklungsstand der Lerner
6. führt zu relevanten mündlichen und schriftlichen Äußerungen
7. verbindet sprachliche Unterweisung und interaktives, kommunikatives Handeln
8. vermeidet mechanischen Sprachgebrauch
9. fördert die Sprachbewusstheit.

(ebd., S. 185)

Die Leitlinien 1 und 2 betonen die Notwendigkeit, dass **Wortschatzarbeit niemals alleinsteht**. Dies wird in mehreren Arbeiten zur Wortschatzarbeit von Sprachen im Allgemeinen unterstützt (vgl. Apeltauer, 2008, S. 114; Guadatiello & Speck-Hamdan, 2013, S. 12; Spitzer, 2006, S. 35; Wildemann & Fornol, 2016, S. 313). Wie im vorherigen Kapitel zu den psycholinguistischen Erkenntnissen beschrieben, ist die Vernetzung von Wörtern und Formulierungen zentral. Wird ein Wort isoliert gelernt, so können Informationen zur Form und zum Inhalt dieses Wortes nicht direkt zugeordnet werden. Das Wort bleibt isoliert. Insofern kann das reine Auswendiglernen und Wiedergeben von einzelnen Wörtern oder auch von Definitionen mit dem Ausdruck **leerer Verbalismus** beschrieben werden (Ulrich, 2013, S. 321). Es geht bei diesem Ausdruck nicht nur darum, dass Wörter nicht alleinstehen (s. Leitlinie 1), sondern durch die Kontextualisierung auch inhaltlich verstanden werden (s. Leitlinie 2). Die Vernetzungen des mentalen Lexikons erklären auch, warum in Leitlinie 3 Wörter und Sprachstrukturen in Wortfeldern abgegrenzt werden sollten. Zur Vermittlung der Formulierungen *Ich vermute, dass...* oder *Ich denke, dass...* könnten diese mit dem Wortfeld der Vermutung in Bezug gesetzt werden. Gleichzeitig werden sie von den Ausdrücken *Man kann beobachten, dass...* und *Man kann feststellen, dass...* (Wortfeld: Beobachtung) klar abgegrenzt. Werden die Formulierungen im Kontext eines eigenen Versuchs genutzt, so finden die ersten drei von Leisen aufgestellten Leitlinien Beachtung.

Die Leitlinien 4 und 5 beziehen sich auf den **Lernstand des Kindes**. Im Fall der Formulierungen *Ich vermute, dass...* und *Man kann beobachten, dass...* kann davon ausgegangen werden, dass sie den meisten Kindern in der 3. Klasse bekannt sind. Das Problem ist nicht das Verfügbarkeitsdefizit. Da Kinder die Formulierungen nicht per se

selbstständig anwenden (dies zeigten beispielsweise die Prä-Tests der ersten Pilotstudie in diesem Forschungsvorhaben), muss von einem **Nutzungsdefizit** ausgegangen werden. Das Nutzungsdefizit resultiert u.a. aus der Unkenntnis der Textsorte (hier: Versuchsprotokoll) und der damit verbundenen Sprache (hier: die Formulierungen, die in einem Versuchsprotokoll im Bildungskontext angemessen erscheinen). Die Einführung mittels Vernetzung und somit die automatisierte Nutzung der Formulierungen würden knapp über dem Entwicklungsstand eines Großteils der Kinder in einer dritten Klasse liegen. Nutzen die Kinder die neu entdeckten Formulierungen, wie z.B. *Ich vermute, dass...* in ihren Versuchsprotokollen, so können in jeder Stunde weitere Wörter und Strukturen hinzukommen, die den Schwierigkeitsgrad erhöhen. In diesem Fall wäre die neue und schwierigere Struktur die unpersönliche Formulierung *Man kann beobachten, dass...*. Durch die Anwendung der neu entdeckten Formulierungen in Bezug auf eine inhaltliche Fragestellung, ist die Äußerung der Kinder relevant (Leitlinie 6). Denn die inhaltliche Frage kann hauptsächlich mithilfe der Formulierungen beantwortet werden. Zusammengefasst: Liegen die Formulierungen knapp über dem Entwicklungsstand, werden schrittweise eingeführt und mit Bezug auf den Inhalt abgestimmt, so entspricht die Wortschatzarbeit den Leitlinien 4, 5 und 6.

Die Leitlinien 7, 8 und 9 zielen auf die **bewusste Nutzung der Formulierungen** ab. Dazu ist es wichtig, dass sie nicht nur von der Lehrkraft vorgestellt, sondern auch von den Kindern in einem interaktiven Setting benutzt werden (s. Leitlinie 7). Das Experimentieren und die damit einhergehende mündliche Kommunikation sind als interaktives *Setting* ideal. Auch wenn Kinder die hier beispielhaft vorgestellten Formulierungen *Ich vermute, dass...* und *Man kann beobachten, dass...* nicht explizit in ihre mündliche Kommunikation einbauen, so werden die Formulierungen spätestens bei der Protokollierung des Versuchs relevant. Das interaktive *Setting* des Experimentierens fördert nicht nur die mündliche und schriftliche Kommunikation, sondern dient auch der Vermeidung des mechanischen Sprachgebrauchs (s. Leitlinie 8). Denn die Formulierungen werden in diesem *Setting* stets mit Inhalt gefüllt. Erst der Inhalt – wie die Beschreibung einer Versuchsbeobachtung – machen die Formulierung *lebendig* und geben ihr jeweils neue Bedeutungen. Um die Formulierungen auch langfristig aus dem mentalen Lexikon abzurufen, ist es schlussendlich wichtig, sich über die Anwendung der Formulierungen bewusst zu sein (s. Leitlinie 9). Erst dann sind die letzten drei von Leisen aufgestellten Leitlinien realisiert.

Ergänzt werden können diese Leitlinien durch eigens auf die Grundschule abgestimmte Prinzipien der Wortschatzarbeit nach Reber und Schönauer-Schneider (2016, S. 68-69):

1. Wortschatzarbeit als grundlegendes Unterrichtsprinzip
2. Projekt- und themenorientiertes Arbeiten
3. Berücksichtigung kindlicher Interessen

4. Lernen mit allen Sinnen
 5. Spielerisches und alltagsintegriertes Lernen
 6. Optimierung der Lehrersprache
 7. Einsatz verschiedener Sprachmodalitäten
- (ebd., S. 68-69)

Demnach sollten Wörter und Formulierungen bestenfalls in jeder Unterrichtsstunde thematisiert oder bereits gelernte Wörter vertieft werden (1). Zur Vertiefung von bereits gelernten Wörtern eignet sich besonders ein themenorientiertes Arbeiten. „Ideal sind Themen des Sachunterrichts“ (ebd., S. 68) (2). Das Interesse der Kinder sollte berücksichtigt werden (3) und sie sollten möglichst mit allen Sinnen lernen (4). Auch hier ist das **Experimentieren als positives *Setting*** hervorzuheben, das die Punkte 3 und 4 berücksichtigt. Weiterhin bietet das Experimentieren eine hervorragende Möglichkeit, um neu entdeckte Wörter und Formulierungen in einem natürlichen und alltagsnahen Kontext anzuwenden (5). Die Sprache der Lehrkraft kann die Wortschatzarbeit „[d]urch allgemeine Gestaltungsmerkmale wie Betonung, Sprechtempo, Pausen, Gestik und Mimik [unterstützen] [...]“ (ebd., S. 69) (6).

Wichtig ist auch der letzte Punkt: Die Kinder sollten Wörter und Formulierungen in allen vier Sprachmodalitäten erfassen (7). Zu den vier Sprachmodalitäten zählen die rezeptiven Modalitäten *Verstehen* und *Sprechen*, sowie die produktiven Modalitäten *Lesen* und *Schreiben* (ebd., S. 69). Reber und Schönauer-Schneider (2016, S. 69) listen diese Modalitäten genau in der Reihenfolge: *Verstehen – Sprechen – Lesen – Schreiben*. Diese Reihenfolge erinnert an ein Prinzip aus der Fremdsprachendidaktik, demzufolge es wichtig ist „[...] nur vorher Gehörtes zu sprechen, nur vorher mündlich Behandeltes zu lesen und nur vorher Gelesenes auch zu schreiben“ (Von Ossietzky, o. D.). Konkret werden beispielsweise englische Vokabeln gerne zunächst genannt oder erklärt (*Verstehen*), daraufhin im Klassenverband mündlich wiedergegeben (*Sprechen*), in einem dritten Schritt an die Tafel geschrieben (*Lesen*) und abschließend von den Kindern entweder in das Vokabelheft übertragen oder im schriftlichen Kontext aktiv angewandt (*Schreiben*). Für das vorliegende Forschungsvorhaben ist die Einführung der Wortspeicher an diese Reihenfolge orientiert. Interessant ist die Frage, inwieweit die Arbeit mit Wortspeichern einen Nutzen hat, wenn nur die beiden erstgenannten Sprachmodalitäten und das Schreiben (*Verstehen – Sprechen – Schreiben*) verwendet oder wenn alle Sprachmodalitäten (*Verstehen – Sprechen – Lesen – Schreiben*) durchlaufen werden. Aufschluss darüber soll das vorliegende Forschungsvorhaben geben.

Neben den allgemeinen Prinzipien der Wortschatzarbeit lässt sich das Konzept des *Scaffolding* auf die Arbeit mit Wortspeichern beziehen.

2.3.4 *Scaffolding* - Lernen mit Gerüst

Der Begriff *Scaffolding* bedeutet in der englischen Sprache *Gerüst* und dient somit als Metapher für eine unterstützende Struktur. Das bedeutet, dass ein Kind durch verschiedene Maßnahmen – die wie ein Baugerüst zusammenkommen – unterstützt werden sollte. Der Begriff *Scaffolding* im Zusammenhang des Unterrichtens wurde erstmals von Wood, Bruner und Ross (1976, S. 90) erwähnt. Sie untersuchten die Rolle eines lernenden Kindes zwischen drei und fünf Jahren und einer lehrenden Person hinsichtlich des Problemlösens genauer und erkannten, dass der Prozess des Problemlösens nicht nur selbstständig oder durch die Imitation der lehrenden Person geschieht (ebd., S. 90). Stattdessen ist das Lösen von Problemen oftmals ein **interaktiver Prozess**, bei dem das lernende Kind mithilfe einer lehrenden Person ein Lernziel erreicht, das es alleine nicht erreicht hätte (ebd., S. 90). Möglich ist dies dadurch, dass die lehrende Person besonders schwierige Elemente des Problemlösens kontrolliert und dem lernenden Kind die Elemente vorzeigt, die es bewältigen kann (ebd., S. 90).

Die Interaktion zwischen einem lernenden Kind und einer unterstützenden lehrenden Person basiert auf einer Theorie Vygotskijs, in der es um das Konzept einer **Zone der nächsten Entwicklung** oder auch Zone der proximalen Entwicklung geht (Vygotsky, 1978). Vorausgesetzt wird: Zwei Kinder sind auf dem gleichen mentalen Entwicklungsstand (ebd., S. 86). Sie sind beide dazu fähig, Probleme bis zu dem Level eines Achtjährigen zu lösen (ebd., S. 86). Mit Unterstützung durch eine weitere Person löst das eine Kind Probleme bis zu dem Level eines neunjährigen Kindes (ebd., S. 86). Das zweite Kind löst mit Unterstützung Probleme bis zu dem Level eines Zwölfjährigen (ebd., S. 86). Die Differenz zwischen acht und neunjährigem-Level sowie die Differenz zwischen acht- und zwölfjährigem Level bezeichnet Vygotsky als Zone der nächsten Entwicklung (ebd., S. 86). Er definiert: „*It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers* [Hervorhebung im Original]” (Vygotsky, 1978, S. 86). Übersetzt handelt es sich um die Distanz zwischen dem aktuellen Entwicklungsstand und einem potenziell möglichen Entwicklungsstand, die mithilfe der Unterstützung von Erwachsenen oder fähigeren Mitschüler:innen überwunden wird. Metaphorisch bezeichnet Vygotsky diese Zone als Blüte, die zu der Frucht, d.h. zu dem angestrebten Entwicklungsstand des eigenständigen Problemlösens, hinführt (ebd., S. 86). Laut Vygotskys Vorstellung ist nicht nur die Kenntnis des aktuellen mentalen Entwicklungsstandes, sondern auch die Kenntnis der Zone der nächsten Entwicklung notwendig (ebd., S. 87). Nur so kann unterschieden werden, ob ein Kind, das sich aktuell auf der mentalen Stufe eines Achtjährigen befindet, mit Unterstützung auch Probleme eines Neunjährigen oder eines Zwölfjährigen lösen kann (ebd., S. 87). In anderen Worten: **Der Unterschied**

zwischen den Kindern in dem oben genannten Beispiel wird nicht durch ihren aktuellen Entwicklungsstand, sondern durch die Kenntnis ihrer jeweiligen Zone der nächsten Entwicklung deutlich (ebd., S. 87).

Das oben beschriebene *Scaffolding* bezieht sich genau auf diese Zone der nächsten Entwicklung. Konkret sieht die unterstützende Struktur **im Unterrichtskontext** so aus, dass eine Lehrperson ihre Hilfe der jeweiligen Zone der nächsten Entwicklung von einzelnen Schüler:innen oder von Schüler:innengruppen anpasst (*contingency*) (van de Pol, Volman, & Beishuizen, 2010, S. 274). Dabei ist eine gute Diagnostik Voraussetzung für die lernstandsbezogene Hilfe (ebd., S. 274). In einem zweiten Schritt wird das helfende Gerüst sukzessive abgebaut (*fading*) (ebd., S. 275). Je nach Kompetenz und Entwicklung der Schüler:innen kann das Gerüst langsamer oder schneller zurückgenommen werden (ebd., S. 275). Durch diesen Schritt des Zurücknehmens von Seiten der Lehrperson wird den Schüler:innen automatisch mehr Selbstständigkeit zugesprochen (*transfer of responsibility*) (ebd., S. 275). Schlussendlich stellt die Idee des *Scaffolding* im Unterrichtskontext metaphorisch zwei Zahnräder dar, die sich gegenseitig bedingen. Wird an dem Zahnrad der Unterstützung in Richtung einer Erhöhung gedreht, so ist das Zahnrad der Eigenständigkeit der Schüler:innen auf einen geringen Stand eingestellt. Umgekehrt erhöht sich die Eigenständigkeit der Schüler:innen durch Abnahme des unterstützenden Gerüsts. Grundlage beider Zahnräder ist die wichtige Interaktion zwischen Lehrperson und Kind, wobei erstere ihr Unterstützungsangebot auf den diagnostizierten Lernstand des Kindes abstimmt.

Der **Vorteil** dieser Vorgehensweise liegt auf der Hand: Ein Kind lernt mithilfe des unterstützenden Gerüsts vor allem WIE es denken muss und nicht nur WAS es denken soll (Gibbons, 2014, S. 14). *Scaffolding* ist nicht zu verwechseln mit einer reinen Hilfestellung. Hammond (2001, S. 20) gibt hierzu das imaginäre Beispiel eines Kindes, das sich unsicher über die Rechtschreibung eines Wortes ist. Bei einer simplen Hilfe würde eine Lehrperson dem Kind sagen, wie das Wort geschrieben wird (ebd., S. 20). Anders bei dem Konzept des *Scaffolding*. Hier würde die Lehrperson das Kind beispielsweise dazu ermuntern, sich die Klänge des Wortes und die mögliche schriftliche Repräsentation der Laute ins Bewusstsein zu rufen (ebd., S. 20). Diese auffordernden Worte der Lehrperson führen das Kind gedanklich hin zur richtigen Schreibweise, sodass das Kind bei einem nächsten Wort die gelernte Vorgehensweise anwenden und selbstständig eine korrekte Rechtschreibung finden kann (ebd., S. 20). Kurzum: Wird ein Kind in seiner Zone der nächsten Entwicklung unterstützt, so lernt es die prozedurale Herangehensweise an ein Problem und kann diese später – auch ohne Unterstützung – auf andere Probleme oder auf andere Aufgaben übertragen (Gibbons, 2014, S. 14).

Scaffolding – und die damit einhergehenden Vorzüge – werden von Hammond und Gibbons (2005) auf den Zweitspracherwerb von Schüler:innen in Australien übertragen.

Dabei geht es nicht nur um die englische Sprache, sondern besonders um das englische akademische Sprachregister (Michell & Sharpe, 2005, S. 35) – dem Pendant zur deutschen Bildungssprache. Da Sprache und Inhalt stark miteinander verbunden sind, liegt der Fokus allerdings nicht nur auf der **sprachlichen Unterstützung**, sondern er schließt die **inhaltliche Bewältigung von Aufgaben** mit ein (ebd., S. 32). Dies wird auch in dem Titel des Werkes *Scaffolding Language, Scaffolding Learning* von Gibbons (2014) deutlich. Das Ziel ist Folgendes: Anstatt curriculare Vorgaben für sprachschwächere Schüler:innen zu vereinfachen (*dumbing-down*), sollen sprachschwächere Schüler:innen mittels *Scaffolding* zum Erreichen der curricularen Vorgaben hingeführt werden (*supporting-up*) (Hammond & Gibbons, 2005, S. 6).

Wie das **Scaffolding in Bezug auf Sprachenlernen** genau vonstatten geht, ist in einem empirisch hergeleiteten und empirisch überprüften Modell zusammengefasst (ebd., S. 7). Demnach findet *Scaffolding* sowohl auf einer Makro-Ebene als auch auf Mikro-Ebene statt (ebd., S. 11). Die Makro-Ebene beschreibt die geplanten Unterstützungsmaßnahmen bevor die Lehrperson in den Unterricht geht (ebd., S. 12). Die Mikro-Ebene beschreibt dahingehend die spontanen Unterstützungsmaßnahmen im Unterricht selbst (ebd., S. 12). Unter Makro-*Scaffolding* zählen laut dem Modell die folgenden Aspekte:

- 1) **Lernstandsanalyse:** Die Lehrperson stellt das inhaltliche und sprachliche Niveau der Kinder fest. Darauf aufbauend passt die Lehrperson ihren Unterricht an das inhaltliche und sprachliche Vorwissen an. So erinnert sie die Kinder beispielsweise an vorangegangene Erkenntnisse und knüpft sprachlich zunächst an die Alltagssprache der Kinder an. Dabei sind stets die Lernziele präsent (ebd., S. 13-14).
- 2) **Auswahl der Aufgaben:** Die Lehrperson macht den Kindern diese Lernziele transparent. Weiterhin wird den Kindern klar, inwieweit die einzelnen Aufgaben zu den großen Lernzielen hinführen. Dadurch verlieren die Kinder nicht die Orientierung und haben immer die Lernziele und den Weg zu den Lernzielen im Blick (ebd., S. 14-15).
- 3) **Reihenfolge der Aufgaben:** Die Lehrperson stellt sicher, dass die Aufgaben so aufeinander aufbauen, dass das Lernergebnis der ersten Aufgabe die Basis für das Lernergebnis der zweiten Aufgabe ist. Aufeinander aufbauende Aufgaben führen zu deutlich bedeutsameren finalen Lernergebnissen als einzelne unabhängige Aufgaben (ebd., S. 15).
- 4) **Sozialformen:** Die Lehrperson entscheidet sich für bestimmte Sozialformen, die zu den jeweiligen Lernzielen passen. Es geht dementsprechend nicht nur um eine willkürliche Auswahl von Sozialformen, sondern um eine bewusste Entscheidung, mit welcher Sozialform ein Lernziel am besten erreicht werden kann (ebd., S. 15-16).
- 5) **Darstellungsformen:** Die Lehrperson nutzt unterschiedliche Darstellungsformen, um den Kindern ein bestimmtes Konzept näherzubringen. So kann die Lehrperson beispielsweise eine durchgeführte Handlung gleichzeitig mit mündlicher Sprache

begleiten, auf die geschriebenen Notizen der Kinder zurückgreifen oder selbst einen Text an die Tafel schreiben. Als Beispiel für weitere Darstellungsformen werden hier Wandkarten, Grafiken, Fotos, Diagramme und Bilder genannt. Sie alle tragen dazu bei, dass den Kindern ähnliche Informationen auf unterschiedlichen Wegen bereitgestellt werden (ebd., S. 16-17).

- 6) **Vermittelnde Texte oder Erzeugnisse:** Die Lehrperson setzt bestimmte Texte oder Erzeugnisse ein, die als Anker beziehungsweise als Bezugspunkte dienen und innerhalb einer Unterrichtssequenz öfters aufgegriffen werden. Ein Beispiel hierfür wäre ein Reflexionsbogen, den Kinder ausfüllen und auf den die Kinder während einer Gruppendiskussion und einer späteren individuellen Erarbeitung stets zurückgreifen (ebd., S. 17-18). Im naturwissenschaftlichen Kontext könnte laut Hammond und Gibbons (2005, S. 17-18) ein Versuch im Mittelpunkt stehen, der als Anker zur Verinnerlichung sprachlich formulierter wissenschaftlicher Konzepte dient. Im vorliegenden Forschungsvorhaben dient das Versuchsprotokoll – anstelle eines einzelnen Versuches – als sprachlicher Anker, auf den in jeder Stunde zurückgegriffen wird.
- 7) **Metalinguistisches und metakognitives Bewusstsein:** Die Lehrperson achtet nicht nur auf die Erarbeitung der Lernziele, sondern gleichzeitig auf die damit einhergehenden sprachlichen Grundlagen. Diese sprachlichen Grundlagen werden sowohl implizit im Kontext der Aufgaben als auch in regelmäßigen Abständen explizit dargestellt (ebd., S. 18). Anlehnend an ein Konzept aus der Mediendidaktik, das zwischen *Lernen mit Medien* und *Lernen über Medien* unterscheidet (Gervé & Peschel, 2013, S. 61-63), könnte an dieser Stelle von ***Lernen mit Sprache*** und ***Lernen über Sprache*** gesprochen werden. Beides ist bedeutsam und gewünscht. Zu dem *Lernen über Sprache* zählt auch, dass die Lehrperson die Aufmerksamkeit der Kinder auf bestimmte Wörter oder grammatikalische Strukturen lenkt. Die systematische Vermittlung von Sprache erfordert eine sogenannte Metasprache – eine Sprache, um über Sprache zu reden. Ist den Kindern beispielsweise der Unterschied zwischen Alltagssprache und Bildungssprache bewusst, so entwickelt sich neben dem neu erworbenen metalinguistischen Bewusstsein auch ein neues metakognitives Bewusstsein (Hammond & Gibbons, 2005, S. 18-19).

Bei Betrachtung des gesamten Modells ist das geplante **Makro-Scaffolding die Grundlage** für das interaktionale Mikro-Scaffolding im Unterricht, sodass Lehrkräfte und Schüler:innen innerhalb der Zone der nächsten Entwicklung lernen (ebd., S. 20). Hammond und Gibbons (2005, S. 20) formulieren: „*Thus we see the designed-in level of Scaffolding [= Makro-Scaffolding] as enabling the interactional level [= Mikro-Scaffolding], which, in turn, enables teachers and students to work within the ZPD [Zone of Proximal Development] [Hervorhebung d. Verf.]*.”

Die weiteren Ausführungen von Gibbons (2014, S. 18) zeigen, dass *Scaffolding* explizit den Einsatz von Wortspeichern umfasst. Sie gibt hierzu das Beispiel einer

Unterrichtsstunde, in der Kinder einen argumentativen Text schreiben sollen. *Scaffolding* findet in diesem Beispiel durch das Bereitstellen von Verbindungswörtern, wie *zuerst* oder *auf der anderen Seite* statt (ebd., S. 18). Anstelle von einzelnen Verbindungswörtern kann die Lehrperson den Kindern auch eine Liste mit Wörtern oder Formulierungen an die Hand geben (ebd., S. 18). Der Gebrauch von Wortspeichern ist somit stark in das Konzept des *Scaffolding* integriert. Dies wird auch von Götze (2018, S. 104) unterstrichen, die den Einsatz von Wortspeichern explizit dem Makro-*Scaffolding* zuordnet. Sie schreibt: „Typische Makro-Scaffolding-Elemente, die sich in unterrichtspraktischen Erprobungen und kleinen Studien als hilfreich herausgestellt haben, sind z.B. Wortspeicher [...] oder auch mit Sprachmitteln und schriftlichen Sprachvorbildern aufbereitete Arbeitsblätter [...]“ (ebd., S. 104). Es bleibt festzuhalten, dass Wortspeicher als Teil des *Scaffolding* einen hohen Stellenwert einnehmen.

Für das Forschungsvorhaben ist von Bedeutung, dass eine Lehrkraft laut dem *Scaffolding*-Modell von Hammond und Gibbons (2005) **unterschiedliche Darstellungsformen** zur Vermittlung von Sprache einsetzen sollte (Aspekt 5). Da das Konzept *Scaffolding* den Einsatz von Wortspeichern einschließt, sollten Wortspeicher entsprechend unterschiedlich dargestellt werden. Im Unterrichtsalltag ist das Ziel eine generell abwechslungsreiche Darstellungsart. Im Rahmen dieses Forschungsvorhabens sollen zwei Darstellungsformen untersucht werden: die rein auditive Vermittlungsform durch mündliche Nennung der sprachlichen Hilfen und die visuelle Vermittlungsform in Form eines Wortspeicher-Plakats. Der Begriff *Darstellungsform* ist mit dem Begriff *Sprachmodalität* aus dem letzten Kapitel verknüpft. Während *Darstellungsform* die unterschiedliche Darbietung der Wortspeicher aus der Perspektive der Lehrkraft bezeichnet, sind die Sprachmodalitäten (*Verstehen – Sprechen – Lesen – Schreiben*) die entsprechende Bezeichnung für die Aktion aus Sicht der Schüler:innen. Unter Rückbezug auf die *psycholinguistischen Erkenntnisse* wird mit der auditiven Vermittlungsform die Basis einer Wortschatzarbeit (phonologische und semantische Information) geprüft, während mit der visuellen Vermittlungsform (zusätzliche graphemische Information) eine weiterführende Darstellungsform in den Blick genommen wird.

Einige weitere Aspekte des *Scaffolding* können im Rahmen des Forschungsvorhabens nicht umgesetzt werden. So ist die notwendige Diagnose der *Zone der nächsten Entwicklung* bezogen auf jedes individuelle Kind aus zeitökonomischen Gründen nicht zu leisten. Auch das Zurücknehmen des helfenden Sprachgerüsts (*fading*) ist grundsätzlich langfristig angelegt und innerhalb einer Unterrichtssequenz kaum realisierbar.

Wie *Scaffolding* und Wortspeicher in der bisherigen Forschung eingesetzt wurden und welchen Nutzen ihr Einsatz erbrachte, wird im folgenden Kapitel deutlich.

2.3.5 Forschungsstand zur Wirksamkeit von *Scaffolding* und Wortspeichern im naturwissenschaftlichen Lernen

Empirisch gilt es als wichtiges Forschungsfeld, die **allgemeinen *Scaffolding* Ansätze an spezifischen Inhalten und fachspezifisch zu konkretisieren** (Kempert et al., 2018, S. 15). Gewünscht sind Ansätze, die von oben nach unten reichen (*top-down approach*) (van de Pol et al., 2010, S. 278). Ein Beispiel für einen *top-down approach* und inhaltliche Konkretisierung bezüglich des Sachunterrichts findet sich in einer Arbeit von Merz (2015). Er entwickelte eine sprachensible Unterrichtskonzeption zum Thema Wärmeisolation in der vierten Klasse (ebd., S. 164). Hierbei experimentieren die Kinder mithilfe eines Forscherbuches an verschiedenen Experimentierstationen (ebd., S. 164). Im Forscherbuch stehen den Kindern nicht nur die Experimentieranleitungen, sondern auch lexikalische Hilfsmittel und ein Glossar „[...] zum Nachschlagen wichtiger Fachbegriffe [...]“ zur Verfügung (ebd., S. 167). Die gesamte Unterrichtskonzeption basiert auf dem *Scaffolding*- und dem *Focus-on-Meaning*-Ansatz (ebd., S. 161-163). Allerdings liegt der Fokus der Arbeit auf der Entwicklung und Durchführung der Unterrichtskonzeption. Das Ergebnis der Unterrichtskonzeption wurde lediglich aus eigener Betrachtung beschrieben. Zusammenfassend schreibt Merz (2015, S. 174): „Der Unterricht kann als erfolgreich betrachtet werden. So nahmen viele Kinder das Angebot des Glossars und der lexikalischen Hilfsmittel wahr. Es gelang den Schüler/-innen ihre Gedanken im Forscherbuch adäquat zu verschriftlichen und zu präsentieren.“ Interessant wäre die Frage, ob die Kinder auch ohne Zusatz von lexikalischen Hilfsmitteln ihre Gedanken adäquat hätten verschriftlichen können. Weiterhin muss beachtet werden, dass die ausführlich entwickelte Unterrichtskonzeption lediglich innerhalb eines einzelnen Tages im Rahmen von fünf bis sechs Zeitstunden umgesetzt wurde. Von daher werden im Folgenden größer angelegte Studien mit *top-down approach* und inhaltlicher Konkretisierung aufgeführt.

Hierzu zählt eine Studie, die an einer Gesamtschule in Duisburg als Teil des Modellprojekts *Deutsch als Zweitsprache in allen Fächern* (proDaZ) stattfand (Agel et al., 2012). In dieser Studie geht es konkret um den Forscherkreislauf.¹⁰ In sogenannten **Sprachforscherkursen** wurden ausgewählte mehrsprachige Kinder aus der fünften Klasse auf das Schreiben und Präsentieren von Versuchsprotokollen vorbereitet (ebd., S. 37). Dabei basierten die Kurse auf dem Prinzip des *Scaffolding* und waren handlungsorientiert angelegt (ebd., S. 37). Als Gerüst konnten die Kinder auf Sprachmuster zurückgreifen, die ihnen in den Sprachforscherkursen gezielt an die Hand gegeben wurden (ebd., S. 37). Mithilfe der Sprachmuster und mithilfe von Worterklärungen lernten die Kinder eigene kleine Versuche zu protokollieren (ebd., S. 37). Im Vergleich zu einer Kontrollgruppe

¹⁰ vgl. Kapitel 2.3.4

von mehrsprachigen Kindern, die den 25-stündigen Sprachforscherkurs über 6 Monate lang nicht besuchten, stach die Interventionsgruppe positiv hervor (ebd., S. 37, 41). Bei ihnen wurde die fachliche Sprache bezüglich der geförderten Aspekte: Textsorte, Textaufbau, Wortschatz und Syntax, „[...] signifikant häufiger, angemessener und variantenreicher [angewandt]“ (ebd., S. 41). Die Studie zeigt, dass gerade die explizite fachsprachliche Einführung zur sprachlichen Verbesserung führt (ebd., S. 41). Außerdem konnten die Kinder der Interventionsgruppe häufiger die Protokoll-Bereiche *Durchführung*, *Beobachtung* und *Auswertung* inhaltlich voneinander trennen (ebd., S. 42). Dies lässt darauf schließen, dass die Sprachforscherkurse auch einen wesentlichen Beitrag zu einem wissenschaftlichen Verständnis des Forscherkreislaufes leisten (ebd., S. 42). Die Kenntnis, dass ein additiver Sprachforscherkurs den Kindern sowohl sprachlich als auch methodisch den Forscherkreislauf näherbringen kann, ist für das vorliegende Forschungsvorhaben von Bedeutung. So könnte Sprachbildung ggf. auch im regulären Unterricht bezüglich schriftlicher Versuchsprotokolle einen sprachlichen und inhaltlichen Nutzen haben. In Abgrenzung zu den Sprachforscherkursen wird der Faktor Mehrsprachigkeit in diesem Forschungsvorhaben nicht weiter untersucht. Grund hierfür ist im Fokus auf Sprachbildung zu sehen: Da Kinder mit unterschiedlichen Sprachereferenzen bezüglich Bildungssprache in die Schule kommen (Gogolin & Duarte, 2016, S. 479), ist die Vermittlung von Bildungssprache für alle Kinder notwendig. Insgesamt liefert die Studie zu den Sprachforscherkursen ein Indiz dafür, dass **Sprachförderung und Experimentieren im Sachunterricht gut vereinbar sind.**

Die Vereinbarkeit von Sprachförderung und Experimentieren ist auch das Ergebnis einer explorativen Studie, die in der ersten Jahrgangsstufe von zwei Sprachförderschulen durchgeführt wurde (Gottwald, 2016, S. 95). In dieser Studie bestand das Treatment aus dem Experimentieren mit Fokus auf der Nutzung von Sprechanschlüssen (ebd., S. 93-94). Im Prä-Post-Design wurden der Wortschatz getestet und die naturwissenschaftlichen Kenntnisse mithilfe einer Befragung erhoben (ebd., S. 98). **Im Ergebnis wirkte sich das Experimentieren positiv auf die Sprache der Kinder aus und naturwissenschaftliche Kenntnisse wurden aufgebaut** (ebd., S. 151-152). In dem Studiendesign kamen keine Wortspeicher zum Einsatz und die Ergebnisse wurden nicht mit einer Kontrollgruppe verglichen. Das Ergebnis wird dennoch angeführt, um die prinzipielle Vereinbarkeit von Sprachförderung und Experimentieren zu unterstreichen. Das vorliegende Forschungsprojekt grenzt sich u.a. durch den Fokus auf Sprachbildung (für alle Kinder) anstelle einer spezifischen Sprachförderung ab.

Im Rahmen des Projekts *Promotion of Migrants in Science Education* (PROMISE) wurde eine explorative Studie durchgeführt, die die **Arbeit mit lexikalischen Hilfsmitteln** in den Fokus rückte (Tajmel, 2017b, S. 367). Anhand von 13 Teilnehmenden der Klasse 7 untersuchte Tajmel (2017b, S. 321) die Auswirkung von lexikalischen Hilfsmitteln auf die Sprachhandlungsfähigkeit der Schüler:innen, die in Deutschland geboren

waren und mehrsprachig mit Türkisch und Deutsch aufwuchsen. In Abgrenzung zum vorliegenden Forschungsvorhaben wurde den Schüler:innen ein Versuch vorgeführt, den sie nach einer Beobachtungsphase auf einem Arbeitsblatt *ohne* lexikalische Hilfsmittel oder *mit* lexikalischen Hilfsmitteln beschreiben sollten (ebd., S. 322). Jede:r Lernende schrieb zwei Texte: Zuerst einen Text *ohne* lexikalische Hilfsmittel und im Anschluss einen Text *mit* Hilfen (ebd., S. 322). Im zweiten Fall standen den Schüler:innen die lexikalischen Mittel aufgrund des erwähnten Forschungsdesigns *während des Schreibprozesses* zur Verfügung. Unter den fünf Fragestellungen der Studie bezog sich eine Frage darauf, ob die angebotenen lexikalischen Mittel tatsächlich genutzt wurden (ebd., S. 321). Weiterhin war das Ziel der Untersuchung, ob die Texte durch Verwendung der Hilfsmittel sprachlich und fachlich besser werden (ebd., S. 322). Das vorliegende Forschungsvorhaben ist an diese Fragestellungen angelehnt.

Tajmel (2017b, S. 345) konnte zeigen, **dass die Schüler:innen tatsächlich die lexikalischen Hilfsmittel in ihren Texten nutzten** (ebd., S. 332, 345). Durchschnittlich wurden „[...] 4,7 (33%) der 14 angegebenen lexikalischen Mittel verwendet“ (ebd., S. 345). Durch Grobanalyse und Expert:innen-Beurteilung konnte festgestellt werden, dass **sowohl die sprachliche als auch die fachliche Qualität durch die Zuhilfenahme der lexikalischen Hilfsmittel stieg** (ebd., S. 346). Konkret wiesen die Texte, in denen lexikalische Hilfsmittel verwendet wurden, vermehrt passivische – und somit bildungssprachliche – Strukturen auf (ebd., S. 346). Bezogen auf die fachliche Qualität zeigte sich in den Urteilen von Expert:innen in 12 von 13 Fällen eine Übereinstimmung dahingehend, dass die Texte mit lexikalischen Hilfsmitteln fachlich besser waren (ebd., S. 346). Ein Text, der eine kurze Länge aufwies und nicht übereinstimmend beurteilt werden konnte, wurde nach einer intensiven Analyse nach fachlichen, sprachlichen Kriterien und weiterer Expert:innenbeurteilung aufgrund seiner höheren Präzision (im Gegensatz zur Beschreibung ohne lexikalische Hilfsmittel) ebenfalls als fachlich besser eingestuft (ebd., S. 346). Da es sich um eine Studie in der siebten Klasse handelt, stellt sich die Frage, ob auch Kinder der Grundschule mithilfe von sprachlichen Mitteln qualitativ hochwertigere Texte verfassen.

Indizien hierfür ergeben sich aus einer Studie für den Mathematikunterricht (Götze, 2018). In dieser Studie wurden insgesamt vierzehn dritte Klassen und eine vierte Klasse zum Thema *Rechenhaus* und dem Übergang vom Erklären-Wie (Wie ist das Kind bei der Rechnung von Feldern des Rechenhauses vorgegangen?) zum Erklären-Warum (Versteht das Kind, warum es so vorgegangen ist?) untersucht (ebd., S. 105). In einer vierstündigen Unterrichtseinheit bot jede Unterrichtsstunde ein sprachliches Unterstützungselement im Sinne des Makro-*Scaffolding*. In der ersten Stunde handelte es sich um die Unterstützung durch Wortspeicher. „[...] [Z]entrale Sprachmittel (Wörter und auch Satzphrasen) [wurden] auf einem Plakat notiert und an einem Beispiel visualisiert und dienten als Orientierungsrahmen für Sprachhandlungen“ (ebd., S. 105-106). In den

weiteren drei Unterrichtsstunden fand die sprachliche Unterstützung durch andere Elemente des Makro-*Scaffolding* statt (ebd., S. 105). Die Unterrichtsreihe mit sprachlicher Unterstützung wurde in zwei Treatmentgruppen durchgeführt, von denen sich eine Treatmentgruppe durch die zusätzliche Aufmerksamkeit der Lehrkräfte auf Mikro-*Scaffolding*, unterschied (ebd., S. 107-108). In einer Kontrollgruppe wurde die Unterrichtsreihe ohne Unterstützung auf Makro-Ebene durchgeführt (ebd., S. 108). Das Ergebnis der beschriebenen Studie zeigte, dass **Wortspeicher gemeinsam mit weiteren Unterstützungsmaßnahmen im Sinne des Makro-*Scaffolding* durchaus bereits in der dritten Klasse das mathematische Erklären unterstützen** (ebd., S. 116). Zusammenfassend hält Götze (2018) fest: „Durch die Unterstützung auf der Makro-Ebene bekommen die Kinder Sprachmittel zur Verfügung gestellt, mit Hilfe derer sie die operativen Muster generalisierend und vollständig beschreiben können (EG1 und EG2), was bei ausbleibender sprachlicher Unterstützung bei nur sehr wenigen Kindern zu beobachten ist (KG)“ (S. 116).

In Bezug auf den Sachunterricht ist der Nutzen von Wortspeichern in der Schulausgangsphase in Bezug auf den Forscherkreislauf – soweit bekannt – nicht in größerem Rahmen erforscht. Unter Berücksichtigung der Studienergebnisse von Tajmel (2017b) und Götze (2018), die der Arbeit mit Wortspeichern beziehungsweise lexikalischen Hilfsmitteln einen positiven Effekt im Physik- und Mathematikunterricht bescheinigen, liegt die starke Vermutung eines hohen Nutzens im Sachunterricht nahe. Eine Überprüfung, ob dies der Fall ist, wäre wichtig. Denn:

Im Sachunterricht sollen Kinder vielfältige Kompetenzen erwerben, die ohne sprachliche Bildung nicht zu erwerben sind, wie bspw. Fragen formulieren, beschreiben, argumentieren, Hypothesen äußern. Die sich aus der pädagogisch-didaktischen Perspektive ergebende Frage ist, wie die Bildungsprozesse erfolgen sollen, damit dieser Kompetenzerwerb möglich ist. (Jeretin-Kopf, 2014)

Eine Studie, die den Nutzen von Wortspeichern im naturwissenschaftlichen Sachunterricht untersucht und somit auch die Sprachhandlungen des Beschreibens, Argumentierens und des Äußerns von Vermutungen umfasst, könnte Aufschluss über den hier angestrebten Kompetenzerwerb liefern. Dieses Forschungsdesiderat wird in dieser Forschungsarbeit angegangen.

2.4 Zusammenfassung

In der Einleitung zu diesem Forschungsvorhaben wird Bildungssprache als Schlüssel zum Bildungserfolg dargestellt. Da Sprache stets im Kontext verwendet wird und der naturwissenschaftliche Erkenntnisweg bildungssprachliche Sprachhandlungen – wie das *Erklären* – erfordert, bietet sich die Verknüpfung der Sprachbildung mit der naturwissenschaftlichen Perspektive des Sachunterrichts an. Insbesondere Wortspeicher bieten den Kindern die Möglichkeit, bereits bekannte Wörter und Formulierungen schrittweise im Sinne des *Scaffolding* in Erinnerung zu rufen und diese im neuen Kontext – wie bei der Niederschrift von Versuchsprotokollen – aktiv anzuwenden.

Diese Schlussfolgerung wird empirisch untermauert. Zusammenfassend aus dem Kapitel *Forschungsstand zur Wirksamkeit von Scaffolding und Wortspeichern im naturwissenschaftlichen Lernen* weist eine Studie auf den Umstand hin, dass Sprachförderung und das Experimentieren im Sachunterricht gut vereinbar sind (Gottwald, 2016). Weitere Untersuchungen weisen auf den Nutzen lexikalischer Hilfsmittel hin (vgl. Götz, 2018; Tajmel, 2017b). Allerdings beziehen sich diese Untersuchungen entweder auf Sprachförderung anstatt Sprachbildung, auf weiterführende Schulen, andere Schulfächer oder Wortspeicher werden als Teil eines gesamtsprachlichen Konzepts untersucht.

Es ist demnach dringend erforderlich, Wortspeicher als isolierte Variable zu betrachten. Da Wortspeicher in unterschiedlichen Darstellungsformen zum Einsatz kommen können, sollte der Nutzen verschiedener Arten der Wortspeicher-Vermittlung näher untersucht werden. Dadurch wird gleichzeitig der Nutzen verschiedener Sprachmodalitäten geprüft. Das Kapitel *Psycholinguistische Erkenntnisse* zeigt, dass *phonologische* und *semantische Informationen* als Basis zur Aufnahme neuer Sprachmittel dienen. Aufgrund dessen bietet sich eine Studie zur Untersuchung des Nutzens von auditiv vermittelten Wortspeichern an. Zur Untersuchung einer weiteren Darstellungsform eines Wortspeichers ist die visuelle Vermittlung von Wortspeichern interessant. Diese stellt den Kindern zusätzlich zu phonologischen und semantischen Informationen *graphemische Informationen* bereit. Außerdem stellt sie eine im Unterricht bereits etablierte Form der Wortspeicher-Arbeit dar.¹¹

Nach aktuellem Wissensstand gibt es keine Studie zum Nutzen von isoliert betrachteten Wortspeichern in auditiver und in visueller Form auf das Verfassen bildungssprachlicher Versuchsprotokolle im naturwissenschaftlichen Sachunterricht. Die Forschungsfragen, die im nächsten Kapitel vorgestellt werden, und die daran anknüpfende empirische Untersuchung sollen zu einem tieferen Verständnis der Sachlage beitragen.

¹¹ vgl. Kapitel 2.3.1

3 FORSCHUNGSFRAGEN

Aus dem theoretischen Rahmen ergibt sich die übergreifende Forschungsfrage.

Übergreifende Forschungsfrage:

Welchen Nutzen hat die Arbeit mit Wortspeichern für das Erstellen von Versuchsprotokollen im naturwissenschaftlichen Sachunterricht?

Zur Untersuchung der übergreifenden Fragestellung sind vier detaillierte Forschungsfragen von Interesse, deren jeweilige Herleitung auf die zusammenfassende Darstellung folgt.

F1 – Nutzung der Wortspeicher: Wie häufig verwenden Grundschulkinder die visuell oder auditiv vermittelten Sprachmittel aus Wortspeichern kontextuell richtig in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts?

F2 – Sprachregister: Lässt sich durch visuell oder auditiv angebotene Wortspeicher die Häufigkeit bildungssprachlicher Wörter und Formulierungen in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts erhöhen?

F3 – Inhalt: Lässt sich durch visuell oder auditiv angebotene Wortspeicher der Inhalt in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts verbessern?

F4 – Datenintegration: Fördert die visuelle oder auditive Vermittlung von Wortspeichern das Verfassen von bildungssprachlichen und inhaltlich zutreffenden Versuchsprotokollen im naturwissenschaftlichen Sachunterricht?

Die **erste Forschungsfrage** ergibt sich aus dem logischen Schluss, dass Wortspeicher ausschließlich bei Verwendung einen möglichen Nutzen erzielen können. Voraussetzung für den Einsatz der Wortspeicher von Seiten der Kinder ist zum einen die Erinnerung an die Sprachmittel und zum anderen die aktive Entscheidung zum Einbezug der Sprachmittel in das individuell geschriebene Versuchsprotokoll (vgl. Morek & Heller, 2012, S. 95). Unabhängig der möglichen beschriebenen Ursachen wird mit dieser Forschungsfrage die tatsächliche Verwendung der Sprachmittel untersucht. Da der Wortspeicher eine subjektive Selektion an allgemein bekannten bildungssprachlichen Wörtern und Formulierungen darstellt, ist die Anwendung dieser Sprachmittel auch bei Kindern ohne explizite vorherige Unterrichtung des Wortspeichers denkbar. Von Interesse ist, ob Kinder durch die Arbeit mit visuell oder auditiv angebotenen Wortspeichern **verstärkt** zur Nutzung der mutmaßlich bekannten Sprachmittel beim Schreiben des voraussichtlich unbekanntem Textformats des Versuchsprotokolls animiert werden.

Die Unterscheidung zwischen visuell und auditiv angebotenen Wortspeichern beruht auf den in Kapitel 2.3 vorgestellten unterschiedlichen Gestaltungsoptionen der Wortschatzarbeit. Beide Gestaltungsoptionen sollten unterrichtsauthentisch sein, d.h. die Sprachbildung sollte im Unterricht wenig aktive Lernzeit einnehmen, keiner überaus aufwendigen Vorbereitungszeit von Seiten der Lehrkraft bedürfen und im regulären Unterricht bereits zum Einsatz kommen. Bei der **auditiven Vermittlung** von Wortspeichern werden die sprachlichen Hilfen von Seiten der Lehrkraft vor der Erarbeitungsphase, in der die Kinder einen Versuch durchführen und protokollieren, mündlich vorgestellt und von den Schüler:innen mündlich nachgesprochen. Bei der **visuellen Vermittlung** werden die Wortspeicher zusätzlich mithilfe eines Plakats an der Tafel vorgestellt. In den beschriebenen *Grundzügen der Wortschatzarbeit* und dem empirisch entwickelten Modell zum *Scaffolding* wird prinzipiell ein Wechsel verschiedener Sprachmodalitäten bzw. der Einsatz verschiedener Darstellungsformen befürwortet. Das Kapitel *Psycholinguistische Erkenntnisse* zeigt, dass verschiedene Sprachmodalitäten bzw. verschiedene Darstellungsformen auch zu vermehrten Verknüpfungen im mentalen Lexikon führen. Durch die vermehrten Verknüpfungen könnte die Erinnerung und der Abruf von Sprachmitteln erleichtert werden. Es ist entsprechend davon auszugehen, dass Lernende vermehrt Wörter und Formulierungen anwenden, wenn diese zuvor sowohl in auditiver als auch in visueller Form (hier: in Form eines Wortspeicher-Plakats) vermittelt werden, als wenn sie lediglich in auditiver Form (hier: durch mündliche Nennung der Wortspeicher) angeboten werden.¹² Aufgrund dieser Feststellung und dem empirischen Ergebnis von Tajmel (2017b, S. 345) zur tatsächlichen Anwendung lexikalischer Hilfsmittel (mit der Einschränkung, dass das vorliegende Forschungsdesign von ihrem

¹² Für eine bessere Lesbarkeit wird die Unterscheidung *visuell* vs. *auditiv* in allen vier Forschungsfragen verwendet, obwohl die visuelle Unterrichtung in Plakatform die auditive Vermittlung miteinschließt.

Forschungsdesign abweicht) lassen sich **zwei einseitig gerichtete Unterschiedshypothesen (H1 und H2) für Forschungsfrage 1** formulieren.

H1 – Nutzung der Wortspeicher: In den Versuchsprotokollen der Gruppe, die mit Wortspeichern in Plakatform lernte, befinden sich nach der Unterrichtssequenz signifikant mehr – korrekt angewandte – Wörter und Formulierungen aus den Wortspeichern als bei den Versuchsprotokollen der Gruppe, die die Wortspeicher auditiv vermittelt bekam.

H2 – Nutzung der Wortspeicher: In den Versuchsprotokollen der Gruppe, die mit Wortspeichern in auditiver Form lernte, befinden sich nach der Unterrichtssequenz signifikant mehr – korrekt angewandte – Wörter und Formulierungen aus den Wortspeichern als bei den Versuchsprotokollen der Gruppe, die die Wortspeicher nicht vermittelt bekam.

Der Zusatz der angemessenen Kontextualisierung in der ersten Forschungsfrage lässt sich u.a. mit Morek und Heller (2012, S. 94) begründen, die der kontextuell korrekten Anwendung bildungssprachlicher Wörter und syntaktischer Formen eine hohe Bedeutung beimessen.

Die **zweite Forschungsfrage** ist insofern wichtig, da die Nutzung der bildungssprachlichen Mittel aus den Wortspeichern nicht zwangsläufig mit einer Verbesserung des Versuchsprotokolls einhergeht. Es liegt die Vermutung nahe, dass Kinder durch Nutzung der Wortspeicher ein insgesamt bildungssprachlich geschriebenes Versuchsprotokoll darbieten. Demgegenüber ist das Verfassen eines bildungssprachlich geschriebenen Versuchsprotokolls auch ohne Anwendung der Wortspeicher möglich. Der Fokus auf Bildungssprache anstelle von weiteren möglichen Sprachregistern ist im theoretischen Rahmen ausführlich begründet.¹³ Aufgrund der Tatsache, dass es bezüglich visueller und auditiver Wortspeicher in Bezug auf das Sprachregister keine vorliegenden theoretischen oder empirischen Indikatoren für einen Unterschied zwischen den Gruppen gibt, werden zu dieser Fragestellung keine Hypothesen formuliert.

Ebenso wie die zweite Forschungsfrage geht die **dritte Forschungsfrage** dem Nutzen der Wortspeicherarbeit nach. Da Sprache und Inhalt stark miteinander verknüpft sind, ist ein Einfluss der Wortspeicher auf den Inhalt der Versuchsprotokolle naheliegend. Auf der anderen Seite könnte der Fokus auf Sprachbildung dazu führen, dass die inhaltliche Lernzeit etwas kürzer ist und somit die inhaltliche Auseinandersetzung mit einem

¹³ vgl. Kapitel 2.1.2

Versuch weniger tief greift. Eine empirische Untersuchung zum Einsatz lexikalischer Hilfsmittel weist auf eine höhere fachliche Qualität nach Nutzung der Hilfsmittel hin (Tajmel, 2017b, S. 346). Da die Rahmenbedingungen dieser Studie (u.a. Schulform und Schulfach) von den Rahmenbedingungen des vorliegenden Forschungsvorhabens abweichen und eine inhaltliche Verschlechterung durch den Fokus auf Sprachbildung möglich wäre, werden für diese Fragestellung keine Hypothesen festgelegt.

Die **vierte Forschungsfrage** wird zur Verflechtung der vorangestellten drei Forschungsfragen formuliert, denn für ein umfassendes Verständnis zum Nutzen von Wortspeichern ist eine isolierte Betrachtung einzelner Komponenten nicht ausreichend. Stattdessen ermöglicht die Betrachtung der Zusammenhänge zwischen Wortspeicher, Sprachregister und Inhalt einen Einblick in die übergreifende Fragestellung. Die Annahme, dass Wortspeicher einen positiven Effekt auf Schulergebnisse haben, ist aus der Theorie des *Scaffolding* und den empirischen Untersuchungen zum *Scaffolding* abgeleitet. Mit der vierten Forschungsfrage wird geprüft, ob sich dieser positive Effekt konkret auf Versuchsprotokolle im naturwissenschaftlichen Sachunterricht übertragen lässt. Die Unterscheidung zwischen den angesprochenen Sinnen (visuell oder auditiv) bei der Wortspeicher-Vermittlung lässt eine nuancierte Überprüfung der oben genannten Annahme zu. Durch Verknüpfung quantitativer und qualitativer Ergebnisse ermöglicht die Untersuchung der vierten Forschungsfrage differenzierte Rückschlüsse in einem lückenhaft erforschten Gebiet.

4 METHODE

Das Forschungsdesign ist auf die Beantwortung der vier vorgestellten Forschungsfragen ausgerichtet. Im ersten Kapitel wird die *Stichprobe* vorgestellt, die in dieser empirischen Studie untersucht wurde. Im Kapitel *Rahmenbedingungen und Verlauf* werden, teils unvorhergesehene, äußere Bedingungen und der Verlauf der Untersuchung beschrieben. Zentral für dieses Kapitel ist weiterhin die Art der *Datenerhebung* und die besondere Methode der *Datenauswertung*, die dieser Untersuchung zugrunde liegt.

4.1 Stichprobe

Ab dem neunten Lebensjahr können Kinder fachsprachliche – und somit auch bildungssprachliche – Elemente verstehen und anwenden (Trautmann & Reich, 2008, S. 47). Aus diesem Grund bot sich die Untersuchung der Fragestellungen an Kindern ab dem dritten Schuljahr (Alter: 8-9 Jahre) an. Da Kinder in der vierten Klasse ggf. Vorerfahrungen zum Verfassen von Versuchsprotokollen mitbringen, wurden für die Stichprobe Lernende aus **stufenbezogenen dritten Klassen** herangezogen. Zur Vermeidung einer starken Unterscheidung der Klassen aufgrund unterschiedlicher Einzugsgebiete, konzentrierte sich die Auswahl der Stichprobe auf eine Schule. Die drei parallelen Klassen wurden per Losverfahren einer Kontrollgruppe (KG), Treatmentgruppe 1 (TG1) und Treatmentgruppe 2 (TG2) zugewiesen. Die Bedeutung dessen wird im nächsten Kapitel *Rahmenbedingungen und Verlauf* erläutert. Die finale Zusammensetzung der Stichprobe ist Tab. 3 zu entnehmen.

Die Gesamtanzahl der Versuchspersonen (Vpn) pro Klasse lag zwischen 24 und 26, sodass es sich um drei Klassen mit vergleichbarer Klassengröße handelte. Aufgrund erschwerter Rahmenbedingungen durch Corona-Quarantänen und Corona-Impfterminen war die tatsächlich auswertbare Stichprobe deutlich kleiner. Sie wird im Folgenden als *Panel* bezeichnet. Zum *Panel* zählen alle Vpn mit unterschriebenem Elternbrief, die an allen Haupttestungen (Prä-, Post- und Follow-Up-Test) anwesend waren. Ebenfalls nicht berücksichtigt im *Panel* ist eine Vp, die aus emotionalen Gründen keinen Post-

Test abgeben konnte. Das *Panel* umfasst insgesamt 46 Vpn, davon 20 Vpn in der KG, 12 Vpn in TG1 und 14 Vpn in TG2. Die unterschiedliche Anzahl an Vpn pro *Panel*-Gruppe wurde bei der Ergebnisdarstellung berücksichtigt. Für eine vergleichbare Basis aller Ergebnisse zu allen vier Forschungsfragen wurde stets das *Panel* und keine weitere Vp außerhalb des *Panel*s untersucht. Die Bezeichnungen KG, TG1 und TG2 werden in diesem Forschungsvorhaben zur Bezeichnung der *Panel*-Gruppen verwendet.

Tab. 3: Zusammensetzung der Stichprobe

	Anzahl der Vpn		Geschlecht		Alter	LRS	ndH
	(n)		(n)				
	Gesamt	<i>Panel</i>	männlich	weiblich	(M)	(n)	(n)
	Prä/Post/ Follow-Up						
KG	26	20	10	10	8;9	2	2
TG1	24	12	4	8	8;7	1	1
TG2	24	14	7	7	8;7	0	0
GESAMT	74	46	21	25	8;8	3	3

Vpn = Versuchspersonen, KG = Kontrollgruppe, TG1 = Treatmentgruppe 1, TG2 = Treatmentgruppe 2; LRS = Lese-Rechtschreib-Schwäche; ndH = Kinder mit nicht-deutscher Herkunftssprache

Da das *Panel* die Stichprobe darstellt, beziehen sich die Angaben in den weiteren Spalten der Tabelle auf diese Vpn. Die Zugehörigkeit zu einem Geschlecht war in KG und TG2 ausgeglichen. In TG1 war das weibliche Geschlecht mit 8 Vpn etwas stärker vertreten als das männliche Geschlecht mit 4 Vpn. Die Gesamtanzahl an männlichen Vpn (21) und weiblichen Vpn (25) unterschied sich nicht stark, sodass keine weiteren Untersuchungen zur Überprüfung des Geschlechts als Störvariable vorgenommen wurden. Das durchschnittliche Alter wurde auf Grundlage der individuellen Geburtsmonate und -jahre berechnet und war in den drei Klassen mit einem Wert zwischen 8 Jahren und 7 Monaten (8;7) und 8 Jahren und 9 Monaten (8;9) vergleichbar. Vpn mit einer Lese-Rechtschreibschwäche (LRS) waren in TG1 und der KG präsent. Ebenso waren Vpn mit nicht-deutscher Herkunftssprache (ndH) in TG1 und der KG vertreten. In der KG handelte es sich um zwei Vpn, die zuhause mit ihren Familien Vietnamesisch und im zweiten Fall teils Deutsch und teils Vietnamesisch sprechen. In TG1 handelte es sich um eine Vp mit polnischer Herkunftssprache. Die Zusammensetzung der Stichprobe hinsichtlich LRS und ndH ist insofern relevant, da das Forschungsvorhaben thematisch auf sprachlichen Kenntnissen und Fähigkeiten beruht. Die beschriebene Verteilung von

Kindern mit LRS und ndH wurde mithilfe eines allgemeinen Sprachstandsvergleiches zwischen den Gruppen kontrolliert.

4.2 Rahmenbedingungen und Verlauf

Zentrales Ziel des Forschungsvorhabens war die Untersuchung des Nutzens von Wortspeichern im natürlichen Kontext, d.h. im regulären Schulleben. Durch die Felduntersuchung lassen sich die Ergebnisse direkt auf das Unterrichten im Sachunterricht übertragen. Allerdings geht eine Felduntersuchung auch mit Schwierigkeiten einher, wie „[...] die zufällige Zuweisung von Versuchspersonen zu den Gruppen, die Anforderung, dass sich die Treatments von Versuchs- und Kontrollgruppe nur in einem klar definierten Aspekt unterscheiden, die gleichbleibende Durchführung des Treatments und andere mehr [...]“ (Becker-Mrotzek et al., 2015, S. 201).

Die zufällige Gruppenzuweisung ist im Schulkontext aus organisatorischen Gründen schlichtweg nicht oder kaum durchzuführen. Insgesamt stellte daher der Klassenverband von drei dritten Klassen jeweils eine abgegrenzte Gruppe der Untersuchung dar. Bei dem vorliegenden Forschungsvorhaben handelt es sich somit um ein **quasi-experimentelles Untersuchungsdesign**. Wie Abb. 11 zu entnehmen, fand in den drei Gruppen entweder kein Treatment (KG), ein auditives Wortspeicher-Treatment (TG1) oder ein visuelles Wortspeicher-Treatment (TG2) statt. Bei dem auditiven Wortspeicher-Treatment wurden in den Treatment-Stunden jeweils fünf relevante Sprachmittel mündlich vorgestellt und von den Vpn chorisch wiederholt, bevor die Vpn einen Versuch durchführten und protokollarisch festhielten. Bei dem visuellen Wortspeicher-Treatment wurden in den Treatment-Stunden dieselben relevanten Sprachmittel visuell in Form eines Plakates gezeigt und in gleicher Weise besprochen. Die Zuordnung der Gruppen zu den Treatments bzw. zur Kontrolle fand per Losverfahren statt, sodass eine **randomisierte Gruppenzuordnung** gegeben war.

KG	TG1	TG2
KEIN TREATMENT	AUDITIVE WORTSPEICHER	WORTSPEICHER IN PLAKATFORM
<ul style="list-style-type: none"> • Sprachtest • Prä-Test • Post-Test • Follow-Up-Test 	<ul style="list-style-type: none"> • Sprachtest • Prä-Test • Post-Test • Follow-Up-Test 	<ul style="list-style-type: none"> • Sprachtest • Prä-Test • Post-Test • Follow-Up-Test

Abb. 11: Forschungsdesign

Um die Vergleichbarkeit der Klassenverbände zumindest sprachlich festzustellen und auf die unterschiedliche Zusammensetzung der Stichprobe einzugehen, wurde zu Beginn der Unterrichtssequenz ein **Sprachtest** vorgenommen. An einem weiteren Tag führten die Vpn einen Versuch durch und begleiteten diesen durch ein entsprechendes Versuchsprotokoll. Die Versuchsprotokolle wurden eingesammelt und stellten den **Prä-Test** dar. Von den Vpn erstellte Versuchsprotokolle waren auch das Erhebungsinstrument für **Post-Test** und **Follow-Up-Test**. Der konkrete Verlauf der Hauptstudie wird in Kapitel 4.2.3 vorgestellt.

Um einen möglichen Einfluss der Lehrkraft auf die Unterrichtssequenz zu vermindern, wurden **alle Unterrichtsstunden von der Autorin durchgeführt**. Durch die Konstanz der Lehrkraft war die Art und Weise des Unterrichts in allen drei Klassen vergleichbar. In dem vorliegenden Forschungsvorhaben ist dies von besonderer Bedeutung, denn Sprachbildung findet nicht nur im Zuge der Unterrichtsvorbereitung, sondern auch spontan im Unterricht statt. Würden die Unterrichtsstunden von verschiedenen Lehrkräften durchgeführt, so könnten diese in unterschiedlichem Maße auf Maßnahmen des *Mikro-Scaffolding* zurückgreifen und dadurch Unterschiede der Vergleichsgruppen induzieren. Ein möglicher Einfluss der Lehrkraft auf die Unterrichtssequenz wurde weiterhin dadurch vermindert, dass für jede Unterrichtsstunde ein detailliert ausgearbeiteter Stundenverlauf vorlag, der sich in den drei verschiedenen Klassen nur durch die unterschiedliche oder fehlende Bereitstellung von Wortspeichern unterschied. Dadurch, dass die Autorin die Vpn vor Beginn der Unterrichtssequenz nicht kannte, waren Erwartungseffekte von Seiten der Lehrkraft gegenüber individuellen Schüler:innen abgeschwächt.

Zur Feststellung der Durchführbarkeit des Forschungsvorhabens wurden sowohl eine Vorstudie als auch zwei Pilotstudien vorangestellt. Der Verlauf des Forschungsvorhabens spiegelt sich in Abb. 12 wider.

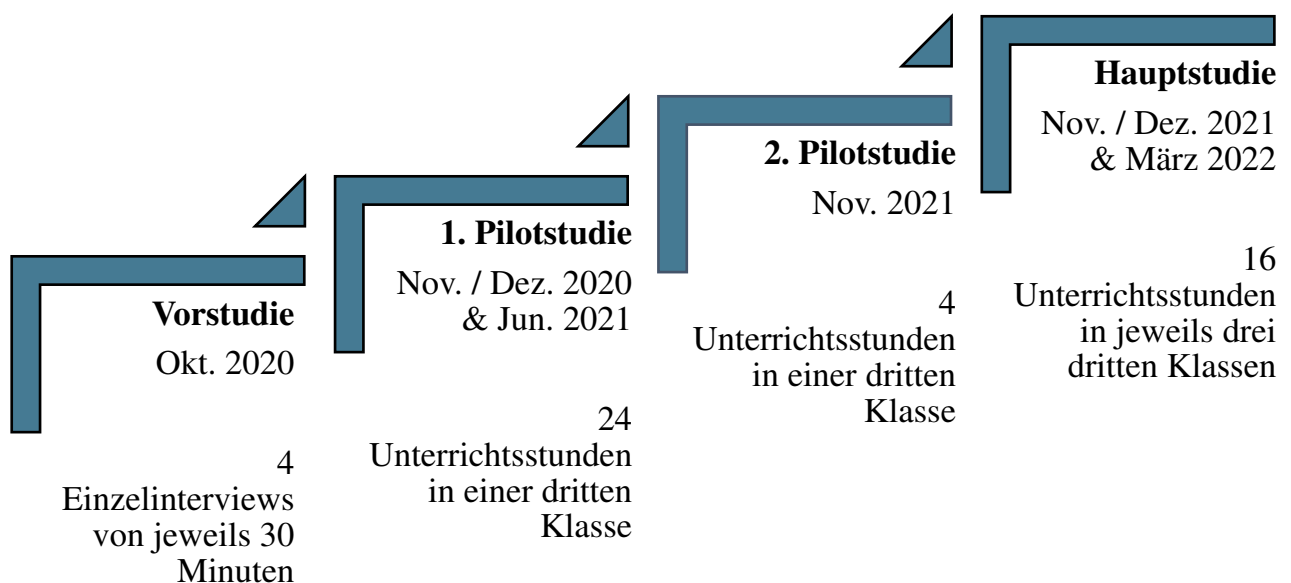


Abb. 12: Verlauf des Forschungsvorhabens

4.2.1 Vorstudie

Ziel der Vorstudie war die Ermittlung der inhaltlichen, sprachlichen und praktischen Grundlagen der Kinder bezogen auf die Teststunden. In den Teststunden war das Durchführen und Protokollieren von einem Versuch vorgesehen. Mögliche Schwierigkeiten bei der selbstständigen Durchführung der Versuche sollten in der Vorstudie eindeutig sichtbar werden. Darum wurden die Versuchsaufgaben der Teststunden in Einzelinterviews mit jeweils vier Kindern – zwei Mädchen und zwei Jungen – der dritten Grundschulklasse durchgeführt. Jedes der vier Interviews wurde von der Autorin geleitet, dauerte etwa 30-35 Minuten und wurde zwecks nachträglicher Betrachtung der Handlungen und verbalen Äußerungen videographiert.

Im Allgemeinen orientierten sich die Interviews an der Methode des **Lauten Denkens** (Konrad, 2010). Dabei werden die Kinder aufgefordert, grundsätzlich alle Gedanken und Wahrnehmungen auszusprechen, die bei der Bearbeitung der Aufgaben aufkommen (ebd., S. 2). Durch den Einblick in die handlungsbegleitenden oder retrospektiven Gedanken wird der Verarbeitungsprozess im Gehirn sichtbar (ebd., S. 2). Verdeutlicht wird dies in Abb. 13 unter Hinzunahme des Mehrspeichermodells. Da die Methode des Lauten Denkens bereits in der Grundschule angewandt (Schnell, 2016, S. 32) und inhaltlich im Zusammenhang mit dem naturwissenschaftlichen Erkenntnisweg im Biologieunterricht genutzt wurde (Konrad, 2010, S. 11), war von einer ähnlichen Anwendungsmöglichkeit der Methode in Interviews zur Versuchsdurchführung im Sachunterricht auszugehen.

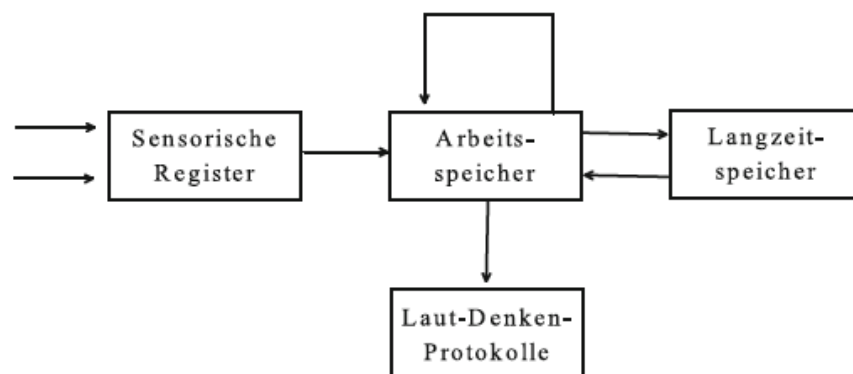


Abb. 13: Gedächtnismodell zum Lauten Denken (Konrad, 2010, S. 6)

Ein Grund für den Einsatz der Methode war der Erhalt konkreter Informationen, die in einem nächsten Schritt der Test-Optimierung dienen (Sandmann, 2014, S. 254). Darüber hinaus wurden nicht nur Informationen über praktische und inhaltliche Grundlagen der Kinder, sondern vor allem auch sprachliche Grundlagen offenbart. Durch das laute Vorlesen der Aufgabenstellungen wurden weiterhin „[...] missverständliche Formulierungen und unerwünschte Verstehensprobleme [...] identifiziert“ (ebd., S. 254). In den

Interviews wurde deshalb insgesamt (a) **zum lauten Vorlesen der Aufgabenstellungen** aufgefordert und es wurden (b) **handlungsbegleitende Verbalisierungen** gewünscht, sowie die (c) **Verbalisierung nachträglicher Gedanken** durch gezielte Fragestellungen bezüglich der Vermutung, des Materials, der Durchführung, Beobachtung und Erklärung des Versuches angeregt.

Aus der Betrachtung des Filmmaterials und den Transkripten ergaben sich im Nachhinein konkrete Anpassungsbedarfe. Die Änderungen bezogen sich auf die **Lesbarkeit** (z.B. die Formatierung der Versuchsanweisung in ein Zeilenformat anstelle des ursprünglich gedachten Spaltenformats), auf die **Praktikabilität** (z.B. das Auffüllen eines Messbechers bei einem Stromversuch mit 150 ml Wasser anstelle von 100 ml Wasser zwecks vereinfachter Eintauchmöglichkeit für kurze Nägel), auf das **inhaltliche Verständnis** (z.B. das Umformulieren der Aufgabenstellung von *Beobachte das Ei* zu *Beobachte, in welcher Höhe sich das Ei befindet* zwecks eindeutigerer Anweisung) oder auf das **sprachliche Verständnis** (z.B. das Umformulieren der Aufgabenstellung von *Schütte* zu der weniger missverständlichen Aufforderung *Gieße*).

4.2.2 Pilotstudien

Um die gesamte Unterrichtssequenz inklusive der veränderten Versuchsanweisungen in den Teststunden im natürlichen Setting des Unterrichts zu erproben, wurde eine **erste Pilotstudie** geplant und umgesetzt. Der Verlauf der ersten Pilotstudie (Abb. 14) spiegelt bereits den Verlauf der späteren Hauptstudie wider.

Die erste Pilotstudie wurde von der Autorin durchgeführt und fand in einer dritten Klasse mit 24 Vpn statt. In einer ersten Doppelstunde wurde der **Sprachstand** getestet und anschließend in das Thema der Unterrichtssequenz eingeführt. Im **Prä-Test** und im **Post-Test** gab es Versuchsanweisungen, die in Einzelarbeit ausgeführt und in Form eines strukturierten Versuchsprotokolls festgehalten und eingesammelt wurden. Ebenso verlief der **Follow-Up-Test**, der aus Gründen der ersten Änderungsüberlegungen im praktischen Teil auf Partnerarbeit setzte. Die visuellen Wortspeicher in den Treatment-Stunden unterschieden sich von den Wortspeichern der späteren Hauptstudie. Während der **Treatment-Stunden** der ersten Pilotstudie wurden in jeder Doppelstunde jeweils fünf Wörter oder Formulierungen mündlich genannt, an der Tafel festgehalten und zusätzlich als schriftliche Wortspeicher ausgeteilt. Die schriftlichen Wortspeicher bestanden aus (a) den Wörtern und Formulierungen inklusive ihres Einsatzes in Beispielsätzen, (b) einer Aufgabe zur ersten Anwendung der neu erlernten bzw. aufgefrischten Wörter und Formulierungen, (c) eines Reflexionsnetzes zur Selbsteinschätzung und (d) einer Möglichkeit eigene neue Wörter und Formulierungen in einer zeichnerisch dargestellten *Wortschatzkiste* hinzuzufügen. Die Wortspeicher-Dokumente wurden von den

Vpn in jeder Treatment-Stunde bearbeitet und in einem persönlichen Wortspeicher-Hefter abgelegt.

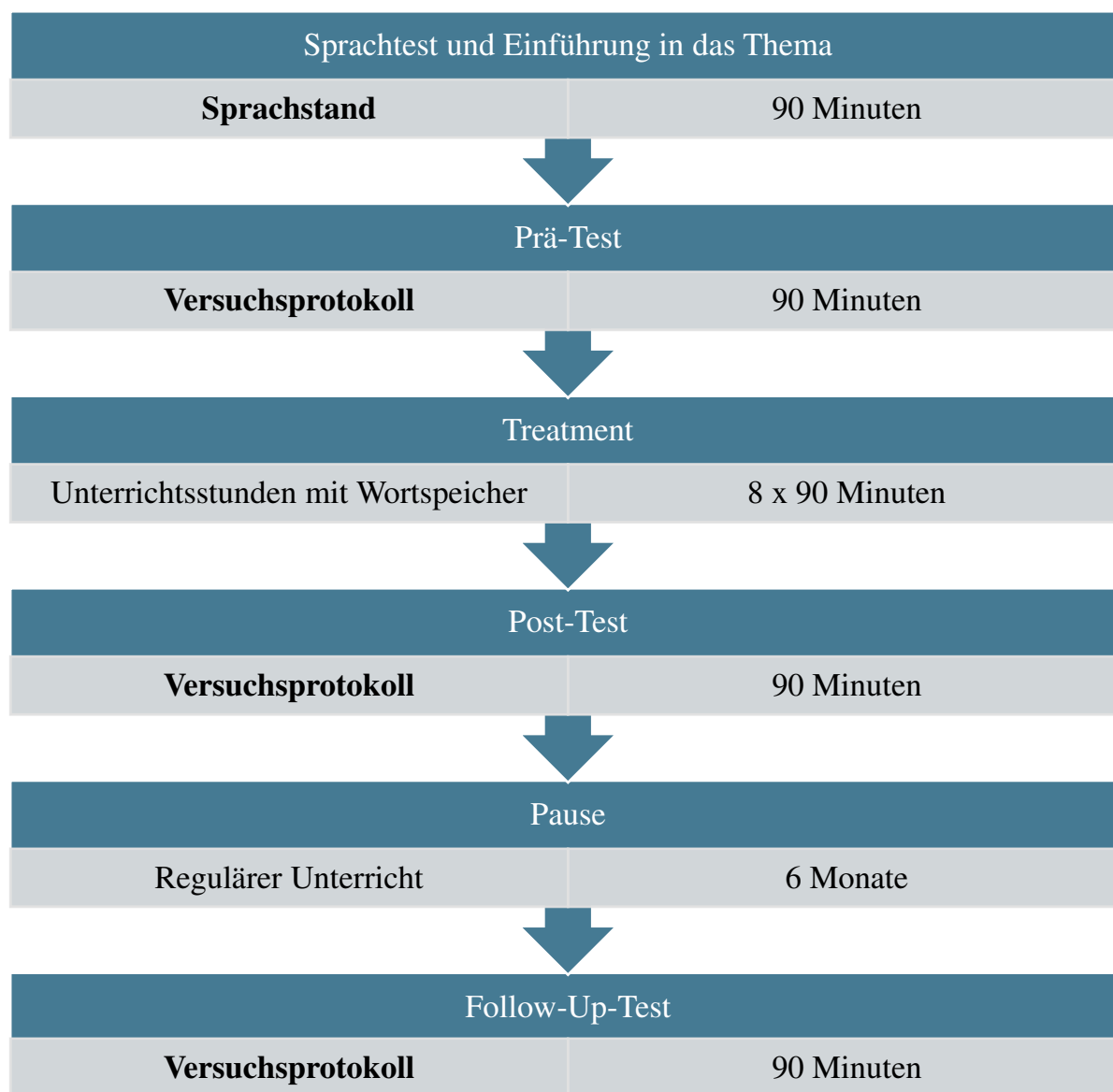


Abb. 14: Verlauf der ersten Pilotstudie

Der Test des Sprachstands fand in zwei Klassen statt, sodass insgesamt 46 Testergebnisse vorlagen. Aus organisatorischen Gründen erfolgten alle weiteren Tests und Treatment-Stunden ausschließlich in einer Klasse. Somit waren jeweils 24 Versuchsprotokolle aus der Prä- und der Follow-Up-Teststunde sowie 15 Versuchsprotokolle aus der Post-Teststunde zu verzeichnen. Die geringe Anzahl an Post-Test-Ergebnissen ist darauf zurückzuführen, dass die Vpn am Tag der zweiten Erhebung aus Gründen der Corona-Pandemie nur zur Notbetreuung in die Schule kamen. Die Corona-Pandemie

war auch Grund dafür, dass der Follow-Up-Test unerwartet erst sechs Monate nach Abschluss der Unterrichtssequenz stattfinden konnte.

Die Auswertung der ersten Pilotstudie erfolgte auf mehreren Ebenen. Der **Sprachtest** bestand aus fünf verschiedenen Lückentexten. Die 46 Testergebnisse des Sprachtests wurden u.a. hinsichtlich des Schwierigkeitsgrades der einzelnen Texte untersucht, sodass diese in schwierigkeitsaufsteigender Reihenfolge in der Hauptstudie angeordnet werden konnten. Mithilfe der **Reliabilitätsanalyse in SPSS** wurden ungeeignete Items im Sprachtest aufgedeckt und für die Hauptstudie neu angepasst.

Die **Versuchsprotokolle** stellten die Testergebnisse der drei Erhebungen dar. In Anlehnung an den naturwissenschaftlichen Forschungskreislauf der Grundschule waren diese wie folgt gegliedert: (1) Die Frage (wurde vorgegeben), (2) die Vermutung, (3) das Material, (4) die Durchführung, (5) die Beobachtung, (6) die Erklärung. Die Auswertung der Versuchsprotokolle war mit der späteren Auswertung der Hauptstudie vergleichbar, sodass die Praktikabilität der Auswertung eingehend getestet werden konnte. Zudem fanden die Kategoriensysteme zur Kodierung von Bildungssprache und Inhalt in der Auswertung der ersten Pilotstudie eingehende Verwendung und wurden in Bezug auf die Hauptstudie verbessert.

Die erste Pilotstudie führte zu zahlreichen **Erkenntnissen**. Eine zentrale Änderung ist im Bereich der Treatment-Stunden – genauer, in der Einführung und Nutzung des Wortspeichers – zu sehen. In der ersten Pilotstudie lag der Wortspeicher jeder Vp in schriftlicher Form inklusive Übung und Reflexion vor. Dies führte dazu, dass sich die Vpn zu einem großen Teil während der Doppelstunden mit dem Wortspeicher beschäftigten. Da der Aufwand der vorherigen Erstellung der Wortspeicher vonseiten der Lehrkraft als auch die lange Wortspeicherarbeit während der Unterrichtsstunden nicht im natürlichen Unterrichtskontext zu erwarten ist, wurde die **visuelle Wortspeicherarbeit in der Hauptstudie neu angepasst**. Anstelle von Wortspeicher-Heften mit Übungen und Reflexion, wurden in der Hauptstudie Wortspeicher in Plakatform und in auditiver Vermittlungsform angeboten. Weiterhin wurde die ausführlich angelegte erste Pilotstudie mit einem gesamten Stundenaufkommen von 24 Unterrichtsstunden dazu genutzt, eine Auswahl der wichtigsten Unterrichtsstunden für eine **verkürzte Unterrichtssequenz** in der Hauptstudie zu treffen.

Eine **zweite Pilotstudie** diente der thematischen Prüfung zur Auswahl der Versuche in den Test-Stunden und zur Testung von Wortspeichern in Plakatform. Es wurden zwei Unterrichtsstunden von je 90 min erprobt. Auf Grundlage des Vergleichs beider Unterrichtsstunden wurde der Prä-Test ausgewählt. Weiterhin wurden leichte Änderungen an der Plakatform vorgenommen, die hauptsächlich die Lesbarkeit im Klassenraum betrafen. Wie die Unterrichtssequenz der Hauptstudie im Detail aussah, ist Gegenstand des nächsten Kapitels.

4.2.3 Verlauf der Hauptstudie

Aus der Vorstudie und den beiden Pilotstudien ergab sich folgender Verlauf der Hauptstudie (Abb. 15).

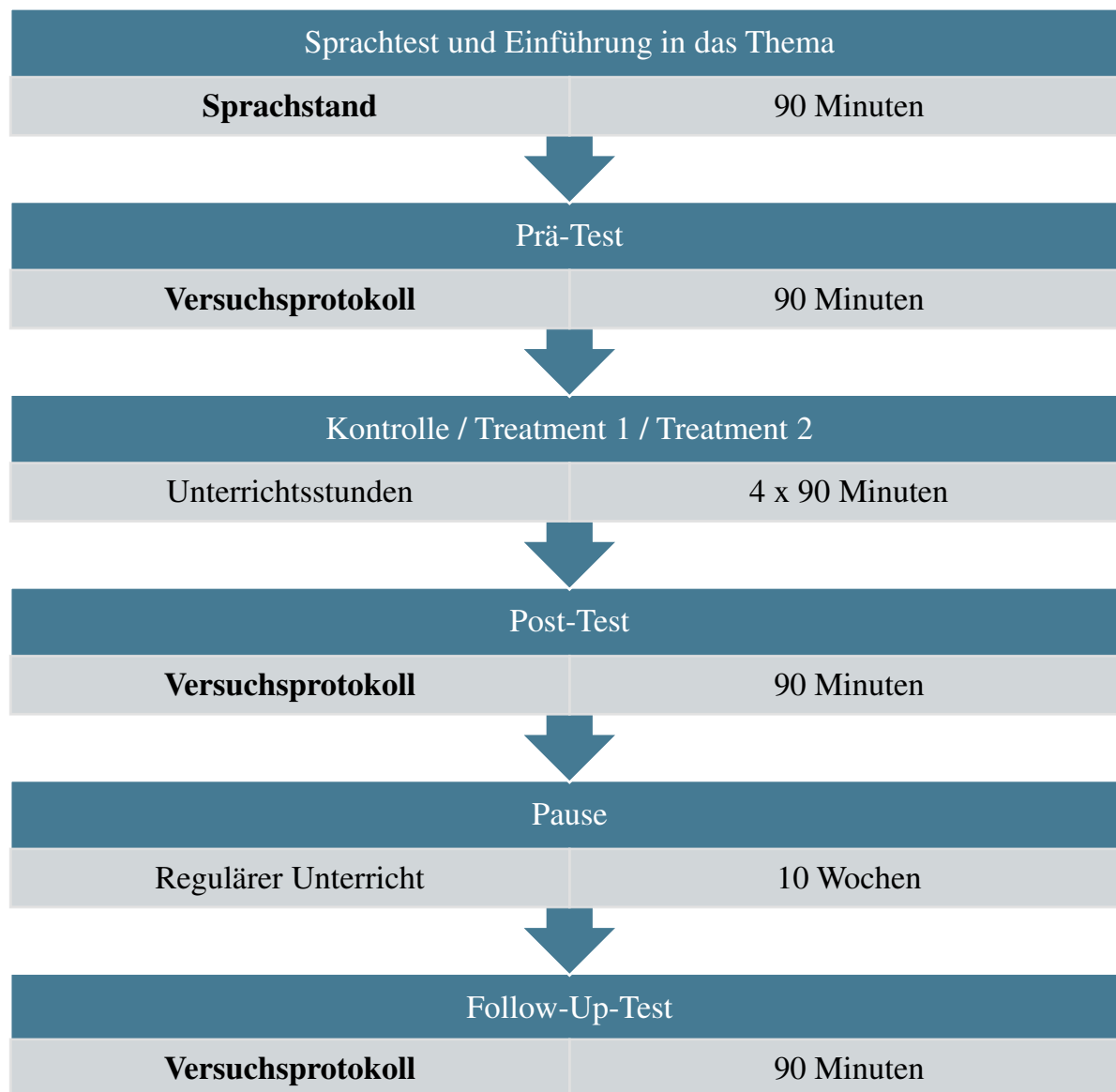


Abb. 15: Verlauf der Hauptstudie

In einem ersten Schritt nahmen die Vpn an einem **Sprachtest** teil, der im Kapitel *Datenerhebung* näher erläutert wird. Anschließend wurde das neue Unterrichtsthema eingeführt. In der darauffolgenden 90-minütigen Unterrichtsstunde fand der **Prä-Test** statt: Die Vpn führten einen unbekanntem Versuch mithilfe einer Versuchsanleitung durch und schrieben ein Versuchsprotokoll. Es folgten vier 90-minütige Unterrichtsstunden, in denen die Vpn entweder ein Treatment erhielten (visuelle oder auditive Wortspeicher)

oder die gleichen Unterrichtsstunden ohne spezielles Treatment durchliefen. Im Anschluss fand eine weitere 90-minütige Unterrichtsstunde statt, in der die Vpn einen Versuch per Versuchsanleitung durchführten und ihn protokollarisch festhielten. Es handelte sich hierbei um den **Post-Test**. Nach einer zehnwöchigen Pause fand eine 90-minütige Unterrichtsstunde mit dem **Follow-Up-Test** statt, der wie die vorherigen beiden Tests aus einem den Vpn unbekanntem Versuch mit Versuchsanweisung und der Aufgabe des protokollarischen Festhaltens bestand. Einen vertieften Einblick in die Unterrichtssequenz bietet das nächste Kapitel.

4.2.4 Thematische und methodische Begründungen zur Unterrichtssequenz

In diesem Kapitel wird die Unterrichtssequenz der Hauptstudie präsentiert. Dabei liegt ein Fokus auf der Begründung, die den einzelnen didaktischen Entscheidungen zu dieser Unterrichtssequenz unterliegen.

Thematisch ging es in der Unterrichtssequenz um das *Salz*. Dies ist mehrfach zu begründen. Ein Grund ist die Vielseitigkeit des Themas. Das Thema *Salz* bietet zahlreiche Möglichkeiten für Versuche, die in der Bereitstellung der Materialien ähnlich zueinander sind und dennoch unterschiedliche Inhalte vermitteln. So ist es möglich, dass alle Versuche des vorliegenden Unterrichtsvorhabens mit zwei Messbechern, Wasser, Salz, einem Teelöffel und nur wenigen weiteren versuchsspezifischen Materialien auskommen. Dennoch sind die vermittelten Phänomene, beziehungsweise Inhalte – sei es die Osmose, die Dichte, die Löslichkeit von Salz oder das Leiten von Strom – verschieden. Durch die wenigen benötigten Materialien sind die Versuche besonders geeignet für die Grundschule. Dies wird auch darin deutlich, dass im Rahmen des Projekts NaWiT AS¹⁴ die Vermittlung des wissenschaftlichen Arbeitens beispielhaft anhand einer Unterrichtssequenz zum Thema *Salz* vorgestellt wird (NaWiT AS, 2023). Die Unterrichtssequenz von NaWiT AS und die darin enthaltenen Versuche zum Thema *Salz* sind erprobt und eignen sich explizit für die dritte und vierte Klasse. Einige Ideen wurden dieser Unterrichtssequenz entnommen, um die Wortspeicher in einem möglichst alltagstauglichen und authentischen Reihenverlauf zu integrieren. Aus dem gleichen Grund wurden Anregungen aus der Grundschulzeitschrift *Sachunterricht Weltwissen* zum Thema *Salz* mit erprobten Versuchen aufgegriffen (Schmeinck, Dölle, Dreher, Oldenburg, & Lüpkes, 2017). Zu den Anregungen, die in dieser Unterrichtssequenz in veränderter Weise aufgenommen wurden, gehörten die Fragestellungen *Verschwundet Salz für immer? Warum geht man im toten Meer nicht unter? Warum wird im Winter Salz gestreut?* (NaWiT AS,

¹⁴ Das Projekt NaWiT AS steht für *Naturwissenschaftlich-technische Kompetenzen und Anregung schulbezogener Unterrichtsentwicklung*. In einem eigens für NaWiT AS eingerichteten Internetportal der Universität zu Köln werden in regelmäßigen Abständen Unterrichtsmaterialien zur Weiterentwicklung des Sachunterrichts zur Verfügung gestellt.

2023), sowie Versuchsanregungen zur *Siedesalzgewinnung*, *Stromleitung* und zum *Schmelzen von Eiswürfeln* (Schmeinck et al., 2017). Neben der hohen Eignung zur Durchführung von Versuchen ist das Thema *Salz* aufgrund des Lebensweltbezuges verwendbar. Dieser ist explizit gefordert, denn der Sachunterricht soll die Kinder darin unterstützen „[...] Phänomene und Zusammenhänge der Lebenswelt wahrzunehmen und zu verstehen [...]“ (GDSU, 2013, S. 9). Alle ausgewählten Versuche zeigen alltagsbezogene Phänomene auf – wie beispielsweise das schnellere Schmelzen von Eis durch Hinzugabe von Salz als Verweis auf das Salzstreuen im Winter. Der Bezug zum Alltag am Ende jeder Stunde ist darauf zurückzuführen, dass der *produktive Umgang mit Erkenntnissen* einen hohen Stellenwert im allgemeinen Forschungsprozess innehat (Aeppli, Gasser, Gutzwiller, & Tettenborn, 2014, S. 65). Mit anderen Worten: „Zum Verstehen gehört [...] auch, dass man in der Lage ist, das gelernte Wissen zu verwenden, um *Alltagsphänomene zu erklären* oder um Verbindungen zwischen verschiedenen Phänomenen herzustellen“ (Grygier & Hartinger, 2009, S. 8). Die Auseinandersetzung mit den alltagsbezogenen und inhaltlich vielseitigen Versuchen zum Thema *Salz* soll den Kindern insgesamt ein vertieftes Verständnis ihrer Lebenswelt vermitteln.

Prozedural stand in der Unterrichtssequenz das Experimentieren auf Grundlage des Forschungskreislaufes im Vordergrund. Die Verbindung zwischen Wortspeichern und experimentell ausgerichtetem Unterricht ist in der besonderen Eignung des Experimentierens zur Sprachbildung begründet.¹⁵ Empirisch untermauern Studien die Aussage, dass Sprachbildung und Experimentieren im Sachunterricht gut vereinbar sind.¹⁶ Um den handlungsbegleitenden und reflektierenden sprachlichen Austausch zu fördern, führten die Kinder die Versuche stets in Partnerarbeit durch. Eine Gruppenarbeit wurde aufgrund der Corona-Situation in diesem Forschungsvorhaben vermieden. Ein beispielhaftes Versuchsprotokoll ist im Anhang A.5 einsehbar. Die Versuchsprotokolle wurden in der Unterrichtssequenz als Forschungsprotokolle bezeichnet. Zur erhöhten Vergleichbarkeit der Versuchsprotokolle erhielten die Kinder in den Test-Stunden einen vorgegebenen zeitlichen Rahmen zum Schreiben von Teil I des Versuchsprotokolls (15 min), zur Durchführung des Versuchs (10 min) und zum Schreiben von Teil II des Versuchsprotokolls (10 min). Die Einteilung der gesamten Unterrichtssequenz in 90-minütige Unterrichtsstunden ist auf die höhere Zeitintensivität bei Versuchen und die didaktisch möglichst zeitnahe und intensive Besprechung der Versuche zurückzuführen. In der Wahl zwischen offenem Experimentieren und dem Durchführen angeleiteter Versuche lassen die angeleiteten Versuche eine höhere Vergleichbarkeit der geschriebenen Versuchsprotokolle zu. Hinzu kommt die Erkenntnis aus einer empirischen Studie im Elementarbereich, dass Versuche unter konkreter Anleitung oder die Kombination aus angeleiteten Versuchen und offenem Experimentieren – zumindest bei den jüngeren

¹⁵ vgl. Kapitel 2.2.3

¹⁶ vgl. Kapitel 2.3.5

Kindern – zu einem höheren Wissenszuwachs als das ausschließlich offene Experimentieren in einer Forscher-Ecke führt (Windt, Scheuer, & Melle, 2010, S. 226). Die Kombination aus angeleiteten Versuchen und offenem Experimentieren in der Unterrichtssequenz der ersten Pilotstudie zeigte auf, dass die Kinder in der dritten Klasse vorab eines umfassenden Trainings zum offenen Experimentieren bedürfen. Um diese Zeit vollends auf die intensive Auseinandersetzung mit den Versuchsprotokollen zu bündeln, waren alle Versuche des vorliegenden Unterrichtsvorhabens angeleitete Versuche. Die jeweiligen Versuchsanweisungen sind im Anhang A.2 bis A.4 abgebildet. Da in einer Studie gezeigt wurde, dass „[...] lediglich maximal ein Fünftel der neun- bis zehnjährigen Schüler in der Lage [waren], selbst eine naturwissenschaftlich relevante Frage zu formulieren“ (Jeretin-Kopf, 2011, S. 145), wurde die Fragestellung für die einzelnen Versuche jeweils vorgegeben.

Das Verfassen eines **Versuchsprotokolls** sollte die Kinder darin unterstützen, ihre sprachlichen und inhaltlichen Gedanken zu einem Versuch in langsamerer Geschwindigkeit zu verarbeiten. Denn durch die zeitliche Drosselung der Sprachverarbeitung beim Schreibprozess wird letztendlich eine „[...] vertiefte kognitive Verarbeitung von Wissen ermöglicht“ (Becker-Mrotzek et al., 2015, S. 200). Weitere Vorteile des Schreibens sind u.a. darin zu sehen, dass sich die Kinder in die Perspektive eines zeitlich und örtlich getrennten Lesers hineinversetzen und ihre Sätze mithilfe bestimmter Wörter und Formulierungen der Kohärenz sinnvoll strukturieren müssen (vgl. Becker-Mrotzek et al., 2015). Daneben wird durch das Schreiben eines Versuchsprotokolls die Anwendung von Bildungssprache induziert. Da ein Schreibprozess im Allgemeinen aufgeteilt werden sollte, um Schüler:innen zu entlasten, bietet sich bei Versuchsprotokollen ein schrittweises Vorgehen an (Peschel, 2013, S. 106, 109). In diesem Sinn wurden in der vorliegenden Unterrichtssequenz die folgenden Abschnitte des Versuchsprotokolls – a) die Vermutung und das Material, b) die Durchführung, c) die Beobachtung und d) die Erklärung – in 90-minütigen Treatment-Stunden jeweils aufeinander aufbauend angeleitet.¹⁷ Die Auflistung zeigt, warum es sich in der vorliegenden Unterrichtssequenz um mindestens vier 90-minütige Treatment-Stunden handeln musste. Eine kürzere Unterrichtssequenz würde der Zerlegung des Versuchsprotokolls mit Fokus auf die einzelnen Abschnitte nicht gerecht. Eine längere Unterrichtssequenz würde zusammen mit den jeweils 90-minütigen Sprach-, Prä- und Post-Tests eine Unterrichtssequenz von mehr als sieben Doppelstunden (> 14 Einzelunterrichtsstunden) ergeben. Eine Einheit länger als 14 Einzelunterrichtsstunden würde aus zeitökonomischen Gründen keiner realen Unterrichtssequenz entsprechen. Außerdem ließe sich in dem Fall ein Abfallen der Schüler:innenmotivation vermuten. Der zeitliche Aspekt war ebenso der Grund dafür, dass das *fading* als wichtiger Bestandteil des *Scaffolding* in dieser Unterrichtssequenz nicht

¹⁷ Da die Auflistung des Materials keiner 90-minütigen Übung bedarf, bildeten die Aufstellung einer Vermutung und die Auflistung des Materials zusammengekommen den Fokus einer Stunde.

durchgeführt wurde. Ein graduelles Zurücknehmen der Wörter und Formulierungen aus dem Wortspeicher würde mehrerer auf dem *Scaffolding*-Konzept basierender Unterrichtssequenzen bedürfen. Die vorliegende Unterrichtssequenz war somit ausschließlich auf die Arbeit mit Wortspeichern als Bestandteil des Makro-*Scaffolding* fokussiert.

Die **Wortspeicher** wurden in den drei Klassen entweder visuell in Plakatform (TG2), auditiv (TG1) oder nicht (KG) eingesetzt. Um eine vergleichbare aktive Lernzeit in allen drei Klassen trotz der unterschiedlichen Varianten sicher zu stellen, war jeder Unterrichtsstunde der vorliegenden Unterrichtssequenz eine didaktische Reserve zur inhaltlichen Vertiefung beigefügt. Da die Inhalte in den Teststunden stark von den anderen Unterrichtsstunden abwichen, war eine Auswirkung der didaktischen Reserve auf das inhaltliche Ergebnis in den Teststunden nicht zu erwarten. Die Auswahl der Sprachmittel für die Wortspeicher der vorliegenden Unterrichtssequenz beruhte u.a. auf den Erkenntnissen der Pilotstudie: Wörter und Formulierungen, die in den Prä-Tests der ersten Pilotstudie bereits vorkamen, wurden ausgeschlossen. Hauptsächlich ist die Auswahl der Sprachmittel orientiert an Anregungen aus Agel et al. (2012, S. 37) Beese et al. (2017, S. 177-178), Both et al. (2013, S. 269), Tajmel und Hägi-Mead (2017, S. 37), sowie Quehl und Scheffler (2008, S. 73), die im Sinne des Forschungsvorhabens eingegrenzt bzw. weiterentwickelt wurden. Abb. 16 zeigt, welche Wörter und Formulierungen nach entsprechender Auswahl in TG1 und TG2 zum Einsatz kamen. Im Anhang A.6 bieten die *Impressionen aus der Unterrichtssequenz* einen Einblick in das final erstellte Wortspeicher-Plakat und dessen Nutzung im Unterricht.

1. Treatment	2. Treatment	3. Treatment	4. Treatment
Die Vermutung	Die Durchführung	Die Beobachtung	Die Erklärung
<ul style="list-style-type: none"> • Meine Vermutung ist, dass ... • Ich vermute, dass ... • Ich denke, dass ... • ..., denn ... 	<ul style="list-style-type: none"> • Zunächst ... • Danach ... • Daraufhin ... • Anschließend ... • Schließlich ... 	<ul style="list-style-type: none"> • Man kann beobachten, dass ... • Es lässt sich feststellen, dass ... • Wenn ..., dann ... • Nach kurzer Zeit • unmittelbar 	<ul style="list-style-type: none"> • Die Beobachtung lässt sich wie folgt erklären: ... • Das liegt daran, dass ... • Dadurch, dass ... • ..., deshalb ... • Die Vermutung war richtig / falsch.
Das Material			
<ul style="list-style-type: none"> • Man benötigt: ... 			

Abb. 16: Verwendeter Wortspeicher in der Hauptstudie

Die bisher genannten Überlegungen mündeten in die Planung der **Unterrichtssequenz**, die im Folgenden zusammengefasst ist (Abb. 17).

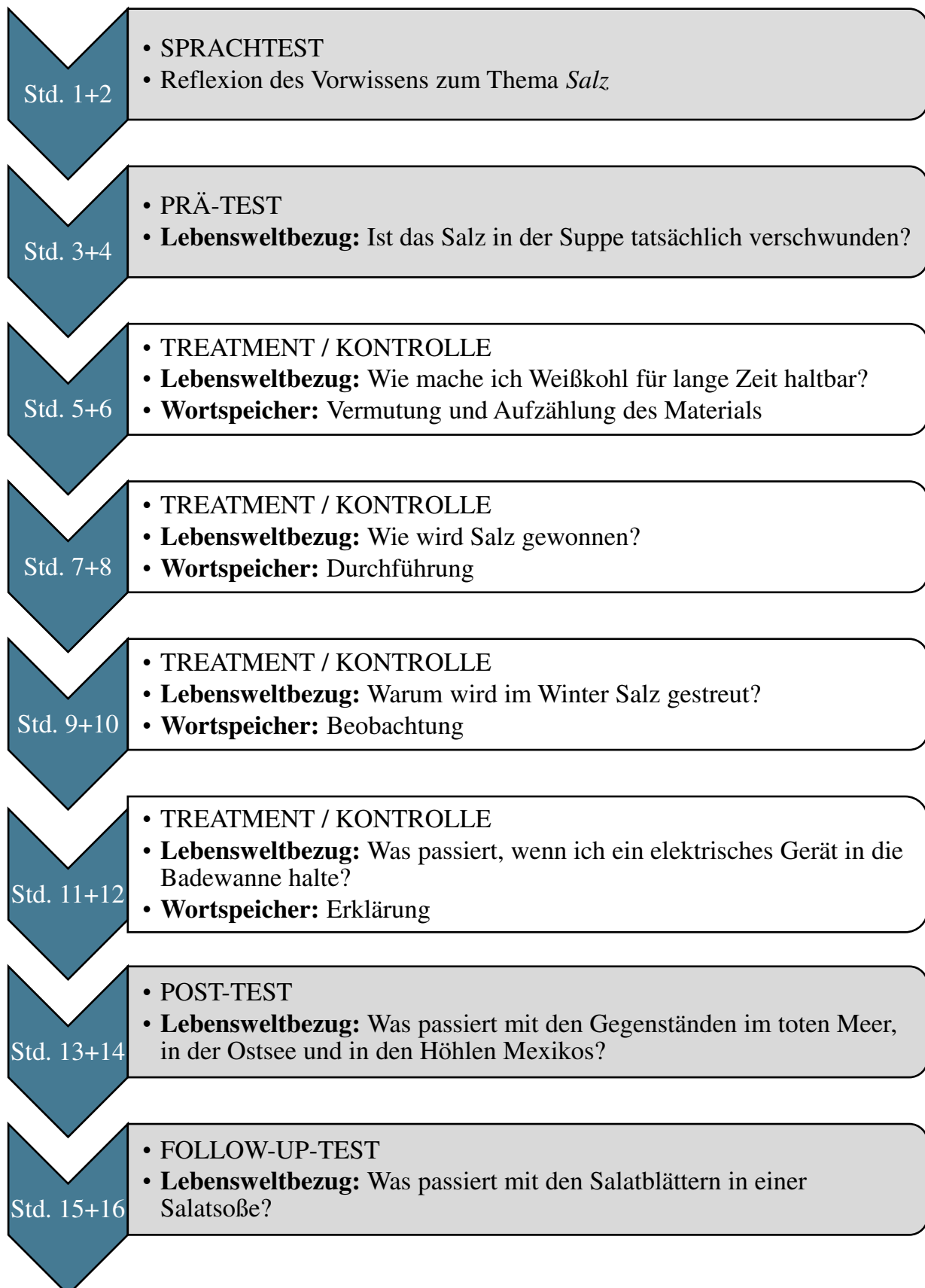


Abb. 17: Thematische Unterrichtssequenz der Hauptstudie

4.3 Datenerhebung

4.3.1 Sprachstandserhebung (C-Test)

In Kapitel 4.2 ist die Notwendigkeit eines Sprachtests zur Vergleichbarkeit der Klassenverbände im quasi-experimentellen Versuchsdesign erwähnt. Zur Ermittlung des Sprachstandes bietet sich der sogenannte **C-Test** an. Er eignet sich vor allem deshalb, weil es „[...] ein ökonomisches und reliables Verfahren [...]“ ist, das „[...] insbesondere dann erfolgreich einsetzbar [ist], wenn das Ziel eine vom vorangehenden Unterricht bzw. von der individuellen Lerngeschichte weitgehend unabhängige globale Sprachstandsfeststellung ist“ (Grotjahn, 2002, S. 211). Die hohe Reliabilität scheint selbst dann gegeben, wenn ein C-Test erstmalig zum Einsatz kommt (ebd., S. 214). Ebenso zeichnet sich der C-Test durch eine hohe Durchführungsobjektivität aus (ebd., S. 214). Auch die Validität wird durch Studien zur Korrelation von C-Test-Ergebnissen und anderweitiger Sprachkriterien unterstützt (ebd., S. 215). Im vorliegenden Forschungsvorhaben eignet sich der C-Test besonders wegen der erwähnten **Praktikabilität** und der anzunehmenden **hohen Validität**. Zudem eignet sich der C-Test als **normorientiertes bzw. bezugsgruppenorientiertes Messverfahren**, d.h. die Ergebnisse der einzelnen Vpn werden in Relation zueinander betrachtet und interpretiert (ebd., S. 218). Da für das Forschungsvorhaben die Parallelisierung der Versuchsgruppen im Vordergrund steht, ist vor allem der Bezug der einzelnen Sprachstände untereinander und weniger der kriteriengeleitete Sprachstand von Bedeutung. Ein weiterer Grund für den Einsatz des C-Tests ist darin zu sehen, dass es sich um einen **sehr gut untersuchten Sprachtest** handelt (ebd., S. 211), der bereits in anderen Studien zur allgemeinen Bestimmung der Sprachkompetenz zum Einsatz kam (u.a. Ahrenholz et al., 2018, S. 173; Hiller, Kleinbub, & Nafz, 2017, S. 63).

Der C-Test besteht im Allgemeinen aus kurzen Lückentexten mit jeweils 60-80 Wörtern, die sich thematisch voneinander unterscheiden (Grotjahn, 2002, S. 211-212). Ein Auszug aus dem C-Test des Forschungsvorhabens ist in Abb. 18 dargestellt. Der C-Test in voller Länge befindet sich im Anhang A.1. Der erste Satz und das Textende sind komplett vorgegeben. Ansonsten ist ab dem zweiten Satz die zweite Hälfte jedes zweiten Wortes entfernt (ebd., S. 212). Jeder Text sollte mindestens 20 Lücken, d.h. Items, aufweisen. Meistens werden vier Texte mit 25 getilgten Worthälften oder fünf Texte mit 20 Lücken verwendet (ebd., S. 212). Für eine möglichst große Themenbandbreite wurden in diesem Forschungsvorhaben **fünf Texte** ausgewählt, die jeweils genau **20 Items** aufweisen. Da es sich um authentische und kindgemäße Texte handeln sollte, wurden Texte zu verschiedenen Themen aus einem Sachunterrichtsbuch der Klassen 3 und 4 (Blasek et al., 2010) gewählt.

Die Feuerwehr

Herr Weiß ist Hausmeister in einer Schule und arbeitet für die Feuerwehr. Für d_____ ¹⁾ Feuerwehr arbe_____ ²⁾ er ab_____ ³⁾ ehrenamtlich, d_____ ⁴⁾ heißt, e_____ ⁵⁾ bekommt ke_____ ⁶⁾ Geld da_____ ⁷⁾. Er i_____ ⁸⁾ Mitglied ei_____ ⁹⁾ freiwilligen Feue_____ ¹⁰⁾. In al_____ ¹¹⁾ Gemeinden u_____ ¹²⁾ kleinen Stä_____ ¹³⁾ gibt e_____ ¹⁴⁾ eine sol_____ ¹⁵⁾ Feuerwehr. D_____ ¹⁶⁾ Feuerwehr h_____ ¹⁷⁾ viele schwi_____ ¹⁸⁾ Aufgaben. Da_____ ¹⁹⁾ müssen d_____ ²⁰⁾ Feuerwehrleute in ihrer Freizeit immer wieder üben. Mitglied in der Feuerwehr zu sein, bedeutet viel Arbeit.

Abb. 18: Auszug aus dem C-Test

Zur Erstellung der Lücken wurde ein *Online Didaktisierungstool* verwendet, das einen eingegebenen Text automatisch mit Lücken in gewünschter Länge ausstattet (Deutsche Welle & Lingofox, 2021). Die Aufgabe der Vpn war die Nennung der getilgten Worthälften. Für die Länge der Lücken gibt es verschiedene Möglichkeiten, von a) einer gleichen Lückenzahl bei jeder Tilgung, b) unterschiedliche Lückenzahl je nach Länge der getilgten Worthälften, zu c) die Lücke zeigt die exakte fehlende Buchstabenanzahl an (Grotjahn, 2002, S. 212). Um den Test etwas leichter als Option (a) zu gestalten und auf der anderen Seite zu vermeiden, dass sich die Validität durch das ermöglichte Zählen von Buchstaben und somit Raten von Wörtern wie in Option (c) verringert, wurde Option (b) gewählt und die **Lückenzahl entsprechend der fehlenden Wortlänge** angepasst (ebd., S. 212). Zur Erhöhung der Reliabilität gab es eine Zeitvorgabe (ebd., S. 216). Nach jeweils fünf Minuten wurden die Vpn zur Bearbeitung des nächsten Textes aufgefordert. Dadurch wurde sichergestellt, dass die Vpn alle Texte bearbeiteten – selbst, wenn diese teilweise unvollständig blieben. In der ersten Pilotstudie hatte sich die Zeitvorgabe von fünf Minuten pro Text bewährt. Insgesamt bekamen die Vpn **25 Minuten Zeit** zur Bearbeitung des C-Tests. Um einer Demotivation der Vpn vorzubeugen, wurden die Texte des C-Tests in aufsteigender Schwierigkeit zusammengesetzt (ebd., S. 212).

4.3.2 Versuchsprotokolle als Erhebungsinstrumente

Neben dem C-Test wurden Daten der Vpn zu den **Messzeitpunkten (MZP) 1, 2 und 3** erhoben. Hierbei handelte es sich um **Versuchsprotokolle**, die in der jeweiligen 90-minütigen Schulstunde von den Vpn geschrieben und eingesammelt wurden. Ziel des **Prä-Tests** (MZP 1) war die Erkenntnis, inwieweit die Vpn bereits zum Schreiben eines Protokolls in der Lage sind. Hierzu gehörte das Interesse, ob die Vpn ohne vorheriges Treatment bereits (i) Wörter und Formulierungen, die im Wortspeicher vorkommen, anwenden, (ii) ein bildungssprachliches Register verwenden und (iii) ein inhaltlich sinn-ergebendes Versuchsprotokoll verfassen. Der **Post-Test** (MZP 2) prüfte demgegenüber, inwieweit sich die Versuchsprotokolle in den drei genannten Punkten i, ii und iii durch ein bestimmtes Treatment bzw. in der KG von den vorherigen Versuchsprotokollen abgrenzten. Dabei war nicht nur der Vergleich von Prä-Test und Post-Test, sondern vor allem der Vergleich der Post-Tests untereinander, d.h. zwischen KG, TG1 und TG2, von Bedeutung. Der **Follow-Up-Test** (MZP 3) fand 10 Wochen nach Beendigung der Unterrichtssequenz statt und sollte Hinweise auf die langfristige Wirkung der Treatments geben. Eine Zusammenstellung der Datenerhebung ist in Tab. 4 ersichtlich. In der Tabelle stehen die Zahlen und Buchstaben für die unterschiedlichen eingesammelten Versuchsprotokolle. Dabei zeigen die Zahlen die Zuordnung zum Prä-, Post- und Follow-Up-Test an. Eine gleiche Zahl bedeutet, dass das Versuchsprotokoll **auf Grundlage der gleichen Versuchsanweisung** geschrieben wurde. Die Buchstaben zeigen die Zugehörigkeit der Versuchsprotokolle zu den verschiedenen Gruppen mit unterschiedlichem Treatment bzw. zur KG an. Bei einem Versuchsprotokoll mit gleichem Zahlen-Buchstaben-Code handelte es sich um dasselbe Versuchsprotokoll, das je nach Erhebungsfokus (i, ii, iii) unterschiedliche Daten lieferte.

Tab. 4: Überblick über die Datenerhebung

		KG	TG1	TG2
Prä-Test	i) Nutzung der Wortspeicher	1 A	1 B	1 C
	ii) Sprachregister	1 A	1 B	1 C
	iii) Inhalt	1 A	1 B	1 C
Post-Test	i) Nutzung der Wortspeicher	2 A	2 B	2 C
	ii) Sprachregister	2 A	2 B	2 C
	iii) Inhalt	2 A	2 B	2 C
Follow-Up-Test	i) Nutzung der Wortspeicher	3 A	3 B	3 C
	ii) Sprachregister	3 A	3 B	3 C
	iii) Inhalt	3 A	3 B	3 C

Aus Gründen der Lesbarkeit wurden Punkt- und Kommata in den Versuchsprotokollen vor der Auswertung sinnergebend eingesetzt und Rechtschreibfehler korrigiert. Durch die **Korrektur der Rechtschreibfehler** sind ursprünglich falsch geschriebene Wörter und Formulierungen Teil der Auswertung. Grundlage für dieses Vorgehen ist im Fokus auf die **Sprachhandlungsfähigkeit** der Vpn zu sehen.¹⁸ Hierbei wird die gelingende Kommunikation als sprachliches Ziel im Fachunterricht genannt, während die sprachliche Korrektheit zweitrangig ist (Tajmel & Hägi-Mead, 2017, S. 16). Orientiert an dem Ziel der Sprachhandlungsfähigkeit stand in diesem Forschungsvorhaben nicht die korrekte Rechtschreibung, sondern die beabsichtigte und erkennbar umgesetzte Anwendung der Sprachmittel im Fokus.

Alle erhobenen Materialien und Daten sind im Institut für Didaktik des Sachunterrichts an der Universität zu Köln archiviert und über die Institutsleitung zugänglich.

4.4 Datenauswertung

4.4.1 C-Test

Die Auswertung des **C-Tests** kann auf verschiedene Arten erfolgen. Zumeist werden bei C-Tests nur die Richtig-Falsch-Werte (R/F-Werte) ausgewertet und interpretiert (Baur, Goggin, & Wrede-Jackes, 2013, S. 8). Dabei handelt es sich um Items, bei denen Semantik, Orthografie als auch die grammatikalische Anpassung der jeweiligen Worthälfte einwandfrei korrekt angegeben sind. Würden – umgekehrt – Orthografiefehler und Varianten als korrekt berücksichtigt, könnte dies die Auswertungsreliabilität beeinträchtigen (Grotjahn, Klein-Braley, & Raatz, 1992, S. 283). Um dies zu vermeiden, wurde in diesem Forschungsvorhaben auf Grundlage der ersten Pilotstudie eine Auswertungsvorschrift mit den original einzusetzenden Items und den zugelassenen Varianten erstellt. Beispielsweise wurde im ersten Text des C-Tests *Eine Schule in Australien* bei dem Item *de_____*⁽¹⁰⁾ sowohl das Original *deshalb* als auch die Variante *deswegen* akzeptiert (siehe Anhang A.1). Varianten außerhalb der Auswertungsvorschrift, Orthografiefehler, sowie grammatikalisch falsche Anpassungen wurden bei der Auswertung nicht berücksichtigt. Schrieb eine Vpn *Für die Feuerwehr *arbeiten er* oder **dass heißt* anstelle von *das heißt*, so erhielt die Vpn für das spezielle Item keinen Punkt. Die Auswertung des C-Tests, in dem Rechtschreibfehler aus Gründen der Auswertungsreliabilität zum Ausschluss des jeweiligen Items führten, ist nicht zu verwechseln mit den Versuchsprotokollen, in denen leichte Rechtschreibfehler zur Vorbereitung auf die Auswertung korrigiert wurden. Da jeder der fünf Texte des C-Tests 20 Items umfasst, war eine maximale Gesamtpunktzahl von 100 möglich. Zum Vergleich der Punktzahl-

¹⁸ vgl. Kapitel 2.1.3

Mittelwerte zwischen KG, TG1 und TG2 wurde die Durchführung einer einfaktoriellen Varianzanalyse (ANOVA) angestrebt. Die inferenzstatistische Auswertung mittels ANOVA wird im Kapitel *Erste Teilstudie (F1)* in Bezug auf die Auswertung der ersten Forschungsfrage genauer vorgestellt.

4.4.2 Mixed-Methods als übergreifendes Forschungsdesign

Die Frage nach dem Nutzen von Wortspeichern im naturwissenschaftlichen Sachunterricht wurde in diesem Forschungsvorhaben auf verschiedenen Ebenen untersucht. Insgesamt handelt es sich um einen **Mixed-Methods** Ansatz, d.h. quantitative als auch qualitative Daten werden im gleichen Forschungsvorhaben erhoben, ausgewertet und schlussendlich miteinander verknüpft (Kuckartz, 2014, S. 33). Die Kombination beider Methoden birgt ein großes Potenzial, das vor allem in dem höheren Grad der Differenziertheit und somit dem besseren Verständnis des Untersuchungsgegenstandes zu sehen ist (Creswell, 2014, S. 18-19; Kuckartz, 2014, S. 54). Zur Eingrenzung der Mixed-Methods Forschung stehen verschiedene Designtypen zur Verfügung. Da das Mixed-Methods-Design in diesem Forschungsvorhaben zur Ermittlung der Wirkung einer Treatmentmaßnahme angewandt wird, lässt es sich grundsätzlich als **Evaluationsdesign** bezeichnen (Kuckartz, 2014, S. 95). Dieses erfordert mindestens zwei MZP (ebd., S. 95). Das Ziel der übergreifenden Forschungsfrage ist ein besseres Verständnis über den Nutzen von Wortspeichern. Um diesbezüglich detaillierte Einsichten zu erlangen, wurde neben den zwei MZP des Prä- und Post-Tests auch ein 10 Wochen später stattfindender Follow-Up-Test durchgeführt. Da die Auswertung zu allen drei MZP am gleichen Dokument und sowohl qualitativ als auch quantitativ stattfand, ist die Bezeichnung **Pre-Post-Follow-Up Paralleldesign** zutreffend (ebd., S. 95).

Abb. 19 skizziert dieses Paralleldesign. Im vorliegenden Forschungsprozess handelt es sich bei der ersten Forschungsfrage (F1) um eine quantitativ ausgerichtete Erhebung und Analyse. Deskriptive Ergebnisse, die Voraussetzungen zur Inferenzstatistik und inferenzstatistische Ergebnisse werden im Ergebnisteil für die erste Forschungsfrage detailliert dargestellt. Um Erkenntnisse zur zweiten und dritten Forschungsfrage (F2 & F3) zu erlangen, wurden die gleichen Dokumente (hier: Versuchsprotokolle) hinsichtlich qualitativer Merkmale (hier: Verwendung der Bildungssprache und Beurteilung des Inhalts) untersucht. Für ein genaueres Verständnis wurden die qualitativ erhobenen Daten zur zweiten und dritten Forschungsfrage ebenfalls quantitativ analysiert. Die vierte Fragestellung (F4) dient der Verknüpfung von qualitativen und quantitativen Ergebnissen und ist somit Grundlage für das Mixed-Methods-Verfahren. Im Schaubild ist dies durch die Meta-Inferenzen gekennzeichnet. Meta-Inferenz ist „[...] die vergleichende Betrachtung der Schlussfolgerungen (Inferenzen) beider Stränge [...]“ (Kuckartz, 2014, S. 101). Im Fall des vorliegenden Forschungsvorhabens werden die Ergebnisse vergleichend

betrachtet, während die Schlussfolgerungen Teil der späteren Diskussion darstellen. Um Wiederholungen zu vermeiden, fand die eingehende Analyse und Ergebnisdarstellung der qualitativ erhobenen Daten erst bei Verknüpfung aller Ergebnisse in der vierten Teilstudie (F4) statt.

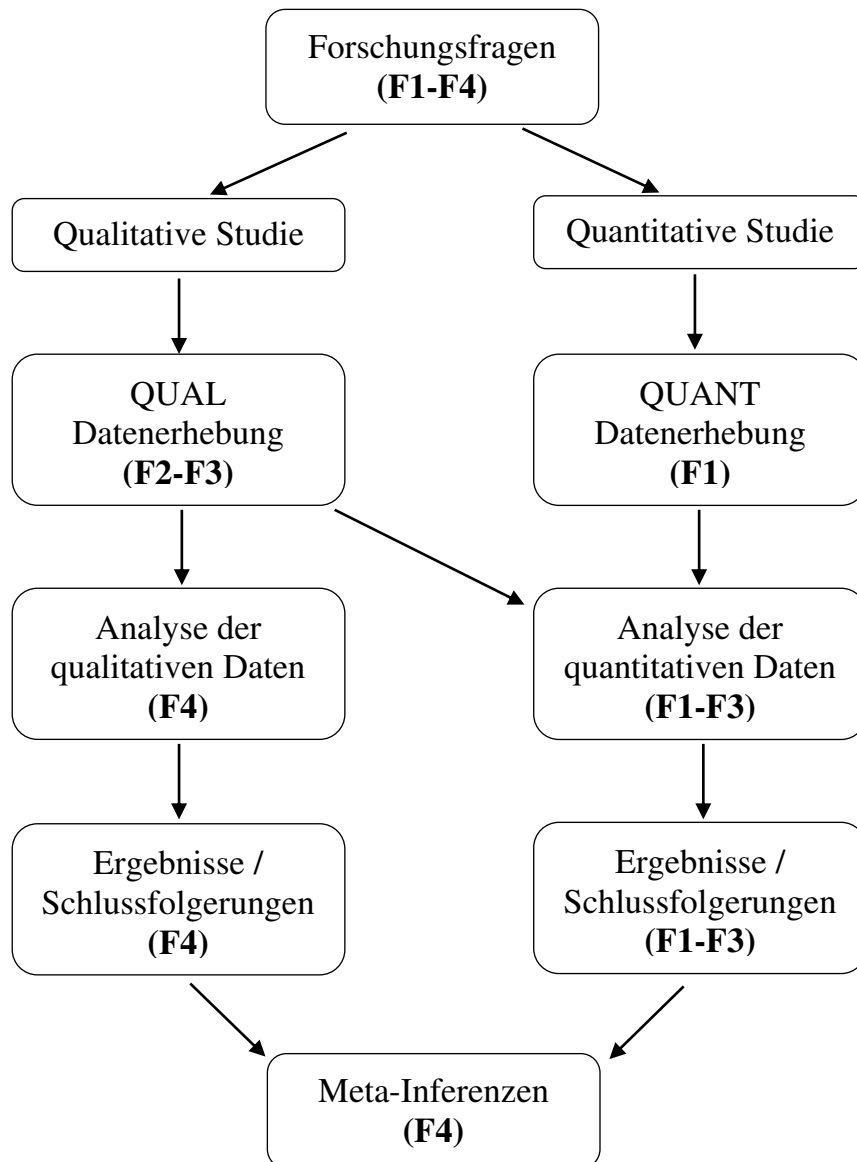


Abb. 19: Paralleles Design in Anlehnung an Kuckartz (2017, S. 166)

Insgesamt wurden zur Auswertung der übergreifenden Forschungsfrage die Programme **Excel 2022**, **SPSS 29.0** und **MAXQDA 2020** (VERBI Software, 2020) herangezogen. Die genaue Beschreibung der Methodik wird im Folgenden deutlich. Ausgehend von jeder einzelnen Forschungsfrage wird die Datenauswertung im Detail erläutert.

4.4.3 Erste Teilstudie (F1)

In der ersten Teilstudie geht es darum, inwieweit die Vpn den Wortspeicher nutzen, wenn dieser zuvor in Plakatform (TG 2), auditiv (TG 1) oder nicht (KG) zur Verfügung gestellt wurde. Zur Untersuchung der Forschungsfrage wurden die Wörter und Formulierungen aus dem Wortspeicher in den einzelnen Versuchsprotokollen mithilfe der **Worthäufigkeitsfunktion in MAXQDA** ermittelt. Die in MAXQDA festgestellte Anzahl der verwendeten Sprachmittel pro Versuchsprotokoll wurde in **SPSS** ausgewertet.

Im Kapitel *Forschungsfragen* wurden für die erste Teilstudie zwei Unterschiedshypothesen vorgestellt. Der ersten Hypothese zufolge ist die Nutzung der Wortspeicher in TG2 höher als in TG1. Der zweiten Hypothese zufolge ist die Nutzung der Wortspeicher in TG1 höher als in der KG.

Zur Prüfung der Hypothesen wurde die Durchführung einer einfaktoriellen Varianzanalyse (ANOVA) im Post-Test und im Follow-Up-Test angestrebt. Mithilfe der ANOVA soll geprüft werden, „[...] ob sich mindestens zwei der drei untersuchten Gruppen zufällig voneinander unterscheiden“ (Döring & Bortz, 2016, S. 709). In anderen Worten war eine **3x1 faktorielle Varianzanalyse** jeweils für MZP 2 und MZP 3 geplant, in der die Gruppenzugehörigkeit (KG, TG1, TG2) als Zwischensubjektfaktor (*between*) untersucht wurde. Die Gruppenunterschiede stellen dabei die unabhängige Variable (UV) und die Anzahl der verwendeten Wortspeicher-Sprachmittel in den Versuchsprotokollen die abhängige Variable (AV) dar. Das Signifikanzniveau wurde auf 5 % festgelegt.

Zu den Voraussetzungen für die spätere Interpretation des 3x1 faktoriellen Designs zählen neben der allgemein erforderten Additivität, Linearität und Unabhängigkeit, vor allem **Normalverteilung** und **Varianzhomogenität** (Field, 2018, S. 230, 534). Diese galt es vorab zu prüfen. Da die ANOVA bei Verletzung der Normalverteilung und unter weiteren variierenden Bedingungen robust zu sein scheint (Blanca, Alarcón, Arnau, Bono, & Bendayan, 2017, S. 556) wurde die ANOVA auch bei Voraussetzungs-Verletzung durchgeführt und zur Kontrolle mit dem Ergebnis des nicht-parametrischen Kruskal-Wallis Tests verglichen.

Bei Annahme der Varianzhomogenität wurde eine Bonferroni-korrigierte Post-hoc-Analyse durchgeführt. Mithilfe der Bonferroni-Korrektur wird die Alphafehlerkumulierung überwacht (Hemmerich, 2016). Bei ungleichen Varianzen in den Gruppen wurde auf den dafür besser geeigneten Games-Howell-Test zurückgegriffen (Brosius, 2018, S. 599).

Neben der Hypothesenprüfung war die Durchführung einer ANOVA zur Prüfung von Unterschieden zwischen den drei MZP innerhalb einer Versuchsgruppe geplant. In anderen Worten sollte mithilfe einer **1x3 faktoriellen Varianzanalyse** das Ergebnis über die Zeit (PRÄ, POST, FOLLOW-UP) als Innersubjektfaktor (*within*) betrachtet werden. Dabei bildet das jeweilige Treatment erneut die UV und die Anzahl der verwendeten

Sprachmittel aus dem Wortspeicher stellen die AV dar. Neben den bereits genannten Bedingungen zur Durchführung der ANOVA ist für das 1x3 faktorielle Design zusätzlich **Sphärizität** erforderlich (Field, 2018, S. 230, 663). Sphärizität ist bei Varianzhomogenität zwischen den MZP gegeben. Eine Testung auf Varianzhomogenität zwischen den Gruppen entfällt, da in dieser Analyse ein Innersubjektfaktor anstelle von unabhängigen Stichproben getestet wird. Bei Verletzung einer oder mehrerer Voraussetzungen wurde das Resultat der 1x3 ANOVA mit dem Ergebnis des nicht-parametrischen **Friedman-Tests** kontrolliert.

4.4.4 Zweite Teilstudie (F2)

Die zweite Teilstudie ist vor dem Hintergrund zu betrachten, dass der angebotene Wortspeicher eine künstliche Auslese an bildungssprachlichen Wörtern und Formulierungen darstellt. Zur Ermittlung, inwieweit die untersuchten Vpn andere bildungssprachliche Sprachmittel außerhalb des Wortspeichers in ihren Versuchsprotokollen anwandten, eignet sich eine „**qualitativ orientierte kategoriengeleitete Textanalyse** [Hervorhebung d. Verf.]“ (Mayring, 2019, S. 2). Diese Form der Textanalyse beinhaltet sowohl qualitative als auch quantitative Schritte und bezeichnet insofern einen Mixed-Methods-Ansatz innerhalb der zweiten Teilstudie. Konkret wurde zur Beantwortung der zweiten Forschungsfrage Bildungssprache qualitativ anhand der Versuchsprotokolle ermittelt. In einem weiteren Schritt wurden die Kategorienhäufigkeiten quantitativ analysiert.

Um eine Konfundierung zwischen den Begriffen *Erhebung*, *Analyse* und *Auswertung* zu vermeiden, werden diese Begriffe für das vorliegende Forschungsvorhaben wie folgt definiert: *Erhebung* bezeichnet die qualitative Ermittlung von Bildungssprache aus den Versuchsprotokollen, der Begriff *Analyse* wird für die quantitative Untersuchung der Kategorienhäufigkeiten verwendet und der Begriff *Auswertung* fasst die qualitativen und quantitativen Schritte zusammen. Der Begriff *Inhaltsanalyse* stellt eine Ausnahme dar, da er sich auf die qualitative Erhebung (d.h. Textanalyse) bezieht.

Die Art und Weise der qualitativen Erhebung ist mithilfe verschiedenster Techniken der **qualitativen Inhaltsanalyse** möglich. Um eine flexible Vorgehensweise zu ermöglichen, in der jeder einzelne Auswertungsschritt passend zum Forschungsinteresse und zu den gegebenen Auswertungsmöglichkeiten gewählt werden kann, wurde auf das Konzept des **Werkzeugkastens** nach Schreier (2014, S. 17) zurückgegriffen, das die verschiedenen Varianten der qualitativen Inhaltsanalyse miteinander vereinbart.

In diesem Konzept werden die einzelnen Optionen des qualitativen Auswertungsverfahrens je nach Analysephase vorgestellt. Tab. 5 bietet eine erste Übersicht über die gewählten Optionen bzw. *Werkzeuge* in jeder Phase der Inhaltsanalyse.

Tab. 5: Inhaltsanalyse zu F2 nach dem Werkzeugkasten von Schreier (2014, S. 17-18)

Schritte der Inhaltsanalyse	„Werkzeuge“ / Optionen
(a) Festlegen der Forschungsfrage	Lässt sich durch visuell oder auditiv angebotene Wortspeicher die Häufigkeit bildungssprachlicher Wörter und Formulierungen in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts erhöhen?
(b) Auswahl des Materials	Versuchsprotokolle der Vpn
(c) Erstellen des Kategoriensystems	Deduktiv-induktiv <ul style="list-style-type: none"> • deduktiv: <ul style="list-style-type: none"> ○ <i>Sprachhandlungen, Lexikon, Morphosyntax</i> (RaBi-Skala) • induktiv: <ul style="list-style-type: none"> ○ offenes Kodieren von bildungssprachlichen Wörtern und Formulierungen, die der RaBi-Skala nicht zugeordnet werden konnten; mögliche Änderung der RaBi-Skala → Ergebnis des induktiven Kodierens: <ul style="list-style-type: none"> • Löschung der Unterkategorie <i>Komplexes Verbgefüge</i> • Hinzufügen einzelner Wörter zur Kategorie <i>Lexikon - Sonstiges</i> • Ermittlung von Ankerbeispielen
(d) Unterteilung des Materials in Einheiten	Kontexteinheit: <ul style="list-style-type: none"> • das gesamte Versuchsprotokoll Kodiereinheit: <ul style="list-style-type: none"> • <i>Sprachhandlungen</i>: Wort, Formulierung oder Satz • <i>Lexikon</i>: Wort • <i>Morphosyntax</i>: Formulierung
(e) Probekodierung	Kontinuierliche Überarbeitung und Anpassung <ul style="list-style-type: none"> • anhand 100 % des Materials der ersten Pilotstudie
(f) Evaluation und Modifikation des Kategoriensystems	Kodiervergleich mit Intercoder <ul style="list-style-type: none"> • anhand eines Versuchsprotokolls je MZP aus der ersten Pilotstudie <ul style="list-style-type: none"> → Kodierbesprechung mit Intercoder und Anpassung des Kategoriensystems
(g) Hauptkodierung	Durchführung der Kodierung durch Versuchsleitung <ul style="list-style-type: none"> • 100 % des Materials Kodiervergleich mit Intercoder <ul style="list-style-type: none"> • 10 % des Materials • Berechnung des Intercoder-Koeffizienten in MAXQDA • Kodierbesprechung mit Intercoder, Anpassung des Kategoriensystems und Überprüfung der Neuerungen anhand 100 % des Materials Besprechung von diskussionswürdigen Kodierungen mit Deutschlehrkraft
(h) Weitere Auswertung und Ergebnisdarstellung	<ul style="list-style-type: none"> • Beschreibung von Kodier-Häufigkeiten • Statistische Auswertung von Häufigkeitsunterschieden

Die einzelnen Schritte der Inhaltsanalyse werden im Folgenden ausführlicher beschrieben. Die Buchstaben in Klammern kennzeichnen die aufeinander aufbauenden Phasen des Auswertungsverfahrens in Anlehnung an Schreier (2014, S. 17-18).

Nach Betrachtung der Forschungsfrage (a) ist die Auswahl des Materials (b) entscheidend. In diesem Forschungsvorhaben handelte es sich um verschriftlichte Versuchsprotokolle von Kindern einer dritten Grundschulklasse. Die Erstellung des Kategoriensystems (c) zur Erfassung des Sprachregisters erfolgte **deduktiv-induktiv**. **Deduktiv**, d.h. theoriegeleitet, wurden die Ober- und Unterkategorien vorgegeben. Die deduktive Kategorienbildung ist darauf zurückzuführen, dass bereits empirisch hergeleitete Kriterien der kindlichen Bildungssprache existieren, sodass diese verwendet werden konnten. In Kapitel 2.1.4 wurden die RaBi-Skala (Tietze et al., 2016) und das WuW-Modell (Juskabacher & Beckert, 2015) vorgestellt. Die beschriebene kognitiv-reflexive und die entwicklungspsychologische Dimension der Bildungssprache aus dem WuW-Modell sind bei der Bewertung der schriftlichen Protokolle schwer zu erfassen. Im Gegensatz zur Analyse medial mündlicher Sprache würde ein Kind im medial Schriftlichen beispielsweise seltener die eigene Formulierung für den Lesenden offenkundig reflektieren oder korrigieren. Deshalb wurde für dieses Forschungsvorhaben auf die gut analysierbaren Oberkategorien *Sprachhandlungen*, *Lexikon* und *Morphosyntax* und auf die jeweiligen Unterkategorien der RaBi-Skala zurückgegriffen. Mayring (2019, S. 7) betont die Bedeutung der Kodierregeln zur Abgrenzung zwischen den einzelnen deduktiven Kategorien, sodass die Kodierregeln auch in dem vorliegenden Kategoriensystem explizite Anweisungen zur Kodierung vorgaben. Neben dem deduktiv erstellten Kategoriensystem wurde das Kategoriensystem in diesem Forschungsvorhaben nach einer ersten Pilotstudie **induktiv**, d.h. durch Entwicklung am Material, angepasst. Damit ließ sich die Erfassung des Sprachregisters flexibel gestalten und unerwartete Kriterien wurden mitberücksichtigt. Die induktive Kategorienbildung geschah durch offenes Kodieren der Wörter und Formulierungen, die als bildungssprachlich angesehen werden konnten und keiner der deduktiven Kategorien zuzuordnen waren.

Im Ergebnis konnten einige bildungssprachliche Wörter keiner Kategorie aus der Oberkategorie *Lexikon* zugeordnet werden. Da es sich um nur wenige Wörter handelte, wurden diese dem Bereich *Sonstiges* zugeordnet, sodass keine neue Kategorie entstand. Weiterhin wurde das deduktiv erstellte Kategoriensystem in der ersten Pilotstudie bzgl. seiner Eignung überprüft. Die Prüfung ergab, dass die Unterkategorie *Komplexes Verbgefüge* in den Versuchsprotokollen der Vpn kein bildungssprachliches Element darstellte und somit aus dem deduktiven Kategoriensystem entfernt wurde. Da durch das offene Kodieren keine neuen Kategorien entstanden, ist das final erstellte Kategoriensystem hauptsächlich als deduktiv erstelltes Kategoriensystem zu betrachten. Bereits in der ersten Pilotstudie wurde das Kategoriensystem an 100 % des Materials getestet. Auch die induktive Anpassung erfolgte an dem Gesamtmaterial. Dies war zur Erstellung

eines möglichst genauen Kategoriensystems angebracht. Das finale Kategoriensystem bestand nach der ersten Pilotstudie – in Anlehnung an Tietze et al. (2016) – insgesamt aus den in Abb. 20 ersichtlichen Kategorien:

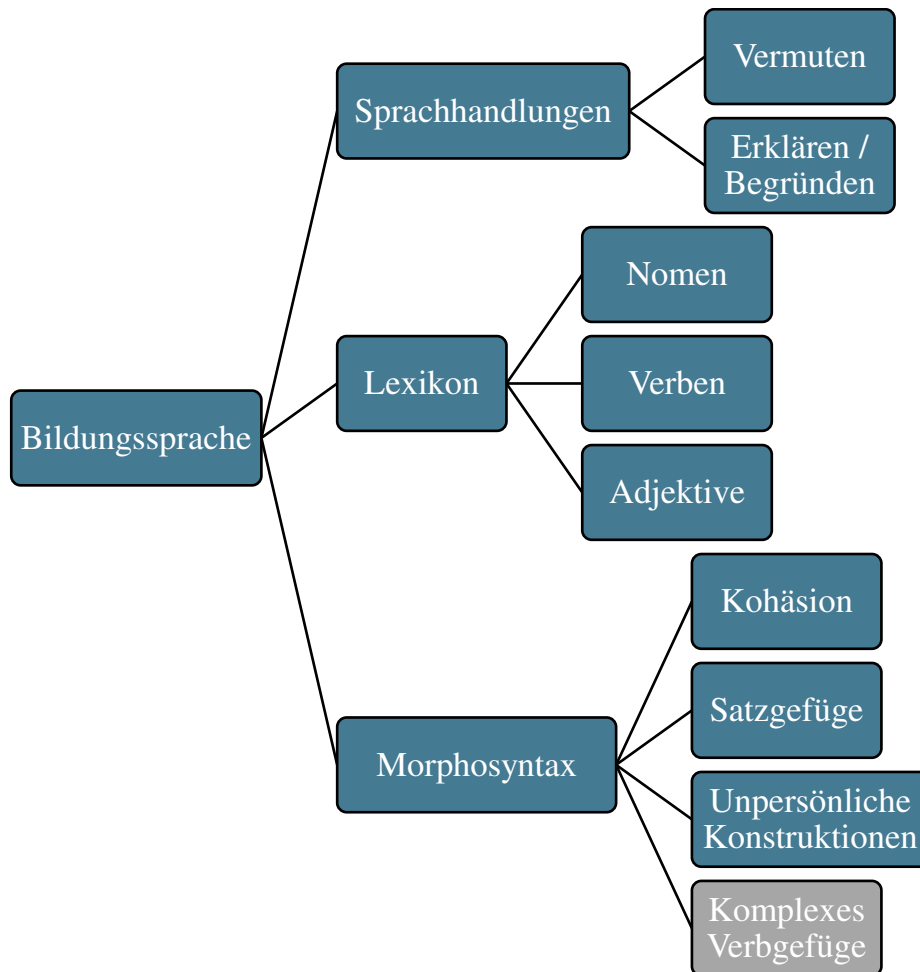


Abb. 20: Ober- und Unterkategorien zur Kodierung von Bildungssprache

(hier: ursprüngliches Kategoriensystem; im induktiv geänderten Kategoriensystem für die Hauptstudie entfiel die Kategorie *komplexes Verbgefüge*)

Abb. 21 zeigt einen Ausschnitt aus dem deduktiv erstellten Kategoriensystem zur Oberkategorie *Sprachhandlungen*. Die Ankerbeispiele wurden anhand von Versuchsprotokollen der ersten Pilotstudie induktiv ermittelt. Das gesamte Kategoriensystem ist im Anhang B.1 einsehbar.

Sprachhandlungen	Die Sprachhandlungen sind eng mit den im Unterricht verwendeten Operatoren – wie beispielsweise Benenne, Beschreibe, Erkläre – verknüpft. Da das Aufstellen einer Vermutung sowie das Erklären und Begründen anspruchsvolle Sprachhandlungen darstellen und diese oft mit bildungssprachlichen Mitteln einhergehen (Tietze et al., 2016, S. 13), wird das Vorkommen dieser zwei Sprachhandlungstypen in den Versuchsprotokollen genauer untersucht und entsprechend kodiert. Je nach Kontext werden Wörter, Phrasen oder ganze Sätze kodiert. Wichtig ist, dass die Sprachhandlung in den Kodierungen erkennbar ist.		
Unterkategorien	Zuordnung	Ankerbeispiele	Abgrenzung
<i>Vermuten</i>	<p>Diese Kategorie ist zu kodieren, wenn in den Versuchsprotokollen eine sprachlich ersichtliche Vermutung aufgestellt wird.</p> <p>Die Formulierung ist sprachlich korrekt und wird im Kontext richtig angewandt.</p> <p>Explizit ausgedrückte Vermutungen werden auch außerhalb der Rubrik <i>Vermutungen</i> im Versuchsprotokoll markiert.</p>	<ul style="list-style-type: none"> ✓ Ich glaube, die Lampe wird leuchten. ✓ Die Lampe könnte leuchten. ✓ Ich vermute die Lampe leuchtet. 	<p>Die folgenden Formulierungen aus dem Wortspeicher werden <u>nicht</u> kodiert:</p> <ul style="list-style-type: none"> ✗ Meine Vermutung ist, dass... ✗ Ich vermute, dass... ✗ Ich denke, dass... <p>Das folgende Wort drückt nur unzureichend eine Vermutung aus und wird deshalb nicht kodiert:</p> <ul style="list-style-type: none"> ✗ würde
<i>Sonstiges</i>	Diese Kategorie ist zu kodieren, wenn in den Versuchsprotokollen weitere Sprachhandlungen erkennbar sind, die offensichtlich mit Bildungssprache einhergehen. Zur weiteren Differenzierung ist an dieser Stelle beispielsweise die Einleitung einer Materialliste (<i>Man braucht..., Ich benutze...</i>), die Einleitung einer Durchführung (<i>So wird der Versuch durchgeführt: ...</i>) und die Einleitung einer Beobachtung (<i>Ich beobachte, dass...</i>) zu kodieren.		

Abb. 21: Auszug aus dem Kategoriensystem zur Kodierung von Bildungssprache

Wichtig für die Intercoder-Übereinstimmung ist die Unterteilung des Materials in Einheiten (d) und, damit verbunden, die Festlegung der Kontext- und Kodiereinheiten. Als Kontexteinheit ist in diesem Forschungsvorhaben das gesamte Versuchsprotokoll eines Kindes inklusive der Kenntnis, ob das Kind mit einer bestimmten Form von Wortspeichern gelernt hat, zu nennen. Kodiert wurde in jeder Kontexteinheit entweder ein Wort oder eine Formulierung. Die Kodiereinheit des einzelnen Wortes bezog sich vor allem auf die Oberkategorie *Lexikon*. Demgegenüber wurden bei der Oberkategorie *Morphosyntax* hauptsächlich Formulierungen kodiert, die unterschiedliche Längen aufweisen konnten. In der Oberkategorie *Sprachhandlungen* wurden entweder einzelne Wörter (z.B. *damit*) als auch Formulierungen (z.B. *Ich glaube, dass...*) oder ganze Sätze (z.B. *Die Lampe könnte leuchten.*) kodiert. Die Größe der Kodiereinheit hing davon ab, bei welcher Kodierlänge eine Sprachhandlung erst ersichtlich war. Während die Formulierung *Ich glaube, dass...* bereits einen ausreichenden Hinweis auf eine Vermutung liefert, so ist dies bei *Die Lampe könnte leuchten* erst im ganzen Satz erkennbar. Wichtig war außerdem, dass nicht mehr kodiert wurde, als zur Erfassung einer Ober- bzw. Unterkategorie notwendig war. Da einige Wörter oder Formulierungen zur Zusammenfügung von Sätzen genutzt wurden, konnte eine Kodierung auch satzübergreifend sein. Dies war vor allem bei Kodiereinheiten zur Unterkategorie *Kohäsion* der Fall. An dieser Stelle wurde das letzte Wort des vorangehenden Satzes und das nächste Wort oder die nächste Formulierung des folgenden Satzes kodiert (z.B. *passiert. Aber...*). Demgegenüber wurde eine Kodierlänge von mehr als zwei Sätzen vermieden.

Anstelle einer weiteren Probekodierung (e) wurde das Kategoriensystem unter Einbezug aller Versuchsprotokolle der ersten Pilotstudie kontinuierlich angepasst. Die Evaluation und Modifikation des Kategoriensystems (f) fand bereits bei der Erstellung des Kategoriensystems statt. Zur Erhöhung der Reliabilität des Kategoriensystems wurden drei ausgeloste Versuchsprotokolle in der ersten Pilotphase von einer zweiten Person gegenkodiert und die Kodierregeln entsprechend diskutiert und angepasst.

In der darauffolgenden Hauptkodierung (g) wurde das Kategoriensystem an 100 % des Materials der Hauptstudie angewandt. Um Doppelkodierungen zu vermeiden, wurde den Oberkategorien eine Priorität zugesprochen. Da vor allem die Sprachhandlungen für das Verfassen der Versuchsprotokolle von Interesse sind, wurden die Versuchsprotokolle zuerst in Hinblick auf die Unterkategorien *vermuten* und *erklären/begründen* untersucht. Wurde das Wort *damit* bereits an dieser Stelle in Bezug auf die Sprachhandlung des *Erklärens und Begründens* kodiert, so war keine erneute Kodierung unter der Oberkategorie *Lexikon* möglich. Da Überschneidungen zwischen der Oberkategorie *Lexikon* und *Morphosyntax* aufgrund ihrer definierten Abgrenzung nicht oder kaum möglich waren, fiel eine weitere Hierarchisierung aus. Die Hauptkodierung wurde von der Autorin vorgenommen. 10 % des Materials wurden von einer zweiten Person kodiert. Die zu überprüfende Prozentzahl an Daten (10 %) wurde in Anlehnung an eine

empirische Studie mit Intercodierung von Krepf (2019, S. 175) festgelegt. Die Berechnung des Intercoder-Koeffizienten in MAXQDA gab Aufschluss über die Übereinstimmungsquote der Kodierungen. Bei Nicht-Übereinstimmung wurden ausgewählte Kodierungen mit einer Deutschlehrkraft besprochen.

Bezüglich der weiteren Auswertung und Ergebnisdarstellung (h) wurde zunächst eine beschreibende Darstellung des kodierten Materials vorgenommen. Dadurch konnten die Ergebnisse der Kodierhäufigkeiten pro Ober- und Unterkategorie detailliert abgebildet werden. In einem zweiten Schritt wurden die Kodierhäufigkeiten zusammengefasst, um die Anzahl der bildungssprachlichen Wörter und Formulierungen außerhalb des Wortspeichers in ihrer Gesamtheit zu analysieren. Um den gesamten Wert bei Mehrfachkodierungen abzusichern, wurde zusätzlich die Gesamtanzahl bei Einfachkodierungen ermittelt. Dabei werden bildungssprachlich kodierte Wörter und Formulierungen nur einmal pro Dokument gezählt. Die Gesamtanzahl bei Einfachkodierungen diente der Absicherung, dass die Ergebnisse bei Mehrfachkodierungen nicht zum großen Teil auf wiederholten Sprachmitteln beruhten.

Nach der Erhebung mittels qualitativer Inhaltsanalyse wurde die Gesamtanzahl an bildungssprachlichen Mitteln bei Mehrfach- und Einfachkodierungen mittels 3x1 und 1x3 faktorieller Varianzanalyse quantitativ analysiert. Die Vorgehensweise der quantitativen Analysen richtete sich an die Beschreibung, die im Kapitel *Erste Teilstudie (F1)* nachzulesen ist.

4.4.5 Dritte Teilstudie (F3)

Im Folgenden wird die Auswertung hinsichtlich der dritten Teilstudie erörtert. Da die möglichen inhaltlichen Antworten der Kinder in den Versuchsprotokollen nicht von vornherein vorhersehbar waren, wurde auch für diese Fragestellung eine „**qualitativ orientierte kategoriengeleitete Textanalyse** [Hervorhebung d. Verf.]“ (Mayring, 2019, S. 2) angesetzt. Dies bedeutet, dass die dritte Teilstudie – ebenso wie Teilstudie 2 – den Mixed-Methods-Ansatz mit den aufeinanderfolgenden Schritten der qualitativen Erhebung und der quantitativen Analyse beinhaltet.

Wie in der beschriebenen Auswertung der zweiten Forschungsfrage, fand die Auswertung der dritten Teilstudie durch eine individuelle Auswahl von Optionen aus dem **Werkzeugkastenmodell** der qualitativen Inhaltsanalyse nach Schreier (2014, S. 17-18) statt. Es werden im Folgenden erneut die aufeinander aufbauenden Phasen und die jeweiligen begründet ausgewählten Werkzeuge vorgestellt. Tab. 6 bietet eine Übersicht über die gewählten Optionen für die dritte Teilstudie.

Tab. 6: Inhaltsanalyse zu F3 nach dem Werkzeugkasten von Schreier (2014, S. 17-18)

Schritte der Inhaltsanalyse	„Werkzeuge“ / Optionen
(a) Festlegen der Forschungsfrage	Lässt sich durch visuell oder auditiv angebotene Wortspeicher der Inhalt in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts verbessern?
(b) Auswahl des Materials	Versuchsprotokolle der Vpn
(c) Erstellen des Kategoriensystems	Deduktiv-induktiv <ul style="list-style-type: none"> • deduktiv: <ul style="list-style-type: none"> ○ <i>korrekt, unzureichend, fehlerhaft</i> • induktiv: <ul style="list-style-type: none"> ○ genaue Abgrenzung zwischen den Kategorien; mögliche Änderungen <p>→ Ergebnis des induktiven Kodierens:</p> <ul style="list-style-type: none"> • Hinzufügen einer neuen Kategorie <i>Keine Antwort</i> zur genaueren Betrachtung der Kategorie <i>fehlerhaft</i> • Ermittlung von Ankerbeispielen
(d) Unterteilung des Materials in Einheiten	Kontexteinheit: <ul style="list-style-type: none"> • das gesamte Versuchsprotokoll Kodiereinheit: <ul style="list-style-type: none"> • gesamter Abschnitt (z.B. <i>Die erste Vermutung; die zweite Beobachtung</i>)
(e) Probekodierung	Kontinuierliche Überarbeitung und Anpassung <ul style="list-style-type: none"> • anhand 100 % des Materials der ersten Pilotstudie
(f) Evaluation und Modifikation des Kategoriensystems	Kodiervergleich mit Intercoder <ul style="list-style-type: none"> • anhand eines Versuchsprotokolls je MZP aus der ersten Pilotstudie <p>→ Kodierbesprechung mit Intercoder und Anpassung des Kategoriensystems</p>
(g) Hauptkodierung	Durchführung der Kodierung durch Versuchsleitung <ul style="list-style-type: none"> • 100 % des Materials Kodiervergleich mit Intercoder <ul style="list-style-type: none"> • 10 % des Materials • Berechnung des Intercoder-Koeffizienten in MAXQDA • Kodierbesprechung mit Intercoder, Anpassung des Kategoriensystems und Überprüfung der Neuerungen anhand 100 % des Materials
(h) Weitere Auswertung und Ergebnisdarstellung	<ul style="list-style-type: none"> • Beschreibung von Kodier-Häufigkeiten • Statistische Auswertung von Häufigkeitsunterschieden

Zur Untersuchung der dritten Forschungsfrage (a) boten sich die Versuchsprotokolle der Vpn als Auswertungsmaterial an (b). Es handelte sich um dieselben Versuchsprotokolle, die auch in der ersten und zweiten Teilstudie unter einer anderen Perspektive beleuchtet wurden.

Das Kategoriensystem wurde **deduktiv-induktiv** erstellt (c). Zunächst wurde der *Inhalt* deduktiv als Oberkategorie definiert. Die Erstellung der Unterkategorien erfolgte deduktiv-induktiv. Im Interesse der vorliegenden Forschung ist die Einschätzung der inhaltlichen Qualität der Versuchsprotokolle. Bei der Analyse des Inhalts geht es somit um eine Bewertung. Voraussetzung dafür sind **bewertende Codes** (Kuckartz, 2014, S. 112). Um eine zu große Bandbreite an bewertenden Kategorien und eine damit einhergehende diffuse Auswertung zu vermeiden, wurden drei Kategorien zur inhaltlichen Beurteilung gewählt: *korrekt*, *unzureichend* und *fehlerhaft*. Eine Kategorie zur Spezifizierung der inhaltlichen Antwort war nicht im Forschungsinteresse. Die genaue Abgrenzung der Unterkategorien wurde induktiv ermittelt. Sie konnte je nach Thema des Versuchsprotokolls und somit je nach MZP variieren. Es ergaben sich folgende finale Kategorien (Abb. 22):

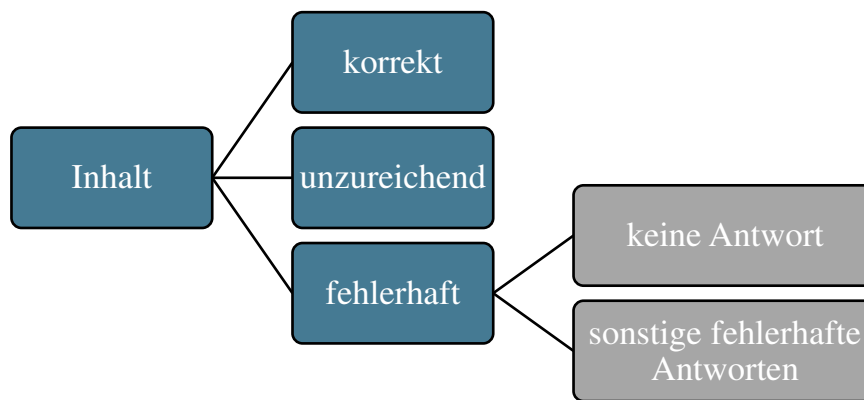


Abb. 22: Ober- und Unterkategorien zur Kodierung des Inhalts

Die Beschreibung der Kategorien ist in dem ausführlichen Kategoriensystem im Anhang B.2 einzusehen. Da die Kategorie *fehlerhaft* sowohl falsche als auch fehlende Antworten einschließt, wurden innerhalb der Kategorie *fehlerhaft* die fehlenden Antworten zusätzlich mit der Subkategorie *keine Antwort* kodiert. Diese induktive Änderung des Kategoriensystems ermöglichte eine vertiefte Untersuchung, wie oft Vpn z.B. aus Gründen der Unkenntnis oder aus zeitlichen Gründen nichts schrieben. Die beiden Kategorien *keine Antwort* und *sonstige fehlerhafte Antworten* sind in Abb. 22 grau markiert. Dies verdeutlicht, dass die zusätzliche Kodierung nicht Teil des Kategoriensystems war, sondern **ergänzend** innerhalb der bestehenden Kategorie durchgeführt wurde.

Die Kontexteinheit setzte sich erneut aus dem gesamten Versuchsprotokoll eines Kindes, inklusive der Kenntnis, ob das Kind mit einer bestimmten Form von Wortspeichern gelernt hat, zusammen (d). Im Gegensatz zur zweiten Fragestellung wurde in dieser Teilstudie nicht ein Wort oder eine Formulierung, sondern jeweils der gesamte

Abschnitt kodiert (e). In jedem Versuchsprotokoll gab es insgesamt neun Abschnitte, die es inhaltlich zu kodieren galt:

- a) Die *Vermutung* (3x),
- b) die *Durchführung* (1x),
- c) das *Material* (1x) → Gesonderte Kodierung: Zählung der genannten Materialien
- c) die *Beobachtung* (3x) und
- d) die *Erklärung* (1x).

Die Zahl in Klammern zeigt die Anzahl der jeweiligen Abschnitte an. Bei der *Vermutung* handelte es sich um insgesamt drei Vermutungen, die von den Vpn festgehalten werden sollten und entsprechend als individuelle Abschnitte bewertet wurden. Im Anhang A.5 ist das vorstrukturierte Versuchsprotokoll mit der jeweiligen Aufteilung in einzelne Abschnitte erkennbar. Die große Kodiereinheit ist damit zu begründen, dass jeweils der gesamte aufgeschriebene Abschnitt einer der bewertenden Codes zugeordnet werden konnte. Eine Aufspaltung der Antworten in einzelne Kodiereinheiten wurde aus Gründen der Übersichtlichkeit vermieden. Der Abschnitt zur Aufzählung des *Materials* (1x) wurde gesondert kodiert. Zur besseren Vergleichbarkeit wurde die Anzahl der korrekt angegebenen Materialien gezählt.

Das Kategoriensystem wurde in der ersten Pilotstudie an 100 % des Materials getestet und überarbeitet (e). Weiterhin wurden drei ausgeloste Versuchsprotokolle in der ersten Pilotstudie von einer weiteren Person gegenkodiert. Die Kodierungen wurden diskutiert und das Kategoriensystem entsprechend modifiziert (f).

In der Hauptkodierung (g) wurde das Kategoriensystem zur inhaltlichen Kodierung aller Versuchsprotokolle aus der Hauptstudie verwendet. Die Hauptkodierung wurde von der Autorin durchgeführt. 10 % des Materials wurden zusätzlich von einer zweiten Person kodiert. Der mittels MAXQDA berechnete Intercoder-Koeffizient gab Aufschluss über weitere notwendige Anpassungen.

Die Auswertung und Ergebnisdarstellung (h) fand – wie in Teilstudie 2 – in zwei Schritten statt: Zunächst beschreibend und in einem zweiten Schritt statistisch vergleichend. Zur detaillierten Beschreibung gehörte die Angabe, wie viele *Materialien* in den Versuchsgruppen zu den verschiedenen MZP jeweils genannt wurden. Weiterhin wurden die Kodierhäufigkeiten für die einzelnen Unterkategorien (*korrekt, unzureichend, fehlerhaft*) beschrieben, um eine ausführliche Darstellung der Ergebnisse zu garantieren. Im Anschluss wurden die Kodierhäufigkeiten quantitativ ausgewertet.

Eine Besonderheit war hierbei der Umgang mit den bewertenden Codes. Da es sich um ein dreistufiges Kategoriensystem handelt, wurde zur Erfassung gradueller Unterschiede ein Punktecode vergeben. Null Punkte gab es für eine falsche oder fehlende

Antwort (*fehlerhaft*), einen Punkt für eine unvollständige Antwort (*unzureichend*) und zwei Punkte für die richtige Antwort (*korrekt*). Die Bepunktung ergab eine Rangfolge von positiv zu negativ beurteilten Antworten. Die vorgenommenen Kodierungen konnten somit, in Analogie zu einem Test, als erreichte Punkte gewertet werden. Da sie quantitative Ausprägungen der Kategorien darstellen, war mit diesen Werten eine quantitative Auswertung möglich.

Analog zur zweiten Teilstudie wurde die inhaltliche Gesamtpunktzahl abschließend mittels 3x1 und 1x3 faktorieller Varianzanalyse auf inferenzstatistische Signifikanzen geprüft. Die Vorgehensweise ist im Kapitel *Erste Teilstudie (F1)* dargestellt.

4.4.6 Vierte Teilstudie (F4)

Die Auswertung der vierten Teilstudie basierte auf der Datenauswertung zu den vorherigen Forschungsfragen 1, 2 und 3. Als wesentlicher Bestandteil des Mixed-Methods-Verfahrens wurden die quantitativ und qualitativ gewonnenen Daten im Rahmen der vierten Forschungsfrage miteinander verbunden (Schreier & Odağ, 2020, S. 14).

Für einen ersten Überblick über die Verknüpfung der qualitativen und quantitativen Daten wurden *Kreuztabellen* und (*interaktive*) *Segment-Matrizen* in den Mixed-Methods Funktionen von MAXQDA genutzt (Kuckartz, 2014, S. 140). Darauf aufbauend wurden einzelne Aspekte eingehender untersucht. Dazu gehörte die Analyse der quantitativen Daten bezogen auf die einzelnen Abschnitte des Versuchsprotokolls (Vermutung, Material, Durchführung, ...). Die qualitativ erhobenen Daten aus der zweiten und dritten Teilstudie wurden zur Beantwortung der vierten Forschungsfrage im Wortlaut vorgestellt, sofern ihre Auswertung zur Beantwortung der Forschungsfrage beitragen konnte. Überschneidungen der Verwendung von Wortspeichern und Bildungssprache mit den bewertenden inhaltlichen Codes wurden mithilfe des *Code-Relations-Modells* in MAXQDA dargestellt.

Da einige Vpn aus TG1 und TG2 keine Wörter oder Formulierungen aus dem Wortspeicher nutzten, wurde eine weitere Teiluntersuchung zur Unterscheidung von Vpn mit oder ohne Verwendung der Wortspeicher veranlasst. Für diese Teiluntersuchung wurden nur die Vpn aus TG1 und TG2 zum MZP 2 und 3 herangezogen, sodass die Wortspeicher den Vpn bekannt waren. Auch für diese Teiluntersuchung wurden qualitative und quantitative Ergebnisse mithilfe von MAXQDA verknüpft und entsprechend dargestellt.

5 ERGEBNISSE

In diesem Kapitel werden die Ergebnisse bezogen auf die einzelnen Forschungsfragen in den jeweiligen Kapiteln *F1 – Nutzung der Wortspeicher*, *F2 – Sprachregister*, *F3 – Inhalt* und zum Schluss *F4 – Datenintegration* vorgestellt. Da die Kontrolle des Sprachstandes aller drei Klassen sowie die Intercoder-Übereinstimmung der Kodierungen für mehrere oder alle Forschungsfragen von Relevanz ist, werden diese Ergebnisse im Kapitel *Gütekriterien* vorab beschrieben.

5.1 Gütekriterien

5.1.1 Sprachstandsvergleich mithilfe des C-Tests (Validität)

Ist der Sprachstand von KG, TG1 und TG2 vergleichbar, lassen sich Gruppenunterschiede bezüglich der Forschungsfragen mit höherer Wahrscheinlichkeit auf das jeweilige Treatment – und nicht auf einen unterschiedlichen Sprachstand – zurückführen. Durch die Prüfung, ob sich die KG, TG1 und TG2 sprachlich voneinander unterscheiden, trägt der C-Test somit zur Inhaltsvalidität des Forschungsvorhabens bei.

Die individuellen C-Tests konnten einen Wert zwischen 0 und 100 aufweisen. Der Wert 0 bedeutet, dass keine sprachliche Lücke korrekt ausgefüllt war und somit keine Punkte vergeben wurden. C-Tests mit vollständig korrekt gegebenen Antworten, erhielten die bestmögliche Punktzahl 100. Für den Vergleich der Mittelwerte zwischen den drei Versuchsgruppen wurde jeweils eine einfaktorielle Varianzanalyse (ANOVA) berechnet. Da sich die Ergebnisse aus der Untersuchung der Forschungsfragen auf das *Panel* beziehen, wurde der Sprachstand zwischen den *Panel*-Gruppen mit $n = 20$ in KG, $n = 12$ in TG1 und $n = 14$ in TG2 verglichen. Aus Gründen der Übersichtlichkeit sind alle Ergebnisse der Auswertung des C-Tests (deskriptive Statistik, Prüfung der Voraussetzungen zur Berechnung der ANOVA und die ANOVA) tabellarisch im Anhang C.1 festgehalten.

In TG2 ($M = 60.36$, $SD = 18.67$) erzielten die Kinder im Mittel etwas mehr Punkte als in TG1 ($M = 57.33$, $SD = 24.06$) und der KG ($M = 46.10$, $SD = 16.92$).

Die Bedingungen einer **Normalverteilung** der Punkte im C-Test waren laut Shapiro-Wilk-Test innerhalb jeder Gruppe erfüllt, mit $p = .370$ in KG, $p = .383$ in TG1 und $p = .525$ in TG2. Die Analyse des Boxplots ergab einen Ausreißer in der KG: Mit 87 Punkten im C-Test lag die Vpn oberhalb des 1,5fachen Interquartilsabstandes. Aufgrund der Tatsache, dass es sich um einen leichten Ausreißer handelte, wurde der Wert nicht von der weiteren Analyse ausgeschlossen. Die Prüfung auf Varianzhomogenität, getestet mit dem Levene-Test, ergab keine signifikanten Unterschiede zwischen den Varianzen von KG, TG1 und TG2, basierend auf dem Median ($p = .530$). Folglich waren die Bedingungen zur Durchführung der ANOVA erfüllt. Das Signifikanzniveau wurde bei 5% festgelegt.

Der Gruppenvergleich mittels ANOVA zeigte, dass sich die Gruppen nicht signifikant in den Sprachtest-Ergebnissen unterschieden, $F(2, 43) = 2.53$, $p = .091$. Somit ist auch der Einbezug der Kinder mit Lese-Rechtschreibschwäche (LRS) und nicht-deutscher Herkunftssprache (ndH) zu legitimieren. Die Legitimation der Inklusion von Kindern mit LRS ergibt sich ebenfalls aus dem Umstand, dass Rechtschreibfehler in diesem Forschungsdesign nicht beachtet werden. Der angenommene vergleichbare Sprachstand des *Panels* bildet den Ausgangspunkt für die weiteren Ergebnisse.

5.1.2 Intercoder-Übereinstimmung (Objektivität und Reliabilität)

Die qualitativ erhobenen Daten beziehen sich auf die sprachliche und inhaltliche Auswertung der naturwissenschaftlichen Versuchsprotokolle mittels Kodierungen in MAXQDA. Um die Intersubjektivität dieser Kodierungen zu prüfen, wurde die Intercoder-Übereinstimmung bestimmt (Rädiker & Kuckartz, 2019, S. 287). Dabei wurden die Kodierungen der Autorin (erste Messperson) mit den Kodierungen einer zweiten Messperson auf Grundlage des gleichen Kategoriensystems verglichen. Die Kodierungen erfolgten unabhängig voneinander. Anlehnend an Rädiker und Kuckartz (2019, S. 289) wird an dieser Stelle der Begriff *Intercoder-Übereinstimmung* anstelle von *Intercoder-Reliabilität* verwendet. Diese Begriffsnuance verdeutlicht, dass neben der Berechnung eines Übereinstimmungswertes (Intercoder-Reliabilität) zusätzlich die Diskussion über Unstimmigkeiten und die darauffolgende Anpassung des Kategoriensystems einen hohen Stellenwert im qualitativ ausgerichteten Forschungsteil einnimmt (ebd., S. 289). Aus Gründen der Übersichtlichkeit werden die Diskussionspunkte und Anpassungen nicht im Einzelnen aufgeführt. Es sei darauf hingewiesen, dass diese Diskussionen und Anpassungen prozessbegleitend stattfanden. Insgesamt liegen die Ergebnisse der Zweitkodierung für **14 Versuchsprotokolle** vor. Bezogen auf das *Panel*, das

insgesamt 138 Versuchsprotokolle umfasst, stellt die Zweitkodierung von 14 Versuchsprotokollen 10 % des Gesamtmaterials dar.

Tab. 7 zeigt die Übereinstimmungswerte bezüglich der Oberkategorien *Sprachhandlungen*, *Lexikon*, *Morphosyntax* und *Inhalt*, sowie die differenzierten Übereinstimmungswerte bezogen auf die jeweiligen Unterkategorien.

Tab. 7: Intercoder-Übereinstimmung in 14 Versuchsprotokollen

Kategorien	Übereinstimmung (absolut)	Nicht-Übereinstimmung (absolut)	Kappa nach Brennan und Prediger (1981)
Sprachhandlungen	32	3	0.83 / 0.89
Vermuten	10	1	0.88
Erklären/Begründen	22	2	0.89
Lexikon	32	17	0.48 / 0.57
Nomen	26	4	0.82
Verben	2	11	-0.13
Adjektive	4	2	0.56
Morphosyntax	46	17	0.60 / 0.67
Kohäsion	36	10	0.71
Satzgefüge	10	5	0.56
Unpersönliche Konstruktionen	0	2	-0.33
Inhalt	194	58	0.74
Anzahl der genannten Materialien	26	2	0.92
Korrekt	98	20	0.77
Unzureichend	46	19	0.61
Fehlerhaft	24	17	0.45
INSGESAMT	304	95	0.75

Die Übereinstimmungswerte und Nicht-Übereinstimmungen, sowie die Angabe von Kappa nach Brennan und Prediger (1981), beziehen sich auf Code-Überlappungen auf

Segmentebene mit einer minimalen Überlappung von 90 %, berechnet in MAXQDA. Bei den Oberkategorien *Sprachhandlungen*, *Lexikon* und *Morphosyntax* wurden zwei Werte ermittelt: Der jeweils zweite Wert ist ein korrigierter Wert, bei dem die Wahrscheinlichkeit der zufälligen Übereinstimmung rechnerisch durch die Hinzunahme einer weiteren Kategorie x erhöht wurde. Da die Vergabe mehrerer Codes innerhalb einer Oberkategorie an einem Segment vermieden wurde und um den Effekt der Überschätzung zu verringern, bezieht sich die Interpretation von Kappa auf den erstgenannten (geringeren) Kappa-Wert.

Tab. 7 ist zu entnehmen, dass insgesamt 304 Codes an der gleichen Stelle vergeben wurden. Da die Codierungen beider Personen in MAXQDA separat überprüft wurden, handelt es sich um insgesamt 152 gemeinsam kodierte Textstellen in den Versuchsprotokollen (Rädiker & Kuckartz, 2019, S. 300-301). Demgegenüber stehen 95 Textstellen, die keine übereinstimmende Kodierung aufwiesen.

Bezogen auf die Oberkategorien sind Kappa-Werte zwischen 0.48 und 0.83 zu verzeichnen. Zur **Interpretation der Kappa-Werte** wurden die Vergleichsmaßstäbe von Landis und Koch (1977, S. 165) verwendet, die sich auf die Interpretation von Cohens Kappa beziehen. Demnach gelten Werte ab 0.41 als moderat (*moderate*), Werte ab 0.61 als substantiell (*substantial*) und Werte zwischen 0.81 und 1 als nahezu perfekt (*almost perfect*) (ebd., S. 165). Da es sich bei den ermittelten Werten in Tab. 7 nicht um Cohens Kappa, sondern um Kappa nach Brennan und Prediger (1981) handelt, muss die Interpretation etwas kritischer ausfallen (Rädiker & Kuckartz, 2019, S. 303).¹⁹ Die Oberkategorien *Sprachhandlungen* (κ_n 0.83) und *Inhalt* (κ_n 0.74) zeigen auch unter strengerer Betrachtung eine gute Übereinstimmung. Die Oberkategorien *Lexikon* (κ_n 0.48) und *Morphosyntax* (κ_n 0.60) können lediglich als moderate Übereinstimmung gelten.

Bei Betrachtung der einzelnen Kappa-Werte in den beiden Oberkategorien fallen die Kodierungen der *bildungssprachlichen Verben* (κ_n -0.13) und die Kodierungen der *unpersönlichen Konstruktionen* (κ_n -0.33) im negativen Zahlenbereich auf. Unter Berücksichtigung der absoluten Übereinstimmungshäufigkeiten lassen sich diese Werte genauer analysieren. Der Kappa-Wert von -0.33 basiert beispielsweise auf lediglich zwei Kodierungen. Mit Blick in die Daten von MAXQDA wird ersichtlich, dass eine Person zwei unpersönliche Konstruktionen ermittelte und entsprechend kodierte, während die zweite Person diese nicht entsprechend auswies. Die fehlende Ermittlung zweier unpersönlicher Konstruktionen in 14 Versuchsprotokollen ist kein unrealistisches Szenario. Der Wert -0.33 darf insofern nicht überinterpretiert werden. Genauso muss der Wert -

¹⁹ Die tendenziell kritischere Bewertung ist darauf zurückzuführen, dass Kappa nach Brennan und Prediger (1981) sehr wohl einen Wert von 1 annehmen kann, während dies bei Cohens Kappa rechnerisch oftmals nicht möglich ist (Rädiker & Kuckartz, 2019, S. 303). Als Konsequenz müssen die Schwellenwerte zur Interpretation von Cohens Kappa bei der Interpretation von Kappa nach Brennan und Prediger (1981) leicht erhöht werden (ebd., S. 303).

0.13 bei der Ermittlung bildungssprachlicher Verben in Hinblick auf die geringe Gesamtzahl der Verb-Kodierungen betrachtet werden. Die differenzierte Darstellung der Kappa-Werte bezogen auf die Unterkategorien zeigt, dass der niedrige Kappa-Wert der Oberkategorien Lexikon und Morphosyntax hauptsächlich auf fehlende Übereinstimmungswerte zurückzuführen sind, die bei genauerer inhaltlicher Betrachtung zum Teil relativiert werden müssen.

Insgesamt ergibt sich aus der Zweitkodierung ein Kappa von 0.75. Auch unter tendenziell strengerer Interpretation dürfte dieser Wert als gut betrachtet werden. Eine Nachvollziehbarkeit des Kategoriensystems und eine damit verbundene Objektivität konnten insgesamt angenommen werden. Auf Grundlage der Intercoder-Übereinstimmungswerte wurden die fehlenden Übereinstimmungen diskutiert und eine abschließende Kontrolle mithilfe einer Gesamtansicht aller Kodierungen vorgenommen. Daraufhin schloss sich die Auswertung der vier Forschungsfragen an. Die Ergebnisse werden im Folgenden jeweils fragebezogen dargestellt.

5.2 F 1 – Nutzung der Wortspeicher

Die erste Forschungsfrage lautet wie folgt:

F1 – Nutzung der Wortspeicher: Wie häufig verwenden Grundschul Kinder die visuell oder auditiv vermittelten Sprachmittel aus Wortspeichern kontextuell richtig in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts?

5.2.1 Deskriptive Ergebnisse

Zur Beantwortung der Fragestellung wurden die verwendeten Sprachmittel aus den Wortspeichern in MAXQDA ermittelt und die Anzahl der jeweiligen Kodierungen in SPSS übertragen. Alle vorgestellten Ergebnisse beziehen sich auf das *Panel*, mit $n = 20$ in der KG, $n = 12$ in TG1 und $n = 14$ in TG2. Alle deskriptiven und inferenzstatistischen Ergebnisse sind im Anhang C.2 dargestellt und werden im Folgenden beschrieben. Abb. 23 bietet eine erste Übersicht darüber, wie viele Wörter und Formulierungen durchschnittlich in den einzelnen Gruppen und zu den einzelnen MZP eruiert wurden.

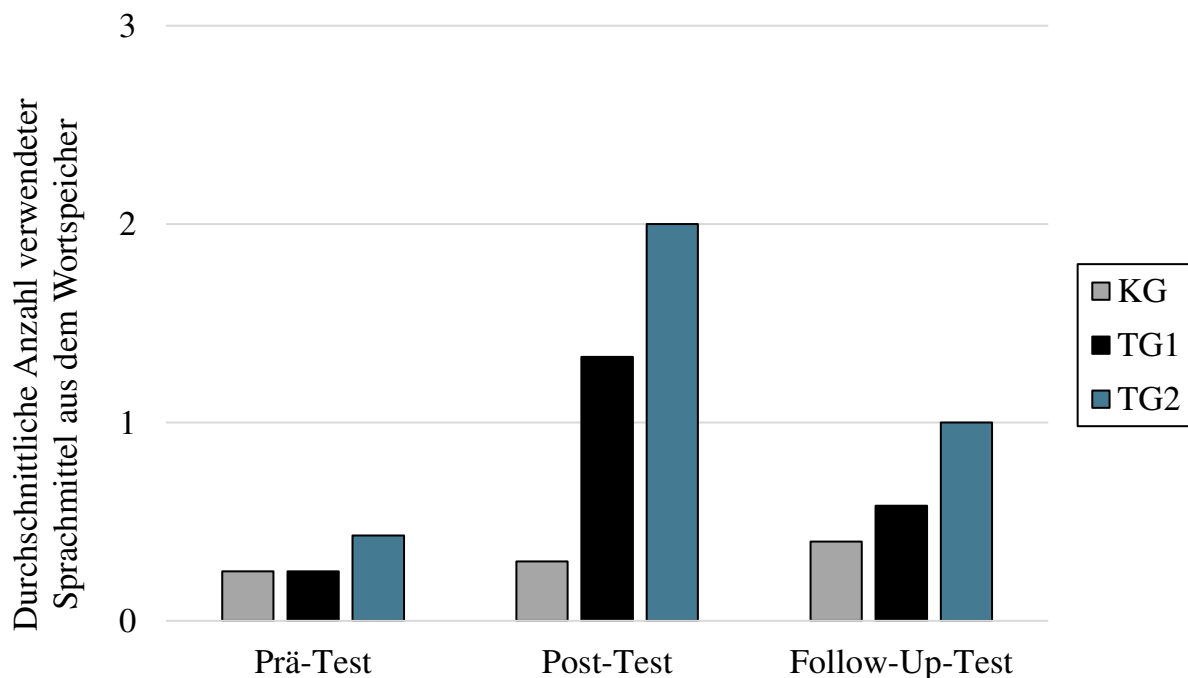


Abb. 23: Gruppenunterschiede bezogen auf die Verwendung der Wortspeicher (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Im **Prä-Test** benutzten alle Vpn im Durchschnitt eine sehr geringe Anzahl an Sprachmitteln aus dem – für die Vpn unbekanntem – Wortspeicher. In der KG und TG1 war die durchschnittliche Anzahl exakt gleich, $M_{KG\&TG1} = 0.25$, $SD_{KG} = 0.55$, $SD_{TG1} = 0.45$. Die Versuchsprotokolle der Vpn aus TG2 zeichneten sich durch eine leicht höhere durchschnittliche Verwendung der unbekanntem Sprachmittel vor Beginn der Unterrichtssequenz aus, $M = 0.43$, $SD_{TG2} = 0.85$.

Nach der Unterrichtssequenz (im **Post-Test**) verwendeten die Vpn aus TG2 mit durchschnittlich 2 Sprachmitteln pro Versuchsprotokoll ($SD = 2.04$) die meisten Wörter und Formulierungen aus dem Wortspeicher, gefolgt von TG1 mit durchschnittlich 1.33 Sprachmitteln ($SD = 1.88$) und zuletzt der KG mit durchschnittlich 0.30 Sprachmitteln pro Versuchsprotokoll ($SD = 0.57$).

Der zehn Wochen später stattgefundenen **Follow-Up-Test** zeigte die gleiche Tendenz in geringerem Ausmaß. Die Vpn aus TG2 verwendeten im Durchschnitt 1 Sprachmittel pro Versuchsprotokoll ($SD = 1.41$), während der Mittelwert in TG1 bei 0.58 ($SD = 1.17$) und in der KG bei 0.40 Sprachmitteln pro Versuchsprotokoll ($SD = 0.94$) lag.

Auffallend ist die jeweils hohe Standardabweichung bei erhöhtem Mittelwert. Dies ist darauf zurückzuführen, dass die meisten Vpn in jedem Test und aus jeder Gruppe keine Sprachmittel aus dem Wortspeicher verwendeten (Modalwert $Mo = 0$ in allen Gruppen und zu jedem Testzeitpunkt). Die hohe Standardabweichung bei höherem Mittelwert spiegelt die Differenz zwischen der Verwendung der Sprachmittel und dem Großteil der Versuchsprotokolle ohne Sprachmittel wider.

Um zu testen, ob der deskriptiv beschriebene Gruppenunterschied inferenzstatistisch signifikant ist, wurde die Berechnung einer 3x1 ANOVA bezogen auf die jeweiligen MZP angestrebt. Da sich die Ergebnisse deskriptiv auch im zeitlichen Verlauf unterscheiden, sollte eine 1x3 ANOVA die inferenzstatistische Signifikanz der unterschiedlichen Werte zu den drei MZP innerhalb derselben Versuchsgruppe prüfen. Die Prüfung der Voraussetzungen für die Berechnung der 3x1 und der 1x3 ANOVA wird im nächsten Kapitel vorgestellt.

5.2.2 Voraussetzungen für Inferenzstatistik

Aufgrund der ad-hoc-Stichprobe und des quasi-experimentellen Versuchsdesigns war zunächst die **Vergleichbarkeit von KG, TG1 und TG2** zu prüfen. Die Analyse des kontrollierenden Sprachtests ergab, dass von einem vergleichbaren Sprachstand auszugehen war.²⁰ Weiterhin wurde die Vergleichbarkeit der Versuchsgruppen anhand der Mittelwertsunterschiede im Prä-Test geprüft. Das Ergebnis wird im Anschluss an die

²⁰ vgl. Kapitel 5.1.1

notwendige Betrachtung der Voraussetzungen zur Berechnung der ANOVA beschrieben.

Zur **Prüfung der Normalverteilung** wurden der Shapiro-Wilk- und der Kolmogorov-Smirnov-Test herangezogen. Die Ergebnisse sind im Anhang C.2 abgebildet. Bei allen drei Versuchsgruppen war die Verwendung der Sprachmittel im Prä-, Post- und Follow-Up-Test laut Shapiro-Wilk-Test nicht normalverteilt, mit $p = .036$ im Post-Test von TG2, $p = .004$ im Post-Test von TG1 und $p < .001$ in allen weiteren Gruppen und zu allen anderen MZP. Laut Kolmogorov-Smirnov-Test mit Signifikanzkorrektur nach Lilliefors war lediglich die Verwendung der Sprachmittel in TG2 im Post-Test normalverteilt, $p = .081$. Da dem Shapiro-Wilk-Test eine höhere Power als dem Kolmogorov-Smirnov-Test bescheinigt wird (Razali & Wah, 2011, S. 32), wurde eine Verletzung der Normalverteilung in allen Gruppen und allen Tests angenommen.

Die Verletzung der Normalverteilung ist dadurch zu begründen, dass die meisten Vpn keine oder wenige der Sprachmittel in ihren Versuchsprotokollen anwandten. Dies ist den deskriptiven Ergebnissen zu entnehmen. Dementsprechend handelte es sich bei der Untersuchung auf Normalverteilung bei allen Tests und allen Gruppen jeweils um eine linksgipflige Verteilungsform (Schiefe > 0), die mit Ausnahme der TG1 im Prä-Test und der TG2 im Post-Test durch einen schmalen Gipfel (Kurtosis > 0) gekennzeichnet war. Die Verletzung der Normalverteilung war aufgrund fehlender Anwendung der Sprachmittel in allen Prä-Tests und in allen Tests der KG erwartungskonform. Die Histogramme im Anhang C.2 verdeutlichen die jeweils linksgipfligen Verteilungsformen.

Als **Ausreißer** sind anhand der Boxplot-Analyse (Abb. 24) insgesamt sieben Werte festzustellen. Bei den sieben abweichenden Werten, die in Abb. 24 als Kreise dargestellt sind, handelt es sich um eine erhöhte Verwendung der Wortspeicher. Die Anzahl der verwendeten Sprachmittel aus dem Wortspeicher liegt bei den Ausreißern mindestens um das 1,5 fache des Interquartilsabstandes oberhalb der Box, die die mittleren 50 % der Daten repräsentiert (25 % bis 75 %). Das X innerhalb jeder Box zeigt den Mittelwert an, der Strich innerhalb einer Box visualisiert den Median. Abb. 24 zeigt, dass der Abstand zwischen den Ausreißern und der Box maximal eine Differenz von drei Sprachmitteln beträgt.

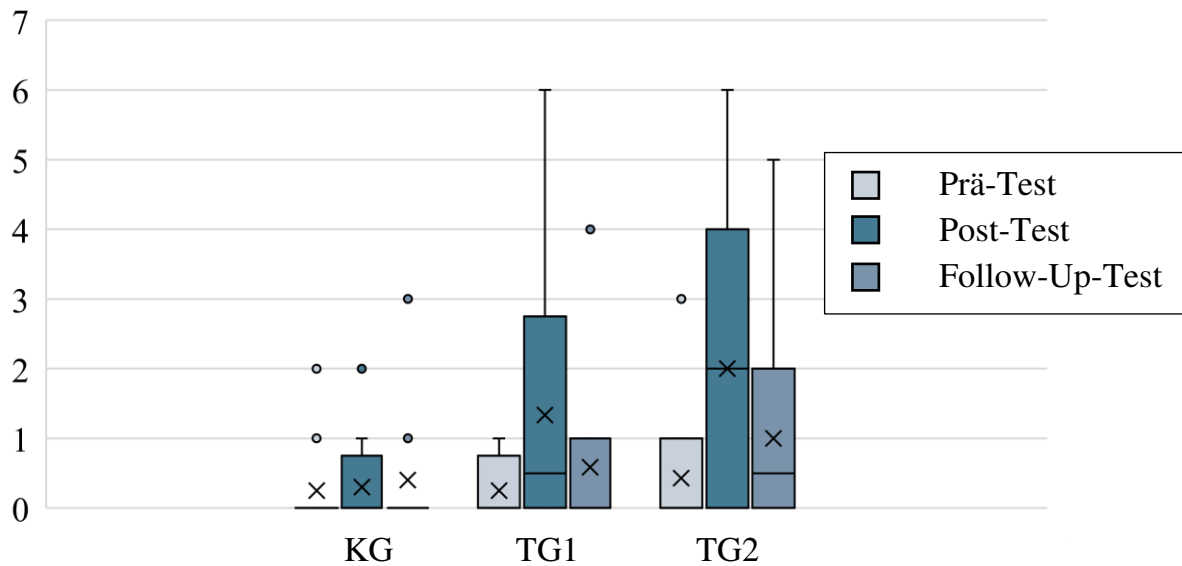


Abb. 24: Boxplot zur Analyse von Ausreißern bezogen auf die Verwendung der Wortspeicher (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Die **Varianzhomogenität bzw. Homoskedastizität** wurde mithilfe des Levene-Tests geprüft. Im Prä-Test ($p = .684$) und im Follow-Up-Test ($p = .266$) konnte eine Gleichheit der Varianzen gemäß dem Levene-Test basierend auf dem Median angenommen werden. Im Gegensatz dazu war Varianzhomogenität zwischen den Gruppen laut Levene-Test basierend auf dem Median nicht für den Post-Test erfüllt ($p = .003$). Die Voraussetzung der **Sphärizität** wurde anhand des Mauchly-Tests geprüft. Sie stellte sich gleichermaßen als nicht erfüllt dar, mit $p_{KG} = .029$, $p_{TG1} < .001$ und $p_{TG2} = .031$.

Aufgrund der (wenigen) extremen Ausreißer, der Verletzungen von Normalverteilung, Varianzhomogenität und Sphärizität wurden zusätzlich zu der angestrebten 3x1 und 1x3 faktoriellen Varianzanalyse jeweils die korrespondierenden nicht-parametrischen Tests durchgeführt, d.h. der **Kruskal-Wallis-Test** zur Kontrolle der 3x1 faktoriellen Varianzanalyse und der **Friedman-Test** zur Kontrolle der 1x3 faktoriellen Varianzanalyse.

Um zu testen, ob die drei Gruppen ohne vorheriges Treatment vergleichbar in ihrer Anwendung von Sprachmitteln aus dem Wortspeicher sind, wurde eine **3x1 faktorielle Varianzanalyse des Prä-Tests** durchgeführt. Das Ergebnis der ANOVA zeigt, dass sich die drei Gruppen im Prä-Test nicht signifikant unterscheiden, $F(2, 43) = 0.38$, $p = .684$. Der kontrollierende nicht-parametrische Kruskal-Wallis-Test bestätigte das Ergebnis, $H(2) = 0.35$, $p = .838$, sodass die Vergleichbarkeit der Ausgangsgruppen angenommen wurde.

5.2.3 Gruppenunterschiede – Wortspeicher

Im Folgenden werden die Ergebnisse der 3x1 ANOVA zur Testung von Gruppenunterschieden (*between*) präsentiert.

Im **Post-Test** unterschieden sich die Versuchsprotokolle der drei Gruppen hinsichtlich der Verwendung der Wortspeicher-Sprachmittel signifikant voneinander, $F(2, 43) = 5.39, p = .008, f = 0.5^{21}$. Nach Cohen (1988, S. 285-287) handelt es sich um einen starken Effekt ($f > 0.4$).

Aufgrund ungleicher Varianzen im Post-Test wurde zur weiteren Analyse und zur Prüfung der Hypothesen der Games-Howell Post-hoc Test durchgeführt. **Demnach unterschieden sich die Versuchsprotokolle von KG und TG2 hinsichtlich der Wortspeicher-Nutzung im Post-Test signifikant, $p = .007$.** Die weiteren Paarvergleiche waren nicht signifikant, mit $p = .192$ für den Unterschied zwischen KG und TG1 und $p = .665$ für den Unterschied zwischen TG1 und TG2.

Zusätzlich zum Games-Howell Post-hoc Test wurden zur genaueren Betrachtung des signifikanten Unterschieds im Post-Test geplante Kontraste ermittelt, die u.a. den Vergleich zwischen der Vermittlung der Wortspeicher-Sprachmittel im Allgemeinen und der fehlenden expliziten Vermittlung (TG1 & TG2 vs. KG) vornahmen. Das Ergebnis des beschriebenen Kontrasts ergab, dass die **Wortspeicher-Gruppen (TG1 & TG2) nach der Unterrichtsreihe signifikant mehr Wortspeicher-Sprachmittel verwendeten als die KG, $t(28.99) = 3.378, p = .002, r = 0.46$.** Nach Cohen (1988, S. 285-287) handelt es sich um einen mittleren Effekt ($0.3 < r < 0.5$).

Nach zehn Wochen regulärem Unterricht (**Follow-Up-Test**) unterschieden sich die Versuchsprotokolle der drei Gruppen hinsichtlich der Verwendung der Wortspeicher-Sprachmittel nicht signifikant voneinander, $F(2, 43) = 1.12, p = .336^{22}$.

5.2.4 Zeitliche Unterschiede – Wortspeicher

Im Folgenden werden die Ergebnisse der 1x3 ANOVA zur Testung des Innersubjektfaktors (*within*) dargestellt.

Abb. 25 dient als erster Überblick über die Unterschiede innerhalb der Gruppen im zeitlichen Verlauf bezüglich der durchschnittlichen Wortspeicher-Anwendung.²³ Sichtbar ist der Anstieg in TG1 und TG2 direkt nach der Unterrichtssequenz (Post-Test) und die

²¹ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 8.33, p = .015$.

²² Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 3.24, p = .198$.

²³ Die Liniendiagramme in dieser Studie basieren auf denselben Daten der jeweils vorangestellten Säulendiagramme (Gruppenunterschiede). Sie dienen zur besseren Veranschaulichung der zeitlichen Unterschiede.

Abnahme der verwendeten Wortspeicher 10 Wochen nach der Unterrichtssequenz (Follow-Up-Test). In der KG ist diese Tendenz nicht erkennbar.

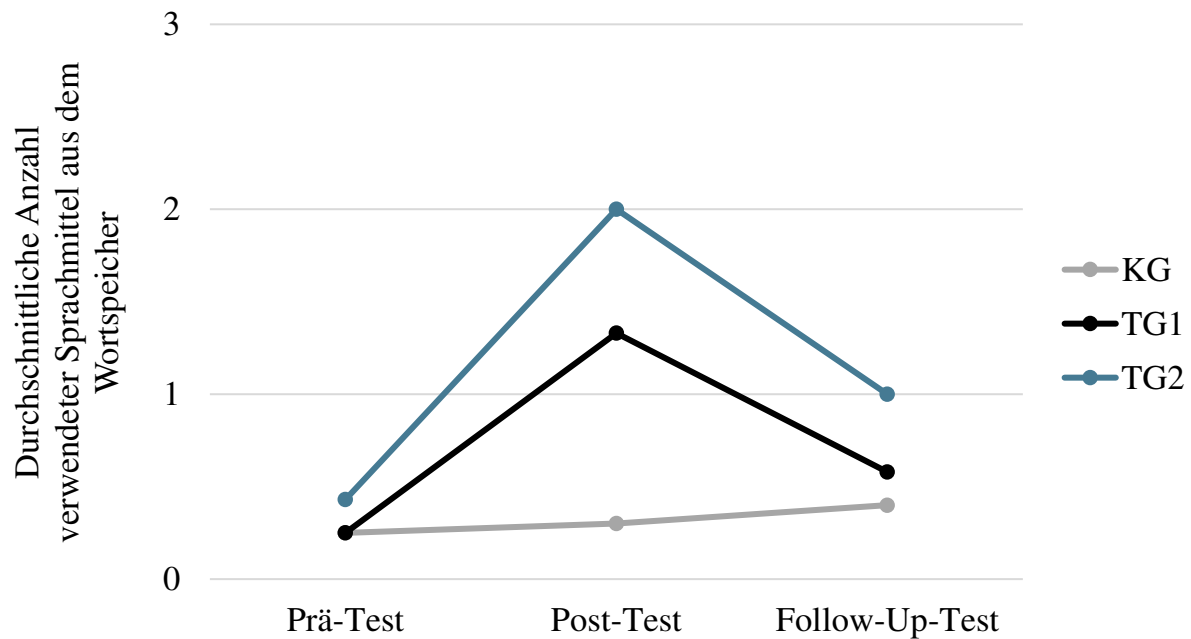


Abb. 25: Zeitlicher Verlauf bezogen auf die Verwendung der Wortspeicher (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Aufgrund der verletzten Sphärizität wurde bei allen drei Gruppen eine Korrektur der Freiheitsgrade vorgenommen. Während die Greenhouse-Geisser Korrektur eine teils zu strenge Korrektur darstellt, ist die Huynh-Feldt Korrektur eine teils zu schwache Korrektur (Field, 2018, S. 658, 670). Da bei einem Greenhouse-Geisser-Epsilon ϵ über 0.75 die Huynh-Feldt Korrektur empfohlen wird, wurden die Ergebnisse unter diesem Gesichtspunkt zur späteren Interpretation ausgewählt (ebd., S. 658, 671).

In der **KG** lag die Huynh-Feldt Schätzung zur Sphärizitätsabweichung bei $\epsilon = 0.81$. Die Verwendung der Wortspeicher-Sprachmittel unterschied sich nicht signifikant zwischen Prä-, Post- und Follow-Up-Test, $F(1.61, 30.61) = 0.32, p = .68^{24}$.

Bei den Vpn, die die Wortspeicher auditiv vermittelt bekamen (**TG1**), lag die Greenhouse-Geisser Schätzung zur Sphärizitätsabweichung bei $\epsilon = 0.55$. Auch hier gab es über die drei Messzeitpunkte hinweg keine signifikante Änderung in der Anzahl der verwendeten Sprachmittel aus dem Wortspeicher $F(1.1, 12.13) = 3.22, p = .095^{25}$.

Ein signifikanter Effekt konnte in der Gruppe festgestellt werden, in der die Wortspeicher in Plakatform vermittelt wurden (**TG2**). In dieser Gruppe lag die Greenhouse-

²⁴ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 0.21, p = .902$.

²⁵ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 4.36, p = .113$.

Geisser Schätzung zur Sphärizitätsabweichung bei $\varepsilon = 0.69$. Laut Ergebnis mit korrigierten Freiheitsgraden nach Greenhouse-Geisser hing die Verwendung der Wortspeicher-Sprachmittel signifikant mit dem jeweiligen Messzeitpunkt zusammen, $F(1.39, 18.05) = 5.12$, $p = .027$, $f = 0.63$. Nach Cohen (1988, S. 285-287) handelt es sich um einen starken Effekt. Die Signifikanz ist ebenfalls bei der strengeren Korrektur der Untergrenze ($p = .041$) gegeben. Allerdings konnte das Ergebnis nicht durch den Friedman-Test bestätigt werden, $\chi^2(2) = 5.84$, $p = .054$, sodass keine weiteren Analysen vorgenommen wurden.

5.2.5 Zusammenfassung

Auf Grundlage der vorgestellten Ergebnisse sind in Bezug auf die Hypothesen folgende Erkenntnisse festzuhalten:

Die deskriptive Statistik zeigt sowohl im Post- als auch im Follow-Up-Test Tendenzen, die mit H1 (Hypothese 1: TG2 verwendet mehr Sprachmittel aus dem Wortspeicher als TG1) und H2 (Hypothese 2: TG1 verwendet mehr Sprachmittel aus dem Wortspeicher als die KG) vereinbar sind.

Erwartungswidrig sind die Unterschiede laut inferenzstatistischer Prüfung zum MZP 2 und MZP 3 nicht signifikant, sodass **H1 und H2 verworfen werden**. In beiden Fällen und für beide MZP wird die Nullhypothese angenommen. Dennoch sollten die Erkenntnisse zu den ermittelten Signifikanzen außerhalb der Hypothesen Beachtung finden.

Insgesamt konnte für den Post-Test ein signifikanter Unterschied in der Wortspeichernutzung festgestellt werden. Der Unterschied ist vor allem auf den Vergleich zwischen KG und TG2 zurückzuführen. Auf Basis von deskriptiven Ergebnissen und ANOVA kann festgehalten werden, dass die **TG2 im Post-Test signifikant mehr Sprachmittel aus dem Wortspeicher in ihren Versuchsprotokollen integrierte als die KG**. Das gleiche Ergebnis trifft auf den Vergleich zwischen den Treatmentgruppen zusammengekommen und der KG zu (TG1 & TG2 vs. KG). Das bedeutet für den Post-Test: **Nach einem Wortspeicher-Treatment wurden im Allgemeinen mehr Sprachmittel aus dem Wortspeicher verwendet als ohne Wortspeicher-Treatment**.

Im zeitlichen Verlauf nutzten beide Treatmentgruppen die meisten Wortspeicher-Sprachmittel im Post-Test, gefolgt vom Follow-Up- und dem Prä-Test. Diese Tendenz wurde inferenzstatistisch für TG1 nicht bestätigt. Ein inferenzstatistischer Unterschied innerhalb von TG2 wurde aufgrund des negativen Abgleichs mit dem Friedman-Test verworfen.

Wie viele bildungssprachliche Mittel die Vpn *außerhalb* des Wortspeichers verwendeten, wird im nächsten Kapitel dargestellt.

5.3 F 2 – Sprachregister

Die zweite Forschungsfrage lautet wie folgt:

F2 – Sprachregister: Lässt sich durch visuell oder auditiv angebotene Wortspeicher die Häufigkeit bildungssprachlicher Wörter und Formulierungen in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts erhöhen?

Um die Häufigkeit bildungssprachlicher Anwendungen zu bestimmen, wurden Kodierungen qualitativ anhand eines Kategoriensystems vorgenommen, das im Anhang B.1 in vollständiger Länge abgebildet ist.

Die Häufigkeiten der Kodierungen sind in den Tab. 8 bis 12 ersichtlich, die sich jeweils am Ende der folgenden Unterkapitel befinden. Da die Häufigkeit der Kodierungen in Relation zur Stichprobengröße n betrachtet werden muss, weisen alle Tabellen die errechneten **Mittelwerte** auf. Die Angabe der **Mediane** und die **Anzahl der Dokumente ohne Kodierungen** bieten ein besseres Verständnis darüber, ob es sich um vereinzelte Versuchsprotokolle mit Mehrfachkodierungen oder um eine breitere Streuung von Versuchsprotokollen mit bildungssprachlichen Elementen handelt. Da sich die inferenzstatistische Analyse nur auf das Gesamtergebnis aller Kodierungen bezieht, wird die jeweilige **Standardabweichung** im Kapitel *Kodierungen der Bildungssprache insgesamt* angegeben.

Zunächst werden die Häufigkeiten der Kodierungen zu den einzelnen Kategorien der Bildungssprache beschreibend vorgestellt.

5.3.1 Kodierungen der Bildungssprache (Sprachhandlungen)

Tab. 8 zeigt die kodierten Sprachhandlungen, zu denen sprachlich erkennbare Vermutungen und sprachlich erkennbare Erklärungen / Begründungen zählen. Beispiele für sprachlich erkennbare Vermutungen aus der Hauptstudie waren: *Das Salz könnte sich auf den Boden legen* oder *dann ist das Salz vermutlich weg*. In der Tabelle sind die Werte bei Mehrfachkodierungen pro Dokument angegeben. Die Ergebnisse zu den Einfachkodierungen pro Dokument sind dem Anhang D.1 zu entnehmen.

Bei Betrachtung der kodierten sprachlich erkennbaren Vermutungen lässt sich kein auffälliger Unterschied zwischen den Gruppen im Prä-, Post- und Follow-Up-Test feststellen. Anders sieht es bei der Nutzung sprachlich erkennbarer Erklärungen aus. Auffallend ist TG2, die mit Wortspeichern in Plakatform lernte und nach der Unterrichtssequenz insgesamt 24 Begründungen schrieb, $M_{TG2} = 1.71$. Der Mittelwert muss unter der Einschränkung betrachtet werden, dass Mehrfachkodierungen möglich waren. Die Anzahl

der Versuchsprotokolle OHNE Kodierungen lag bei 7, sodass etwas mehr als die Hälfte der Klasse mindestens eine Erklärung sprachlich korrekt einleitete oder erkenntlich machte. Keine der beiden anderen Gruppen schrieb im Post-Test annähernd so viele Begründungen, $M_{KG} = 0.15$; $M_{TG1} = 0.08$. Die hohe Anzahl an sprachlich markierten Erklärungen im Post-Test der TG2 ist besonders vor dem Hintergrund relevant, dass vor der Unterrichtssequenz (Prä-Test) keine einzige Begründung festgehalten wurde.

Das **Gesamtergebnis (Sprachhandlungen)** fasst die Unterkategorien *Vermuten und Begründen* zusammen und spiegelt dementsprechend die hohe Anzahl an Begründungen in TG2 im Post-Test wider. Insgesamt war die Anzahl der Sprachhandlungen in TG2 nach der Unterrichtssequenz das höchste Ergebnis, sowohl in Bezug auf die absolute Häufigkeit (29 Vermutungen und Begründungen) als auch relativ, bezogen auf die Stichprobengröße ($M_{TG2} = 2.07$). Der Median ($Md_{TG2} = 1$) und die Anzahl der Versuchsprotokolle ohne Kodierungen (5) zeigen, dass die Sprachhandlungen nicht alle in einem Versuchsprotokoll vorkamen. Insgesamt schrieben 9 von 14 Vpn (TG2: 64 %) ²⁶ im Post-Test sprachlich erkennbare *Vermutungen und/ oder Erklärungen*. Im Gegensatz dazu verwendeten aus den anderen Gruppen nach der Unterrichtssequenz lediglich 3 von 20 Vpn (KG: 15 %) und 2 von 12 Vpn (TG1: 17 %) Sprachhandlungen, die in dieser Studie als solche kodiert wurden.

²⁶ Dies ist abzulesen anhand der Anzahl OHNE Kodierungen im Post-Test, unter Berücksichtigung von n . Hier: TG2: $n = 14$, Anzahl der Dokumente OHNE Kodierungen = 5. Deshalb gab es 9 Dokumente MIT Kodierungen.

Tab. 8: Häufigkeitsanalyse bildungssprachlicher Sprachhandlungen (Mehrfachkodierungen)

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Vermuten												
KG	4	0	1	0.2	0	0.05	0	0	0	18	20	19
TG1	13	3	4	1.08	0.25	0.33	0	0	0	7	11	9
TG2	1	5	3	0.07	0.36	0.21	0	0	0	13	10	12
Begründen												
KG	2	3	1	0.1	0.15	0.05	0	0	0	18	17	19
TG1	3	1	10	0.25	0.08	0.83	0	0	0	9	11	8
TG2	0	24	8	0	1.71	0.57	0	1	0	14	6	10
Gesamt												
KG	6	3	2	0.3	0.15	0.1	0	0	0	16	17	18
TG1	16	4	14	1.33	0.33	1.17	0.5	0	0	6	10	7
TG2	1	29	11	0.07	2.07	0.79	0	1	0	13	5	9

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = absolute Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

5.3.2 Kodierungen der Bildungssprache (Lexikon)

Tab. 9 präsentiert die Ergebnisse der Oberkategorie *Lexika* mit den Unterkategorien Nomen, Verben, Adjektive und sonstige Wörter. Kodiert wurden Wörter, die im Grundschulalter als bildungssprachlich gelten können. In Bezug auf die Unterkategorie *Nomen* ist der Mittelwert von TG1 im Post-Test hervorzuheben ($M_{TG1} = 1.25$). Insgesamt wurden 15 bildungssprachliche Nomen bei einer Stichprobe von 12 Vpn kodiert. Der Median ($Md_{TG1} = 0$) deutet an, dass nicht die Mehrheit der Vpn mehr als ein Nomen pro Versuchsprotokoll schrieb. Tatsächlich stammen die 15 Nomen aus insgesamt 5 Dokumenten. Im Gegensatz zur TG1 verwendeten die Vpn aus TG2 in absoluten Zahlen und auch bezogen auf die Stichprobe weniger bildungssprachliche Nomen im Post-Test ($M_{TG2} = 0.93$). Auf der anderen Seite ist hier der höhere Median hervorzuheben ($Md_{TG2} = 1$), da sich die 13 bildungssprachlichen Nomen auf 8 Dokumente verteilten. Bezogen auf die Verwendung von bildungssprachlichen Verben, Adjektiven und sonstigen Wörtern ist kein hervorstechendes Ergebnis zu vermerken.

Dem **Gesamtergebnis (Lexikon)** ist zu entnehmen, dass die TG1 im Post-Test die meisten bildungssprachlichen Wörter verwendete ($M_{TG1} = 2$), gefolgt von TG2 ($M_{TG2} = 1.14$) und zum Schluss der KG ($M_{KG} = 0.45$). Das Ergebnis ist hauptsächlich auf die Verwendung von Nomen zurückzuführen. Der Median und die Anzahl der Dokumente ohne Kodierungen zeigen, dass im Post-Test der TG2 die Verwendung bildungssprachlicher Wörter auf einer breiteren Basis stattfand ($Md_{TG2} = 1$) als in der TG1 ($Md_{TG1} = 0.5$) oder der KG ($Md_{KG} = 0$). Insgesamt nutzten in TG2 im Post-Test 9 von 14 Vpn (64 %) bildungssprachliche Wörter. In TG1 waren es dahingehend weniger, 6 von 12 (50 %) und in der KG nutzten lediglich 5 von 20 Vpn (25 %) bildungssprachliche Wörter im Post-Test. Die Tabelle zeigt ebenfalls, dass die meisten bildungssprachlichen Wörter von allen Gruppen im Prä-Test benutzt wurden. Da es sich bei den drei MZP um inhaltlich unterschiedliche Versuche handelt, ist der Vergleich von verwendeten Wörtern im zeitlichen Verlauf nicht uneingeschränkt möglich. Stattdessen fallen die Gruppenunterschiede zu den drei MZP stärker ins Gewicht.

Tab. 9: Häufigkeitsanalyse bildungssprachlicher Lexika (Mehrfachkodierungen)

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Nomen												
KG	13	5	3	0.65	0.25	0.15	0	0	0	13	17	18
TG1	7	15	7	0.58	1.25	0.58	0	0	0	7	7	9
TG2	8	13	2	0.57	0.93	0.14	0	1	0	9	6	12
Verben												
KG	3	3	10	0.15	0.15	0.5	0	0	0	17	17	11
TG1	11	7	5	0.92	0.58	0.42	0	0	0	8	9	9
TG2	1	2	3	0.07	0.14	0.21	0	0	0	13	12	13
Adjektive												
KG	5	0	4	0.25	0	0.2	0	0	0	17	20	18
TG1	9	0	1	0.75	0	0.08	0	0	0	8	12	11
TG2	10	1	1	0.71	0.07	0.07	0	0	0	9	13	13
Sonstiges												
KG	1	1	0	0.05	0.05	0	0	0	0	19	19	20
TG1	0	2	1	0	0.17	0.08	0	0	0	12	10	11
TG2	0	0	4	0	0	0.29	0	0	0	14	14	11
Gesamt												
KG	22	9	17	1.1	0.45	0.85	0.5	0	1	10	15	9
TG1	27	24	14	2.25	2	1.17	1	0.5	0	4	6	7
TG2	19	16	10	1.36	1.14	0.71	1	1	0	6	5	8

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

5.3.3 Kodierungen der Bildungssprache (Morphosyntax)

Tab. 10 stellt die Kodierungen der Oberkategorie *Morphosyntax* mit den Unterkategorien *Kohäsion*, *Satzgefüge* und *unpersönliche Konstruktion* dar. Auffällig ist TG 1, die zehn Wochen nach der Unterrichtssequenz (im Follow-Up-Test) 20 kohäsive Strukturen verwendete ($M_{TG1} = 1.67$). Von 12 Vpn der TG1 verknüpften 9 Vpn (75 %) im Follow-Up-Test ihre Sätze, indem sie beispielsweise die Formulierung *kurz darauf* benutzten. In TG2 waren es 8 von 14 Vpn (57 %) und in der KG 9 von 20 Vpn (45 %). Das Ergebnis spiegelt sich auch im Median zur Kategorie *Kohäsion* im Follow-Up-Test wider. Die Hälfte von TG1 verwendete mindestens zwei kohäsive Strukturen ($Md_{TG1} = 2$). Im Gegensatz dazu benutzte die Hälfte der TG2 mindestens eine kohäsive Struktur ($Md_{TG2} = 1$) und in der KG waren bei weniger als die Hälfte der Gruppe sprachlich verknüpfte Sätze erkennbar ($Md_{KG} = 0$). Bezogen auf Satzgefüge und unpersönliche Konstruktionen waren keine deutlichen Unterschiede in den Häufigkeiten erkennbar.

Das **Gesamtergebnis (Morphosyntax)** spiegelt die hohe Anzahl an kohäsiven Strukturen in TG1 zehn Wochen nach der Unterrichtssequenz wider. Darüber hinaus sind bei der beschreibenden Ergebnisdarstellung keine auffälligen Unterschiede in den Kodierungen zur Morphosyntax feststellbar.

Tab. 10: Häufigkeitsanalyse bildungssprachlicher Morphosyntax (Mehrfachkodierungen)

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Kohäsion												
KG	25	19	19	1.25	0.95	0.95	0	1	0	12	9	11
TG1	9	13	20	0.75	1.08	1.67	0.5	1	2	6	5	3
TG2	15	17	10	1.07	1.21	0.71	0.5	1	1	7	6	6
Satzgefüge												
KG	7	8	11	0.35	0.4	0.55	0	0	0	14	14	13
TG1	8	7	8	0.67	0.58	0.67	0	0	0	7	7	7
TG2	5	7	5	0.36	0.5	0.36	0	0	0	12	9	10
Unpersönliche Konstruktion												
KG	5	1	2	0.25	0.05	0.1	0	0	0	17	19	19
TG1	1	0	2	0.08	0	0.17	0	0	0	11	12	10
TG2	1	3	1	0.07	0.21	0.07	0	0	0	13	11	13
Gesamt												
KG	37	28	32	1.85	1.4	1.6	1	1	1	9	7	9
TG1	18	20	30	1.5	1.67	2.5	1	1	2.5	4	4	2
TG2	21	27	16	1.5	1.93	1.14	0.5	2	1	7	4	5

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

5.3.4 Kodierungen der Bildungssprache insgesamt

Von besonderem Interesse zur Beantwortung der Fragestellung ist das **Gesamtergebnis (Sprachhandlungen, Lexikon, Morphosyntax bei Mehrfachkodierungen)**, dargestellt in Tab. 11 und bildlich verdeutlicht in Abb. 26.

Es ist deutlich zu erkennen, dass im **Post-Test** die Vpn aus TG2 die meisten bildungssprachlich kodierten Wörter oder Formulierungen verwendeten ($M = 5.14$, $SD = 3.90$), gefolgt von TG1 ($M = 4$, $SD = 3.74$) und der KG ($M = 2$, $SD = 2.13$). Ausschlaggebend für die höhere Anzahl bildungssprachlicher Kodierungen gegenüber den anderen Gruppen ist u.a. die Häufigkeit der Erklärungen, die die Vpn in TG2 sprachlich kennzeichneten (vgl. Tab. 8).

Im **Follow-Up-Test** ist das Ergebnis von TG1 hervorzuheben ($M = 4.83$, $SD = 4.17$), das sich von den Ergebnissen der beiden anderen Gruppen abgrenzt ($M = 2.55$, $SD = 2.44$ in der KG; $M = 2.64$, $SD = 2.79$ in TG2). Die häufige Verwendung kohäsiver Strukturen scheint das Ergebnis maßgeblich mit hervorzurufen (vgl. Tab. 10).

Tab. 12 und Abb. 27 zeigen das **Gesamtergebnis bei Einfachkodierungen** und bieten aufgrund der einmalig gezählten Kodierungen pro Dokument Hinweise auf die Vielseitigkeit der bildungssprachlichen Anwendungen. Wörter mit gleichem Wortstamm, wie *erkennen* und *erkannt* oder *trüb* und *trüber* oder *Oberfläche* und *Wasseroberfläche* wurden bei der Analyse der Einfachkodierungen mit einer Kodierung pro Dokument versehen. Dadurch wird vermieden, dass das Gesamtergebnis wiederholend verwendete Sprachmittel anzeigt.

Auch bei Einfachkodierungen ist der Mittelwert der verwendeten Sprachmittel im **Post-Test** in TG2 am höchsten ($M = 3.79$, $SD = 2.58$), gefolgt von TG1 ($M = 2.5$, $SD = 2.36$) und der KG ($M = 1.65$, $SD = 1.79$). Im Gegensatz zu dem Ergebnis bei Mehrfachkodierungen ist im Post-Test in TG2 auch ein höherer Median ($Md = 3.5$) zu verzeichnen als in TG1 ($Md = 2$).

Im **Follow-Up-Test** benutzten die Vpn aus TG1 die meisten bildungssprachlichen Mittel ($M = 3.33$, $SD = 3.06$), gefolgt von TG2 ($M = 2.21$, $SD = 2.19$) und der KG ($M = 2$, $SD = 1.95$). Die Hälfte der Gruppe aus TG1 benutzte im Schnitt im Follow-Up-Test mehr als 2,5 bildungssprachliche Mittel in den Versuchsprotokollen ($Md_{TG1} = 2.5$). In der KG benutzte die Hälfte der Gruppe im Follow-Up-Test mehr als 2 bildungssprachliche Mittel in den Versuchsprotokollen ($Md = 2$) und in der TG2 wurde von der Hälfte der Gruppe mehr als 1 bildungssprachliches Wort oder eine bildungssprachliche Formulierung verwendet ($Md = 1$). Um zu prüfen, ob der deskriptiv beschriebene Unterschied zwischen den Gruppen im Post- und im Follow-Up-Test statistisch signifikant ist, wurden die Gruppenunterschiede für beide Messzeitpunkte per 3x1 ANOVA genauer untersucht. Das Ergebnis liefert das nächste Kapitel.

Tab. 11: Häufigkeitsanalyse bildungssprachlicher Kodierungen insgesamt (Mehrfachkodierungen)

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Sprachhandlungen, Lexika, Morphosyntax												
KG	65	40	51	3.25	2	2.55	2.5	1	2.5	4	6	6
TG1	61	48	58	5.08	4	4.83	4.5	4	3.5	0	1	2
TG2	41	72	37	2.93	5.14	2.64	3	4	1.5	3	1	3

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

Tab. 12: Häufigkeitsanalyse bildungssprachlicher Kodierungen insgesamt (Einfachkodierungen)

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Sprachhandlungen, Lexika, Morphosyntax												
KG	48	33	40	2.4	1.65	2	2	1	2	4	6	6
TG1	41	30	40	3.42	2.5	3.3	4	2	2.5	0	1	2
TG2	29	53	31	2.07	3.79	2.21	1.5	3.5	1	3	1	3

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

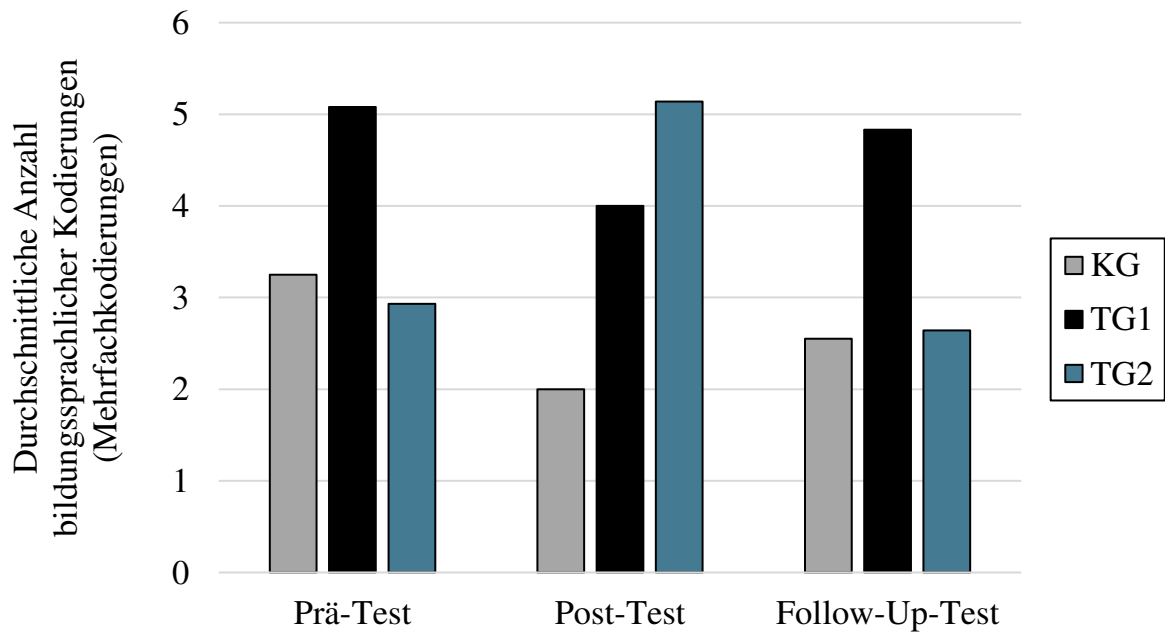


Abb. 26: Gruppenunterschiede bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV)

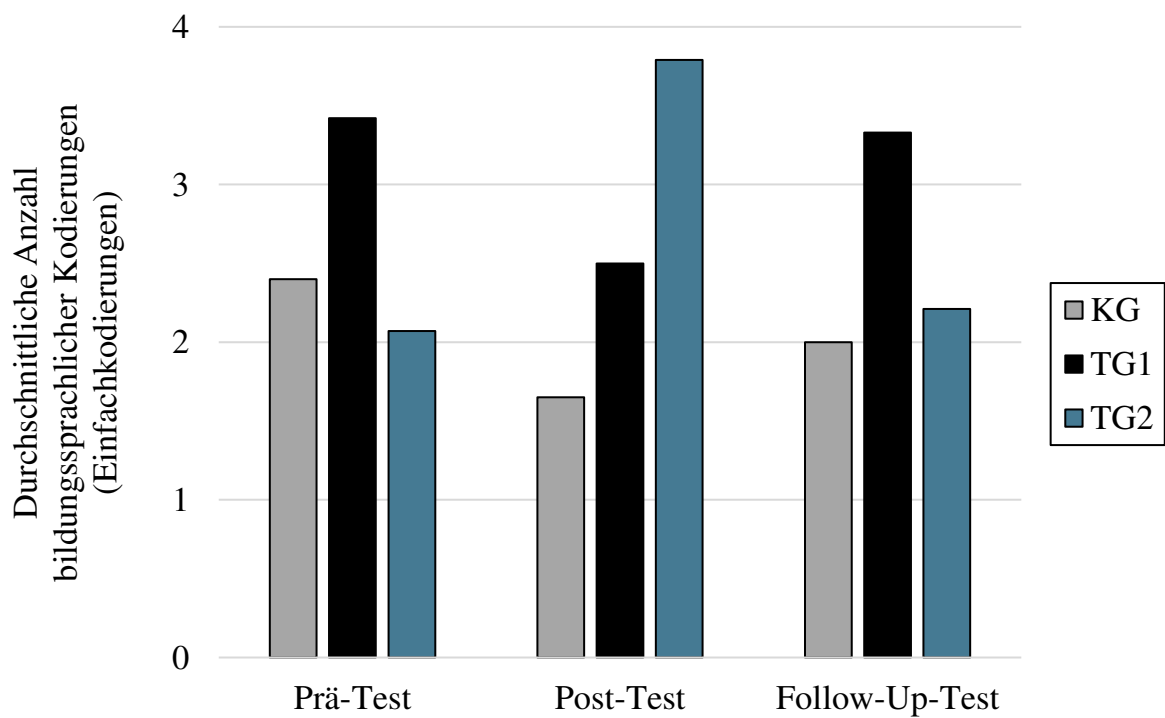


Abb. 27: Gruppenunterschiede bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV)

5.3.5 Gruppenunterschiede – Bildungssprache

Zunächst wurden die Voraussetzungen zur Berechnung der 3x1 ANOVA geprüft. Die exakten Werte des Kolomogorov-Smirnov- und des Shapiro-Wilk-Tests sind dem Anhang C.3 zu entnehmen. Grundsätzlich zeigte der Shapiro-Wilk-Test in den meisten Fällen keine Normalverteilung bezüglich der Verwendung von Bildungssprache: $p_{KG-Post-Mk} = .005$; $p_{TG1-Post-Mk} = .012$; $p_{TG2-Post-Mk} = .045$; $p_{KG-Follow-Mk} = .025$; $p_{TG2-Follow-Mk} = .016$; $p_{KG-Prä-Ek} = .039$; $p_{KG-Post-Ek} = .002$; $p_{TG1-Post-Ek} = .006$; $p_{KG-Follow-Ek} = .011$; $p_{TG2-Follow-Ek} = .026$ (Mk = Mehrfachkodierungen; EK = Einfachkodierungen). Aufgrund der verletzten Normalverteilung in fast allen Fällen war ein Abgleich der ANOVA mit dem Kruskal-Wallis-Test notwendig. Der Boxplot zeigte keine extremen Ausreißer (siehe Anhang C.3). Die Varianzhomogenität war laut Levene-Test in allen Tests bei Mehrfach- und Einfachkodierungen erfüllt, mit Ausnahme der Varianzen im Follow-Up-Test bei Mehrfachkodierungen, $p = 0.45$. Der Gruppenvergleich im Prä-Test zeigte keinen signifikanten Unterschied, sodass die Vergleichbarkeit der Gruppen bezogen auf die allgemeine Verwendung von Bildungssprache angenommen wurde, $F(2, 43) = 2.04$, $p = .142^{27}$ bei Mehrfach- und $F(2, 43) = 1.60$, $p = .215^{28}$ bei Einfachkodierungen.

Für den **Post-Test** zeigte die 3x1 ANOVA einen signifikanten Unterschied in der Verwendung von Bildungssprache zwischen den drei Gruppen sowohl bei Mehrfachkodierungen, $F(2, 43) = 4.21$, $p = .021$, $f = .44^{29}$, als auch bei Einfachkodierungen, $F(2, 43) = 3.89$, $p = .028$, $f = .43^{30}$. In beiden Fällen handelt es sich nach Cohen (1988, S. 285-287) um einen starken Effekt.

Aufgrund der gegebenen Varianzhomogenität im Post-Test wurde der Post-hoc-Test mit Bonferroni-Korrektur gewählt. Dieser zeigte für den Post-Test einen signifikanten Unterschied zwischen KG und TG2 bezüglich der Verwendung von Bildungssprache bei Mehrfachkodierungen, $p = .021$. Auch bei Einfachkodierungen unterschieden sich die Versuchsprotokolle signifikant, $p = .024$. Bei den anderen Paarvergleichen (KG vs. TG1 und TG1 vs. TG2) konnte weder bei Einfach- noch bei Mehrfachkodierungen ein signifikanter Unterschied festgestellt werden.

Im **Follow-Up-Test** unterschied sich die Verwendung von Bildungssprache nicht signifikant zwischen den drei Gruppen TG2, TG1 und KG. Dies galt für Mehrfachkodierungen $F(2, 43) = 2.372$, $p = .105^{31}$, und für Einfachkodierungen, $F(2, 43) = 1.282$, $p = .288^{32}$.

²⁷ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 3.62$, $p = .164$.

²⁸ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 3.40$, $p = .182$.

²⁹ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 9.21$, $p = .010$.

³⁰ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 7.98$, $p = .018$.

³¹ Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 2.55$, $p = .28$.

³² Der nicht-parametrische Kruskal-Wallis-Test unterstützt das Ergebnis, $H(2) = 1.77$, $p = .413$.

5.3.6 Zeitliche Unterschiede – Bildungssprache

Einen ersten Überblick über die Verwendung von Bildungssprache im zeitlichen Verlauf bieten Abb. 26 und Abb. 27.

Aus beiden Diagrammen sticht der Verlauf von TG2 hervor. Während im Post-Test von TG2 durchschnittlich mehr bildungssprachliche Wörter und Formulierungen angewandt wurden, fiel die Anzahl im Follow-Up-Test auf einen Wert unter dem des Prä-Tests (bei Mehrfachkodierungen) bzw. auf einen Wert etwas über dem des Prä-Tests (bei Einfachkodierungen).

Im Gegensatz zur TG2 verwendeten die Vpn aus der KG und TG1 im Post-Test durchschnittlich weniger Bildungssprache als zuvor im Prä-Test oder danach im Follow-Up-Test.

Mithilfe der 1x3 ANOVA wurden die beschriebenen Unterschiede auf inferenzstatistische Signifikanzen geprüft.

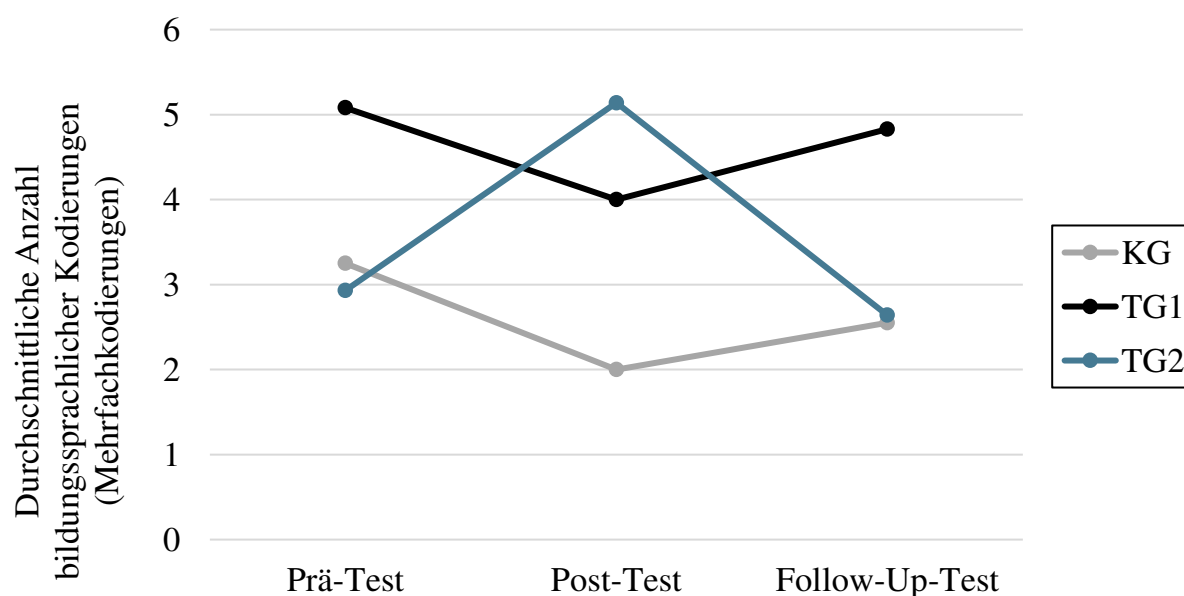


Abb. 28: Zeitlicher Verlauf bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

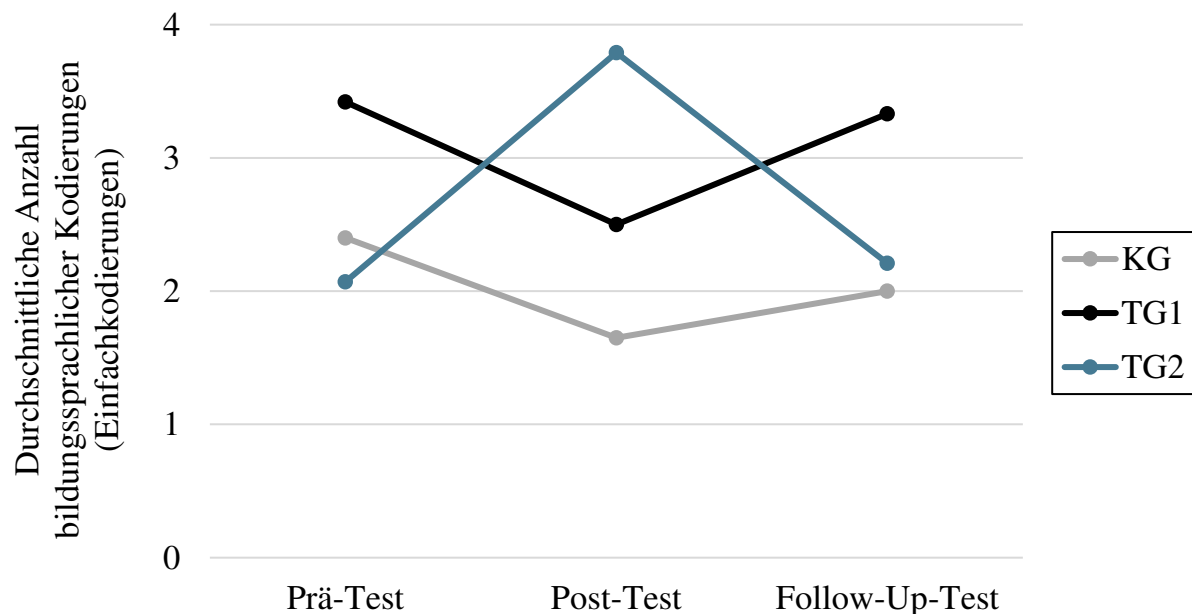


Abb. 29: Zeitlicher Verlauf bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Aufgrund der im vorherigen Kapitel beschriebenen fehlenden Normalverteilung in vielen Gruppen bezüglich der Verwendung von Bildungssprache, wurde die 1x3 ANOVA unter Kontrolle des Friedman-Tests durchgeführt. Für die 1x3 ANOVA war zusätzlich das Kriterium der Sphärizität zu prüfen. Der Mauchly-Test zeigte, dass Sphärizität in TG2 bei Mehrfachkodierungen nicht gegeben war, $p = .006$. Das Greenhouse-Geisser-Epsilon ϵ war in dieser Gruppe unter $.75$, sodass die Greenhouse-Geisser Schätzung zur Sphärizitätsabweichung interpretiert wurde (Field, 2018, S. 658). In allen anderen Gruppen und Arten der Kodierung konnte Sphärizität angenommen werden.

Die folgende Ergebnisdarstellung ist mithilfe der tabellarischen Daten im Anhang C.4 nachzuvollziehen.

Sowohl bei Mehrfachkodierungen als auch bei Einfachkodierungen konnte **in der KG** kein signifikanter Unterschied zwischen Prä-, Post- und Follow-Up-Test bezüglich der Verwendung von Bildungssprache festgestellt werden, $F(1, 19) = 2.04$, $p = .144$ ³³ bei Mehrfachkodierungen und $F(1, 19) = 1.16$, $p = .325$ ³⁴ bei Einfachkodierungen.

³³ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 3.22$, $p = .200$.

³⁴ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 5.05$, $p = .080$.

In **TG1** war ebenfalls kein signifikanter Unterschied zwischen den drei MZP erkennbar, sowohl bei Mehrfachkodierungen, $F(1,11) = .43, p = .653^{35}$ als auch bei Einfachkodierungen, $F(1,11) = .84, p = .447^{36}$.

Einen signifikanten Unterschied gab es in **TG2**: In dieser Gruppe verwendeten die Vpn unterschiedlich viele bildungssprachliche Mittel im Prä-, Post- und Follow-Up-Test. Dies galt bei Mehrfachkodierungen von Bildungssprache, $F(1, 13) = 5.43, p = .026, f = 0.65^{37}$ (mit Greenhouse-Geisser-Schätzung zur Sphärizitätsabweichung: $\epsilon = .64$) und bei Einfachkodierungen, $F(1, 13) = 4.02, p = .030, f = 0.56^{38}$. Nach Cohen (1988, S. 285-287) handelt es sich jeweils um einen starken Effekt.

Dieser Effekt wurde anhand paarweiser Vergleiche mit Bonferroni-Korrektur genauer geprüft. Demnach gab es innerhalb von TG2 einen signifikanten Unterschied zwischen Post-Test und Follow-Up-Test bei Mehrfachkodierungen von Bildungssprache, $p = .003$. Wurden bildungssprachliche Mittel nur einmal pro Dokument gezählt, so war der Unterschied mit $p = .059$ leicht nicht signifikant.

5.3.7 Zusammenfassung

Im Anschluss an die Unterrichtssequenz (Post-Test) verwendeten die drei Gruppen signifikant unterschiedlich viele bildungssprachliche Wörter und Formulierungen – sowohl bei Mehrfachkodierungen als auch unter Berücksichtigung von Einfachkodierungen je Wort und Formulierung pro Dokument. Unter Hinzunahme der deskriptiven Ergebnisse und einer genaueren Vergleichsanalyse kann zusammengefasst werden, dass **die Gruppe mit visuellem Wortspeicher-Treatment im Post-Test mehr bildungssprachliche Mittel nutzte als die Gruppe ohne Treatment**.

10 Wochen nach der Unterrichtssequenz (Follow-Up-Test) unterschieden sich die drei Gruppen nicht signifikant in ihrer Verwendung von bildungssprachlichen Wörtern und Formulierungen außerhalb des Wortspeichers.

Bei der Analyse des zeitlichen Verlaufs zeigte sich, dass **die Gruppe mit visuellem Wortspeicher-Treatment direkt nach der Unterrichtssequenz signifikant mehr bildungssprachliche Mittel anwandte als 10 Wochen später**. Wurden bildungssprachliche Mittel nur einmal pro Dokument gezählt, gab es inferenzstatistisch keinen Unterschied.

³⁵ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 1.90, p = .386$.

³⁶ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 3.35, p = .187$.

³⁷ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 10.08, p = .006$.

³⁸ Der nicht-parametrische Friedman-Test unterstützt das Ergebnis, $\chi^2(2) = 6.17, p = .046$.

5.4 F 3 – Inhalt

Die dritte Forschungsfrage lautet wie folgt:

F3 – Inhalt: Lässt sich durch visuell oder auditiv angebotene Wortspeicher der Inhalt in Versuchsprotokollen des naturwissenschaftlichen Sachunterrichts verbessern?

Die Beurteilung des Inhalts ist methodisch auf differenzierte Art möglich. In diesem Forschungsvorhaben fiel die Wahl der Beurteilungsmethode – aus Gründen der Übersichtlichkeit – auf ein Kategoriensystem mit *korrekten*, *unzureichenden* und als *fehlerhaft* zu bewertenden Antworten. Die Häufigkeiten der korrekt, unzureichend oder fehlerhaft zu bewertenden Antworten werden im folgenden Kapitel vorgestellt. Daran schließt sich die inferenzstatistische Analyse bezüglich möglicher Gruppenunterschiede oder zeitlicher Unterschiede innerhalb der Gruppen an.

5.4.1 Kodierungen des Inhalts

In jedem Versuchsprotokoll wurden neun Abschnitte bewertet: Die drei Vermutungen, die Auflistung des Materials, die Durchführung, die drei Beobachtungen und die Erklärung. Das Kategoriensystem (*korrekt*, *unzureichend*, *fehlerhaft*) ist nur schwer auf die Auflistung des Materials übertragbar. Um eine klare inhaltliche Beurteilung der Materialaufzählung zu erhalten, wurde die Häufigkeit der genannten Materialien direkt kodiert und ausgewertet. Das Ergebnis der Analyse *Materialaufzählung* ist in Tab. 13 ersichtlich. Im oberen Teil der Tabelle steht die Anzahl der genannten Materialien von 0 bis 7. Interessant ist die Betrachtung, wann die Vpn kein einziges Material (Anzahl der Materialien = 0) aufschrieben. Während es im Prä-Test aus allen drei Gruppen mindestens ein oder zwei Vpn gab, die kein einziges Material nannten, wurden im Post-Test von jeder Vp aus jeder Gruppe mindestens zwei oder drei benötigte Materialien aufgezeigt. Zehn Wochen nach der Unterrichtssequenz gab es in der KG erneut zwei Vpn, die keine Materialien nannten.

Der untere Teil der Tab. 13 gibt einen besseren Überblick über die dargestellten Ergebnisse des oberen Teils. Werden alle genannten Materialien pro Gruppe und pro Test aufsummiert, so zeigt sich eine höhere Anzahl an genannten Materialien in der KG im Post- und im Follow-Up-Test. Auch unter Berücksichtigung der größeren Stichprobe ist der Unterschied im Post-Test erkennbar ($M_{KG} = 5.05$, $M_{TG2} = 4.79$, $M_{TG1} = 3.83$). Die zusammengefasste Anzahl der Dokumente ohne Kodierungen unterstreicht, dass im Post-Test – im Gegensatz zum Prä-Test – jede Vpn eine gewisse Anzahl an Materialien auflistete.

Tab. 13: Häufigkeitsanalyse der genannten Materialienanzahl

Anzahl der genannten Materialien von 0 bis 7												
	0	1	2	3	4	5	6	7				
Prä-Test												
KG	2	0	1	8	7	2	0	0				
TG1	1	0	2	3	4	2	0	0				
TG2	2	0	0	4	8	0	0	0				
Post-Test												
KG	0	0	0	1	2	13	3	1				
TG1	0	0	2	2	5	2	1	0				
TG2	0	0	0	2	3	5	4	0				
Follow-Up-Test												
KG	2	0	0	2	1	5	4	6				
TG1	0	0	1	2	0	2	7	0				
TG2	0	0	1	2	0	8	3	0				
Anzahl der genannten Materialien insgesamt												
	PRÄ Anzahl	POST Anzahl	FOLLO W Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOL- LOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
KG	64	101	101	3.2	5.05	5.05	3	5	5.5	2	0	2
TG1	39	46	60	3.25	3.83	5	3.5	4	6	1	0	0
TG2	44	67	66	3.14	4.79	4.71	4	5	5	2	0	0

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe (*n* = 20), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / *n* = 12), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / *n* = 14)

Abb. 30 verdeutlicht das Ergebnis durch die visuelle Darstellung der Mittelwerte.

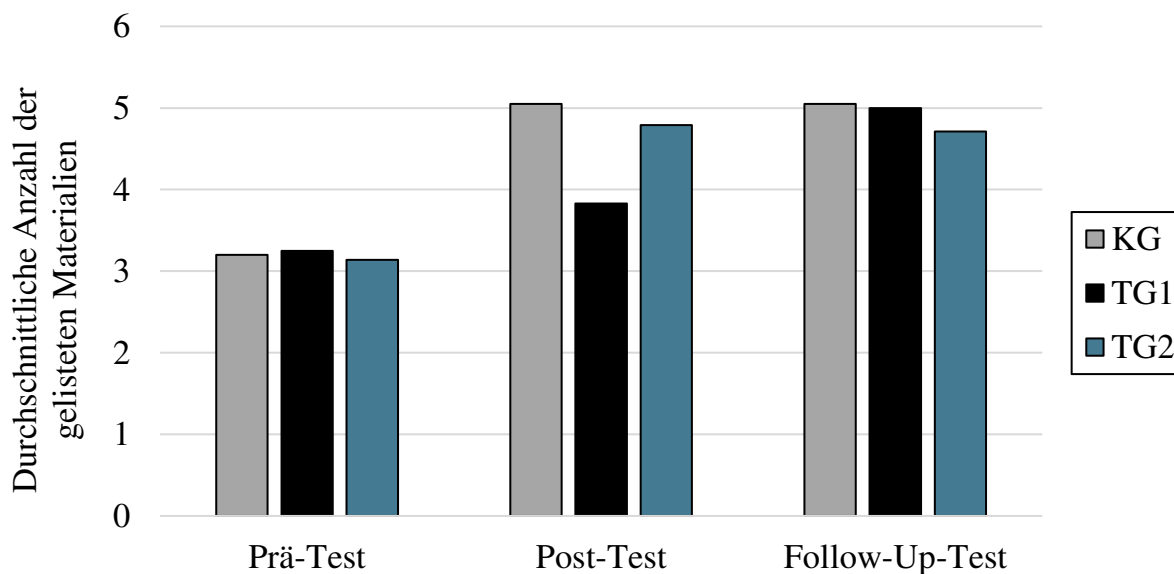


Abb. 30: Gruppenunterschiede bezogen auf die Nennung der Materialien (AV)

Neben der bewerteten Auflistung von Materialien wurden die anderen Abschnitte, d.h. drei *Vermutungen*, die *Durchführung*, drei *Beobachtungen* und die *Erklärung*, per Kategoriensystem bewertet. In Tab. 14 ist die Häufigkeit dieser korrekten, unzureichend oder fehlerhaft kodierten Abschnitte zusammengefasst.

Auffällig ist der hohe Mittelwert korrekter Antworten im Post-Test der KG und der TG2 ($M_{KG} = 4.8$, $M_{TG2} = 4.5$, $M_{TG1} = 3.5$). Die gleiche Tendenz ist bereits vor der Unterrichtssequenz (im Prä-Test) erkennbar. Der Median der korrekten Antworten im Post-Test gibt weiterhin Aufschluss darüber, dass sich die korrekt bewerteten Antworten der KG und TG2 nicht nur auf einzelne Versuchsprotokolle beziehen.

Bei den unzureichenden Antworten sind es im Post-Test die Versuchsprotokolle der Vpn der Treatmentgruppen, die im Mittel die meisten Kodierungen aufwiesen ($M_{TG2} = 2.43$, $M_{TG1} = 2.42$, $M_{KG} = 1.5$).

Die meisten fehlerhaften Antworten wurden – bezogen auf die Stichprobengröße – im Post-Test von TG1 aufgeschrieben, während die Versuchsprotokolle aus TG2 die wenigsten fehlerhaften Antworten aufwiesen ($M_{TG1} = 2.08$, $M_{KG} = 1.7$, $M_{TG2} = 1.07$). Interessant bei der Betrachtung fehlerhafter Antworten ist nicht nur der Mittelwert, sondern auch die Anzahl der Versuchsprotokolle, in denen keine einzige fehlerhafte Antwort vorkam. Im Post-Test der Treatmentgruppen war dies jeweils bei 6 von 14 Vpn (TG2: 43 %) und 4 von 12 Vpn (TG1: 33 %) der Fall. In der KG gab es im Gegensatz dazu lediglich 3 von 20 Versuchsprotokolle (KG: 15 %), in denen im Post-Test keine einzige fehlerhafte Antwort zu verzeichnen war.

Da die inhaltliche Unterkategorie *fehlerhaft* sowohl inkorrekte als auch fehlende Antworten umfasst, war eine Spezifikation der fehlerhaften Antworten für ein umfassenderes Ergebnisbild erforderlich. So wurden fehlende Antworten zusätzlich erfasst. Da es sich bei den fehlenden Antworten um keine eigenständige Kategorie handelt, sind die Häufigkeitsangaben der fehlenden Antworten bereits in der Unterkategorie *fehlerhaft* integriert. Die Häufigkeitsangabe der fehlenden Antworten zeigt, dass im Post-Test die Vpn aus TG1 im Schnitt die meisten Felder leer ließen, gefolgt von TG2 und zuletzt der KG, $M_{TG1} = 0.83$, $M_{TG2} = 0.36$, $M_{KG} = 0.1$. Im Follow-Up-Test wiederum waren die meisten fehlenden Antworten in der KG und kaum fehlende Antworten in TG2 zu verzeichnen, $M_{KG} = 0.75$, $M_{TG1} = 0.5$, $M_{TG2} = 0.07$. Die Anzahl der Versuchsprotokolle ohne fehlende Antworten zeigt, dass im Follow-Up-Test 10 von 20 Vpn (50 %) aus der KG mindestens zu einem Abschnitt nichts schreiben konnten, in TG 1 und 2 waren es jeweils 5 von 12 (TG1: 42 %) beziehungsweise 5 von 14 Vpn (TG2: 36 %).

Tab. 14: Häufigkeitsanalyse der inhaltlichen Kodierungen

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Korrekt												
KG	92	96	45	4.6	4.8	2.25	5	5	2.5	0	0	5
TG1	46	42	32	3.83	3.5	2.67	4	2.5	2.5	0	0	1
TG2	59	63	36	4.21	4.5	2.57	4.5	4	2	0	1	0
Unzureichend												
KG	33	30	54	1.65	1.5	2.7	1.5	1	3	3	7	1
TG1	21	29	38	1.75	2.42	3.17	1	2.5	3	1	3	0
TG2	17	34	41	1.21	2.43	2.93	1.5	2.5	3	4	1	1
Fehlerhaft												
KG	35	34	61	1.75	1.7	3.05	1.5	2	3	4	3	1
TG1	29	25	26	2.42	2.08	2.17	2	2	2	0	4	0
TG2	36	15	35	2.57	1.07	2.5	2	1	3	0	6	2
Spezifikation Fehlerhaft: Fehlende Antworten												
KG	6	2	15	0.3	0.1	0.75	0	0	0.5	18	18	10
TG1	11	10	6	0.92	0.83	0.5	0	0	0	7	7	7
TG2	15	5	10	1.07	0.36	0.07	0	0	0	10	12	9

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

5.4.2 Inhaltliche Punktzahl insgesamt

Um die Kodier-Ergebnisse zu quantifizieren, wurden den einzelnen Ausprägungen (*korrekt, fehlerhaft, unzureichend*) Punkte zugewiesen. Es konnte insgesamt ein Maximalwert von 16 Punkten erreicht werden. Eine hohe Gesamtpunktzahl lässt auf eine hohe Anzahl korrekter bzw. unzureichender Antworten schließen.

Im Anhang C.4 sind die deskriptiven Ergebnisse dargestellt.

In Abb. 31 werden die Gruppenunterschiede bezüglich der durchschnittlichen inhaltlichen Punktzahl visuell vorgestellt.

Im Post-Test wurden die Vpn aus der Wortspeicher-Plakat-Gruppe durchschnittlich mit der höchsten Punktzahl beurteilt ($M = 11.43, SD = 3.01$), knapp gefolgt von der KG ($M = 11.10, SD = 2.77$). Die geringste Punktzahl ist in der Gruppe zu verzeichnen, die die Wortspeicher auditiv vermittelt bekamen ($M = 9.42, SD = 4.08$).

Im Gegensatz zu den Ergebnissen im Post-Test war **im Follow-Up-Test** die höchste inhaltliche Punktzahl in TG1 ($M = 8.50, SD = 2.20$) zu verzeichnen, gefolgt von TG2 ($M = 8.07, SD = 2.62$) und zuletzt der KG ($M = 7.20, SD = 2.78$).

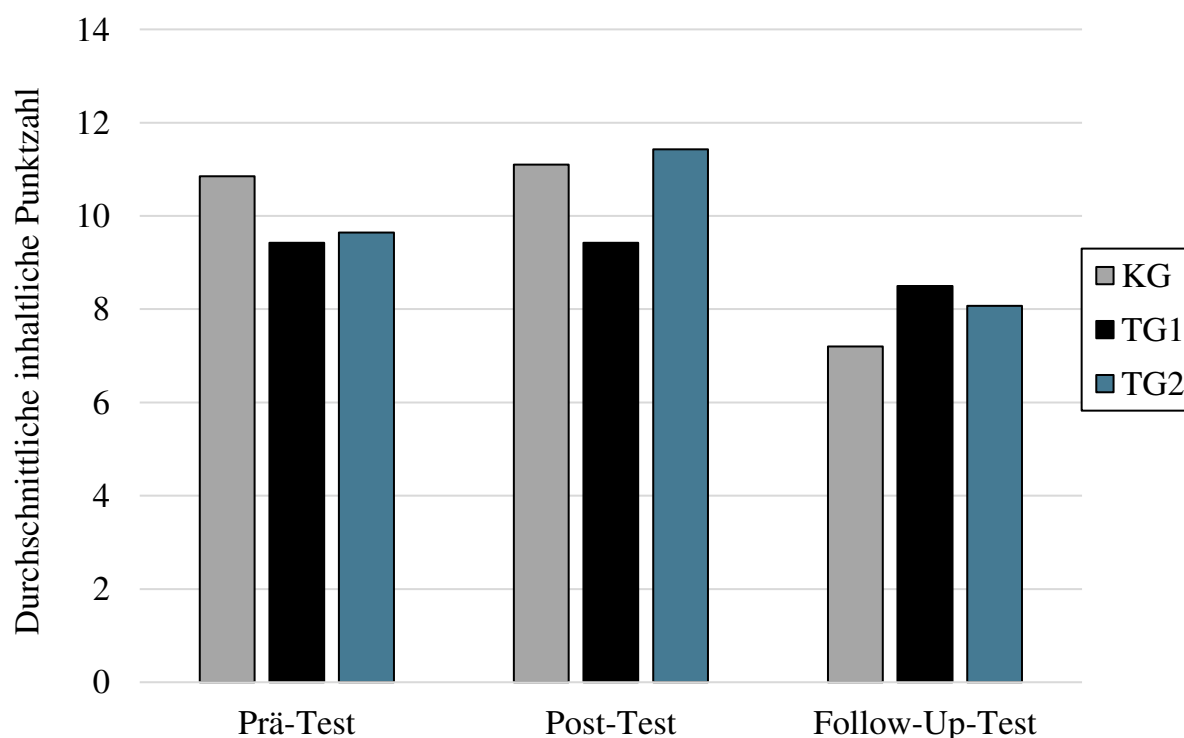


Abb. 31: Gruppenunterschiede bezogen auf die inhaltliche Punktzahl (AV)

5.4.3 Gruppenunterschiede – Inhalt

Zur Durchführung der 3x1 ANOVA wurden die Daten vorab auf Normalverteilung und Varianzhomogenität geprüft. Im Anhang C.4 befinden sich die detaillierten Ergebnisse.

Laut Shapiro-Wilk-Test waren die Daten in allen drei Gruppen im Post- und im Follow-Up-Test normalverteilt, im Post-Test mit $p = .211$ (KG), $p = .241$ (TG1), $p = .191$ (TG2) und im Follow-Up-Test mit $p = .199$ (KG), $p = .609$ (TG1) und $p = .054$ (TG2). Der Boxplot zeigte keine extremen Ausreißer (siehe Anhang C.4). Auch die Varianzhomogenität war gemäß dem Levene-Test basierend auf dem Median gegeben, im Post-Test mit $p = .066$ und im Follow-Up-Test mit $p = .649$. Aufgrund der erfüllten Voraussetzungen entfiel der Abgleich mit dem nicht-parametrischen Kruskal-Wallis-Test. Zur Kontrolle der Vergleichbarkeit von KG, TG1 und TG2 wurden Gruppenunterschiede im Prä-Test geprüft. Da kein signifikanter Unterschied bestand, $F(2, 43) = 1.29$, $p = .285$, wurde die Vergleichbarkeit der Gruppen angenommen.

Im Post-Test unterschieden sich die drei Gruppen nicht signifikant im Inhalt, $F(2, 43) = 1.46$, $p = .244$.

Auch **im Follow-Up-Test** gab es keinen signifikanten Unterschied zwischen KG, TG1 und TG2 bezüglich der inhaltlich erreichten Punktzahl, $F(2, 43) = 1.05$, $p = .359$.

Im Folgenden wurden mithilfe der 1x3 ANOVA mögliche Unterschiede innerhalb der Gruppen bezogen auf den zeitlichen Verlauf analysiert.

5.4.4 Zeitliche Unterschiede – Inhalt

In Abb. 32 ist die durchschnittliche inhaltliche Punktzahl im zeitlichen Verlauf der drei MZP dargestellt. Im Post-Test erhielten die Gruppen tendenziell mehr (KG, TG2) oder gleich viele Punkte (TG1) wie im Prä-Test. Auffällig ist der Follow-Up-Test, in dem alle Gruppen im zeitlichen Verlauf die wenigsten inhaltlichen Punkte erzielten.

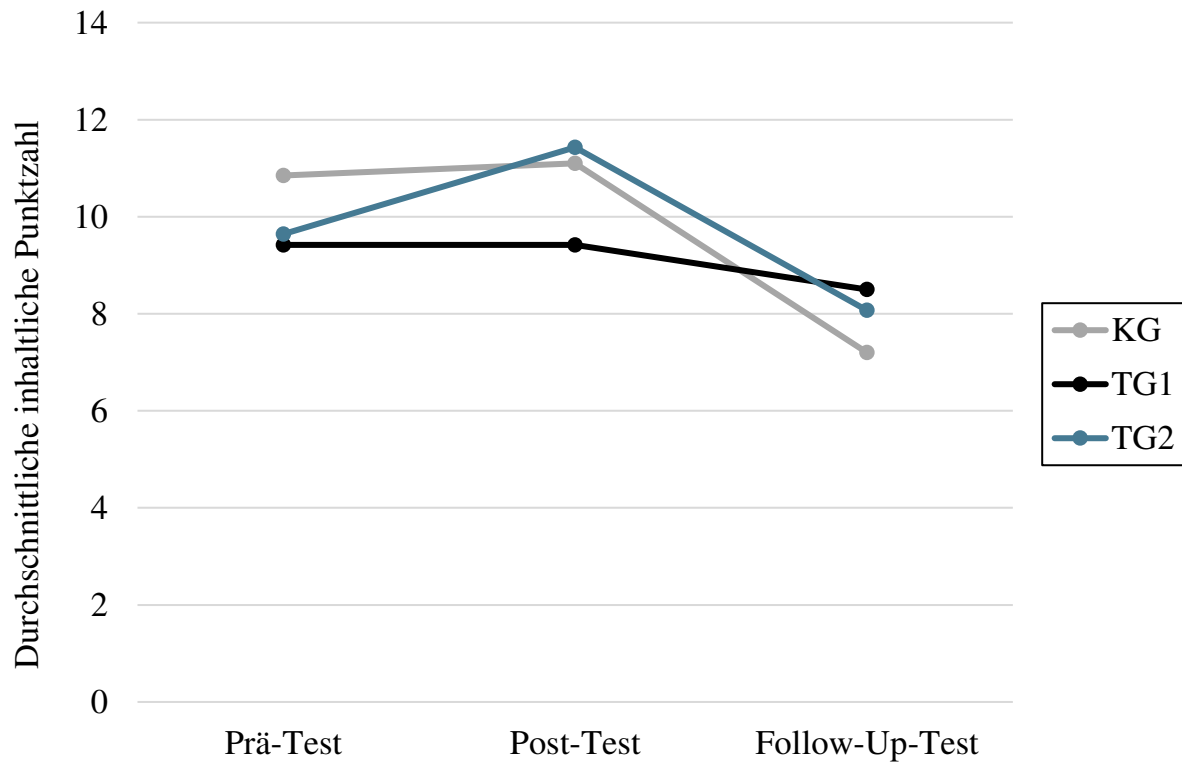


Abb. 32: Zeitlicher Verlauf bezogen auf die inhaltliche Punktzahl (AV)

KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Da die inhaltliche Punktzahl in den einzelnen Gruppen und zu den drei MZP normalverteilt war, konnte die 1x3 ANOVA zur Testung von Unterschieden im zeitlichen Verlauf ohne kontrollierenden Abgleich durchgeführt werden. Der Mauchly-Test ergab, dass auch Sphärizität in allen drei Gruppen gegeben war, mit $p = .979$ in KG, $p = .551$ in TG1 und $p = .753$ in TG2.

Das Ergebnis der 1x3 ANOVA ist im Anhang C.4 dargestellt.

In TG1 gab es keinen signifikanten Unterschied bezüglich der inhaltlichen Punktzahlen im Prä-, Post- und im Follow-Up-Test, $F(1, 11) = 0.61$, $p = .551$.

Anders sieht das Ergebnis **in der KG und in TG2** aus. Sowohl in der KG als auch in TG2 unterschieden sich die inhaltlichen Punktzahlen über die drei Messzeitpunkte hinweg signifikant, mit $F(1, 19) = 11.84$, $p < .001$, $f = 0.79$ in der KG und $F(1, 13) = 6.89$, $p = .004$, $f = 0.73$ in TG2. In beiden Fällen handelt es sich nach Cohen (1988, S. 285-287) um einen starken Effekt.

Für einen detaillierteren Einblick wurden paarweise Vergleiche mit Bonferroni-Korrektur vorgenommen.

In der KG war die inhaltliche Punktzahl signifikant verschieden zwischen dem Prä- und dem Follow-Up-Test, $p = .002$ als auch zwischen dem Post- und dem Follow-Up-Test, $p = .001$.

In TG2 beruhte der signifikante Unterschied auf dem Vergleich zwischen Post- und Follow-Up-Test, $p = .004$.

5.4.5 Zusammenfassung

Zu keinem Messzeitpunkt ließ sich ein inhaltlicher Unterschied zwischen den Gruppen feststellen, die zuvor mit auditiven, visuellen oder ohne Wortspeicher lernten.

Innerhalb der Gruppen wurden Unterschiede aufgedeckt. Unter Einbezug der deskriptiven Ergebnisse kann festgehalten werden, dass **die Gruppe ohne Treatment vor der Unterrichtssequenz (Prä-Test: 10.85 Punkte) und direkt nach der Unterrichtssequenz (Post-Test: 11.10 Punkte) signifikant mehr inhaltliche Punkte erzielte als 10 Wochen nach der Unterrichtssequenz (Follow-Up: 7.20 Punkte)**. Auch in TG2 wurde ein signifikanter Unterschied im zeitlichen Verlauf festgestellt: **Die Gruppe, in der Wortspeicher in Plakatform vorgestellt wurden, erhielt direkt nach der Unterrichtssequenz (Post-Test: 11.43 Punkte) deutlich mehr inhaltliche Punkte als 10 Wochen nach der Unterrichtssequenz (Follow-Up: 8.07 Punkte)**. Die Punktangaben in Klammern sind die jeweils erreichten inhaltlichen Mittelwerte.

5.5 F 4 – Datenintegration

Die vierte Forschungsfrage verbindet die vorherigen Forschungsfragen wie folgt:

F4 – Datenintegration: Fördert die visuelle oder auditive Vermittlung von Wortspeichern das Verfassen von bildungssprachlichen und inhaltlich zutreffenden Versuchsprotokollen im naturwissenschaftlichen Sachunterricht?

5.5.1 Ergebnisse insgesamt

In diesem Kapitel werden die Ergebnisse aus den vorherigen Kapiteln in einem übergeordneten Kontext zusammengefasst und mit qualitativen Daten in Verbindung gebracht. Tab. 15 zeigt einen ersten Überblick über die Gruppenunterschiede im Post-Test. Die drei Gruppen unterschieden sich im Post-Test signifikant in der Verwendung der Wortspeicher und in der Verwendung von Bildungssprache. Inhaltlich war kein signifikanter Unterschied zwischen den Gruppen festzustellen.

Tab. 15: Wortspeicher, Bildungssprache und Inhalt im Post-Test der *Panel*-Gruppen

	KG	TG1	TG2	Sig.
Wortspeicher				
Wortspeicher	0.3 (6)	1.33 (16)	2 (28)	$p = .008$
Bildungssprache				
Sprachhandlungen	0.15 (3)	0.33 (4)	2.07 (29)	
Lexikon	0.45 (9)	2 (24)	1.14 (16)	
Morphosyntax	1.4 (28)	1.67 (20)	1.93 (27)	
Bildungssprache insgesamt	2 (40)	4 (48)	5.14 (72)	$p = .021$
Inhalt				
korrekt	4.8 (96)	3.5 (42)	4.5 (63)	
unzureichend	1.5 (30)	2.42 (29)	2.43 (34)	
fehlerhaft	1.7 (34)	2.08 (25)	1.07 (15)	
Inhaltliche Gesamtpunktzahl	11.10 (222)	9.42 (113)	11.43 (160)	$p = .244$

Angaben bezogen auf Stichprobengröße; Zahlen in Klammern = Absolute Zahlen; Sig. = Signifikanz der 3x1 ANOVA; Mehrfachzählungen möglich

Im Follow-Up-Test war in keiner der untersuchten Kategorien ein statistisch signifikanter Unterschied erkennbar (Tab. 16).

Tab. 16: Wortspeicher, Bildungssprache und Inhalt im Follow-Up-Test der *Panel-*Gruppen

	KG	TG1	TG2	Sig.
Wortspeicher				
Wortspeicher	0.4 (8)	0.58 (7)	1 (14)	$p = .336$
Bildungssprache				
Sprachhandlungen	0.1 (2)	1.17 (14)	0.79 (11)	
Lexikon	0.85 (17)	1.17 (14)	0.71 (10)	
Morphosyntax	1.6 (32)	2.5 (30)	1.14 (16)	
Bildungssprache insgesamt	2.55 (51)	4.83 (58)	2.64 (37)	$p = .105$
Inhalt				
korrekt	2.25 (45)	2.67 (32)	2.57 (36)	
unzureichend	2.7 (54)	3.17 (38)	2.93 (41)	
fehlerhaft	3.05 (61)	2.17 (26)	2.5 (35)	
Inhaltliche Gesamtpunktzahl	7.2 (144)	8.5 (102)	8.07 (113)	$p = .359$

Angaben bezogen auf Stichprobengröße; Zahlen in Klammern = Absolute Zahlen; Sig. = Signifikanz der 3x1 ANOVA; Mehrfachzählungen möglich

Im Folgenden wird dargestellt, in welchen Abschnitten des Versuchsprotokolls die Wortspeicher verwendet wurden. In einem zweiten Schritt wird die Verwendung anderweitiger Bildungssprache pro Abschnitt untersucht. Zuletzt wird die Überschneidung von Wortspeichern und Bildungssprache mit inhaltlichen Beurteilungen innerhalb eines Abschnitts deutlich. Ein Abschnitt umfasst eine Kategorie des Versuchsprotokolls, d.h.: entweder die Vermutung, die Auflistung des Materials, die beschriebene Durchführung, die Beobachtung oder die Erklärung. Die Synthese aus quantitativer und qualitativer Auswertung trägt zu einem tieferen Verständnis der Ergebnisse bei.

Mithilfe des *Code-Relations-Browsers* in MAXQDA wurden zunächst Überschneidungen der Wortspeicher mit den jeweiligen Abschnitten des Versuchsprotokolls detektiert. Die Ergebnisse sind in Tab. 17 ersichtlich.

Tab. 17: Verwendung des Wortspeichers pro Abschnitt

		Vermutung	Material	Durchführung	Beobachtung	Erklärung	Summe
Wortspeicher							
PRÄ	KG	1	0	2	1	1	5
	TG1	1	0	0	0	2	3
	TG2	6	0	0	0	0	6
POST	KG	0	0	5	0	1	6
	TG1	6	1	7	0	2	16
	TG2	15	3	9	0	1	28
FOLLOW	KG	5	0	2	0	1	8
	TG1	0	0	6	1	0	7
	TG2	4	0	10	0	0	14
Summe		38	4	41	2	8	93

Mehrfachkodierungen möglich; die absoluten Zahlen spiegeln nicht den Unterschied der unterschiedlichen Gruppengröße wider (KG: $n = 20$; TG1: $n = 12$; TG2: $n = 14$).

Tab. 17 zeigt die absoluten Kodierungszahlen mit möglichen Mehrfachzählungen pro Dokument. Da die Stichprobengröße innerhalb von KG, TG1 und TG2 unterschiedlich ist, dürfen die Werte nicht zur Gruppenunterscheidung herangezogen werden. Stattdessen geben sie einen Überblick über die allgemeine Nutzung der Wortspeicher in Bezug auf die einzelnen Abschnitte. Die Daten zeigen, dass die Wortspeicher im Post- und im Follow-Up-Test **gruppenübergreifend vor allem zur Aufstellung einer Vermutung oder zur Beschreibung der Durchführung** verwendet wurden. Die Nutzung der Wortspeicher zur Auflistung des Materials, zur Beschreibung der Beobachtung und zur Erklärung tendiert gegen 0, mit vereinzelt Fällen von 1-2 Wörtern oder Formulierungen. Bei der Auflistung des Materials nutzten die Vpn aus TG2 im Post-Test zumindest in drei Fällen den Wortspeicher.

Die qualitativen Daten in Tab. 18 ergänzen die Ergebnisse aus Tab. 17. Zur Aufstellung einer **Vermutung** griffen die Vpn aus allen drei Gruppen auf die Formulierungen *Ich vermute, dass...* oder *Meine Vermutung ist, dass...* zurück. In beiden Treatmentgruppen war das Repertoire mit *Ich denke, dass...* leicht erweitert. In der KG kam zusätzlich das Wortspeicher-Wort *danach* im Abschnitt der Vermutung vor. Die Verwendung war im Kontext richtig, da es ansonsten nicht kodiert worden wäre. Es diente in diesem Fall jedoch nicht der Sprachhandlung des Aufstellens einer Vermutung.

Gleiches gilt für das Wort *zunächst*, dass bei der Auflistung des **Materials** in TG1 sinnvoll eingesetzt wurde und auf der anderen Seite nicht eindeutig den Abschnitt des

Materials widerspiegelt. Nicht abschnittskonforme – aber dennoch korrekt eingesetzte Wörter und Formulierungen aus dem Wortspeicher – sind in Tab. 18 kursiv gestellt. Bei den drei Fällen aus TG2, in denen im Post-Test der Wortspeicher für die Auflistung des Materials genutzt wurde, handelt es sich um drei verschiedene Vpn. Alle drei Vpn leiteten ihr Material mit der Wortspeicher-Formulierung *Man benötigt ...* ein.

Die **Durchführung** wurde am vielseitigsten von TG1 mit den Worten *anschließend, danach, daraufhin* oder *zunächst* beschrieben. In TG2 kamen *anschließend, danach* und *daraufhin* vor. **Auffällig ist die hohe Anzahl des kodierten Wortes *danach* (14x) in TG2.** In der KG wurden zwei Wörter aus dem Wortspeicher für die Durchführung verwendet: *anschließend* und *danach*. Das Wort *danach* fand in der KG in jedem Abschnitt mit Ausnahme des Materials Verwendung.

Bei der **Beobachtung** stand in keinem der Versuchsprotokolle eine abschnittskonforme Formulierung, wie beispielsweise *Man kann beobachten, dass ...*

Bezüglich der **Erklärung** kam nur in der KG und in TG1 das abschnittskonforme Wort *deshalb* vor. Die weiteren Sprachmittel aus dem Wortspeicher wurden bei der Erklärung nicht abschnittskonform verwendet. Dabei handelte es sich hauptsächlich um Sprachmittel, die im Wortspeicher der *Vermutung* oder der *Beobachtung* zugeordnet waren. Tab. 18 zeigt, dass die Verwendung von vermutungsäußernden Sprachhandlungen aus dem Wortspeicher auf die Unkenntnis der richtigen Erklärung zurückzuführen ist. Eigene Erklärungen wurden vorsichtig und insofern mit einer vermutungseinleitenden Formulierung eingeführt. Die Verwendung der *Wenn..., dann*-Struktur ist im Wortspeicher dem Abschnitt der Beobachtung zugeordnet. In den vorliegenden Versuchsprotokollen wurde die Beobachtung mithilfe der *Wenn..., dann*-Struktur im Abschnitt der Erklärung wiederholt um darauf eingehend eine Erklärung zu formulieren.

Tab. 18: Wortspeicher (qualitativ) pro Abschnitt des Versuchsprotokolls

Gruppe	Abschnitt	Wortspeicher
KG	Vermutung	<i>Danach (2x)</i> Ich vermute, dass ... (3x) Meine Vermutung ist, dass ...
TG1		Ich denke, dass ... Ich vermute, dass ... (3x) Meine Vermutung ist, dass ... (3x)
TG2		Ich denke, dass ... Ich vermute, dass ... (14x) Meine Vermutung ist, dass ... (10x)
KG	Material	-
TG1		<i>Zunächst</i>
TG2		Man benötigt (3x)
KG	Durchführung	Anschließend (2x) Danach (7x)
TG1		Anschließend Danach (9x) Daraufhin Zunächst (2x)
TG2		Anschließend (2x) Danach (14x) Daraufhin (3x)
KG	Beobachtung	<i>Danach</i>
TG1		<i>Zunächst</i>
TG2		-
KG	Erklärung	<i>Danach</i> Deshalb <i>Wenn man das Ei in normales Wasser, dann ...</i>
TG1		Deshalb <i>Ich vermute, dass ...</i> <i>Wenn das Wasser neu dazu kommt, dann ...</i> <i>Wenn Salz mit Wasser zusammenstößt, dann wird das höher ...</i>
TG2		<i>Ich denke, dass ...</i>

KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 ($n = 12$), TG2 = Treatmentgruppe 2 ($n = 14$); nicht abschnittskonforme Wörter / Formulierungen sind kursiv hervorgehoben. Die Daten sind zusammengefasst aus Prä-, Post- und Follow-Up-Test.

In Anlehnung an Tab. 17 zeigt Tab. 19 die Anzahl der kodierten Bildungssprache, die pro Abschnitt des Versuchsprotokolls vorlag.

Tab. 19: Verwendung von Bildungssprache pro Abschnitt

		Vermutung	Material	Durchführung	Beobachtung	Erklärung	Summe
		Bildungssprache					
PRÄ	KG	8	0	19	17	19	63
	TG1	19	0	6	15	17	57
	TG2	5	0	12	9	13	39
POST	KG	6	0	16	5	13	40
	TG1	11	0	13	14	10	48
	TG2	9	1	16	12	34	72
FOLLOW	KG	8	0	16	8	16	48
	TG1	10	0	19	7	14	50
	TG2	3	0	11	2	16	32
Summe		79	1	128	89	152	449

Mehrfachkodierungen möglich; die absoluten Zahlen spiegeln nicht den Unterschied der unterschiedlichen Gruppengröße wider (KG: $n = 20$; TG1: $n = 12$; TG2: $n = 14$).

Im Gegensatz zu den Wortspeichern (Tab. 17) zeigt sich bezüglich anderweitiger Bildungssprache ein vielseitigeres Ergebnis. Bildungssprache wurde in allen Abschnitten des Versuchsprotokolls mit Ausnahme des Materials verwendet. Hervorgehoben ist die kodierte Bildungssprache in den Erklärungen der Post-Tests. An dieser Stelle zeigt sich ein großer Unterschied zwischen TG2, in der insgesamt 34 kodierte bildungssprachliche Elemente vorkamen und den beiden anderen Gruppen mit 10 (TG1) bis 13 (KG) kodierten bildungssprachlichen Wörtern oder Formulierungen.

Die qualitative Vertiefung für den konkreten Sachverhalt ist in Tab. 20 dargestellt. Aus der Tabelle wird deutlich, dass der Unterschied vor allem auf einer **Vielzahl an Sprachhandlungen im Post-Test von TG2 im Abschnitt der Erklärungen** beruht. Das Ergebnis korrespondiert mit der zweiten Teilstudie und dessen Erkenntnis, dass TG2 im Post-Test im Allgemeinen sehr viele Sprachhandlungen in den Versuchsprotokollen aufwies. Als neuer Aspekt zur zweiten Teilstudie kommt hinzu, dass es sich zum Großteil um Sprachhandlungen für Erklärungsansätze handelte.

Tab. 20: Bildungssprache (qualitativ) in den Erklärungen der Post-Tests

Gruppe	Bildungssprache im Abschnitt „Erklärungen“ im Post-Test
Sprachhandlungen	
KG	Durch das Salz wird das Ei oben gehalten. weil <i>Wenn es mehr Wasser wird, kann das Salz das Ei nicht tragen.</i>
TG1	das ist meine Erklärung dafür
TG2	<i>Also glaube ich</i> Das ist meine eigene Erklärung. Deswegen (4x) Ich erkläre mir das so: Ich erkläre mir das so, dass Meine Erklärung ist, dass Wegen (3x) Weil (9x) <i>Weil wenn man im Meer badet, geht man, wenn man sich auf den Rücken legt, nicht unter.</i>
Lexikon	
KG	schweben Wasseroberfläche
TG1	Oberfläche (2x) schwebt quasi
TG2	hohl treiben lassen Widerstand
Morphosyntax	
KG	..., aber es ist nicht zerbrochen. Das Salz macht, dass das Ei schwimmt ... Und am Ende ... Und da war dann ... wenn man Salz in den Becher gibt, ... wenn mehr Wasser kommt, schwimmt es in der Mitte. ..., wie eine Strömung und das Wasser wird das Ei oben gehalten.
TG1	Als ich das Ei in das Wasser gelegt habe, ist das Wasser gestiegen. ..., bei Versuch 1 aber nicht Das heißt, dass das Ei mit Salz an der Oberfläche schwimmt. Das Salz drückt das Ei dann nach oben. ..., dass es ...

TG2 ..., aber im anderen Becher war kein Salz.
 Das Ei geht auch nicht unter, wenn ich noch 100 ml Wasser dazu schützte,
 ...
 Dann schützte 4 Teelöffel Salz in den Messbecher zwei.
 ..., dann das Ei in den zweiten Becher tun.
 ..., sondern es ist untergegangen.
 ..., sondern es schwimmt.
 ..., sondern schwimmt.
 ..., sondern unten, ...
 ... wurde es nicht getragen, ...

KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 ($n = 12$), TG2 = Treatmentgruppe 2 ($n = 14$); nicht abschnittskonforme Wörter / Formulierungen sind kursiv hervorgehoben. Die Daten beziehen sich ausschließlich auf den Post-Test.

Mit Ausnahme von *Also glaube ich* und dem *Wenn...-Vergleich* waren alle kodierten Sprachhandlungen abschnittskonform. Gegebenenfalls kann letztere Formulierung auch als abschnittskonform ausgelegt werden, da in diesem Fall die Strukturierung mit *wenn* einen erklärenden Vergleich aus der Lebenswelt nach sich zieht. Alle weiteren Sprachhandlungen im Abschnitt der Erklärung waren eindeutig erklärungsleitende Sprachhandlungen. Häufig wurde *weil* (9x), *deswegen* (4x) oder *wegen* (3x) verwendet. Längere Formulierungen, aus denen eine Erklärung folgt, lauteten: *Ich erkläre mir das so* (2x) oder *Meine Erklärung ist, dass*. Eine Vp schloss die Erklärung mit *Das ist meine eigene Erklärung* ab. Insgesamt ist eine vielseitige Nutzung bildungssprachlicher Formulierungen im Post-Test von TG2 zu verzeichnen, die die eigenen Erklärungsansätze sprachlich hervorheben.

Der Unterschied in der verwendeten Bildungssprache im Abschnitt der Erklärung im Post-Test von TG2 ist auch vor dem zeitlichen Hintergrund innerhalb der Gruppe interessant: **Im Prä-Test kamen in der Erklärung von TG2 insgesamt 13 bildungssprachliche Elemente vor, im Post-Test 34 und im Follow-Up-Test war ein Rückfall auf 16 zu verzeichnen** (vgl. Tab. 19).

Zusammengefasst war die Prüfung der Vorkommen von Wortspeichern und Bildungssprache in den einzelnen Abschnitten der Versuchsprotokolle aufschlussreich, denn das Ergebnis zeigt: Während Wortspeicher fast ausschließlich im Abschnitt der Vermutung oder der beschreibenden Durchführung verwendet wurden, wurde sonstige Bildungssprache abschnittsübergreifend – mit Ausnahme der Materialauflistung – eingesetzt. Weiterhin zeigt der qualitative Exkurs, dass die vermehrt eingesetzten bildungssprachlichen Formulierungen im Abschnitt der Erklärung im Post-Test von TG2 scheinbar bewusst als Hilfestellung für den Erklärungsansatz verwendet wurden.

Aufbauend auf diesen Erkenntnissen wird im zweiten Schritt die Überschneidung von Wortspeichern und Bildungssprache mit den inhaltlichen Beurteilungen innerhalb eines Abschnitts im Post-Test geprüft. Das Ergebnis ist in **Abb. 33 bis 35** dargestellt.

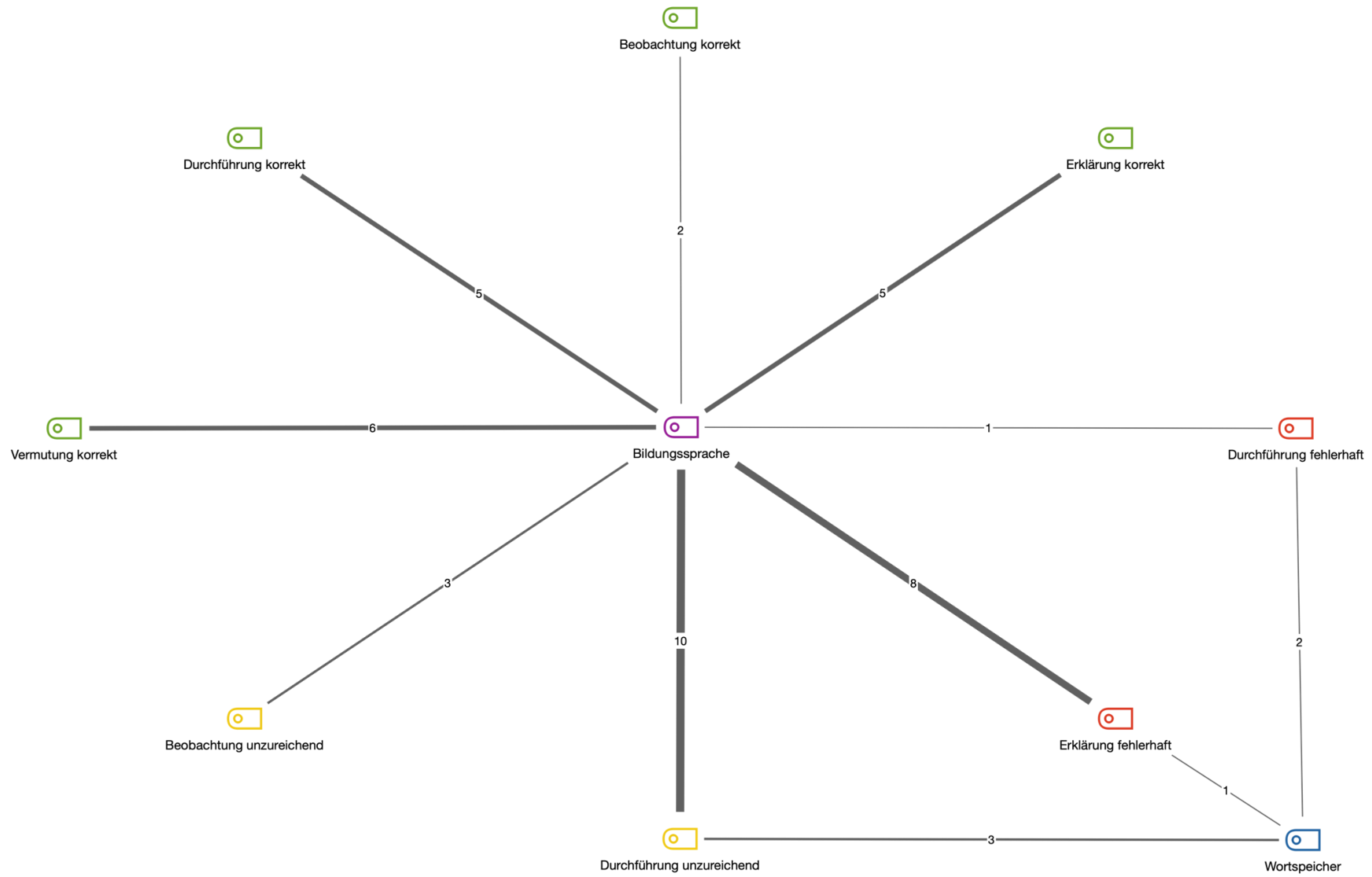


Abb. 33: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der KG ($n = 20$)

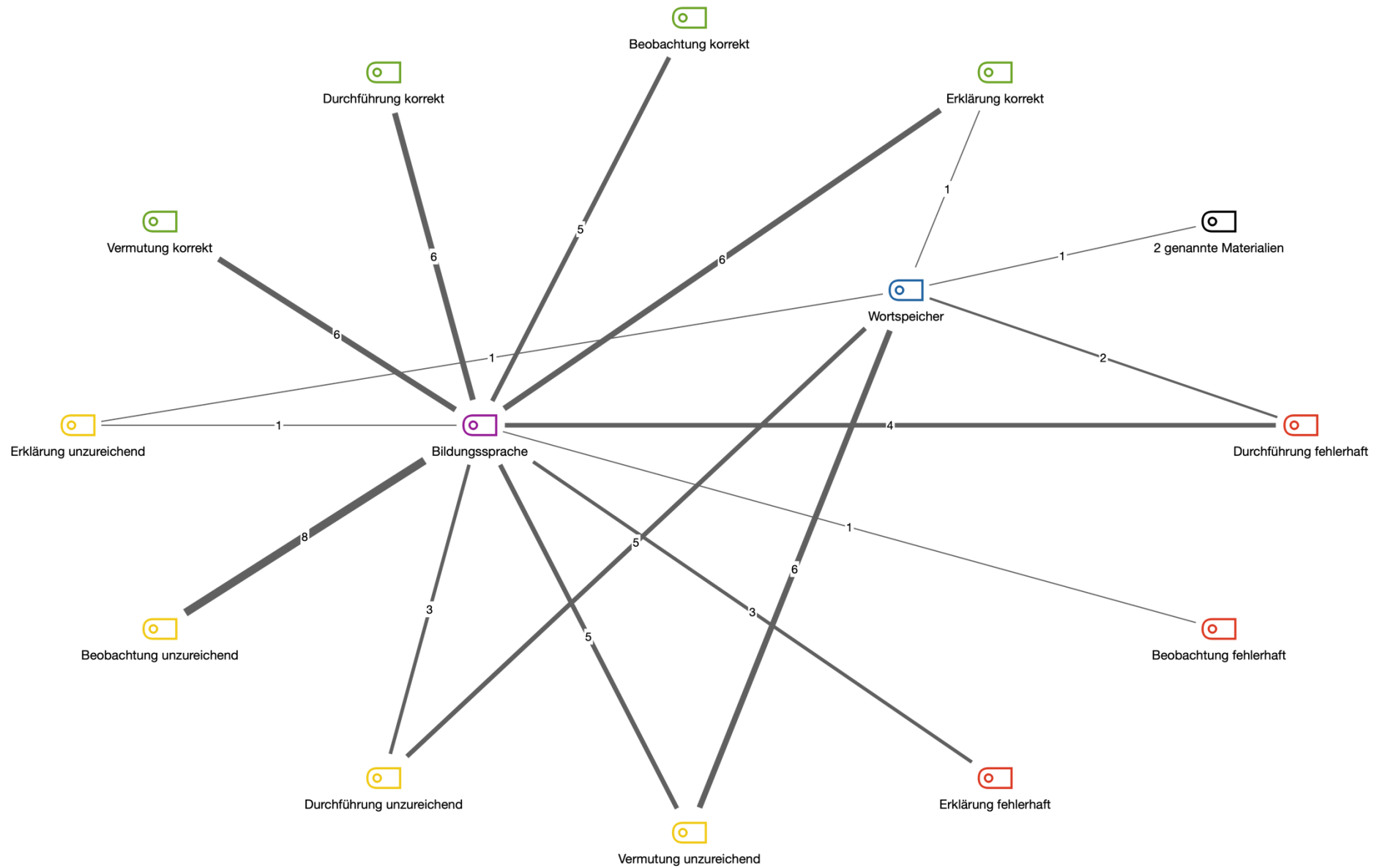


Abb. 34: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der TG1 (n = 12)

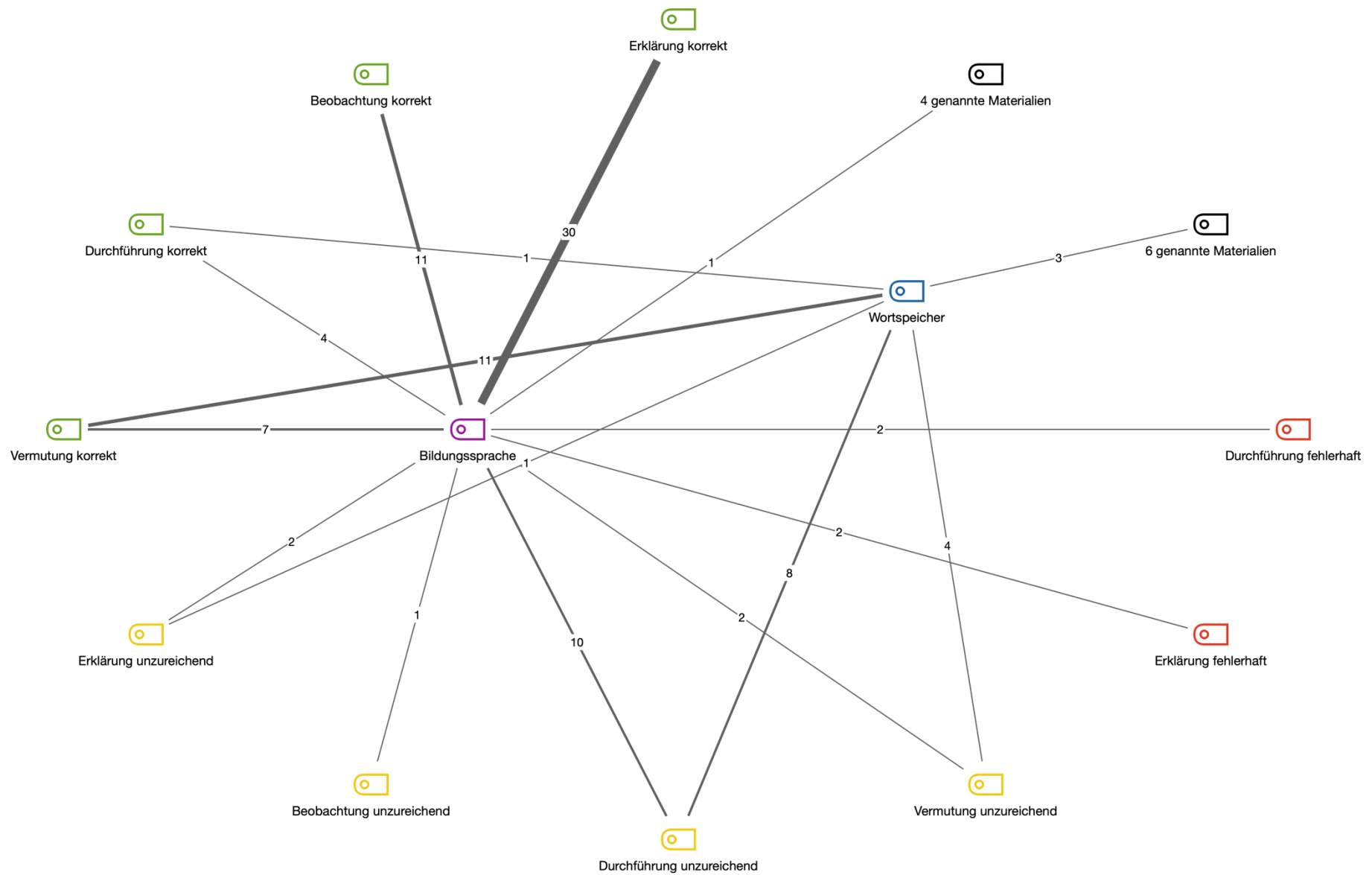


Abb. 35: Überschneidung von Wortspeicher und Bildungssprache mit inhaltlich beurteilten Segmenten im Post-Test der TG2 ($n = 14$)

Zunächst werden die Ergebnisse fokussiert auf die **Wortspeicher** dargestellt.

Abb. 33 zeigt die Überschneidungen der verwendeten Wortspeicher mit den inhaltlich beurteilten Abschnitten im Post-Test der KG. Die folgenden Zahlen in Klammern stehen jeweils für die Anzahl der Überschneidungen. Die 5 Wörter und Formulierungen aus dem Wortspeicher wurden in den Versuchsprotokollen der KG entweder im Zuge einer unzureichend beschriebenen Durchführung (3), einer fehlerhaft beschriebenen Durchführung (2) oder einer fehlerhaften Erklärung (1) verwendet.

In TG1 (**Abb. 34**) wurden die Wortspeicher hauptsächlich bei Vermutungen verwendet, die als unzureichend beurteilt wurden (6), gefolgt von unzureichender Durchführung (5), fehlerhafter Durchführung (2) und unzureichender Erklärung (2). Im Gegensatz zu den Ergebnissen der KG wurde der Wortspeicher in TG1 auch einmal im Zuge einer korrekten Erklärung und bei der Auflistung von zwei Materialien genutzt.

In TG2 (**Abb. 35**) kamen die Wortspeicher vor allem bei korrekt aufgeschriebenen Vermutungen (11) und unzureichender Durchführung (8) vor. Die hohe Anzahl an Durchführungen und Vermutungen deckt sich mit den Ergebnissen aus Tab. 17, wonach die Wortspeicher hauptsächlich in diesen beiden Kategorien verwendet wurden. Während die Vermutungen oftmals korrekt beurteilt sind, mangelt es bei den Durchführungen trotz Verwendung der Wortspeicher an Vollständigkeit und Präzision. Die oftmals korrekte Beurteilung von Vermutungen basiert darauf, dass es laut Kategoriensystem keine falschen Vermutungen gibt. Eine unzureichende Vermutung wurde bei fehlendem Aufgabenbezug und eine fehlerhafte Vermutung bei ausbleibender Aufstellung einer Vermutung kodiert.

Die drei Abb. 29 bis 31 werden im Folgenden unter dem Fokus der **Bildungssprache** beschrieben. Die Bildungssprache kam in der KG (**Abb. 33**) hauptsächlich im Rahmen einer unzureichend beschriebenen Durchführung (10x) oder einer fehlerhaften Erklärung (8x) vor. Erwähnenswert sind ebenso die Überschneidungen mit korrekter Vermutung (6x), korrekter Durchführung (5x) und korrekter Erklärung (5x).

In TG1 (**Abb. 34**) überschneidet sich die Bildungssprache am häufigsten mit unzureichenden Beobachtungen (8x), gefolgt von korrekter Vermutung (6x), korrekter Durchführung (6x) und korrekter Erklärung (6x).

Interessant ist TG2 (**Abb. 35**) in Bezug auf Bildungssprache. In dieser Gruppe kamen insgesamt 30 bildungssprachliche Elemente in Erklärungen vor, die als korrekt bewertet wurden. Zudem gab es in dieser Gruppe 11 Überschneidungen mit Bildungssprache und korrekten Beobachtungen. Die 30 Überschneidungen zwischen Bildungssprache und korrekter Erklärung traten in insgesamt 7 von 14 Versuchsprotokollen (50 %) auf. Da es sich um eine besondere Größenordnung handelt, wird der Anteil der Überschneidungen je Dokument in Tab. 21 qualitativ und im Detail dargestellt.

Tab. 21: Überschneidungen von Bildungssprache und korrekter Erklärung (qualitativ)

Vpn	Überschneidung Bildungssprache & korrekte Erklärung	Bildungssprache
Vp 1	wegen	Sprachhandlung
Vp 2	deswegen	Sprachhandlung
	..., wurde es nicht getragen, ...	Morphosyntax
	..., aber im anderen Becher war kein Salz.	Morphosyntax
	..., sondern es ist untergegangen.	Morphosyntax
Vp 3	Weil	Sprachhandlung
	deswegen	Sprachhandlung
	..., sondern es schwimmt.	Morphosyntax
Vp 4	Also glaube ich ...	Sprachhandlung
	Meine Erklärung ist, dass ...	Sprachhandlung
	weil	Sprachhandlung
	weil	Sprachhandlung
	wegen	Sprachhandlung
	Ich erkläre mir das so:	Sprachhandlung
	Weil wenn man im Meer badet, geht man, wenn man sich auf den Rücken legt, nicht unter.	Sprachhandlung
	wegen	Sprachhandlung
	deswegen	Sprachhandlung
	..., sondern unten, ...	Morphosyntax
Vp 5	weil	Sprachhandlung
	deswegen	Sprachhandlung
	Widerstand	Lexikon
	..., sondern schwimmt.	Morphosyntax
Vp 6	treiben lassen	Lexikon
Vp 7	Ich erkläre mir das so, dass ...	Sprachhandlung
	weil	Sprachhandlung
	weil	Sprachhandlung
	weil	Sprachhandlung
	Das ist meine eigene Erklärung.	Sprachhandlung
	hohl	Lexikon
	Das Ei geht auch nicht unter, wenn ich noch 100 ml Wasser dazu schützte, ...	Morphosyntax

Die Angaben beziehen sich auf den Post-Test von TG2 ($n = 14$)

Aus Tab. 21 wird deutlich, dass Vp 4 insgesamt 10 bildungssprachliche Elemente in einer korrekten Erklärung verwendete. Bei Vp 7 waren es insgesamt 7 bildungssprachliche Kodierungen. Bei dieser Anzahl besteht die Gefahr, dass die Versuchsprotokolle im übertriebenen Maße bildungssprachliche Elemente aufweisen. Die Erklärungen dieser zwei Vpn werden deshalb an dieser Stelle in ihrer Gesamtheit vorgestellt.

Vp 4:

Meine Erklärung ist, dass das Ei in Messbecher 2 oben schwimmt, weil das Salz es nach oben lässt. Weil das Ei ja auch nicht in Messbecher 1 oben schwimmt, sondern unten, also wegen dem Salz. Ich erkläre mir das so: Weil wenn man im Meer badet, geht man, wenn man sich auf den Rücken legt, nicht unter. Im Schwimmbad schon. Also glaube ich wegen dem Salz, deswegen ist das Ei oben.

Vp 7:

Ich erkläre mir das so, dass das Ei im Salzwasser das Ei nicht untergeht, weil das Salz das Ei nach oben drückt. Das Ei im Wasser geht unter, weil das Ei nicht hohl ist. Das Ei geht auch nicht unter, wenn ich noch 100 ml Wasser dazu schüttele, weil das Salz ja immer noch da ist. Das ist meine eigene Erklärung.

Die Verwendung der vielen bildungssprachlichen Elemente in diesen zwei beispielhaften Ausführungen wird in der Interpretation der vierten Teilstudie diskutiert.

Die vielen Kodierungen von erklärenden Sprachhandlungen im Post-Test der TG2 sind vor dem Hintergrund auffällig, dass Sprachhandlungen im Allgemeinen in dieser Studie wenig verwendet wurden. Dies wird in **Abb. 36** deutlich. Für das vorliegende Balkendiagramm wurden alle Versuchsprotokolle des *Panels* ($n = 138$) bezüglich ihrer bildungssprachlichen Kodierungen analysiert. Da die Sprachmittel aus dem Wortspeicher zur Bildungssprache zählen, wurden die Wortspeicher in die Analyse der Bildungssprache einbezogen. Das Ergebnis zeigt, dass die Sprachhandlungen mit 15 % von allen bildungssprachlichen Kodierungen am seltensten auftraten.

In ähnlicher Weise wurde der Inhalt aller Versuchsprotokolle des *Panels* ($n = 138$) untersucht. Das Ergebnis ist in **Abb. 37** dargestellt. Das Diagramm zeigt, dass die Vpn besonders häufig fehlerhafte Erklärungen schrieben (16 % aller inhaltlichen Kodierungen). Im Gegensatz dazu konnten nur wenige Vpn eine korrekte Erklärung formulieren (4 % aller inhaltlichen Kodierungen). Ebenfalls wenige Vpn formulierten eine korrekte Durchführung (3 % aller inhaltlichen Kodierungen). Im nächsten Kapitel wird der Unterschied zwischen den *Versuchsprotokollen mit Wortspeicher-Nutzung* und den *Versuchsprotokollen ohne Sprachmittel aus dem Wortspeicher* genauer betrachtet.

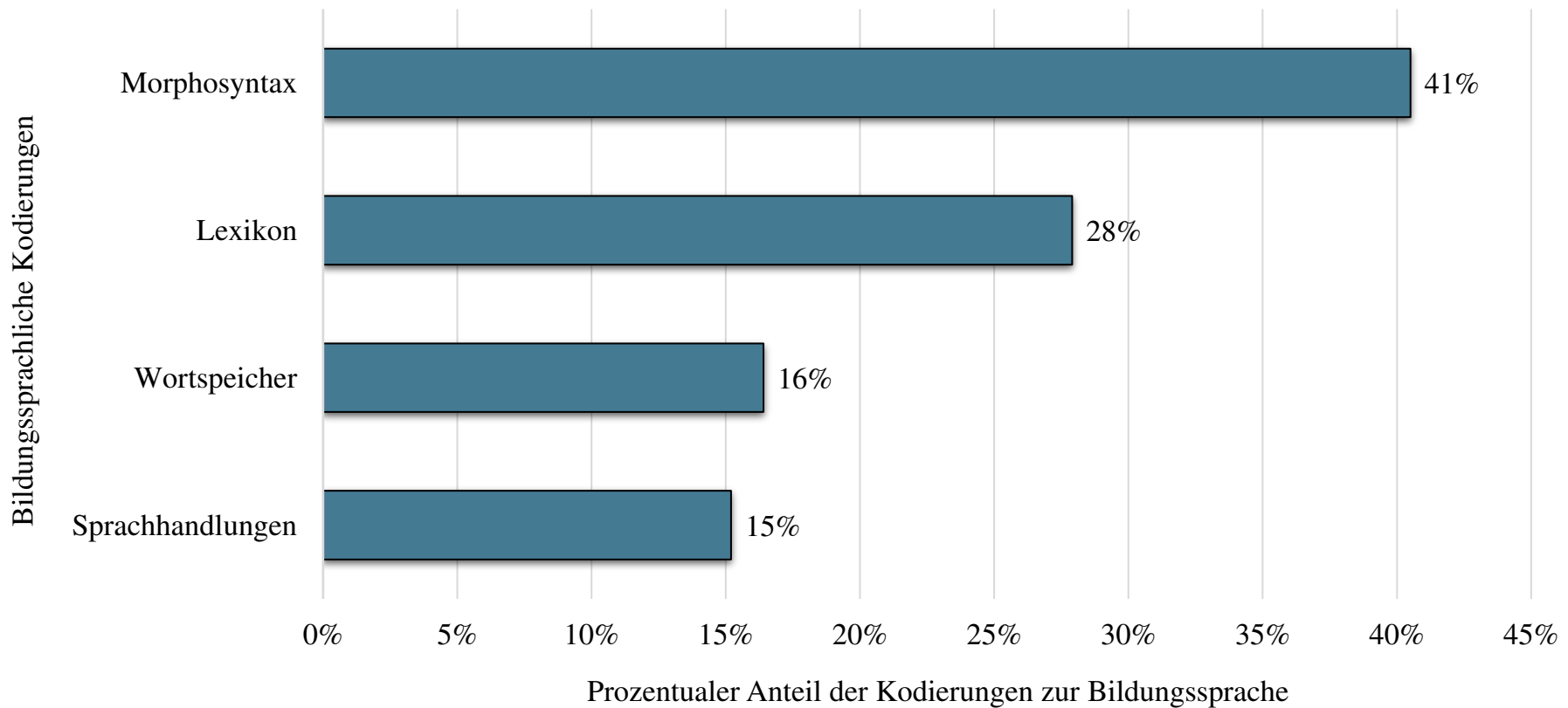


Abb. 36: Prozentualer Anteil aller Kodierungen der Bildungssprache bezogen auf *Sprachhandlungen*, *Lexikon*, *Morphosyntax* und *Wortspeicher*

Die Angaben beziehen sich auf die Anzahl aller Dokumente des *Panels* ($n = 138$). Durch Mehrfachkodierungen pro Dokument liegt die Gesamtzahl der kodierten Segmente bei 566 (hier = 100 %).

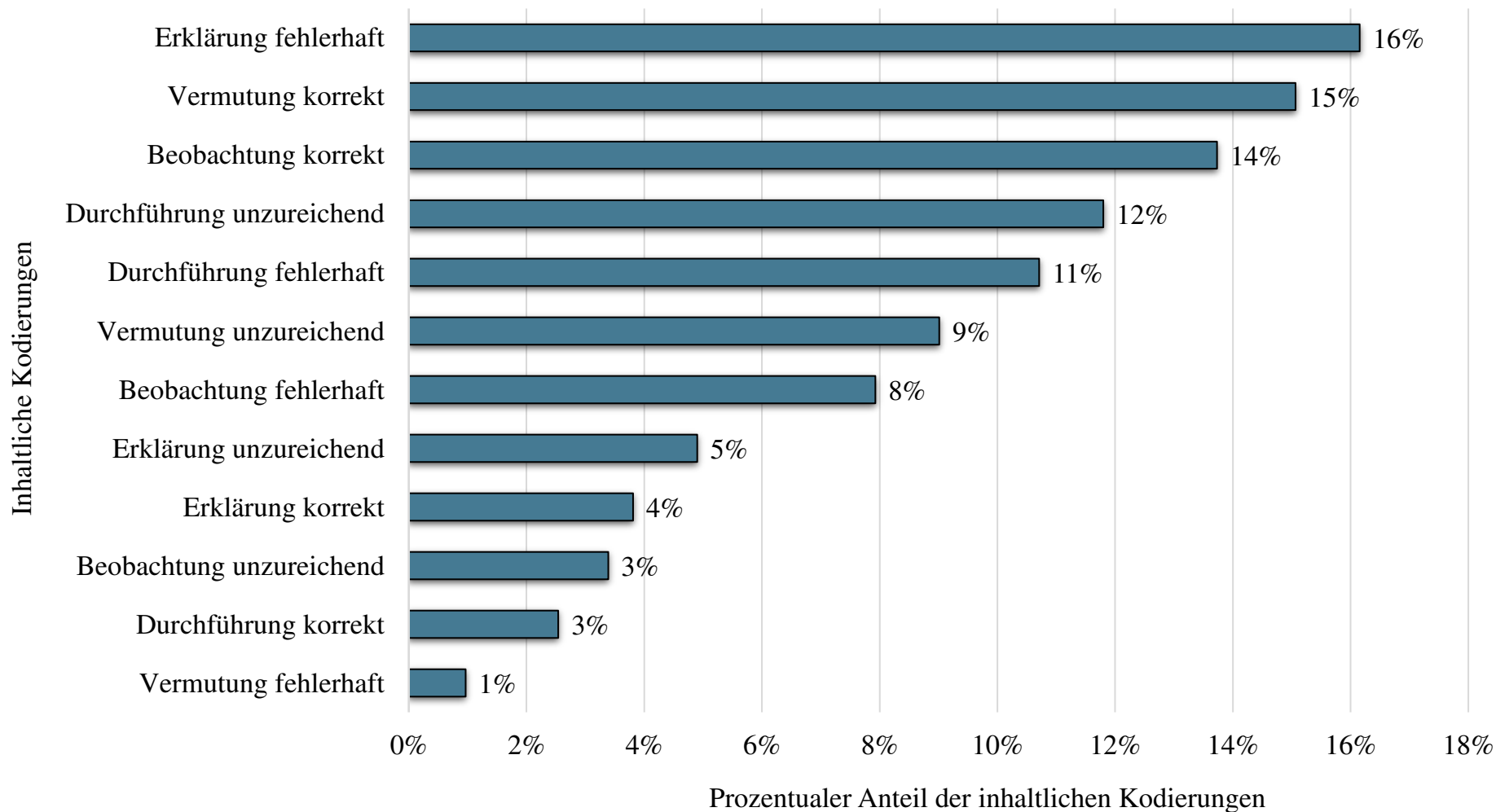


Abb. 37: Prozentualer Anteil aller inhaltlichen Kodierungen bezogen auf *korrekte*, *unzureichende* und *fehlerhafte* Abschnitte

Die Angaben beziehen sich auf die Anzahl aller Dokumente des *Panels* ($n = 138$). Durch Mehrfachkodierungen pro Dokument liegt die Gesamtzahl der kodierten Segmente bei 1.103. Aufgrund der dreifach vorkommenden Vermutungen und Beobachtungen pro Dokument, wurden die Rohdaten entsprechend bereinigt. Die Gesamtzahl der kodierten Segmente nach Bereinigung liegt bei 552 (hier = 100 %).

5.5.2 Vergleich von Vpn mit/ ohne Nutzung der Wortspeicher

In den Teilstudien 1 bis 3 wurde die Verwendung von Wortspeichern, einem bildungssprachlichen Register und den inhaltlichen Beurteilungen jeweils bezogen auf die KG, TG1 und TG2 und die jeweiligen MZP vorgestellt. In diesem Kapitel bezieht sich die Untersuchung nicht auf die Gruppen, sondern auf die tatsächliche Verwendung der Wortspeicher. Die Versuchsprotokolle mit Wortspeichern wurden mit den Versuchsprotokollen derer, die die Wortspeicher trotz Treatment nicht verwendeten, verglichen. Für eine aussagekräftige Analyse wurden nur die Vpn einbezogen, die zuvor ein Wortspeicher-Treatment erhielten und dementsprechend die Möglichkeit der Wortspeicher-Nutzung hatten ($n_{TG1, TG2, Post-Tests, Follow-Up-Tests} = 52$). Zufallsbedingt handelt es sich bei beiden Gruppen (Wortspeicher-Nutzung vs. Wortspeicher-Nicht-Nutzung) um exakt 26 Versuchsprotokolle, sodass die Kodierhäufigkeiten in absoluten Zahlen vergleichbar sind. Tab. 22 zeigt den Überblick über kodierte Wortspeicher, Bildungssprache und beurteilte Inhalte in den Versuchsprotokollen ohne Wortspeicher (Wortspeicher = 0) und den Versuchsprotokollen mit mindestens einem Wort oder einer Formulierung aus dem Wortspeicher (Wortspeicher > 0) in den Post- und Follow-Up-Tests der TG1 und TG2.

Tab. 22: Anzahl der Kodierungen von Wortspeicher, Bildungssprache und Inhalt (AV) in Abhängigkeit der tatsächlichen Wortspeicher-Nutzung

	Wortspeicher = 0	Wortspeicher > 0	Total
Wortspeicher			
Wortspeicher	0	65	65
Bildungssprache			
Sprachhandlungen	14	44	58
Lexikon	15	48	63
Morphosyntax	40	52	92
Bildungssprache insgesamt	69	144	213
Inhalt			
korrekt	72	101	173
unzureichend	71	71	142
fehlerhaft	65	36	101
Inhaltliche Gesamtpunktzahl	215	273	488
N = Dokumente	26	26	52

Mehrfachzählungen möglich; Ergebnisse aus Post- und Follow-Up-Tests von TG1 und TG2

In Anbetracht der Ergebnisse aus Tab. 22 lassen sich sowohl in der Bildungssprache als auch in den inhaltlichen Beurteilungen höhere Werte in der Gruppe ablesen, die die Wortspeicher in ihren Versuchsprotokollen anwandten. Im Bereich der Bildungssprache ist der höhere Wert auf alle drei Kategorien – Sprachhandlungen, Lexikon, Morphosyntax – zurückzuführen. Besonders bei Sprachhandlungen und Lexikon wurden in der Wortspeicher-Nutzungs-Gruppe 30 (Sprachhandlungen) bis 33 (Lexikon) Elemente mehr kodiert als in der Referenzgruppe. Insgesamt wurden in der Wortspeicher-Nutzungsgruppe 75 bildungssprachliche Elemente mehr kodiert als in den Versuchsprotokollen, in denen keine Wortspeicher vorkamen. Bezogen auf die inhaltliche Gesamtpunktzahl erzielte die Wortspeicher-Nutzungs-Gruppe mit 273 gegenüber 215 Punkten eine um 58 Punkte höhere Beurteilung.

Zur detaillierten Untersuchung der Beziehung zwischen Wortspeicher und Bildungssprache bot sich die Analyse der bildungssprachlichen Unterkategorien Sprachhandlungen, Lexikon und Morphosyntax im Zusammenhang mit der Verwendung von Wortspeichern an. Der *Code-Relations-Browser* (Tab. 23) zeigt u.a., welche Unterkategorie von Bildungssprache vermehrt mit der Verwendung der Wortspeicher im gleichen Dokument einhergeht. Daneben werden auch die übereinstimmenden Vorkommen der bildungssprachlichen Unterkategorien untereinander dargestellt.

Tab. 23: Gemeinsames Vorkommen von Wortspeichern und bildungssprachlichen Kategorien in Versuchsprotokollen mit & ohne Wortspeicher-Nutzung

Codesystem	Wortspeicher	Sprachhandlungen	Lexikon	Morphosyntax
Wortspeicher > 0 (n = 26)				
Wortspeicher	-	15	18	20
Sprachhandlungen	15	-	10	13
Lexikon	18	10	-	15
Morphosyntax	20	13	15	-
SUMME	53	38	43	48
Wortspeicher = 0 (n = 26)				
Wortspeicher	-	-	-	-
Sprachhandlungen	-	-	3	5
Lexikon	-	3	-	5
Morphosyntax	-	5	5	-
SUMME	-	8	8	10

Code-Relations-Browser für TG1, TG2, Post- und Follow-Up-Tests, Mehrfachkodierungen möglich

Laut den Ergebnissen aus Tab. 23 weisen die 26 Versuchsprotokolle, in denen Wortspeicher verwendet wurden, meist auch bildungssprachliche Morphosyntax (20 Versuchsprotokolle), bildungssprachliches Lexikon (18 Versuchsprotokolle) und / oder entsprechend kodierte Sprachhandlungen (15 Versuchsprotokolle) auf.

Der untere Teil von Tab. 23 bezieht sich auf die Versuchsprotokolle, in denen keine Sprachmittel aus dem Wortspeicher vorkamen. Der Vergleich der beiden Tabellen zeigt, dass das gemeinsame Vorkommen von verschiedenen bildungssprachlichen Elementen erhöht ist, wenn auch Wortspeicher zum Einsatz kamen. In den 26 Versuchsprotokollen mit Wortspeichern wurde beispielsweise in 15 Fällen sowohl bildungssprachliche Morphosyntax als auch bildungssprachliche Lexik im gleichen Dokument verwendet. Im Gegensatz dazu wiesen die Versuchsprotokolle ohne Wortspeicher weniger gemeinsame Vorkommen bildungssprachlicher Unterkategorien auf. Konkret gab es von den 26 Versuchsprotokollen ohne Wortspeicher nur 5 Versuchsprotokolle, in denen sowohl bildungssprachliche Morphosyntax als auch bildungssprachliche Lexik vorkamen.

In Abb. 38 ist der inhaltliche Vergleich zwischen der Gruppe ohne Verwendung der Wortspeicher (Wortspeicher = 0) und der Gruppe mit Verwendung der Wortspeicher (Wortspeicher > 0) mittels *Zwei-Fälle-Modell* dargestellt. Beide Gruppen bestehen zufallsbedingt aus genau 26 Dokumenten. Die Darstellung hebt hervor, dass nur in der Gruppe ohne Wortspeicher inhaltlich vier fehlerhafte Vermutungen vorkamen.

Eine genauere Analyse der als fehlerhaft kodierten Vermutungen – dargestellt in Tab. 24 – zeigt, dass es sich um insgesamt 3 Vpn handelt: Zwei Vpn aus TG1 und eine Vp aus TG2. Sie schrieben entweder keine Vermutung auf, gaben Ihre Unwissenheit preis (*Kein Plan*) oder drückten ihre Vermutung nicht konkret aus (*Funktioniert (nicht)*). In den Dokumenten, die auch Wortspeicher beinhalteten, waren keine fehlerhaft kodierten Vermutungen zu verzeichnen.

Tab. 24: Fehlerhafte Vermutungen in den Dokumenten ohne Wortspeicher

Dokumentgruppe	Segment
TG1 / Follow-Up-Test / Vpn 1	2.1) Kein Plan.
TG1 / Post-Test / Vpn 2	2.2) Funktioniert.
TG1 / Post-Test / Vpn 2	2.3) Funktioniert nicht.
TG2 / Follow-Up-Test / Vpn 3	2.3) -

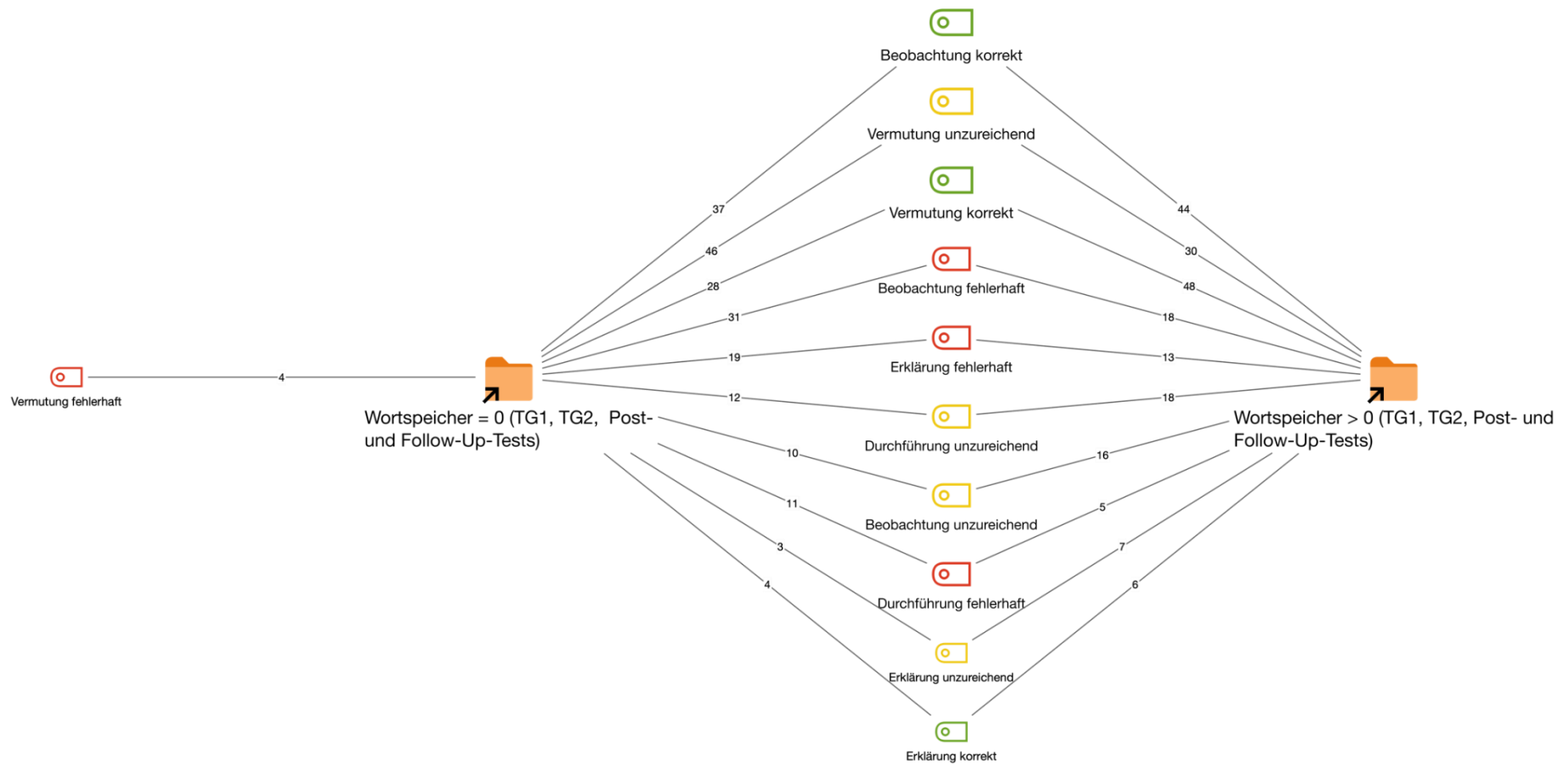


Abb. 38: Inhaltliche Beurteilung (AV) in Abhängigkeit der tatsächlichen Wortspeicher-Nutzung (Mehrfachkodierungen)

Abb. 38 zeigt weiterhin die gemeinsamen inhaltlichen Kodierungen beider Gruppen. Am häufigsten konnten insgesamt korrekte Beobachtungen ermittelt werden. Dagegen stehen korrekte Erklärungen in dieser Abbildung an letzter Stelle.

In die Abbildung nicht eingetragen sind korrekt beschriebene Durchführungen. Da in beiden Gruppen genau drei korrekte Durchführungen kodiert wurden, heben sich die Werte gegenseitig auf und sind in der Abbildung nicht zu sehen. Sie müssten noch unter den korrekten Erklärungen stehen.

Korrekte Durchführungen und korrekte Erklärungen zu formulieren, kam bei beiden Gruppen insgesamt am seltensten vor. Obwohl die Erklärungen häufig (insgesamt 32x in 52 Dokumenten) fehlerhaft waren, konnten die Vpn mit Wortspeicher-Nutzung in 6 Dokumenten korrekte Erklärungen formulieren. In der Gruppe ohne Wortspeicher-Nutzung waren es 4 Dokumente. Wie bei den korrekten Erklärungen waren auch bei den korrekten Beobachtungen und den korrekten Vermutungen mehr Kodierungen in der Gruppe zu verzeichnen, die mit Wortspeichern arbeiteten.

6 DISKUSSION

Abschließend werden die vorgestellten Ergebnisse unter Berücksichtigung der zu Beginn geschilderten Theorie interpretiert und die gefundenen Antworten auf die Forschungsfragen zusammengefasst. Das Kapitel *Methodische Kritik* weist auf Limitationen des Forschungsprojekts und mögliche methodische Alternativen hin. Der *Ausblick* rundet die Diskussion mit weiterführenden Forschungsmöglichkeiten und dem praktischen Nutzen des vorgestellten Forschungsprojekts ab.

6.1 Interpretation der Ergebnisse

Die Interpretation der Ergebnisse gliedert sich nach den vier Teilstudien. Der Fokus liegt auf den Ergebnissen der themenübergreifenden vierten Teilstudie (F4). Um die Interpretation von der Ergebnisdarstellung abzugrenzen, werden Vpn im Folgenden erneut als Kinder oder Schüler:innen bezeichnet.

6.1.1 Interpretation der Ergebnisse zu F1 – Nutzung der Wortspeicher

Es konnte gezeigt werden, dass ein achtstündiges Wortspeicher-Treatment in Plakat- und in auditiver Form in dieser Studie kurzfristig mit einer vermehrten Anwendung der gelernten Wörter und Formulierungen einhergeht. Konkret war die durchschnittliche Nutzung der Sprachmittel aus den Wortspeichern in TG2 und TG1 zusammengenommen höher als die durchschnittliche Nutzung der Sprachmittel in der KG (TG2 & TG1 > KG). Das Ergebnis korrespondiert mit der Erwartung, dass die Vermittlung von Wortspeichern zu mehr Vernetzungen im mentalen Lexikon führt. Je vielfältiger die Vernetzungen im mentalen Lexikon, desto höher ist die Chance der Aktivierung dieser Wörter und Formulierungen und somit zur Anwendung der Wortspeicher.³⁹ In diesem Fall dürften die Vernetzungen im mentalen Lexikon hauptsächlich auf semantischen, phonologischen (TG1) und zusätzlich graphemischen (TG2)

³⁹ vgl. Kapitel 2.3.2

Informationen beruhen. Darüber hinaus ordnet das Treatment die gelernten Wörter und Formulierungen in den kontextuellen Bezug des Versuchsprotokolls ein. Gerade der kontextuelle Bezug könnte für die Verwendung der Wortspeicher maßgeblich sein. Dies lässt sich daraus ableiten, dass in der KG die Wortspeicher kaum Verwendung fanden, obwohl die semantischen, phonologischen und teils graphemischen Informationen über die einzelnen Wörter und Formulierungen auch in der KG als Vorwissen vorausgesetzt werden können. In anderen Worten: Den Kindern ist in der dritten Klasse die Bedeutung, die Aussprache und teilweise auch die Schreibweise der Formulierung *Ich vermute, dass* bekannt. Dennoch verwenden sie diese Formulierung erheblich seltener, wenn sie nicht unmittelbar zuvor durch die Arbeit mit Wortspeichern aufgegriffen wurde. Es fehlt die kontextuelle Verknüpfung der Wörter und Formulierungen mit dem Schreiben eines Versuchsprotokolls.

Die Plakatform ist laut den inferenzstatistischen Ergebnissen eine besonders bedeutsame Vermittlungsform der Wortspeicher. Dies ist aus dem Gruppenunterschied zwischen KG, TG1 und TG2 direkt im Anschluss an die Unterrichtssequenz als auch aus dem gruppeninternen Vergleich abzuleiten. Die Analyse der Gruppenunterschiede im Post-Test zeigte eine signifikant höhere Verwendung der Wortspeicher in TG2 als in der KG. Der gruppeninterne Vergleich wies in TG2 zumindest auf einen signifikanten Unterschied zwischen Prä-, Post- und Follow-Up-Test hin, auch wenn dieser durch den Friedman-Test nicht bestätigt werden konnte. Der positive Effekt schriftlicher Sprachhilfen korrespondiert mit dem Ergebnis von Tajmel (2017b), da auch sie zeigen konnte, dass Schüler:innen der 7. Klasse die schriftlich vermittelten lexikalischen Hilfsmittel zur Beschreibung eines physikalischen Versuchs nutzten.⁴⁰ Zu ähnlichen Ergebnissen gelangt die Studie von Götze (2018), in der die schriftliche Handreichung von Sprachmitteln in der dritten Klasse einen unterstützenden Effekt bei der Formulierung mathematischer Erklärungen erzielte.⁴¹ Als neues Element der vorliegenden Studie ist die Unterscheidung zwischen auditiv und visuell dargebotenen Sprachmitteln anzusehen. Die hohe Wirksamkeit der Plakatform weist darauf hin, dass gerade die graphemischen Informationen im mentalen Lexikon die kurzfristige Behaltensleistung und den späteren Abruf der Wortspeicher begünstigen. Weiterhin konnte in dieser Studie in Abgrenzung zur Studie von Tajmel (2017b) gezeigt werden, dass Sprachmittel aus Wortspeichern auch dann im naturwissenschaftlichen Lernen verwendet werden, wenn sie **vor** Durchführung des Post-Tests vermittelt wurden und während des Post-Tests nicht zur Verfügung standen.

Die auditive Vermittlung der Sprachmittel ist zur Nutzung der Wortspeicher zwar hilfreich, anscheinend jedoch nicht ausreichend, um einen inferenzstatistisch

⁴⁰ vgl. Kapitel 2.3.5

⁴¹ vgl. Kapitel 2.3.5

belegbaren Unterschied hervorzurufen. Warum der Unterschied inferenzstatistisch nicht signifikant war, lässt sich anhand dieser Studie nicht eindeutig klären. Ein Erklärungsansatz ist die fehlende Sprachbewusstheit. In Anlehnung an das Modell von Bildungssprache in der Grundschule von Juska-Bacher und Beckert (2015) wird Sprachreflexion als wichtige Kompetenz dargestellt.⁴² In den neun Leitlinien zur Wortschatzarbeit nach Leisen (2010) soll Wortschatzarbeit zur Förderung von Sprachbewusstheit beitragen.⁴³ In dem empirisch hergeleiteten Modell zum Konzept *Scaffolding* nach Hammond und Gibbons (2005) wird das metalinguistische und metakognitive Bewusstsein hervorgehoben. Laut dem Modell sollten Lehrkräfte die Aufmerksamkeit der Kinder auf sprachbildende Maßnahmen lenken, um somit ein metalinguistisches und metakognitives Bewusstsein zu fördern.⁴⁴ Die im Theorieteil wiederkehrend beschriebene Relevanz der Sprachreflexion könnte in dieser Studie einen Einfluss gehabt haben. Durch die rein auditive Vermittlung war den Kindern die Bedeutung der dargebotenen Sprachmittel für den Einsatz in verschiedenen Versuchsprotokollen mutmaßlich unzureichend bewusst. Im Gegensatz zur TG1 hatten die Kinder in TG2 die Wortspeicher während der Unterrichtssequenz stets an der Tafel präsent und konnten somit ein Bewusstsein dafür entwickeln, dass die Sprachmittel in thematisch unterschiedlichen Versuchsprotokollen zum Einsatz kommen können. In TG1 fehlte die kontinuierliche Präsenz der Sprachmittel, sodass dieses Verständnis auf Meta-Ebene vermutlich ausblieb.

Insgesamt wurden die aufgestellten Hypothesen inferenzstatistisch abgelehnt. Jedoch zeigen die Mittelwerte der Wortspeicher-Nutzung eine Tendenz, die den Hypothesen entspricht, d.h. die Kinder aus TG2 nutzten nach dem Treatment die meisten Sprachmittel aus dem Wortspeicher, gefolgt von TG1 und der KG. Die Tendenz (TG2 > TG1 > KG) ist sowohl direkt nach der Unterrichtssequenz als auch zehn Wochen später erkennbar, wie das Säulendiagramm zu den jeweiligen Mittelwerten (Abb. 23) verdeutlicht. Sie ist mit den gleichen Argumenten zu erklären, die zur Aufstellung der Hypothesen dienten. Im Prinzip wird deutlich, dass nicht nur die Abwechslung von Sprachmodalitäten bzw. die Abwechslung der Darstellungsformen einen Nutzen für Sprachbildung im Sachunterricht darstellen, wie es in Kapitel 2.3.3 und 2.3.4 beschrieben ist. Stattdessen sollte vor allem auf eine Sprachmodalität bzw. Darstellungsform zurückgegriffen werden, die mehrere Sinne anspricht. Da die Plakatform in dieser Studie auch inferenzstatistisch als besonders vorteilhafte Vermittlungsform bestätigt werden konnte und die Mittelwerte tendenziell den Hypothesen entsprechen, sollte an diesen in größer angelegten Studien – trotz Ablehnung in dieser Studie – festgehalten werden.

⁴² vgl. Kapitel 2.1.3

⁴³ vgl. Kapitel 2.3.3

⁴⁴ vgl. Kapitel 2.3.4

6.1.2 Interpretation der Ergebnisse zu F2 – Sprachregister

Durch Wortspeicher in Plakatform lässt sich die Verwendung von Bildungssprache erhöhen. Im Anschluss an die Unterrichtssequenz nutzten die Kinder, die zuvor die Wortspeicher in Plakatform vermittelt bekamen, signifikant mehr bildungssprachliche Wörter und Formulierungen als die Kinder ohne Zugang zu Wortspeichern. Dieses Ergebnis deckt sich mit dem Ergebnis der Studie zu Sprachforscherkursen in einer Gesamtschule.⁴⁵ In den additiven Sprachforscherkursen führten Sprachmuster und Worterklärungen zu fachsprachlicheren Versuchsprotokollen (Agel et al., 2012, S. 41). Die vorliegende Studie zu Wortspeichern zeigt ein ähnliches Ergebnis für eine kürzere, sprachbildende Unterrichtssequenz. Dies ist insofern von Bedeutung, da ein additiver Unterricht aus personellen, finanziellen und organisatorischen Gründen oftmals nicht möglich ist. Die inferenzstatistische Analyse in der zweiten Teilstudie deutet darauf hin, dass Wortspeicher in Plakatform auch unterrichtsintegriert einen Nutzen aufweisen, so dass Sprachbildung im alltäglichen Unterricht an Bedeutung gewinnt.

Ebenfalls interessant ist das Ergebnis im zeitlichen Verlauf: Die Verwendung von Bildungssprache ist in der Wortspeicher-Plakat-Gruppe unmittelbar nach der Unterrichtssequenz höher als zehn Wochen später. Die Kinder in der Wortspeicher-Plakat-Gruppe nutzten im Post-Test signifikant mehr bildungssprachliche Wörter und Formulierungen als im Follow-Up-Test. Die Inhalte zu den drei Messzeitpunkten sind jeweils verschieden. Insofern kann die jeweilige thematische Fragestellung zum Einsatz von mehr oder weniger bildungssprachlichen Wörtern und Formulierungen animieren. Eine Besonderheit des vorliegenden Ergebnisses beruht auf dem Gruppenvergleich der zeitlichen Verläufe: Im Gegensatz zur Wortspeicher-Plakatgruppe wurde in den anderen beiden Gruppen mehr Bildungssprache im Follow-Up-Test als im Post-Test angewandt (vgl. Abb. 28 und 29). Die unterschiedliche Anzahl an bildungssprachlichen Mitteln im Post- und Follow-Up-Test von TG2 kann deshalb nicht auf die thematische Fragestellung zurückgeführt werden. Stattdessen ist zu vermuten, dass die Kinder im Post-Test aufgrund des visuellen Wortspeicher-Treatments ein verstärktes Sprachbewusstsein entwickelten und Bildungssprache vermehrt einsetzten. Nach zehn Wochen regulärem Unterricht scheint dieses Sprachbewusstsein weniger ausgeprägt – die Kinder verwendeten weniger bildungssprachliche Mittel in ihren Versuchsprotokollen. Warum die vermutete Sprachbewusstheit nicht nach einem auditiven Wortspeicher-Treatment verstärkt wird, lässt sich nur mit der höheren Intensität eines visuellen Wortspeicher-Treatments begründen.

⁴⁵ vgl. Kapitel 2.3.5

6.1.3 Interpretation der Ergebnisse zu F3 – Inhalt

Die Angabe der Materialien ist nach der Unterrichtssequenz in allen Gruppen ergebiger als vorher. Im Abschnitt *Materialien* nannte jedes Kind nach der Unterrichtssequenz (Post-Test) mindestens zwei bis drei Materialien. Vor der Unterrichtssequenz gab es insgesamt fünf Kinder, die keine Materialien nannten. Das Ergebnis kann durch den Übungseffekt der Unterrichtssequenz interpretiert werden. Im Prä-Test waren die Kinder mit dem Schreiben eines Versuchsprotokolls nicht vertraut. Vermutlich führte dies dazu, dass die fünf Kinder keine Materialien aufschrieben. Ein weiterer möglicher Faktor könnte der erhöhte Zeitaufwand bei der Bearbeitung der unbekannteren Textsorte sein. Nach der Unterrichtssequenz konnte jedes Kind mindestens zwei Materialien nennen. Anscheinend führte die Unterrichtssequenz zu einem Wissenszuwachs darüber, wie und welche Materialien aufgelistet werden und/ oder wie das Versuchsprotokoll innerhalb der zeitlichen Vorgabe auszufüllen ist. Da deskriptiv kein auffälliger Unterschied in den Häufigkeiten der genannten Materialien zwischen den Gruppen festzustellen war, ist für die Rubrik *Materialien* ein inhaltlicher Nutzen durch die Arbeit mit Wortspeichern auszuschließen.

Die inhaltliche Beurteilung der Versuchsprotokolle ist in allen Gruppen und zu allen Zeitpunkten vergleichbar. Mithilfe des Beurteilungssystems und inferenzstatistischer Prüfung konnte in Bezug auf den Inhalt zu keinem der Messzeitpunkte ein Unterschied zwischen den Versuchsprotokollen der Kinder aus der Wortspeicher-Plakat-Gruppe, der auditiven Wortspeicher-Gruppe und aus der KG festgestellt werden. Im Gegensatz dazu weist die in Kapitel 2.3.5 vorgestellte Studie von Tajmel (2017b, S. 346) auf eine höhere fachliche Qualität durch die Verwendung lexikalischer Hilfsmittel hin.

Der Grund dafür, dass das Ergebnis nicht dem Ergebnis von Tajmel (2017b) entspricht, könnte mit dem unterschiedlichen Forschungsdesign zusammenhängen. Im Gegensatz zur Studie, die sich auf physikalische Versuche in der 7. Klasse bezog, wurden in der vorliegenden Studie Kinder der 3. Klasse im Sachunterricht untersucht. Zweitens wurden die Sprachmittel in der vorliegenden Studie vor dem Post-Test vermittelt und nicht zeitgleich beim Schreibprozess zur Verfügung gestellt. Dies kann das Ergebnis insofern modifizieren, da die Sprachmittel aus den vorherigen Stunden ggf. im Post-Test der vorliegenden Studie nicht mehr präsent waren und entsprechend nicht genutzt wurden. Dementsprechend hatte das Treatment keinen Nutzen bezüglich der inhaltlichen Qualität der Versuchsprotokolle. An dieser Stelle ist der inhaltliche Vergleich von Versuchsprotokollen mit und ohne Nutzung der Wortspeicher (Kapitel 5.5.2) aufschlussreich. Die Erkenntnisse aus diesem Vergleich werden im nächsten Kapitel *Interpretation der Ergebnisse zu F4 – Datenintegration* diskutiert. Ein letzter wichtiger Unterschied zwischen dem Forschungsdesign von Tajmel (2017b) und der vorliegenden

Studie ist die inhaltliche Bewertung. In der explorativen Studie schrieben die Schüler:innen zweimal den gleichen Text: Einmal ohne und einmal mit lexikalischen Hilfsmitteln. Die inhaltliche Beurteilung bestand in einem Vergleich beider Texte unter Zuhilfenahme der Bewertungskriterien *Genauigkeit*, *Fachsprachlichkeit* und *Verständlichkeit* (Tajmel, 2017b, S. 331). Die Bewertungskriterien zeigen, dass Sprache in der Explorationsstudie eine zentrale Funktion bei der Bewertung von Fachlichkeit einnahm. Dies ist im vorliegenden Beurteilungssystem weniger der Fall. Die unterschiedliche Art der Beurteilung könnte die Divergenz der Ergebnisse beider Studien erklären.

Aus einer anderen Perspektive zeigt die vergleichbare inhaltliche Bewertung in allen Gruppen, dass die Wortspeicher-Treatmentgruppen mit unterrichtsintegriertem Fokus auf Sprache inhaltlich keine schlechteren Versuchsprotokolle schreiben als die Gruppe ohne Treatment. Dies wäre aufgrund der etwas verkürzten Lernzeit ebenso denkbar gewesen. Das Ergebnis dieser Teilstudie verdeutlicht somit die Vereinbarkeit von Sprachbildung und inhaltlichen Lernzielen. Das im Schulkontext kursierende Argument *Sprachbildung nimmt zu viel Zeit in Anspruch* wird durch die Studie entkräftigt.

10 Wochen nach der Unterrichtssequenz sind die Versuchsprotokolle von Kindern aus der Wortspeicher-Plakat-Gruppe und den Kindern, die keine Wortspeicher erhielten, inhaltlich schwächer als zuvor. Dies zeigt die inferenzstatistische Untersuchung im zeitlichen Verlauf. In beiden Gruppen war die inhaltlich erreichte Punktzahl 10 Wochen nach der Unterrichtssequenz (im Follow-Up-Test) besonders niedrig gegenüber der Punktzahl im Post-Test. In der KG hob sich die niedrige Punktzahl im Follow-Up-Test zusätzlich stark von der Punktzahl im Prä-Test ab. Dieses Ergebnis lässt sich unter Berücksichtigung von Abb. 32 erklären. Die Abbildung zeigt den zeitlichen Verlauf bezogen auf alle Gruppen. Es ist klar zu erkennen, dass alle Gruppen im Follow-Up-Test weniger inhaltliche Punkte erzielten. Da die Versuchsprotokolle zu den einzelnen Messzeitpunkten thematisch verschieden sind, könnte der dritte MZP einen besonders schwierigen Versuch darstellen. Im Follow-Up-Test bestand die Aufgabe darin, ein Salatblatt a) in Wasser, b) in Salzwasser und c) aus dem Salzwasser wieder in Wasser zu legen und jeweils zu ertasten, wie sich das Salatblatt anfühlt. Der Versuch hob sich von den anderen Versuchen dadurch ab, dass die Beobachtung auf einer taktilen Wahrnehmung beruhte. Dies legt die Vermutung nahe, dass die Kinder gruppenübergreifend mehr Schwierigkeiten zum Schreiben des dazugehörigen Versuchsprotokolls hatten. Ggf. war das Aufstellen einer *Vermutung* weniger greifbar, da sich die Kinder die fühlbaren Eigenschaften und die damit verbundenen möglichen Adjektive zur Beschreibung eines Salatblattes nicht im Vorhinein vorstellen konnten. Auch die *Beobachtung* und die *Erklärung* könnten aufgrund des involvierten taktilen Sinnes erschwert gewesen sein. Da sich die Schwierigkeiten auf alle drei Gruppen zum dritten Messzeitpunkt bezogen, ist der inferenzstatistisch nachgewiesene Unterschied im zeitlichen Verlauf von KG und TG2 weniger aussagekräftig bezüglich des untersuchten Nutzens von Wortspeichern.

6.1.4 Interpretation der Ergebnisse zu F4 – Datenintegration

Ein interessantes Ergebnis aus der datenintegrierten Analyse der Versuchsprotokolle ist die **Tatsache, dass die Kinder aus der Wortspeicher-Plakatgruppe direkt im Anschluss an das Treatment kaum eine Formulierung aus dem Wortspeicher zu dem Abschnitt *Erklärung* verwenden, dafür aber ihre Erklärung mit zahlreichen bildungssprachlichen Formulierungen unterstützen (vgl. Tab. 17, 19)**. Die Verwendung sprachlich erklärender Ansätze war in keiner anderen Gruppe so hoch. Vermutlich war den Kindern aus der Wortspeicher-Plakat-Klasse durch das Wortspeicher-Treatment bewusst, dass sie Erklärungen sprachlich einleiten sollten.

Warum sie bei ihren Erklärungen nicht auf die Wortspeicher zurückgreifen, kann mehrere Gründe haben. Erstens könnten die Formulierungen aus dem Wortspeicher zur Einleitung der Erklärung nicht zu dem jeweils eigenen Sprachstil der Kinder passen (vgl. *sozialsymbolische Funktion* von Bildungssprache).⁴⁶ Kinder können im Unterricht eine geforderte sprachliche Identität annehmen oder ablehnen. In diesem Fall könnte es sein, dass die gehobenen Formulierungen zur Einleitung der Erklärung nicht die Identität der Kinder abbilden und insofern nicht verwendet werden.

Ein weiterer Grund, warum die Wortspeicher in der Kategorie des Erklärens nicht genutzt werden, könnte die fehlende Erinnerungsfähigkeit an die Formulierungen sein. Gerade die Länge der Formulierungen und der mutmaßlich ungenügende Bekanntheitsgrad erhöhen den Schwierigkeitsgrad der Einprägung und des späteren Abrufs. Die Wörter und Formulierungen aus dem Wortspeicher, die von den Kindern tatsächlich verwendet wurden, sind meist kurz (z.B. *danach*) oder der Bekanntheitsgrad dürfte höher sein (z.B. *Ich vermute, dass ...*) als es bei den Formulierungen zum Abschnitt der Erklärung der Fall ist (z.B. *Die Beobachtung lässt sich wie folgt erklären: ...*).

Ein dritter Grund ist in der Komplexität des Erklärens zu sehen. Laut TIMS-Studie konnten nur 8 % der deutschen Kinder im Jahr 2015 die Kompetenzstufe V erreichen, die damit definiert ist, dass Experimente interpretiert und Schlussfolgerungen gezogen werden können (Steffensky et al., 2016, S. 168). Im Jahr 2019 lag die Zahl bei 7 % (Steffensky et al., 2020, S. 145). Die vorliegende Studie bestätigt durch die Analyse inhaltlicher Kodierungen, dass die Erklärung den Kindern besonders schwerfällt (vgl. Abb. 37). In Bezugnahme auf Rincke (2010, S. 256) könnte ein höherer Schwierigkeitsgrad in den Aufgabenstellungen mit einem Rückfall zur Alltagssprache einhergehen.⁴⁷ Dies würde zwar erklären, warum die Kinder in der schwierigeren Erklärung nicht die Wortspeicher nutzten. Auf der anderen Seite erklärt diese Interpretation nicht, warum sie sich stattdessen vermehrt bildungssprachlicher Formulierungen zur Einleitung der

⁴⁶ vgl. Kapitel 2.1.3

⁴⁷ vgl. Kapitel 2.2.3

Erklärung bedienten. Somit scheint eher plausibel, dass den Kindern aus der Wortspeicher-Plakat-Gruppe die Erwartung an eine sprachlich markierte Erklärung bewusst war. Da die konkreten Formulierungen aus dem Wortspeicher nicht behalten wurden, entschieden sie sich mutmaßlich stattdessen zur anderweitigen sprachlichen Kennzeichnung ihrer Erklärungen.

Im Zusammenhang mit dem erhöhten Vorkommen bildungssprachlicher Formulierungen im Abschnitt der *Erklärung* im Post-Test der Wortspeicher-Plakat-Gruppe ist die Betrachtung der Überschneidungen von Wortspeicher, Bildungssprache und Inhalt interessant (vgl. Abb. 35). Sie verdeutlicht, dass **ein Großteil der bildungssprachlichen Formulierungen im Zusammenhang mit korrekten Erklärungen verwendet werden**. Von den insgesamt 34 verwendeten bildungssprachlichen Formulierungen im Abschnitt der *Erklärung* werden 30 bildungssprachliche Formulierungen in *korrekten Erklärungen*, zwei in *unzureichenden Erklärungen* und zwei in *fehlerhaften Erklärungen* eingesetzt. Das Ergebnis könnte darauf zurückzuführen sein, dass sich die Kinder durch die Befassung mit Sprache zusätzlich intensiver mit dem Versuch auseinandersetzen. Im Kapitel *Thematische und methodische Begründungen zur Unterrichtssequenz* wurde erwähnt, dass die zeitliche Drosselung der Sprachverarbeitung mit einer vertieften kognitiven Auseinandersetzung des Versuches einhergeht. In ähnlicher Weise könnte der Fokus auf die sprachliche Einleitung der *Erklärung* zu einer zeitlichen Drosselung und somit vertieften Auseinandersetzung mit dem Versuch und der jeweiligen Erklärungsmöglichkeiten geführt haben.

Ein weiteres Diskussionsthema ist die **Vielseitigkeit der verwendeten Wortspeicher und die Vielseitigkeit von Bildungssprache**. Die qualitative Analyse der Versuchsprotokolle zeigte, dass u.a. das Wort *danach* aus dem Wortspeicher in der Wortspeicher-Plakat-Gruppe insgesamt zahlreich verwendet wurde (vgl. Tab. 18). Die Verwendung erstreckte sich auf nahezu alle Bereiche des Versuchsprotokolls. Ebenfalls umfangreich verwendet wurde das Wort *weil* (vgl. Tab. 20). Laut der in Kapitel 2.1.2 geschilderten Definition von Bildungssprache, zeichnet sich diese u.a. durch einen großen Wortschatz aus. Da Bildungssprache auf konzeptioneller Schriftlichkeit beruht, kann als weiteres Merkmal von Bildungssprache der differenzierte Wortschatz genannt werden.⁴⁸ Die häufige Verwendung der Wörter *danach* oder *weil* steht dieser Definition scheinbar entgegen.

Allerdings bezieht sich die Analyse auf eine Vielzahl von Versuchsprotokollen, sodass die Wortwahl innerhalb einzelner Versuchsprotokolle nicht zwangsläufig durch wiederholende Bildungssprache ausgezeichnet sein muss. Hinzu kommt, dass in der deskriptiven Häufigkeitsanalyse ein Nutzen von Wortspeichern auch bei Einfachkodierungen bezüglich kodierter Sprachhandlungen erkennbar war (siehe Anhang D.1). Da

⁴⁸ vgl. Kapitel 2.1.1, Tab. 1

Einfachkodierungen redundante Kodierungen innerhalb eines Versuchsprotokolls ausschließen, untermauern die Werte eine gewisse Variabilität in der Bildungssprache – zumindest bezogen auf die vermehrten Sprachhandlungen in der Wortspeicher-Plakat-Gruppe nach der Unterrichtssequenz.

Näheres über die Vielseitigkeit der verwendeten Wortspeicher oder zur Vielseitigkeit der eingesetzten Bildungssprache ist durch den qualitativen Blick einzelner Versuchsprotokolle erfahrbar. Bei einem Erklärungsversuch eines Kindes aus dem Post-Test von TG2, das in Kapitel 5.5.1 in Gänze nachzulesen ist, kommt das Wort *weil* dreimal im Abschnitt vor. Dies könnte als Zeichen der Redundanz gewertet werden. Im gleichen Erklärungsversuch wird mit Formulierungen, wie *Ich erkläre mir das so, dass ...* und *Das ist meine eigene Erklärung* inhaltlich die gleiche Aussage getroffen: Es geht um eine Erklärung. Diese Art der sprachlichen Wiederholung ist grundsätzlich nicht mit dem präzisen Wortgebrauch der Bildungssprache vereinbar. Außerdem würde ein Versuchsprotokoll im naturwissenschaftlichen Berufsfeld ohne die genannten Formulierungen auskommen. Auf der anderen Seite können die gewählten Wörter und Formulierungen im Sinne einer Schulsprache didaktisch zu einer sprachlich angemessenen Erklärung hinführen – ähnlich wie im Rahmen der Schulsprache die didaktische Gattung der Erörterung zum Erwerb des bildungssprachlichen Erörterns hinführt.⁴⁹ Inwieweit die vielen bildungssprachlichen Elemente die Erklärung positiv oder negativ beeinflussen, ist diskussionswürdig. Es bleibt festzuhalten, dass die Kinder sichtlich bemüht ihre Erklärung als solche sprachlich markieren. Dies kann einen ersten Schritt zur erwünschten bildungssprachlichen Formulierung von Erklärungen im Allgemeinen darstellen.

Das stärkere Bewusstsein zur Verwendung angemessener Sprache deckt sich mit dem Ergebnis der empirischen Studie zu Sprachforscherkursen.⁵⁰ In dieser Studie wurde die Verwendung unangemessener fachsprachlicher Verben anstelle von umgangssprachlichen Verben wie folgt interpretiert: „Dies legt nahe, dass sie [die Interventionsschüler:innen] ein gewisses Bewusstsein dafür entwickelt haben, dass im Protokoll ein bestimmtes, von der Alltagssprache unterschiedliches sprachliches Register von ihnen verlangt wird und sie bemüht sind, diesen Anforderungen gerecht zu werden“ (Agel et al., 2012, S. 42). Das Ergebnis der vorliegenden Studie wird ähnlich interpretiert und unterscheidet sich von den Sprachforscherkursen insofern, dass sich die Interpretation nicht auf die Verwendung unangemessener fachsprachlicher Verben, sondern auf die Verwendung korrekter bildungssprachlich markierter Wörter, Formulierungen und Satzstrukturen bezieht.

Hinsichtlich des Follow-Up-Tests ist festzuhalten, dass **die Arbeit mit Wortspeichern nach zehn Wochen anscheinend keinen Nutzen (mehr) aufweist – dies gilt sowohl**

⁴⁹ vgl. Kapitel 2.1.2

⁵⁰ vgl. Kapitel 2.3.5

für die Verwendung der Wortspeicher als auch für die Anwendung sonstiger Bildungssprache und die inhaltliche Beurteilung der Versuchsprotokolle. Das Ergebnis zeigt, dass das Treatment von vier Doppelstunden in diesem Fall keinen (inferenzstatistisch messbaren) langfristigen Unterschied zwischen den Gruppen bezogen auf die drei ersten Fragestellungen hervorruft. Eine weiterführende Diskussion dieses Ergebnisses bezieht sich auf konkrete Vorschläge im Schulkontext und ist daher im *schulrelevanten Ausblick* (Kapitel 5.4.2) zu lesen.

Der Vergleich zwischen Versuchsprotokollen, in denen Wortspeicher vorkommen und Versuchsprotokollen ohne Verwendung der Wortspeicher ist insofern interessant, da **die tatsächliche Nutzung der Wortspeicher offenbar mit einer höheren inhaltlichen Punktzahl einhergeht.** Die Versuchsprotokolle aus den Treatmentgruppen, die direkt nach der Unterrichtssequenz oder 10 Wochen später mindestens ein Wort oder eine Formulierung aus dem Wortspeicher einbezogen, erhielten mit insgesamt 273 Punkten ein besseres inhaltliches Ergebnis als die Versuchsprotokolle, in denen die Wortspeicher nicht vorkamen (215 Punkte). **Die Versuchsprotokolle mit verwendeten Wortspeichern enthielten konkret mehr korrekte Vermutungen, korrekte Beobachtungen und korrekte Erklärungen.** Dieses Ergebnis kann zum einen bedeuten, dass sprachliche Hilfen sehr wohl auch zur inhaltlichen Qualität beitragen können – und zwar dann, wenn sie von den Schüler:innen angenommen und verwendet werden.

Auf der anderen Seite kann das Ergebnis darauf zurückzuführen sein, dass sprachstarke Kinder von vornherein einen inhaltlichen Vorsprung haben. In dem Fall würde das Ergebnis nur auf den Zusammenhang von (Bildungs-)Sprache und inhaltlicher Qualität hinweisen, wie er in Kapitel 2.2 zur naturwissenschaftlichen Bildung beschrieben ist. Unterstützt wird die Interpretation durch empirisch gewonnene Schlussfolgerungen, dass Sprache und Denken in wechselseitiger Beziehung zueinander stehen.⁵¹ Dieser Interpretation steht entgegen, dass das naturwissenschaftliche Fachwissen laut Kulgemeyer und Schecker (2013, S. 235) nicht zwangsläufig mit einer adressatengerechten Kommunikation einhergeht.⁵²

Die Aufmerksamkeit und aktive Teilnahme der Kinder am Unterrichtsgeschehen könnte ebenfalls eine Rolle spielen. Eventuell waren die Kinder, die sich die Wortspeicher merkten, im Unterricht engagierter. Eine engagierte Teilnahme am Unterricht könnte sich zusätzlich positiv auf die Qualität der Versuchsdurchführung und damit einhergehend auf die inhaltliche Qualität der Protokolle ausgewirkt haben.

Eine vierte Interpretation bezieht sich auf die Beurteilung. In der Einleitung der vorliegenden Studie ist erwähnt, dass Sprache einen Effekt auf die Beurteilung schriftlicher Lernprodukte haben kann. Es wäre denkbar, dass die Beurteilung des Inhalts bei

⁵¹ vgl. Einleitung (Kempert et al., 2018, S. 8) und Kapitel 2.1.3 (Ufer et al., 2013, S. 197-198)

⁵² vgl. Kapitel 2.1.1

Betrachtung der verwendeten Wortspeicher positiv beeinflusst wird. Abgeschwächt wird diese Interpretation dadurch, dass es sich bei der Vergabe der Punktzahl um ein grobes dreiteiliges Kategoriensystem handelt. Eine falsche Antwort kann auch bei wohlwollender Betrachtung kaum als korrekt beurteilt werden. Darüber hinaus wurden die beurteilten Antworten pro Kategorie stichprobenartig von einer unabhängigen Person gegencodiert, diskutiert und im Nachhinein von der Autorin listenweise als Gesamtergebnis überprüft.

Unter Einbezug aller Interpretationsmöglichkeiten scheinen die ersten drei Interpretationen besonders plausibel und müssten in weiteren Studien genauer untersucht werden.

Der Vergleich zwischen Versuchsprotokollen mit und ohne Verwendung der Wortspeicher zeigt ebenfalls, dass die Nutzung der Wortspeicher mit einem erhöhten Aufkommen an Bildungssprache einhergeht. Bezogen auf Bildungssprache enthielten die 26 Versuchsprotokolle mit Wortspeichern insgesamt 144 kodierte bildungssprachliche Wörter und Formulierungen. In den 26 Versuchsprotokollen ohne Wortspeichern waren es 69 Wörter und Formulierungen. Die Kodierhäufigkeiten zeigen bei gleicher Stichprobengröße somit eine Differenz von 75 Sprachmitteln, die in der Wortspeicher-Verwendungsgruppe zusätzlich vorkamen. Im Detail konnte gezeigt werden, dass in den Versuchsprotokollen mit Wortspeichern mehr Überschneidungen zwischen verschiedenen Kategorien der Bildungssprache vorkommen. **Das Ergebnis impliziert eine höhere Vielseitigkeit bildungssprachlicher Elemente in den Versuchsprotokollen mit Wortspeichern.**

Eine mögliche Interpretation bezieht erneut den Aspekt *Sprachbewusstheit* ein. Es ist zu vermuten, dass die Kinder, die sich auf die Wortspeicher einließen und diese nutzten, im Allgemeinen ein Verständnis für die Bedeutung von Bildungssprache im Zusammenhang mit den Versuchsprotokollen aufweisen. Die Unterrichtssequenz könnte dieses Verständnis gefördert haben. Entsprechend nutzten die Kinder neben den Wortspeichern auch vielseitige bildungssprachliche Wörter und Formulierungen außerhalb des Wortspeichers.

Eine weitere Interpretation bezieht erneut die identitätsstiftende Funktion von Sprache⁵³ mit ein. Es wäre denkbar, dass die Kinder durch die Verwendung der Wortspeicher in die Rolle einer forschenden Person eintreten und dadurch die Schwelle zur Verwendung weiterer bildungssprachlicher Mittel geringer ist.

⁵³ vgl. Kapitel 2.1.3

6.2 Zusammenfassende Schlussfolgerung

Insgesamt unterstützt die Studie die Annahme, dass Wortspeicher von Kindern im Anschluss an eine diesbezügliche Unterrichtssequenz von vier 90-minütigen Schulstunden angenommen werden.⁵⁴

In Plakatform gezeigte Wortspeicher sind der auditiven Vermittlung der Wortspeicher vorzuziehen. So konnte im Anschluss an die Unterrichtssequenz bei der Verwendung der Wortspeicher als auch bei der Anwendung von Bildungssprache inferenzstatistisch ein Unterschied zwischen der KG und der Wortspeicher-Plakatgruppe festgestellt werden. Sowohl *Wortspeicher* als auch *Bildungssprache außerhalb des Wortspeichers* kam in den Versuchsprotokollen der Wortspeicher-Plakatgruppe vermehrt vor.

Bei Betrachtung der Häufigkeitsbeschreibung zeigt sich bezüglich beider Fragestellungen (F1 und F2) eine – statistisch nicht signifikante – Tendenz, die den Vorzug der Plakatform über die auditive Vermittlung und den Vorzug der auditiven Vermittlung über die fehlende Vermittlung verdeutlicht. Das heißt: Eine auditive Vermittlung von Wortspeichern ist – den deskriptiven Mittelwerten zufolge – für die Verwendung der Wortspeicher und für die Anwendung von Bildungssprache förderlicher als das Auslassen der Wortspeicher. Die beschriebene Tendenz deckt sich mit den Erkenntnissen aus der Psycholinguistik, nach denen die Darbietung unterschiedlicher Informationsformen (graphemisch, phonologisch, semantisch) mit einer höheren Verknüpfung im mentalen Lexikon einhergeht.

Ein inhaltlicher Nutzen konnte im Rahmen dieser Untersuchung für die Arbeit mit Wortspeichern in Plakat- und in auditiver Form nicht nachgewiesen werden. Auf der anderen Seite zeigt sich, dass Sprachbildung ohne inhaltliche Einbußen im Unterricht integriert werden kann.

Als neue Erkenntnis aus der vorliegenden Studie wird eine erhöhte Sprachbewusstheit der Kinder aufgrund der Arbeit mit visuellen Wortspeichern vermutet. Diese Vermutung basiert darauf, dass die Kinder aus der Wortspeicher-Plakatgruppe im Anschluss an die Unterrichtssequenz kaum Sprachmittel aus dem Wortspeicher zur Erklärung einsetzten, dafür jedoch eine Vielzahl eigener – kreativer – erklärender Sprachhandlungen nutzten.

Zur besseren Einschätzung der Ergebnisse werden entscheidende Limitationen der Forschungsmethodik im folgenden Kapitel präsentiert.

⁵⁴ Diese Schlussfolgerung basiert auf dem geplanten Kontrast zwischen TG1 & TG2 (zusammengenommen) gegenüber der KG.

6.3 Methodische Kritik

Limitationen der Forschungsmethodik werden zunächst in Bezug auf die Gütekriterien vorgestellt. In puncto **Validität** sind die erstellten Kategoriensysteme zur Kennzeichnung von Bildungssprache und zur Bewertung des Inhalts hervorzuheben. Das Kategoriensystem zur Bildungssprache ist eine wertvolle Stütze zur Erfassung bildungssprachlicher *Morphosyntax*, *Lexikon* und *Sprachhandlungen*. Es basiert zum Großteil auf der nach Tietze et al. (2016) entwickelten Ratingskala – auch RaBi genannt.⁵⁵ Allerdings ist die Ratingskala auf die Erfassung von Bildungssprache in mündlichen Aussagen der Vorschule konzipiert, während sie in dieser Studie zur Erfassung von Bildungssprache im Schriftlichen in einer dritten Klasse angepasst wurde. Eine Änderung des Kategoriensystems ist die Entfernung der Unterkategorie *Komplexe Verbkonstruktionen* aufgrund ihres zahlreichen Vorkommens im schriftlichen Kontext. Weitere Änderungen erfolgten nach Diskussionen mit zwei unabhängigen interkodierenden Personen (während der ersten Pilotstudie und während der Hauptkodierung), sowie nach Absprache mit einer Deutschlehrkraft. Das daraus resultierende Kategoriensystem könnte durch die notwendigen Abweichungen von der ursprünglichen RaBi-Skala an Validität bezüglich der Identifikation von Bildungssprache eingebüßt haben. Darüber hinaus ist eine valide Erfassung von Bildungssprache insofern problematisch, da es nicht *die* Bildungssprache gibt. In dem WuW-Modell von Bildungssprache in der Grundschule nach Juska-Bacher und Beckert (2015, S. 39) wird beispielsweise nicht nur die sprachliche, sondern auch eine kognitiv-reflexive und entwicklungspsychologische Dimension von Bildungssprache berücksichtigt.⁵⁶

Als mögliche Einschränkung der Validität ist auch das Kategoriensystem zur inhaltlichen Beurteilung der Versuchsprotokolle zu nennen. Die Entscheidung fiel aus Gründen der Übersichtlichkeit auf ein grobes dreigliedriges Bewertungssystem mit den Stufen *korrekt*, *unzureichend* und *fehlerhaft*. Durch die grobe Gliederung ist das Bewertungssystem auf verschiedene Inhalte übertragbar und kann somit im Prä-, Post- und Follow-Up-Test angewandt werden. Der Nachteil besteht darin, dass inhaltliche Abgrenzungen nicht im Detail herausgestellt werden. Ein Beispiel hierzu zeigen zwei Erklärungsansätze, die beide laut Kategoriensystem als *unzureichend* kodiert wurden.

Ein Kind schrieb im Follow-Up-Test: *Ich erkläre mir das so: Wenn das Salatblatt mit Wasser in Berührung kommt, nimmt es das Wasser auf und es wird hubbelig. Wenn das Salatblatt mit Salzwasser in Berührung kommt, wird es weich. Dann, wenn das Salatblatt aus dem Salzwasser ins Wasser kommt, bleibt es immer noch weich, weil das Salz im Salatblatt drinbleibt.*

⁵⁵ vgl. Kapitel 2.1.4

⁵⁶ vgl. Kapitel 2.1.4

Ein anderes Kind erklärte den Versuch des Follow-Up-Tests wie folgt: *Das Salz heftet sich an das Blatt und so wird es labbriger. Im Wasser ist kein Salz, also passiert nichts. Das Salz löst sich im Wasser auf und deshalb wird es wieder hart.*

Beide Versuchsansätze sind inhaltlich nicht korrekt. Da es sich um erklärende Ansätze handelt, wurden sie als *unzureichend* bewertet. Inhaltlich grenzen sich beide Erklärungsansätze stark voneinander ab, sodass ein detaillierteres Bewertungssystem Nuancen dieser Art besser darstellen könnte. Die Validität der inhaltlichen Beurteilung würde dadurch erhöht. Zusammenfassend stellt die Validität der Datenauswertung bezüglich Bildungssprache und Inhalt (F2 und F3) eine Problematik dar, die sich daraus ergibt, dass Bildungssprache und Inhalt latente, d.h. schwer erfassbare, Konstrukte darstellen.

Ein weiterer relevanter Aspekt bezüglich der Validität betrifft die Datenerhebung. Das Schreiben eines Versuchsprotokolls wird bedingt durch verschiedene Faktoren. Zunächst wird das **sprachliche Verständnis der Versuchsanweisungen** implizit vorausgesetzt. Zur Erhöhung der Validität wurden die Anweisungen in einer Vorstudie getestet und kritische Stellen durch vereinfachte Sprache ausgetauscht. Dennoch sind sprachliche Verständnisprobleme bei sprachschwächeren Kindern oder bei Kindern, die die Versuchsanweisungen flüchtig lesen, nicht auszuschließen.

Neben dem sprachlichen Verständnis ist eventuell das **kognitive Verständnis der Arbeitsanweisung** von Bedeutung. Es besteht die Möglichkeit, dass Kinder den Versuchsanweisungen inhaltlich nicht folgen können oder wichtige inhaltliche Schritte übersehen. Um dem entgegenzuwirken, hätte eine kleinschrittige gemeinsame Vorgehensweise unter Anleitung der Lehrkraft hilfreich sein können. Dieses Vorgehen wurde in diesem Forschungsvorhaben zugunsten der Selbstständigkeit der Kinder bewusst abgelehnt, wodurch die Gefahr kognitiver Verständnisprobleme und inhaltlich falscher Durchführungen in Kauf genommen wurde. Teilweise sind inhaltliche Verständnisprobleme in den Versuchsprotokollen offensichtlich. So schrieb ein Kind bei der Beobachtung eines Teilversuchs: *Ich verstehe das nicht*. Ein anderes Kind schrieb bei dem Versuch, in dem es um die Höhe des Eies im Wasser ging: *Das Ei wird härter*. Dies weist auf eine falsche Durchführung hin. Da sich die inhaltliche Bewertung sowohl auf die praktische Durchführung als auch auf die Verschriftlichung der praktisch erworbenen Erkenntnisse bezieht, ist die Validität an dieser Stelle eingeschränkt.

Zuletzt ist in der Datenerhebung einbezogen, **inwieweit die Kinder ihre Gedanken zu den Versuchen verschriftlichen können und wollen**. Ein Kind schrieb als Beobachtung, nachdem es Salz in Wasser geschüttet hatte: *Es entsteht Blobärchen*. Es ist anzunehmen, dass das Kind *Bläschen* oder *Blubberblasen* schreiben wollte und sich nicht entsprechend ausdrücken kann. Ausbleibende Antworten sind ein Indiz dafür, dass die Gedanken nicht verschriftlicht werden konnten oder die Motivation zur Verschriftlichung fehlte. Sie können auch auf die sprachlichen und kognitiven

Verständnisschwierigkeiten beim Lesen und Bearbeiten der Versuchsanleitung zurückzuführen sein. Kurzum: Auf Ebene der Datenerhebung kommen verschiedene Faktoren zusammen, die zur Verschriftlichung eines Versuchsprotokolls notwendig sind und somit die Validität beeinflussen. Konfundierung kann nicht ausgeschlossen werden. Der Mehrwert, der trotz dieser Einschränkungen gegeben ist, liegt in der Erprobung in einer authentischen Unterrichtssituation.

Die **Objektivität** wurde in diesem Forschungsvorhaben durch verschiedene Maßnahmen erhöht. Eine höhere Durchführungsobjektivität wurde durch die konstante Versuchsleitung in allen drei Gruppen erreicht, die den Kindern vorab unbekannt war. Die Auswertung mittels festgelegter Kategoriensysteme und die Überprüfung der Auswertung durch Bestimmung der Intercoder-Übereinstimmung trägt zur Auswertungsobjektivität bei. Auf der anderen Seite ist die Versuchsleitung gleichzeitig die auswertende und interpretierende Person. Es kann nicht ausgeschlossen werden, dass unbewusst unterschiedliche Formen der Sprachbildung durch Mikro-*Scaffolding* in den drei Klassen vorgenommen wurden. Bei den Diskussionen mit den intercodierenden Personen stellten sich weiterhin einige Schwachstellen der Kategoriensysteme heraus, die wiederum auf den Umstand der schwer erfassbaren Variablen Bildungssprache und Inhalt zurückzuführen sind. Diese Schwachstellen wurden durch Anpassungen des Kategoriensystems behoben. Außerdem wurde zur Erhöhung der Auswertungsobjektivität eine abschließende Kontrolle bezüglich der Vergleichbarkeit aller finalen Kodierungen vorgenommen.

Ein weiterer Kritikpunkt stellt die **Stichprobe** dar. Insgesamt wurden 138 Versuchsprotokolle ausgewertet, davon 60 aus der KG, 36 aus der Gruppe der auditiven Wortspeicher-Vermittlung und 42 aus der Gruppe der visuellen Wortspeicher-Vermittlung. Aufgrund der Corona-Pandemie während der Untersuchung ist das *Panel* kleiner als antizipiert. Die quantitative Auswertung ist mit Vorsicht zu betrachten, denn die Reichweite der Interpretation statistischer Effekte ist bei der Stichprobengröße von insgesamt 46 Vpn begrenzt. Die Mixed-Methods-Auswertung hat dennoch ihre Berechtigung, da die statistische Auswertung in diesem Design eine forschungsunterstützende Funktion innehat. Ein weiterer Aspekt der Stichprobe ist die Inklusion von Kindern mit Lese-Rechtsschwäche (LRS) und nicht-deutscher Herkunftssprache (ndH). Da der C-Test die sprachliche Vergleichbarkeit der Gruppen unter Berücksichtigung von Kindern mit LRS und ndH suggeriert, sind diese Kinder in der Auswertung einbezogen. Dadurch werden in dem quasi-experimentellen Forschungsdesign praxisnahe Lerngruppen untersucht. Auf der anderen Seite stellt der Einbezug dieser Kinder unter Umständen eine weitere Störvariable dar.

Zuletzt stellt die **Treatment-Dauer** eine Limitation des Forschungsdesigns dar. In diesem Forschungsdesign ist die Treatment-Dauer auf vier 90-minütige Unterrichtsstunden

beschränkt. Inklusiv der Test-Stunden besteht die Unterrichtssequenz aus acht 90-minütigen Unterrichtsstunden. Eine längere Treatment-Dauer wurde aufgrund potenzieller Ermüdungs-Effekte (*fatigue effect*) bzw. potenziellem Motivationsverlust für dieses Forschungsvorhaben verworfen. Das Ergebnis zeigt, dass die vier 90-minütigen Unterrichtsstunden jedoch keinen nachweisbaren nachhaltigen Nutzen hervorrufen.

6.4 Ausblick

Ausgehend von den genannten Limitationen und der Ergebnisdiskussion werden in diesem Kapitel die Konsequenzen für künftige weiterführende Forschungen zusammengefasst. Abschließend wird auf praktische Implikationen der Forschungsergebnisse für den naturwissenschaftlichen Sachunterricht hingewiesen. Das letztgenannte Kapitel betont die empirische Relevanz der vorliegenden Studie.

6.4.1 Wissenschaftlicher Ausblick

In Kapitel 6.3 werden methodische Schwachstellen bezogen auf die Gütekriterien genannt. In **Replikationsstudien** könnte die Durchführungsobjektivität erhöht werden, indem die Unterrichtsstunden aufgezeichnet würden. Die **Video-Aufnahmen** würden unbewusstes Mikro-*Scaffolding* von Seiten der Lehrkraft aufdecken, sodass das sprachliche Verhalten in den drei Klassen in die Auswertung und Interpretation der Ergebnisse einbezogen werden könnte. Zur Erhöhung der Auswertungsobjektivität würde eine **größer angelegte vorherige Testung der Kategoriensysteme** durch verschiedene auswertende Personen beitragen.

Das Problem der Validität bezieht sich darauf, dass das Versuchsprotokoll auf verschiedenen Faktoren basiert, wie beispielsweise das sprachliche und kognitive Verständnis der Versuchsanleitung oder die Schreibmotivation. Eine Möglichkeit zur Erhöhung der Validität wären **qualitative Interviews**, in denen die Versuchsleitung Nachfragen stellen kann. Das Problem der qualitativen Interviews wäre die medial phonische Realisierung der Antworten, anstelle von medial schriftlichen Lernprodukten. Da die Fragestellungen auf mediale und konzeptionelle Schriftlichkeit abzielen, müssten die Interviews derart gestaltet sein, dass die Kinder unter Aufsicht ein Versuchsprotokoll schreiben. Die Versuchsleitung sollte die Kinder zu Antworten motivieren und dürfte bei inhaltlichen oder sprachlichen Verständnisproblemen nachhelfen.

Die Schwachstelle der kleinen Stichprobe könnte mit einem **quantitativ ausgerichteten Forschungsdesign** behoben werden. Der Vorteil einer quantitativen Studie mit größerer Stichprobe ist in der höheren *Power* zu sehen. Weitere Hintergrundvariablen müssten erfasst werden, um mögliche Störvariablen zu kontrollieren.

Neben Replikationsstudien ergeben sich aus den Ergebnissen der vorliegenden Studie Ansatzpunkte für vertiefende Studien. Zunächst werden Möglichkeiten für **vertiefende Studien zu den Wortspeichern** vorgestellt. In der Interpretation der Ergebnisse zur ersten Fragestellung (F1) wird die mentale Verknüpfung zwischen Wortspeicher und dem Schreiben eines Versuchsprotokolls als mögliche Ursache für die häufigere Verwendung der Wortspeicher genannt. Um die Bedeutung des **kontextuellen Bezugs** empirisch zu untermauern sind Studien notwendig, in denen die Kenntnisse über die Wörter und Formulierungen des Wortspeichers in der KG nach Beendigung des Follow-Up-Tests erfragt werden. Sind die Wörter und Formulierungen allen Kindern bekannt, so wäre der kontextuelle Bezug eine plausible Erklärung für die verstärkte Verwendung von Wortspeichern nach der Wortspeicher-Arbeit. Denkbar wäre auf der anderen Seite ein Studiendesign mit Wortspeichern, die Fantasiewörter (*nonsense words*) umfassen. Dies würde die vorherige Kenntnis der Wortspeicher in allen Gruppen ausschließen, sodass das Memorieren und der Abruf der Wortspeicher im Vordergrund ständen.

Bezogen auf **vertiefende Untersuchungen zur Bildungssprache** sollte die weitere Forschung Aspekte aufgreifen, die im Folgenden aufgeführt werden. In der Ergebnisinterpretation der vierten Fragestellung (F4) wurde diskutiert, inwieweit die Verwendung von redundanter Bildungssprache im Versuchsprotokoll als positiv zu bewerten ist. Es ist zwar unbestreitbar, dass Bildungssprache im Grundschulalter erwünscht ist. Doch die qualitative Betrachtung der Ergebnisse zeigt ein Beispiel, wie ein Kind seine Erklärung durch verschiedene Formulierungen mehrfach ausdrückte (*Meine Erklärung ist, dass ...; ... weil ...; Ich erkläre mir das so ...; ...wegen ...; ... deswegen ...*). Hier stellt sich die Frage der Verhältnismäßigkeit. In der Interpretation von F4 wurde auf das Konzept der Schulsprache als Hinführung zu einem realistischen Versuchsprotokoll hingewiesen und das obige Beispiel positiv im Sinne einer ersten Sprachbewusstheit interpretiert. **Das Thema Sprachbewusstheit sollte in zukünftigen Forschungen einen höheren Stellenwert einnehmen.** Vorstellbar wäre eine Studie, die im Anschluss an ein Treatment mit Wortspeichern das Bewusstsein über den Unterschied von Alltagssprache und Bildungssprache, sowie den sprachlichen Wechsel zwischen den beiden Registern überprüft. Ein solcher Test wäre in Anlehnung an Kulgemeyer und Schecker (2009, S. 142) möglich, die zur Testung des Kommunikationsaspekts *Code / Text* den Kindern Sätze aus der Physik vorgaben. Die Kinder sollten diese Sätze der Alltagssprache oder Fachsprache zuordnen (ebd., S. 142). Dies könnte für den Kontext des Versuchsprotokolls abgewandelt werden. Eine derartige Studie würde zeigen, **ob sich durch auditives oder visuelles Wortspeicher-Treatment das Bewusstsein über die verschiedenen Sprachregister ändern ließe.** Eine solche Studie zur Sprachreflexion würde die vorliegenden Ergebnisse mehrfach bereichern: Zum einen würde die neu aufgestellte Vermutung der höheren Sprachbewusstheit als Folge des Wortspeicher-

Treatments getestet. Zum anderen würde die Sprachreflexion als Bestandteil des WuW-Modells zur Beschreibung von Bildungssprache mitberücksichtigt.

Da Sprachbewusstheit ein wichtiger Baustein in der Wortschatzarbeit ist⁵⁷, könnten weitere Studien die **Sprachbewusstheit im Treatment** mitaufnehmen (*top-down approach*). Die Ergebnisse dieser Studie zeigen, dass der isoliert stehende Einsatz von Wortspeichern sinnvoll ist. Interessant wäre in einem nächsten Schritt die Kombination aus Wortspeicher-Arbeit und Sprachbewusstheit. Beispielsweise könnten die Kinder innerhalb eines Wortspeicher-Treatments zum Markieren der verwendeten Sprachmittel oder zum Markieren von Bildungssprache in ihren eigenen Versuchsprotokollen aufgefordert werden. Nach und nach wären weitere Kombinationen im Treatment möglich, um ein effektvolles Konzept für sprachbildendes *Scaffolding* im naturwissenschaftlichen Sachunterricht zu ermitteln und weiterzuentwickeln.

Im Folgenden werden Möglichkeiten für **vertiefende Studien hinsichtlich des Inhalts** erörtert. In der vorgestellten Studie ließen sich inferenzstatistisch keine Unterschiede zwischen Gruppen, die ein Wortspeicher-Treatment erfahren haben, und der Gruppe ohne Wortspeicher-Treatment hinsichtlich der inhaltlichen Beurteilung feststellen. Dies widerspricht dem Ergebnis der Studie von Tajmel (2017b, S. 346), in der die Verwendung lexikalischer Hilfsmittel mit einer höheren fachlichen Qualität einhergeht. Ausgehend von diesem Widerspruch wären vertiefende Studien interessant, in denen die **Variabilität der Beurteilungen durch Lehrkräfte in Bezug auf die Verwendung von Bildungssprache untersucht wird**. Konkret würden mehrere Lehrkräfte die Versuchsprotokolle mithilfe eigens erstellter Kategoriensysteme bewerten. Dies ist insofern von hohem wissenschaftlichem Interesse, da die Beurteilung im Schulkontext je nach Lehrkraft und persönlich erstelltem Auswertungshorizont variabel ist. Die Kategoriensysteme können anschließend offengelegt und verglichen werden. Die Vertiefung in dieser Hinsicht würde verdeutlichen, welchen Einfluss die Verwendung von Bildungssprache (inklusive Wortspeicher) auf die extern beurteilte inhaltliche Qualität haben kann. Eine tiefergehende Studie könnte zeigen, ob die Arbeit mit Wortspeichern und die damit zusammenhängende Verwendung von Bildungssprache auch zu positiveren Beurteilungen der inhaltlichen Qualität im Schulkontext führt. Die Studie wäre von hoher Relevanz für das Schulsystem, da die Beurteilungen durch Lehrkräfte den schulischen Werdegang von Schüler:innen festlegen.

Die Ergebnisse der vorliegenden Studie bezüglich des Inhalts zeigen weiterhin auf, dass vor allem die Erklärungen für Schüler:innen schwer zu bewältigen sind (vgl. Abb. 37). Eine weitere vertiefende Studie zur Inhaltsebene sollte prüfen, **ob die Wortspeicherarbeit anhand eines genotypischen Phänomenkreises einen Einfluss auf die inhaltliche Qualität der Versuchsprotokolle haben könnte**. Die Bezeichnung der

⁵⁷ vgl. Kapitel 2.3.3

Phänomenkreise mit der Unterscheidung zwischen genotypisch oder phänotypisch analogen Phänomenen geht auf Spreckelsen (1997, S. 117, 118) zurück. In diesem Forschungsprojekt stehen phänotypisch analoge Phänomene im Mittelpunkt. Das bedeutet konkret: Die Unterrichtssequenz der vorliegenden Studie besteht aus verschiedenen Versuchen zum Thema Salz, in denen zwecks Forschungsdesign unterschiedliche Prinzipien thematisiert werden. Während im Prä-Test das Prinzip der Löslichkeit zur Erklärung herangezogen werden muss, handelt es sich bei weiteren Versuchen z.B. um das Konzept der Dichte im Zusammenhang mit dem Thema *Schwimmen und Sinken*, um den Stromkreislauf oder die Osmose.⁵⁸ Im Gegensatz dazu plädiert Spreckelsen (1997, S. 118) für einen Phänomenkreis, bestehend aus mehreren genotypisch analogen Phänomenen. In diesem Fall würden unterschiedliche Versuche das gleiche naturwissenschaftliche Prinzip – z.B. das Prinzip der Löslichkeit – verdeutlichen. Durch dieses Vorgehen „[...] haben die Kinder die Möglichkeit, selber eine Erklärung für das Phänomen zu finden und zu formulieren, ohne dass die Lehrperson ihnen etwas erklären muss“ (Beinbrech, 2013, S. 278). Um in den Teststunden inhaltlich gleiche Voraussetzungen für die Schüler:innen zu schaffen und somit die Vergleichbarkeit der Teststunden zu gewährleisten, wurde der genotypische Phänomenkreis in dieser Studie nicht berücksichtigt. In einer Studie mit längerer Treatmentdauer wäre die intensive Behandlung mehrerer Problemstellungen innerhalb eines Halbjahres möglich. In diesem Fall würde sich die Untersuchung genetisch analoger Phänomene und damit auch der Nutzen von Wortspeichern auf selbst entwickelte Erklärungen anbieten. Die Test-Stunden wären jeweils an das Ende eines genotypischen Phänomenkreises zu legen und würden sich dadurch thematisch unterscheiden. Die inhaltlich vertiefende Studie gäbe wichtige Hinweise zum Einfluss der Wortspeicher, wenn Kinder eigenständig richtige Erklärungen finden können.

Im Allgemeinen bleibt festzuhalten, dass die vier 90-minütigen Unterrichtsstunden keinen nachweisbaren nachhaltigen Nutzen hervorrufen – weder bezüglich der Wortspeicher-Nutzung, noch bezogen auf Bildungssprache oder auf die inhaltliche Qualität. Daraus folgt, dass die **gewählte Treatment-Dauer in zukünftigen Studien neu gewählt werden sollte**. In Anlehnung an die Sprachforscherkurse⁵⁹ sollten die Studien für mindestens 6 Monate angelegt sein. In Abgrenzung zu den empirisch geprüften additiven Sprachforscherkursen, sollten neuere Studien die Sprachbildung und das Schreiben von Versuchsprotokollen in den Unterricht integrieren und die Variable der Wortspeicher genauer in den Fokus nehmen. Durch die Integration der Wortspeicher in den vielperspektivischen Sachunterricht könnten Ermüdungseffekte vermieden und die Nachhaltigkeit des Treatments erneut geprüft werden.

⁵⁸ vgl. Kapitel 4.2.4

⁵⁹ vgl. Kapitel 2.3.5

Ein weiteres allgemeines Ergebnis aus der Wortspeicher-Studie zeigt, dass die auditive Vermittlung der Wortspeicher bezogen auf die ersten zwei Forschungsfragen hilfreich scheint, ohne jedoch statistische Signifikanz zu erreichen. Aufgrund dessen sollte eine neu angelegte Studie den **Unterschied zwischen auditiver und visueller Wortspeicher quantitativ anhand einer größeren Stichprobe** untersuchen. Ein nachweisbarer Effekt bei auditivem Wortspeicher-Treatment könnte dazu beitragen, dass viele Lehrkräfte, denen die Vorbereitung von Wortspeichern in Plakatform im Unterrichtsalltag nicht möglich ist, zumindest auf die auditive Vermittlung von Wortspeichern zurückgreifen.

Die Nähe zwischen Wissenschaft und Schulsystem wird bereits im wissenschaftlichen Ausblick deutlich. Diese Verknüpfung wird im folgenden Kapitel näher beleuchtet.

6.4.2 Schulpraktische Implikationen

Die bisherigen Ausführungen verdeutlichen die Relevanz der Wortspeicher-Arbeit im naturwissenschaftlichen Sachunterricht. **Lehrkräfte sollten bei der Vermittlung des Forschungskreislaufes entsprechende Wortspeicher in Plakatform anbieten.** Diese Empfehlung basiert darauf, dass die Kinder durch die visuelle Wortspeicher-Arbeit in diesem Forschungsprojekt mehr Wörter und Formulierungen der Bildungssprache anwenden. Dies fördert die gewünschte Sprachhandlungsfähigkeit.⁶⁰ Da eine repetitive Verwendung von Bildungssprache bzgl. ihrer Angemessenheit im Versuchsprotokoll diskutiert werden kann, gilt diese Empfehlung zunächst ausschließlich für erstmalig formulierte Versuchsprotokolle der dritten Klasse. In dieser Altersstufe und mit Einführung eines neuen Textformats ist selbst eine Repetition von bildungssprachlichen Elementen als Zeichen einer ersten Sprachbewusstheit positiv zu bewerten. Eine Reduktion der inhaltlichen Qualität ist bei der additiven Vermittlung visueller Wortspeicher nicht zu erwarten. Dies verdeutlicht, dass Wortspeicher sehr gut in den Unterricht integriert werden können und sollten. **Sind Wortspeicher in Plakatform für die Lehrkraft im Unterrichtsalltag nicht zu leisten, könnten auditive Wortspeicher ggf. zielführend sein.** So ist selbst bei einer mündlichen Vermittlung wichtiger Wörter und Formulierungen eine Tendenz hinsichtlich vermehrt eingesetzter Bildungssprache und keine Reduktion der inhaltlichen Qualität zu sehen. Die vorliegende Studie unterstreicht jedoch den weitaus größeren Nutzen von Wortspeichern in Plakatform. Ob der Nutzen auch bei visuellen Wortspeichern in Form von Arbeitsblättern oder digitalen Medien gegeben ist, bleibt in dieser Studie offen.

⁶⁰ vgl. Kapitel 2.1.3

Die Auswahl der Wortspeicher hängt grundsätzlich vom sprachlichen Lernziel von Seiten der Lehrkraft ab. Die Studie zeigt, dass kürzere und bekannte Wörter und Formulierungen eher verwendet werden als es bei längeren oder unbekannteren Sprachmitteln der Fall ist (vgl. Tab. 18). Auf der anderen Seite werden die längeren oder unbekannteren Formulierungen aus dem Abschnitt der Erklärung in vielen Fällen durch ähnliche Formulierungen von Seiten der Kinder ersetzt (vgl. Tab. 20). Möchte die Lehrkraft, dass die gleichen Sprachmittel aus dem Wortspeicher verinnerlicht werden, sollte sie anfangs auf Kürze und einen hohen Bekanntheitsgrad setzen. Steht allerdings die Bildungssprache im Allgemeinen im Fokus, kann ein höherer Schwierigkeitsgrad der Wortspeicher zur kreativen und individuellen Verwendung von Bildungssprache beitragen. Letzteres führt zur Sprachhandlungsfähigkeit und lässt auf einen hohen Grad an individueller Sprachbewusstheit schließen.

Ein weiteres Kriterium bei der Auswahl der Wortspeicher bezieht sich auf die bildungssprachlichen Kategorien *Sprachhandlungen*, *Lexikon* und *Morphosyntax*. Unter Einbezug aller Versuchsprotokolle werden *Sprachhandlungen*, d.h. Vermutungen und Begründungen, in diesem Forschungsprojekt besonders selten von Kindern verwendet (vgl. Abb. 36). **Somit sollten gerade die Sprachhandlungen in Wortspeichern eine stärkere Beachtung finden.** Eine empirische Studie, die im Theorieteil vorgestellt wurde, deutet auf einen Zusammenhang zwischen geforderten Sprachhandlungen und Länge der Schüler:innen-Antworten hin (Pineker-Fischer, 2017, S. 357).⁶¹ Gerade das *Erklären* könnte zu längeren Antworten führen (ebd., S. 357), falls das Ergebnis der Studie auf medial schriftliche Produkte im Sachunterricht übertragbar ist. Ob sich die Versuchsprotokolle durch längere Antworten inhaltlich verbessern, ist unklar. Sprachlich gesehen sind längere Antworten in Versuchsprotokollen der Grundschule wünschenswert – vor allem, wenn die Antworten sprachlich ersichtliche Vermutungen und Begründen enthalten. Auf der Grundlage, dass die Sprachhandlungen in der vorliegenden Studie selten verwendet wurden und ggf. zu längeren Antworten führen könnten, wird Lehrkräften zur gezielten Förderung und Forderung der Sprachhandlungen mittels visueller Wortspeicher geraten.

Zuletzt zeigen die Ergebnisse dieser Studie, dass eine einmalige Unterrichtssequenz mit auditiven oder visuellen Wortspeichern für einen langfristigen Nutzen nicht ausreicht. Es kann festgehalten werden, dass **die Arbeit mit visuellen Wortspeichern regelmäßig stattfinden muss.** Dies wird in der Studie zwar nicht geprüft, durch die Ergebnisse jedoch nahegelegt. Die Forderung nach einer regelmäßigen Integration von Sprachbildung im Unterrichtsalltag korrespondiert mit dem Ergebnis eines Reviews, in dem die Studienergebnisse zur Wirkung additiver und integrativer Sprachförderung zusammengefasst werden (Egert, 2017). Eine Erkenntnis aus dem Review lautet: „Um Ergebnisse

⁶¹ vgl. Kapitel 2.2.3

zu zeigen, muss sprachliche Förderung regelmäßig stattfinden“ (Egert, 2017, S. 43). Zwar bezieht sich das Ergebnis auf spezielle Sprachförderprogramme in Kindertageseinrichtungen, doch die Ergebnisse der vorliegenden Studie unterstützen die Erkenntnis für den allgemeineren Kontext der Sprachbildung im Grundschulalter. Anknüpfende Studien sind notwendig, um einen langfristigen Effekt bei regelmäßiger Wortspeicher-Arbeit tatsächlich zu prüfen.

Zusätzlich zur regelmäßigen Wortspeicher-Arbeit bietet sich die **Einbindung in ein Gesamtsprachenkonzept** an. Auch dies wird durch die vorliegende Studie nicht direkt geprüft. Der Vorschlag ergibt sich aus dem Umkehrschluss, dass die isolierte Wortspeicher-Arbeit nur einen kurz anhaltenden Effekt erzielte. Analog hierzu fassen Kniffka und Roelcke (2016) zusammen: „Konzepte sind umso erfolgversprechender, je stärker sie in ein Gesamtsprachenkonzept eingebettet sind“ (ebd., S. 117). Das Konzept *Scaffolding* bietet bereits ein Gesamtsprachenkonzept, das die Wortspeicher integriert. Der allgemeine Nutzen von *Scaffolding* (in unterschiedlichen Auslegungen) konnte in empirischen Studien gezeigt werden.⁶² Der Mehrwert der vorliegenden Studie besteht in der Erkenntnis, dass die isolierte Wortspeicher-Vermittlung für einen langfristig erkennbaren Unterschied nicht ausreicht. Ob ein ausgearbeitetes *Scaffolding*-Konzept im Kontext des Forschungskreislaufes langfristig zu bildungssprachlich formulierten Versuchsprotokollen führt, muss genauer untersucht werden.

Eine dritte Möglichkeit zur Anregung eines langanhaltenden positiven Effekts von Wortspeichern ist die **durchgängige Sprachbildung**. Die Forderung nach einer durchgängigen Sprachbildung bezieht sich auf das ganze Schulsystem. Darunter zusammengefasst ist u.a. die Sprachbildung in jeder Klassenstufe (vertikale Sprachbildung) und die fachübergreifende Sprachbildung (horizontale Sprachbildung) (Gogolin, 2013, S. 14-15; 2019, S. 1-3). In weiteren Studien müsste erforscht werden, ob die durchgängige Wortspeicherarbeit (horizontal und vertikal) zu einem nachhaltigeren (sprachlichen) Nutzen führt, als es in der vorliegenden Studie der Fall gewesen ist.

Alles in allem ist die regelmäßige Vermittlung von Wortspeichern, ihre Einbindung in ein Gesamtsprachenkonzept und die durchgängige Sprachbildung auf Schulebene eine naheliegende Möglichkeit zur langfristigen Förderung von bildungssprachlich formulierten und qualitativ konstanten Versuchsprotokollen. Der Ansatz auf verschiedenen Ebenen – Klassenzimmer, Klassenübergreifend und Schulebene – ist insofern sinnvoll, da Sprachentwicklung keine lineare Messlatte darstellt. Im Zweitspracherwerb wird die Sprachentwicklung mit einem Brokkoli verglichen "[...] dessen multidimensionales und stark verzweigendes Wachstum nicht vorhersagbar ist, obwohl er eben auch keine vollkommen willkürliche Gestalt hat“ (Dimroth, 2019, S. 28). Der Vergleich verdeutlicht in eindrucklicher Weise, dass auch Wortspeicher nicht als lineares Instrument zur

⁶² vgl. Kapitel 2.3.5

Entwicklung von Bildungssprache betrachtet werden sollten. Stattdessen dürfte ihr Einsatz in allen Klassenstufen, allen Jahrgangsstufen und als Teil eines Gesamtkonzeptes der Komplexität einer (bildungs-)sprachlichen Entwicklung besser gerecht werden.

Abschließend lässt sich festhalten: Die Frage nach dem Nutzen von Wortspeichern kann mit einem klaren Votum für visuelle Wortspeicher beantwortet werden. Ohne inhaltliche Einbußen können Kinder durch Wortspeicher in Plakatform zum Schreiben eines bildungssprachlichen Versuchsprotokolls animiert und so ihre **Sprachhandlungsfähigkeit** gefördert werden. Zudem deutet die vorliegende Studie daraufhin, dass durch die Arbeit mit visuellen Wortspeichern ein **reflektierter Umgang mit Sprache** angeregt werden kann. In diesem Sinne können sie helfen, die Kinder zu befähigen, in der Schule ihre kommunikativen Ziele zu erreichen und sprachlich erfolgreich zu sein.

7 LITERATURVERZEICHNIS

- Aeppli, J., Gasser, L., Gutzwiller, E. & Tettenborn, A. (2014). *Empirisches wissenschaftliches Arbeiten. Ein Studienbuch für die Bildungswissenschaften*. Bad Heilbrunn: Klinkhardt.
- Agel, C., Beese, M. & Krämer, S. (2012). Naturwissenschaftliche Sprachförderung. Ergebnisse einer empirischen Studie. *Der mathematische und naturwissenschaftliche Unterricht*, 65(1), 36-44.
- Ahrenholz, B., Knoblich, L. & Reichel, J. (2018). Sprache im Fachunterricht. Analysen mündlicher und schriftlicher Wissensvermittlung im Schulunterricht. In I. Winkler, A. Gröschner, & M. May (Hrsg.), *Lehrerbildung in einer Welt der Vielfalt: Befunde und Perspektiven eines Entwicklungsprojekts* (S. 167-181). Bad Heilbrunn: Klinkhardt.
- Apeltauer, E. (2008). Das Kieler Modell: Sprachliche Frühförderung von Kindern mit Migrationshintergrund. In B. Ahrenholz (Hrsg.), *Deutsch als Zweitsprache: Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund* (S. 111-133). Freiburg: Fillibach.
- Archie, C., Rank, A. & Franz, U. (2014). Sprachbildung im und durch Sachunterricht. In A. Hartinger & K. Lange (Hrsg.), *Sachunterricht - Didaktik für die Grundschule* (S. 198-206). Berlin: Cornelsen.
- Atkinson, R. C. & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Hrsg.), *Psychology of Learning and Motivation: Advances in research and theory* (Bd. 2, S. 90-197). New York: Academic Press.
- Baur, R. S., Goggin, M. & Wrede-Jackes, J. (2013). Der C-Test: Einsatzmöglichkeiten im Bereich DaZ. *proDaZ*. Zugriff am 12.05.2020 unter https://www.uni-due.de/imperia/md/content/prodaz/c_test_einsatzmoeglichkeiten_daz.pdf
- Becker-Mrotzek, M., Brinkhaus, M., Grabowski, J., Hennecke, V., Jost, J., Knopp, M., Schmitt, M., Weinzierl, C. & Wilmsmeier, S. (2015). Kohärenzherstellung und Perspektivenübernahme als Teilkomponenten der Schreibkompetenz: Von der diagnostischen Absicherung zur didaktischen Implementierung. In A. Redder, J. Naumann, & R. Tracy (Hrsg.), *Forschungsinitiative Sprachdiagnostik und Sprachförderung - Ergebnisse* (S. 177-205). Münster: Waxmann.
- Beese, M., Kleinpaß, A., Krämer, S., Reschke, M., Rzeha, S. & Wiethoff, M. (2017). *Praxishandbuch Sprachbildung Biologie*. Stuttgart: Klett.
- Beinbrech, C. (2013). Mit Kindern in Gesprächen gemeinsam nachdenken: Naturwissenschaftliches Argumentieren im Sachunterricht. In E. Gläser & G. Schönknecht (Hrsg.), *Sachunterricht in der Grundschule. Entwickeln - gestalten - reflektieren* (S. 271-282). Frankfurt am Main: Grundschulverband (Beiträge zur Reform der Grundschule, 136).
- Bernstein, B. (1964). Elaborated and Restricted Codes: Their Social Origins and Some Consequences. *American Anthropologist*, 66(6), 55-69. Verfügbar unter <http://www.jstor.org/stable/668161>
- Beyer, A. (2015). Wenn zwei sich streiten, freut sich dann der Dritte? *Bildungssprache vs. Schulsprache - eine terminologische Untersuchung*. Pegasus-

- Onlinezeitschrift*, 15(2), 1-39. Verfügbar unter http://www.pegasus-onlinezeitschrift.de/2015_2/pegasus_2015-2_beyer_bildschirm.pdf
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- Bildungsministerium NRW (2021). Lehrpläne für die Primarstufe in Nordrhein-Westfalen. Verfügbar unter https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/300/ps_lp_sammelband_2021_08_02.pdf
- Blanca, M. J., Alarcón, R., Arnau, J., Bono, R. & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4), 552-557. <https://doi.org/10.7334/psicothema2016.383>
- Blasek, B., Ferratusco, B., Herb, M., Hillebrand, K., Kiesinger-Jehle, B., Rebenstorff, H., Schäffner, C. & Volk, M. (Hrsg.). (2010). *Frida & Co - Sachunterricht*. (Ausgabe A. 3/4 NRW). München: Oldenbourg Schulbuchverlag.
- Bochnik, K. (2017). *Sprachbezogene Merkmale als Erklärung für Disparitäten mathematischer Leistung: Differenzierte Analysen im Rahmen einer Längsschnittstudie in der dritten Jahrgangsstufe*. Münster: Waxmann.
- Both, S., Pechstein, O. & Siehr, I. (2013). Wortschatzarbeit im naturwissenschaftlichen Unterricht - Biologie, Chemie, Physik. *Sprachsensibler Fachunterricht: Handreichung zur Wortschatzarbeit in den Jahrgangsstufen 5-10 unter besonderer Berücksichtigung der Fachsprache*, 237-304. Verfügbar unter https://bildungserver.berlin-brandenburg.de/fileadmin/bbb/themen/sprachbildung/Durchgaengige_Sprachbildung/Publicationen_sprachbildung/sprachsensibler_fachunterricht/1_Sprachsensibler_Fachunterricht-Vorwort.pdf
- Brennan, R. L. & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687-699. <https://doi.org/10.1177/001316448104100307>
- Brosius, F. (2018). *SPSS: Umfassendes Handbuch zu Statistik und Datenanalyse*. Frechen: mitp-Verlag.
- Budke, A. & Kuckuck, M. (2017). Sprache im Geographieunterricht. In A. Budke & M. Kuckuck (Hrsg.), *Sprache im Geographieunterricht: Bilinguale und sprachensible Materialien und Methoden* (S. 7-35). Münster: Waxmann.
- Christie, F. (1985). Language and schooling. In S. Tchudi (Hrsg.), *Language, Schooling, and Society* (S. 21-40). Upper Montclair: Boynton/Cook.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Routledge.
- Creswell, J. (2014). Die Entwicklung der Mixed-Methods-Forschung: Einleitung von John Creswell, University of Nebraska-Lincoln. In U. Kuckartz (Hrsg.), *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer.
- Deutsche Welle & Lingofox. (2021). Didaktisierungstool: Sprachkompetenz schulen. Zugriff am 01.07.2021 unter <https://lingofox.dw.com/index.php?url=c-test>
- Dimroth, C. (2019). Lernaltersprachen. In S. Jeuk & J. Settinieri (Hrsg.), *Sprachdiagnostik Deutsch als Zweitsprache*. Berlin: De Gruyter.

- Döring, N. & Bortz, J. (2016). Datenanalyse. In N. Döring & J. Bortz (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Bd. 5. Aufl., S. 597-784). Berlin: Springer.
- Eckhardt, A. G. (2008). *Sprache als Barriere für den schulischen Erfolg: Potentielle Schwierigkeiten beim Erwerb schulbezogener Sprache für Kinder mit Migrationshintergrund*. Münster: Waxmann.
- Egert, F. (2017). Sprachförderung: additiv oder integrativ? *Grundschule DEUTSCH*, 54, 42-43.
- Feilke, H. (2012a). Bildungssprachliche Kompetenzen - fördern und entwickeln. *Praxis Deutsch*, 233, 4-13.
- Feilke, H. (2012b). Schulsprache - Wie Schule Sprache macht. In S. Günthner, W. Imo, D. Meer, & J. G. Schneider (Hrsg.), *Kommunikation und Öffentlichkeit: Sprachwissenschaftliche Potenziale zwischen Empirie und Norm* (S. 149-175). Berlin: De Gruyter.
- Feilke, H. (2019). Schulsprache. *Sprache im Fach*. Zugriff am 27.02.2023 unter https://epub.ub.uni-muenchen.de/61964/1/Feilke_Schulsprache.pdf
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Los Angeles: SAGE Publications.
- Fornol, S. L. (2017). Bildungssprache - mehr als konzeptionelle Schriftlichkeit? In F. Heinzel & K. Koch (Hrsg.), *Individualisierung im Grundschulunterricht: Anspruch, Realisierung und Risiken* (S. 178-182). Wiesbaden: Springer.
- Fuchs, E., Haberfellner, C. & Öhlerer, K. (2014). *Sprachsensibler Unterricht in der Grundschule: Fokus Mathematik*. Graz: Österreichisches Sprachen-Kompetenz-Zentrum. Verfügbar unter http://oesz.at/sprachsensiblerunterricht/UPLOAD/Praxisreihe_22_FINAL_WEB.pdf
- GDSU (Hrsg.). (2013). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Gervé, F. & Peschel, M. (2013). Medien im Sachunterricht. In E. Gläser & G. Schönknecht (Hrsg.), *Sachunterricht in der Grundschule: entwickeln - gestalten - reflektieren* (S. 58-77). Frankfurt/M: GSV.
- Gibbons, P. (2014). *Scaffolding Language, Scaffolding Learning: Teaching English Language Learners in the Mainstream Classroom* (2nd ed.). Portsmouth: Heinemann.
- Glück, H. & Rödel, M. (Hrsg.). (2016). *Metzler Lexikon Sprache* (5. Aufl.). Stuttgart: J.B. Metzler.
- Gogolin, I. (2013). Mehrsprachigkeit und bildungssprachliche Fähigkeiten: Zur Einführung in das Buch ‚Herausforderung Bildungssprache - und wie man sie meistert‘. In I. Gogolin, I. Lange, U. Michel, & H. H. Reich (Hrsg.), *Herausforderung Bildungssprache - und wie man sie meistert* (S. 7-18). Münster: Waxmann.
- Gogolin, I. (2019). Durchgängige Sprachbildung. Zugriff am 27.02.2023 unter https://epub.ub.uni-muenchen.de/62290/1/Gogolin_Durchgaengige_Sprachbildung_Stand%208.7.19.pdf
- Gogolin, I. & Duarte, J. (2016). Bildungssprache. In J. Kilian, B. Brouër, & D. Lüttenberg (Hrsg.), *Handbuch Sprache in der Bildung* (Bd. 21, S. 478-499). Berlin: De Gruyter.

- Gogolin, I. & Lange, K. (2011). Bildungssprache und Durchgängige Sprachbildung. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (S. 107-127). Wiesbaden: VS Verlag.
- Gottwald, A. (2016). *Sprachförderndes Experimentieren im Sachunterricht: Wie naturwissenschaftliches Arbeiten die Sprache von Grundschulkindern fördern kann*. Wiesbaden: Springer.
- Götze, D. (2015). *Sprachförderung im Mathematikunterricht (überarb.)*. Berlin: Cornelsen.
- Götze, D. (2018). Schriftliches Erklären operativer Muster fördern. *Journal für Mathematik-Didaktik*, 40(1), 95-121. 10.1007/s13138-018-00138-4
- Gräber, W. (2002). „Scientific Literacy“ - Naturwissenschaftliche Bildung in der Diskussion. In P. Döbrich (Hrsg.), *Qualitätsentwicklung im naturwissenschaftlichen Unterricht. Fachtagung am 15. Dezember 1999*. (S. 1-28). Frankfurt a. M.: Gesellschaft zur Förderung Pädagogischer Forschung e.V. (GFPF) & Deutsches Institut für internationale Pädagogische Forschung (DIPF).
- Gräber, W. & Nentwig, P. (2002). Scientific Literacy - Naturwissenschaftliche Grundbildung in der Diskussion. In W. Gräber, P. Nentwig, T. Koballa, & R. Evans (Hrsg.), *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung* (S. 7-20). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Grotjahn, R. (2002). Konstruktion und Einsatz von C-Tests: Ein Leitfaden für die Praxis. In R. Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (S. 211-225). Bochum: AKS-Verlag.
- Grotjahn, R., Klein-Braley, C. & Raatz, U. (1992). C-Tests in der praktischen Anwendung. Erfahrungen beim Bundeswettbewerb Fremdsprachen. In R. Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (S. 263-296). Bochum: Brockmeyer.
- Grygier, P. (2005). Wissenschaftsverständnis - Schon in der Grundschule? In D. Cech & H. Giest (Hrsg.), *Sachunterricht in Praxis und Forschung - Erwartungen an die Didaktik des Sachunterrichts* (S. 177-189). Bad Heilbrunn: Klinkhardt
- Grygier, P. & Hartinger, A. (2009). *Gute Aufgaben Sachunterricht*. Berlin: Cornelsen Scriptor.
- Guadatiello, A. & Speck-Hamdan, A. (2013). Warum sollte Bildungssprache gefördert werden? In A. Guadatiello & A. Speck-Hamdan (Hrsg.), *Bildungssprache / Deutsch als Zweitsprache fördern. Beiträge zur Reform der Grundschule Band 135* (S. 9-12). Frankfurt a. M.: Grundschulverband e.V.
- Habermas, J. (1978). Umgangssprache, Wissenschaftssprache, Bildungssprache. *Merkur*, 32(359), 327-342.
- Halliday, M. A. K. & Hasan, R. (1985). *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*. Victoria: Deakin University.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar* (4th ed.). New York: Routledge.
- Hammond, J. (Hrsg.). (2001). *Scaffolding: Teaching and Learning in Language and Literacy Education*. Sydney: Primary English Teaching Association.
- Hammond, J. & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect*, 20(1), 6-30.

- Hartinger, A., Grygier, P., Tretter, T. & Ziegler, F. (2013). *Lernumgebungen zum naturwissenschaftlichen Experimentieren (Handreichungen des Programms SINUS an Grundschulen)*. Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN). Verfügbar unter http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Hartinger_et_al_fuer_web.pdf
- Hemmerich, W. (2016). StatistikGuru: Bonferroni-Korrektur. Zugriff am 27.02.2023 unter <https://statistikguru.de/lexikon/bonferroni-korrektur.html>
- Hergovich, A. (2021). *Allgemeine Psychologie: Denken und Lernen*. Stuttgart: UTB.
- Hiller, F., Kleinbub, I. & Nafz, S. (2017). „... oder Niederschlag könnte man auch sagen“ - Bildungssprachlichen Wortschatz fördern mit Erklärvideos. In S. Merten & K. Kuhs (Hrsg.), *Arbeiten am Wortschatz: Sprechen und Zuhören* (S. 55-78). Trier: WVT.
- Hoffmann, L. (2019). Alltagssprache. Zugriff am 27.02.2023 unter https://epub.ub.uni-muenchen.de/61747/1/Hoffmann_Alltagssprache.pdf
- Hoffmann, M. (2007). *Funktionale Varietäten des Deutschen - kurz gefasst*. Potsdam: Universitätsverlag.
- Imhof, M. (2012). *Psychologie für Lehramtsstudierende (3. Aufl.)*. Wiesbaden: Springer.
- Jeretin-Kopf, M. (2011). *Die Bedeutung der Sprache für die Konstruktion naturwissenschaftlicher Kognitionsmuster*. Berlin: Mensch & Buch Verlag.
- Jeretin-Kopf, M. (2014). Sprache als Lernmedium im Sachunterricht. In M. Michalak (Hrsg.), *Sprache als Lernmedium im Fachunterricht: Theorien und Modelle für das sprachbewusste Lehren und Lernen* (S. 72-90). Hohengehren: Schneider.
- Juska-Bacher, B. & Beckert, C. (2015). *Bildungssprache am Schulanfang: Theoretische Herausforderungen - empirische Erkenntnisse - Förderperspektiven*. Baltmannsweiler: Schneider.
- Kaltenbacher, E. & Klages, H. (2008). Deutsch für den Schulstart: Zielsetzungen und Aufbau eines Förderprogramms. In B. Ahrenholz (Hrsg.), *Deutsch als Zweitsprache: Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund* (S. 135-154). Freiburg: Fillibach.
- Kempert, S., Schalk, L. & Saalbach, H. (2018). Sprache als Werkzeug des Lernens: Ein Überblick zu den kommunikativen und kognitiven Funktionen der Sprache und deren Bedeutung für den fachlichen Wissenserwerb. *Psychologie in Erziehung und Unterricht*, 65, 1-20. [10.2378/PEU2018.art19d](https://doi.org/10.2378/PEU2018.art19d)
- Kniffka, G. & Neuer, B. (2008). Wo geht's hier nach ALDI? - Fachsprachen lernen im kulturell heterogenen Klassenzimmer. In A. Budke (Hrsg.), *Interkulturelles Lernen im Geographie-Unterricht* (S. 121-135). Potsdam: Universitätsverlag.
- Kniffka, G. & Roelcke, T. (2016). *Fachsprachenvermittlung im Unterricht*. Paderborn: Schöningh.
- Kniffka, G. & Siebert-Ott, G. (2012). *Deutsch als Zweitsprache: Lehren und lernen (3. Aufl.)*. Paderborn: Schöningh.
- Koch, P. & Österreicher, W. (1986). Sprache der Nähe - Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In O. Deutschmann, H. Flasche, B. König, M. Kruse, W.

- Pabst, & W.-D. Stempel (Hrsg.), *Romanistisches Jahrbuch* (Bd. 36, S. 15-43). Berlin: De Gruyter.
- Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476-490). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kosler, T. (2017). Naturwissenschaftliches Denken mit Kindern? Zur Diskussion um die Möglichkeit, Kinder im Elementar- und Primarbereich an naturwissenschaftliches Denken heranzuführen. *Widerstreit Sachunterricht*. Zugriff am 05.10.2021 unter <http://www.widerstreit-sachunterricht.de/ebeneI/superworte/naturwiss/kosler.pdf>
- Krepf, M. (2019). *Wie analysieren ExperInnen und NovizInnen Unterricht? Validierung des Konstrukts Analysekompetenz mittels Qualitativer Inhaltsanalyse*. Bad Heilbrunn: Klinkhardt.
- Kuckartz, U. (2014). *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*. Wiesbaden: Springer.
- Kuckartz, U. (2017). Datenanalyse in der Mixed-Methods-Forschung. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 69(2), 157-183. 10.1007/s11577-017-0456-z
- Kulgemeyer, C. (2010a). Physikalische Kommunikationskompetenz überprüfen. *Naturwissenschaften im Unterricht. Physik*, 21(116), 9-13.
- Kulgemeyer, C. (2010b). *Physikalische Kommunikationskompetenz. Modellierung und Diagnostik*. Berlin: Logos.
- Kulgemeyer, C. & Schecker, H. (2009). Kommunikationskompetenz in der Physik: Zur Entwicklung eines domänenspezifischen Kommunikationsbegriffs. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 131-153.
- Kulgemeyer, C. & Schecker, H. (2013). Schülerinnen und Schüler erklären Physik - Modellierung, Diagnostik und Förderung von Kommunikationskompetenz im Physikunterricht. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen*. Münster: Waxmann.
- Kultusministerkonferenz. (2019). Empfehlung: Bildungssprachliche Kompetenzen in der deutschen Sprache stärken. Zugriff am 27.02.2023 unter https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2019/2019_12_05-Beschluss-Bildungssprachl-Kompetenzen.pdf
- Landis, J. R. & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. 10.2307/2529310
- Lehmann, A., Pilz, A. & Sarich, T. (2013). Wortschatzarbeit im Deutschunterricht. *Sprachsensibler Fachunterricht: Handreichung zur Wortschatzarbeit in den Jahrgangsstufen 5-10 unter besonderer Berücksichtigung der Fachsprache*, 19-93. Verfügbar unter https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/themen/sprachbildung/Durchgaengige_Sprachbildung/Publicationen_sprachbildung/sprachsensibler_fachunterricht/1_Sprachsensibler_Fachunterricht-Vorwort.pdf
- Leisen, J. (2010). *Handbuch Sprachförderung im Fach: Sprachsensibler Fachunterricht in der Praxis*. Stuttgart: Klett.
- Leisen, J. (2012). Praktische Ansätze schulischer Sprachförderung: Der sprachensible Fachunterricht. In P. Bodensteiner & A. Zöllner (Hrsg.), *Begegnen, Verstehen,*

- Zukunft sichern. Beiträge der Schule zu einem gelungenen kulturellen Miteinander* (S. 29-42). München: Hanns Seidel Stiftung.
- Leisen, J. (2015). Fachlernen und Sprachlernen! Bringt zusammen, was zusammen gehört! *Der Mathematische und naturwissenschaftliche Unterricht (MNU)*, 3, 132-137.
- Leisen, J. (2023). Sprachbildung und Bildungssprache. Zugriff am 27.02.2023 unter <http://www.sprachsensiblerfachunterricht.de/sprachbildung>
- Leisen, J. (Hrsg.). (1999). *Methoden-Handbuch: Deutschsprachiger Fachunterricht (DFU)*. Bonn: Varus.
- Leufer, N. (2016). *Kontextwechsel als implizite Hürden realitätsbezogener Aufgaben: Eine soziologische Perspektive auf Texte und Kontexte nach Basil Bernstein*. Wiesbaden: Springer.
- Levelt, W. (1995). The ability to speak: from intentions to spoken words. *European Review*, 3(1), 13-23. Verfügbar unter https://www.mpi.nl/world/materials/publications/levelt/Levelt_The_Ability_to_speak_1995.pdf
- Loser, F. (2004). Sachunterricht als Sprachunterricht: Das exemplarische Lehren und sein Beitrag zu einer pädagogischen Theorie des Lehren und Lernens. *Widerstreit Sachunterricht*, 2, 1-17.
- Luchtenberg, S. (2011). Zur Rolle von Fachsprachen in Elementar- und Primarstufe. *Grundschule*, 43(10), 24-26.
- Maak, D. (2018). *Sprachliche Merkmale des fachlichen Inputs im Fachunterricht Biologie: Eine konzeptorientierte Analyse der Enkodierung von Bewegung*. Berlin: De Gruyter.
- Marquardt-Mau, B. (2011). Der Forschungskreislauf: Was bedeutet Forschen im Sachunterricht? In D. T. S. und & D. K.-u. Jugendstiftung (Hrsg.), *Wie gute naturwissenschaftliche Bildung an Grundschulen gelingt. Ergebnisse und Erfahrungen aus prima(r)forscher* (S. 32-37). Bonn: Deutsche Telekom Stiftung und Deutsche Kinder- und Jugendstiftung.
- Mayring, P. (2019). Qualitative Inhaltsanalyse - Abgrenzungen, Spielarten, Weiterentwicklungen. *Forum Qualitative Sozialforschung*, 20(3), 1-15. <https://doi.org/10.17169/fqs-20.3.3343>
- Merz, M. (2015). Bildungssprache im naturwissenschaftlichen Sachunterricht. In A. Wildemann & S. Fornol (Hrsg.), *Papers of Excellence - Ausgewählte Arbeiten aus den Fachdidaktiken* (Bd. 6, S. 153-179). Aachen: Shaker.
- Michell, M. & Sharpe, T. (2005). Instructional scaffolding in English as a second language classrooms. *Prospect*, 20(1), 31-58.
- Morek, M. & Heller, V. (2012). Bildungssprache - Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für angewandte Linguistik*, 57(1), 67-101. <https://doi.org/10.1515/zfal-2012-0011>
- NaWiT AS (2023). Salz - so wie Wissenschaftler aus Versuchen lernen. Zugriff am 01.08.2021 unter <https://nawitas.uni-koeln.de/unterricht/salz>
- Peschel, C. (2013). Schreiben in der Zweitsprache Deutsch - Orientierung an Textmustern im Fachunterricht. In C. Röhner & B. Hövelbrinks (Hrsg.), *Fachbezogene Sprachförderung in Deutsch als Zweitsprache: Theoretische Konzepte und empirische Befunde zum Erwerb bildungssprachlicher Kompetenzen*. Weinheim: Beltz Juventa.

- Pineker-Fischer, A. (2017). *Sprach- und Fachlernen im naturwissenschaftlichen Unterricht: Umgang von Lehrpersonen in soziokulturell heterogenen Klassen mit Bildungssprache*. Wiesbaden: Springer.
- Quehl, T. & Scheffler, U. (2008). Möglichkeiten fortlaufender Sprachförderung im Sachunterricht. In C. Bainski & M. Krüger-Potratz (Hrsg.), *Handbuch Sprachförderung* (S. 66-79). Essen: Neue Deutsche Schule.
- Quehl, T. & Trapp, U. (2013). *Sprachbildung im Sachunterricht der Grundschule. Mit dem Scaffolding-Konzept unterwegs zur Bildungssprache*. Münster: Waxmann.
- Rädiker, S. & Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Wiesbaden: Springer.
- Rank, A., Hartinger, A., Wildemann, A. & Tietze, S. (2018). Bildungssprachliche Kompetenzen bei Vorschulkindern mit Deutsch als Erst- und Zweitsprache. *Zeitschrift für Grundschulforschung*(11), 115-129.
<https://doi.org/10.1007/s42278-018-0009-1>
- Rank, A., Wildemann, A., Hartinger, A. & Tietze, S. (2018). Early Steps into Science and Literacy - EASI Science-L Teil 1: Wirkungen sprachlicher Anregungsqualität in naturwissenschaftlichen Bildungsangeboten auf die sprachlichen Fähigkeiten von Vorschulkindern. In Stiftung Haus der kleinen Forscher (Hrsg.), *Wirkungen naturwissenschaftlicher Bildungsangebote auf pädagogische Fachkräfte und Kinder* (S. 140-193). Opladen: Verlag Barbara Budrich.
- Rau-Patschke, S. (2019). (Fach-)sprachliche Unterstützungsmaßnahmen für inklusive Lerngruppen im Sachunterricht. In D. Pech, C. Schomaker, & T. Simon (Hrsg.), *Inklusion im Sachunterricht: Perspektiven der Forschung* (S. 159-168). Bad Heilbrunn: Klinkhardt.
- Razali, N. M. & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Analytics*, 2.
- Reber, K. & Schönauer-Schneider, W. (2016). Alltagsintegrierte Sprachförderung in der Grundschule: Schwerpunkt Wortschatz. In R. Sigel & E. Inckemann (Hrsg.), *Diagnose und Förderung von Kindern mit Zuwanderungshintergrund im Sprach- und Schriftspracherwerb: Theorien, Konzeptionen und Methoden in den Jahrgangsstufen 1 und 2 der Grundschule* (S. 59-73). Bad Heilbrunn: Klinkhardt.
- Rincke, K. (2010). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 235-260.
- Rincke, K. & Leisen, J. (2015). Sprache im Physikunterricht. In E. Kircher, R. Girwidz, & P. Häußler (Hrsg.), *Physikdidaktik: Theorie und Praxis* (3. Aufl.) (S. 635-656). Berlin: Springer.
- Rincke, K. & Markic, S. (2018). Sprache und das Lernen von Naturwissenschaften. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (S. 31-48). Berlin: Springer.
- Roelcke, T. (2019). Fachsprache. *Sprache im Fach*. Zugriff am 27.02.2023 unter https://epub.uni-muenchen.de/61967/1/Roelcke_Fachsprache.pdf
- Sandmann, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.),

- Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179-188). Berlin: Springer.
- Scheuer, R., Kleffken, B. & Ahlborn-Gockel, S. (2010). Experimentieren als neuer Weg der Sprachförderung: Verknüpfung naturwissenschaftlicher und sprachlicher Bildung. In D. Höttecke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik* (S. 248-250). Berlin: Lit Verlag.
- Schleppegrell, M. J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Mahwah: Lawrence Erlbaum.
- Schmeink, D., Dölle, S., Dreher, T., Oldenburg, I. & Lüpkes, J. (Hrsg.) (2017). Faszinierendes Salz. Schmecken, erforschen, verwenden. *Sachunterricht Weltwissen*, 3.
- Schmitt, A. (2017). Im Sachunterricht die Grenzen der Sprache einreißen. *Grundschule*, 8, 30-33.
- Schmölzer-Eibinger, S. (2013). Sprache als Medium des Lernens im Fach. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach. Sprachlichkeit und fachliches Lernen* (S. 25-40). Münster: Waxmann.
- Schnell, C. (2016). „Lautes Denken“ als qualitative Methode zur Untersuchung der Validität von Testitems: Erkenntnisse einer Studie zur Diagnose des ökonomischen Fachwissens von Schülerinnen und Schülern der Sekundarstufe I. *Zeitschrift für ökonomische Bildung*, 5, 26-49.
- Schreier, M. (2014). Varianten qualitativer Inhaltsanalyse: ein Wegweiser im Dickicht der Begrifflichkeiten. *Forum Qualitative Social Research*, 15.
- Schreier, M. & Odağ, Ö. (2020). Mixed Methods. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 1-26). Wiesbaden: Springer.
- Schroeter-Brauss, S., Wecker, V. & Henrici, L. (2018). *Sprache im naturwissenschaftlichen Unterricht*. Münster: Waxmann.
- Schuchardt, K. & Mähler, C. (2019). Arbeitsgedächtnis im Kindesalter im Dorsch Lexikon der Psychologie. Zugriff am 27.02.2023 unter <https://dorsch.hogrefe.com/stichwort/arbeitsgedaechtnis-im-kindesalter>
- Selimi, N. (2014). *Wortschatzarbeit konkret* (Bd. 2. korr. Aufl.). Baltmannsweiler: Schneider Verlag Hohengehren.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.
- Siarova, H., Sternadel, D. & Szönyi, E. (2019). Research for CULT Committee - Science and Scientific Literacy as an Educational Challenge. Zugriff am 27.02.2023 unter [http://www.europarl.europa.eu/RegData/etudes/STUD/2019/629188/IPOL_STU\(2019\)629188_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2019/629188/IPOL_STU(2019)629188_EN.pdf)
- Smirnova, E. & Mortelmans, T. (2010). *Funktionale Grammatik: Konzepte und Theorien*. Berlin: De Gruyter.
- Söll, L. & Hausmann, F. J. (1985). *Gesprochenes und geschriebenes Französisch* (3. Aufl.). Berlin: Schmidt.
- Spitzer, M. (2006). *Lernen: Gehirnforschung und die Schule des Lebens*. Heidelberg: Spektrum [Nachdruck 2014].

- Spreckelsen, K. (1997). Phänomenkreise als Verstehenshilfen. In W. Köhnlein, B. Marquardt-Mau, & M. Schreier (Hrsg.), *Vielperspektivisches Denken im Sachunterricht* (S. 111-127). Bad Heilbrunn: Klinkhardt.
- Steffensky, M., Kleickmann, T., Kasper, D. & Köller, O. (2016). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Hrsg.), *TIMSS 2015: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 137-188). Münster: Waxmann.
- Steffensky, M., Scholz, L. A., Kasper, D. & Köller, O. (2020). Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse. In K. Schwippert, D. Kasper, O. Köller, N. McElvany, C. Selter, M. Steffensky, & H. Wendt (Hrsg.), *TIMSS 2019: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 115-168). Münster: Waxmann.
- Steinmüller, U. & Scharnhorst, U. (1987). Sprache im Fachunterricht - Ein Beitrag zur Diskussion über Fachsprachen im Unterricht mit ausländischen Schülern. *Zielsprache Deutsch*, 18(4), 3-12.
- Sutton, S. S. & Williamson, J. K. (2017). MINT-spezifische Textsortenkompetenz: Ein Plädoyer und Unterrichtsmodul für Protokoll als Textsorte im fachsensiblen DaF-Unterricht. In E. Tschirner, J. Möhring, & K. Cothrun (Hrsg.), *Deutsch als zweite Bildungssprache in MINT-Fächern* (S. 75-102). Tübingen: Stauffenburg.
- Tajmel, T. (2010). DaZ-Förderung im naturwissenschaftlichen Fachunterricht. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (Bd. 2., durchgesehene und aktualisierte Auflage, S. 107-132). Tübingen: Narr.
- Tajmel, T. (2017a). Die Bedeutung von ‚Alltagssprache‘ – eine physikdidaktische Betrachtung. In B. Lütke, I. Petersen, & T. Tajmel (Hrsg.), *Fachintegrierte Sprachbildung: Forschung, Theoriebildung und Konzepte für die Unterrichtspraxis* (S. 253-267). Berlin: De Gruyter.
- Tajmel, T. (2017b). *Naturwissenschaftliche Bildung in der Migrationsgesellschaft: Grundzüge einer Reflexiven Physikdidaktik und kritisch-sprachbewussten Praxis*. Wiesbaden: Springer.
- Tajmel, T. & Hägi-Mead, S. (2017). *Sprachbewusste Unterrichtsplanung: Prinzipien, Methoden und Beispiele für die Umsetzung*. Münster: Waxmann.
- Tietze, S., Rank, A. & Wildemann, A. (2016). Erfassung bildungssprachlicher Kompetenzen von Kindern im Vorschulalter. Grundlagen und Entwicklung einer Ratingskala (RaBi). 42 S. Verfügbar unter urn:nbn:de:0111-pedocs-120766
- Trautmann, C. & Reich, H. H. (2008). Pragmatische Basisqualifikationen I und II. In K. Ehlich, U. Bredel, & H. H. Reich (Hrsg.), *Referenzrahmen zur altersspezifischen Sprachaneignung* (S. 41-48). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Ufer, S., Reiss, K. & Mehringer, V. (2013). Sprachstand, soziale Herkunft und Bilingualität: Effekte auf Facetten mathematischer Kompetenz. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 185-202). Münster: Waxmann.

- Ulrich, W. (2013). Wissenschaftliche Grundlagen der Wortschatzarbeit im Fachunterricht. *Sprachsensibler Fachunterricht: Handreichung zur Wortschatzarbeit in den Jahrgangsstufen 5-10 unter besonderer Berücksichtigung der Fachsprache*, 307-330. Verfügbar unter https://bildungsserver.berlin-brandenburg.de/fileadmin/bbb/themen/sprachbildung/Durchgaengige_Sprachbildung/Publicationen_sprachbildung/sprachsensibler_fachunterricht/1_Sprachsensibler_Fachunterricht-Vorwort.pdf
- van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review*, 22, 271-296. 10.1007/s10648-010-9127-6
- Vasylyeva, T. & Kurtz, G. (2015). Das „Robuste Wortschatztraining“ im schulischen Spracherwerb: Theoretische Begründung und Beispiele aus der Praxis. In H. Klages & G. Pagonis (Hrsg.), *Linguistisch fundierte Sprachförderung und Sprachdidaktik: Grundlagen, Konzepte, Desiderate* (S. 237-264). Berlin: De Gruyter.
- VERBI Software. (2020). MAXQDA 2020 [Software für qualitative Datenanalyse]. Berlin: VERBI Software.
- Verboom, L. (2017). WEGE durch den Sprachförderschlingel - Strukturierung des Fachwortschatz-Lernprozesses. In A. S. Steinweg (Hrsg.), *Mathematikdidaktik Grundschule* (Bd. 7, S. 57-72). Bamberg: University of Bamberg Press.
- Vollmer, H. J. & Thürmann, E. (2010). Zur Sprachlichkeit des Fachlernens: Modellierung eines Referenzrahmens für Deutsch als Zweitsprache. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (2. Aufl.) (S. 107-132). Tübingen: Narr Francke Attempto.
- Von Ossietzky, C. (o. D.). Überblickswissen: Englische Fachdidaktik. Zugriff am 27.02.2023 unter https://uol.de/f/3/inst/anglistik/personen/wolfgang.gehring/download/Ueberblick_Fachdidaktik.pdf
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Weis, I. (2013). *DaZ im Fachunterricht: Sprachbarrieren überwinden - Schüler erreichen und fördern*. Mülheim an der Ruhr: Verlag an der Ruhr.
- Wildemann, A. & Fornol, S. (2016). *Sprachsensibel unterrichten in der Grundschule: Anregungen für den Deutsch-, Mathematik- und Sachunterricht*. Seelze: Klett, Kallmeyer.
- Wildemann, A., Rank, A., Hartinger, A. & Sutter, S. (2016). Bildungssprache im Kontext kindlicher Entwicklung: Eine Studie zur Erfassung früher bildungssprachlicher Fähigkeiten. *DDS - Die Deutsche Schule, Beiheft 13*, 67-83.
- Windt, A., Scheuer, R. & Melle, I. (2010). Naturwissenschaftliches Experimentieren im Elementarbereich - Evaluation verschiedener Methoden. In D. Höttecke (Hrsg.), *Entwicklung naturwissenschaftlichen Denkens zwischen Phänomen und Systematik* (S. 224-226). Berlin: Lit Verlag.
- Wood, D., Bruner, J. & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

8 ANHANG

8.1 Anhang A: Test-Stunden

8.1.1 Anhang A.1: C-Test



Mein Name: _____

Meine Klasse: _____

Da hat wohl jemand Buchstaben geklaut...

Aufgabe:

In den 5 kurzen Texten fehlen bei einigen Wörtern manche Buchstaben.

Schreibe die fehlenden Buchstaben auf die Linie.

Du hast 5 Minuten Zeit für jeden Text.

1) Eine Schule in Australien

Joanne ist in Australien auf einer Farm zu Hause. Ihre
näch_____ ¹⁾ Nachbarn woh_____ ²⁾ 120 km we_____ ³⁾ weg.

Ih_____ ⁴⁾ Schule i_____ ⁵⁾ noch wei_____ ⁶⁾ entfernt. Do_____ ⁷⁾ kann

s_____ ⁸⁾ nicht tag_____ ⁹⁾ hin, des_____ ¹⁰⁾ wird s_____ ¹¹⁾ am

Comp_____ ¹²⁾ unterrichtet. Im_____ ¹³⁾ zur glei_____ ¹⁴⁾ Zeit

mu_____ ¹⁵⁾ sie si_____ ¹⁶⁾ bei ih_____ ¹⁷⁾ Lehrer mel_____ ¹⁸⁾ und

bek_____ ¹⁹⁾ ihre Aufg_____ ²⁰⁾. Ihre Mutter hilft ihr, wenn sie Fragen hat

und passt auf, dass sie ihre Schulzeit einhält.

2) Die Feuerwehr

Herr Weiß ist Hausmeister in einer Schule und arbeitet für die Feuerwehr. Für d____¹⁾ Feuerwehr arbe____²⁾ er ab____³⁾ ehrenamtlich, d____⁴⁾ heißt, e____⁵⁾ bekommt ke____⁶⁾ Geld da____⁷⁾. Er i____⁸⁾ Mitglied ei____⁹⁾ freiwilligen Feue____¹⁰⁾. In al____¹¹⁾ Gemeinden u____¹²⁾ kleinen Stä____¹³⁾ gibt e____¹⁴⁾ eine sol____¹⁵⁾ Feuerwehr. D____¹⁶⁾ Feuerwehr h____¹⁷⁾ viele schwi____¹⁸⁾ Aufgaben. Da____¹⁹⁾ müssen d____²⁰⁾ Feuerwehrleute in ihrer Freizeit immer wieder üben. Mitglied in der Feuerwehr zu sein, bedeutet viel Arbeit.

3) Die Brüder Wright

Die Brüder Wright waren schon als Kinder begeistert vom Fliegen. Die Brü____¹⁾ setzten zunä____²⁾ Fahrräder zusa____³⁾. Doch da____⁴⁾ begannen s____⁵⁾, Segelflugzeuge z____⁶⁾ entwickeln. M____⁷⁾ den Segelfl____⁸⁾ konnten s____⁹⁾ in d____¹⁰⁾ Luft seg____¹¹⁾. Doch d____¹²⁾ reichte d____¹³⁾ Brüdern ni____¹⁴⁾. Sie wol____¹⁵⁾ ein Flug____¹⁶⁾ bauen, d____¹⁷⁾ vom Bo____¹⁸⁾ abheben kon____¹⁹⁾. Am 17. Deze____²⁰⁾ 1903 war es dann so weit. An diesem Tag flog das erste Flugzeug der Brüder Wright.

4) Wasser ist lebensnotwendig

Alle Menschen, Tiere und Pflanzen brauchen Wasser zum Leben. Ein Mensch _____¹⁾ kann im _____²⁾ Notfall ein _____³⁾ Monat lang _____⁴⁾ nichts essen _____⁵⁾, aber ohne _____⁶⁾ Wasser im _____⁷⁾ er schließlich _____⁸⁾ nach wenigen _____⁹⁾ Tagen verdu _____¹⁰⁾. Für _____¹¹⁾ ist _____¹²⁾ selbstverständlich, da _____¹³⁾ immer saubere _____¹⁴⁾ Wasser aus _____¹⁵⁾ der Leitung _____¹⁶⁾ kommt. Des _____¹⁷⁾ vergessen wir _____¹⁸⁾ leicht, wie _____¹⁹⁾ große Bedeutung _____²⁰⁾ Wasser für den Menschen und alles Leben auf der Erde hat. Wir müssen also sorgfältig mit unserem Wasser umgehen.



5) Wir bauen ein Schatzsuchgerät

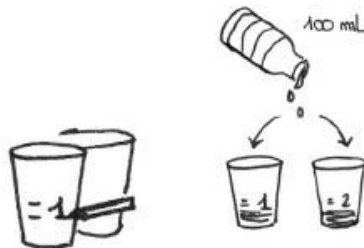
So gehst du vor. Wähle ein _____¹⁾ Forscherfrage aus _____²⁾. Vermute zu _____³⁾, was die _____⁴⁾ Antwort sein _____⁵⁾ könnte. Über _____⁶⁾ dir dann _____⁷⁾, wie du _____⁸⁾ in _____⁹⁾ Versuch die _____¹⁰⁾ Lösung _____¹¹⁾ die Frage _____¹²⁾ herausfinden _____¹³⁾. Hole die _____¹⁴⁾ alle notwendigen _____¹⁵⁾ Materialien für _____¹⁶⁾ deinen Versuch _____¹⁷⁾. Probiere die _____¹⁸⁾ Ideen aus _____¹⁹⁾. Kannst du _____²⁰⁾ damit die Forscherfrage beantworten? Wenn nicht, überlege dir einen anderen Versuch. Entwirf einen Forscherbogen. Beschreibe dein Vorgehen und die Ergebnisse deines Versuches.

8.1.2 Anhang A.2: Versuchsanweisung im Prä-Test

Was passiert mit dem Salz?

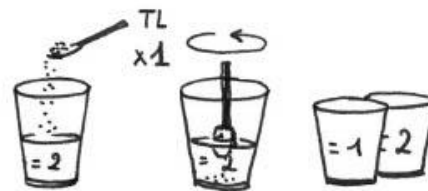
Los geht's:

1. Markiere einen Messbecher mit 1 und den anderen Messbecher mit 2.
Gieße 100 ml Wasser in beide Messbecher.



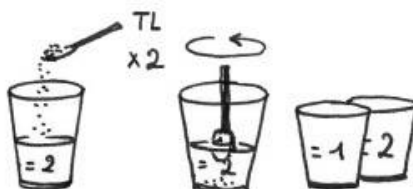
2. Versuch 1:

Schütte einen Teelöffel Salz in deinen Messbecher 2.
Rühre um, Zähle dabei bis 30.
Beobachte die Flüssigkeit.
Vergleiche den Messbecher 2 mit Messbecher 1.



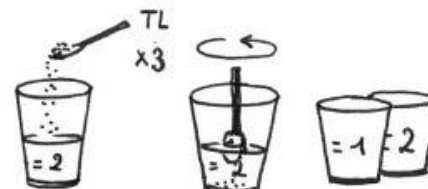
3. Versuch 2:

Trockne den Löffel gut ab.
Schütte zwei weitere Teelöffel Salz in deinen Messbecher 2.
Rühre um, Zähle dabei bis 30.
Beobachte die Flüssigkeit.
Vergleiche den Messbecher 2 mit Messbecher 1.



4. Versuch 3:

Trockne den Löffel gut ab.
Schütte drei weitere Teelöffel Salz in deinen Messbecher 2.
Rühre um, Zähle dabei bis 30.
Beobachte die Flüssigkeit.
Vergleiche den Messbecher 2 mit Messbecher 1.

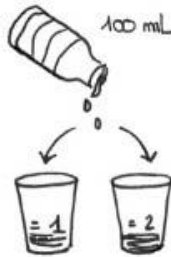


8.1.3 Anhang A.3: Versuchsanweisung im Post-Test

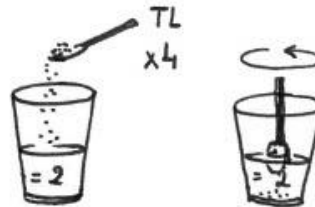
Was passiert mit dem Ei?

Los geht's:

1. Markiere einen Messbecher mit 1 und den anderen Messbecher mit 2. Gieße 100 ml Wasser in beide Messbecher.



2. Schütte 4 Teelöffel Salz in deinen Messbecher 2. Rühre so lange um, bis sich das Salz im Wasser gelöst hat.



3. Versuch 1:
Lege vorsichtig mithilfe des Teelöffels ein Ei in Messbecher 1. Beobachte, in welcher Höhe sich das Ei befindet.



4. Versuch 2:
Lege das Ei danach in Messbecher 2. Beobachte, in welcher Höhe sich das Ei befindet.



5. Versuch 3:
Gieße vorsichtig etwas Leitungswasser aus der Flasche in den Messbecher 2, bis du auf 200 ml kommst. Beobachte, in welcher Höhe sich das Ei befindet.

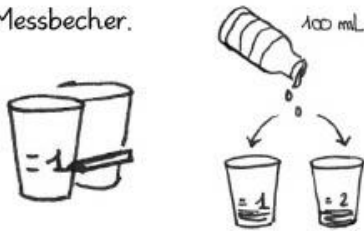


8.1.4 Anhang A.4: Versuchsanweisung im Follow-Up-Test

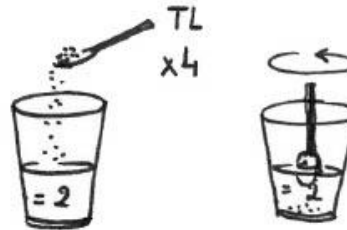
Was passiert mit den Salatblättern?

Los geht's:

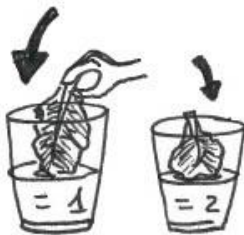
1. Markiere einen Messbecher mit 1 und den anderen Messbecher mit 2.
Gieße 100 ml Wasser in beide Messbecher.



2. Schütte 4 Teelöffel Salz in deinen Messbecher 2.
Rühre so lange um, bis sich das Salz im Wasser gelöst hat.



3. Versuch 1 und 2:
Lege ein Salatblatt in den Messbecher 1.
Lege ein zweites Salatblatt mit ähnlicher Größe in den Messbecher 2.



4. Versuch 1 und 2:
Warte.
Ertaste zwischendurch, wie sich die zwei Salatblätter anfühlen.



5. Versuch 3:
Nehme das Salatblatt aus Messbecher 1 heraus.
Lege das Salatblatt aus Messbecher 2 in deinen Messbecher 1.
Ertaste zwischendurch, wie sich das Salatblatt anfühlt.



8.1.5 Anhang A.5: Vorlage für ein Versuchsprotokoll im Prä-Test



Das Versuchsprotokoll Teil 1

2

Mein Name:

Meine Klasse:



1) Die Frage: Was passiert mit dem Salz?

2) Meine Vermutungen:



Versuch 1 - Wasser mit 1 Teelöffel Salz:

Versuch 2 - Wasser mit 3 Teelöffeln Salz:

Versuch 3 - Wasser mit 6 Teelöffeln Salz:

3) Das Material:

4) Die Durchführung:










ACHTUNG: Führe jetzt den Versuch durch.
Schreibe deine Beobachtung und deine Erklärung auf die Rückseite!



Teil 1

8.1.6 Anhang A.6: Impressionen aus der Unterrichtssequenz

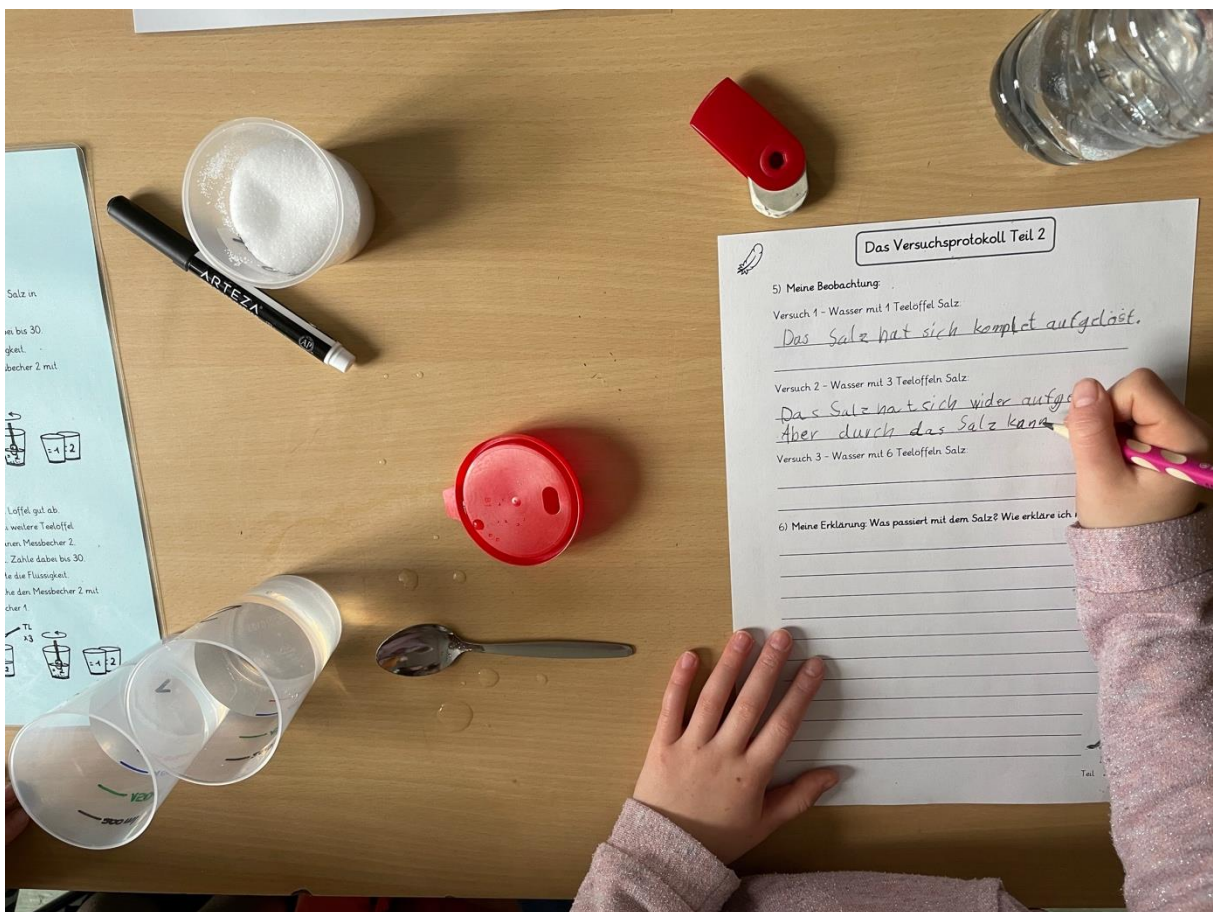
Forschungsprotokoll

<p>1 Die Frage ?</p> <ul style="list-style-type: none"> ➤ Was ...? ➤ Wie ...? ➤ Wieso ...? ➤ Weshalb ...? ➤ Warum ...? 	<p>2 Die Vermutung </p> <ul style="list-style-type: none"> ➤ Meine Vermutung ist, dass ... ➤ Ich vermute, dass ... ➤ Ich denke, dass ... ➤ ..., denn ... 	<p>3 Das Material </p> <ul style="list-style-type: none"> ➤ Man benötigt: ... 
<p>4 Die Durchführung </p> <ul style="list-style-type: none"> ➤ Zunächst ... ➤ Danach ... ➤ Daraufhin ... ➤ Anschließend ... ➤ Schließlich ... 	<p>5 Die Beobachtung </p> <ul style="list-style-type: none"> ➤ Man kann beobachten, dass ... ➤ Es lässt sich feststellen, dass ... ➤ Wenn ..., dann ... ➤ Nach kurzer Zeit ... ➤ ... unmittelbar ... 	<p>6 Die Erklärung !</p>

Der Wortspeicher wurde in der dritten Treatment-Stunde bezogen auf den Abschnitt *Beobachtung* ergänzt.



Nutzung der Tafel in jeder Stunde. Der Ablauf wurde an der linken Tafelseite festgehalten, während an der rechten Tafelhälfte die Abschnitte des Versuchsprotokolls wiederholt wurden. Die Eselsbrücke FVMD-BE wurde in allen drei Gruppen als Rap eingeführt, um mithilfe der Anfangsbuchstaben (**F**rage, **V**ermutung, **M**aterial ...) die Abschnitte des Versuchsprotokolls und deren Reihenfolge zu verinnerlichen.



Eine Vpn füllt den zweiten Teil des Versuchsprotokolls – die *Beobachtung* und *Erklärung* – aus.

8.2 Anhang B: Kategoriensysteme

8.2.1 Anhang B.1: Kategoriensystem – Bildungssprache

Sprachhandlungen	Die Sprachhandlungen sind eng mit den im Unterricht verwendeten Operatoren – wie beispielsweise <i>Benenne</i> , <i>Beschreibe</i> , <i>Erkläre</i> – verknüpft. Da das Aufstellen einer Vermutung sowie das Erklären und Begründen anspruchsvolle Sprachhandlungen darstellen und diese oft mit bildungssprachlichen Mitteln einhergehen (Tietze et al., 2016, S. 13), wird das Vorkommen dieser zwei Sprachhandlungstypen in den Versuchsprotokollen genauer untersucht und entsprechend kodiert. Je nach Kontext werden Wörter, Phrasen oder ganze Sätze kodiert. Wichtig ist, dass die Sprachhandlung in den Kodierungen erkennbar ist.		
Unterkategorien	Zuordnung	Ankerbeispiele	Abgrenzung
<i>Vermuten</i>	<p>Diese Kategorie ist zu kodieren, wenn in den Versuchsprotokollen eine sprachlich ersichtliche Vermutung aufgestellt wird.</p> <p>Die Formulierung ist sprachlich korrekt und wird im Kontext richtig angewandt.</p> <p>Explizit ausgedrückte Vermutungen werden auch außerhalb der Rubrik <i>Vermutungen</i> im Versuchsprotokoll markiert.</p>	<ul style="list-style-type: none"> ✓ Ich glaube, die Lampe wird leuchten. ✓ Die Lampe könnte leuchten. ✓ Ich vermute die Lampe leuchtet. 	<p>Die folgenden Formulierungen aus dem Wortspeicher werden <u>nicht</u> kodiert:</p> <ul style="list-style-type: none"> ✗ Meine Vermutung ist, dass... ✗ Ich vermute, dass... ✗ Ich denke, dass... <p>Das folgende Wort drückt nur unzureichend eine Vermutung aus und wird deshalb nicht kodiert:</p> <ul style="list-style-type: none"> ✗ würde

<p><i>Erklären / Begründen</i></p>	<p>Diese Kategorie ist zu kodieren, wenn in den Versuchsprotokollen eine Erklärung oder Begründung gegeben wird, die sprachlich zu erkennen ist. Darunter zählen Kausalsätze.</p> <p>Die Formulierung ist sprachlich korrekt und wird im Kontext richtig angewandt.</p> <p>Explizit ausgedrückte Begründungen werden auch außerhalb der Rubrik <i>Erklärung</i> im Versuchsprotokoll markiert.</p>	<ul style="list-style-type: none"> ✓ Wegen dem Salz geht die Lampe an. ✓ ... darum ... <p>Weitere Wörter, die oftmals mit einer Begründung einhergehen:</p> <ul style="list-style-type: none"> ✓ weil, wegen, so, meine Erklärung..., deswegen..., durch... ✓ Die Phrase <i>Wenn...</i> wird lediglich kodiert, falls diese im Kontext eine Erklärung darstellt. 	<p>Die folgenden Formulierungen aus dem Wortspeicher werden <u>nicht</u> kodiert:</p> <ul style="list-style-type: none"> ✗ Der Versuch lässt sich wie folgt erklären: ... ✗ Das liegt daran, dass ... ✗ Dadurch, dass ... ✗ ..., deshalb... <p>Die folgenden Wörter dienen meist nur unzureichend einer Erklärung:</p> <ul style="list-style-type: none"> ✗ damit ✗ je mehr..., desto... ✗ wenn...
<p><i>Sonstiges</i></p>	<p>Diese Kategorie ist zu kodieren, wenn in den Versuchsprotokollen weitere Sprachhandlungen erkennbar sind, die offensichtlich mit Bildungssprache einhergehen. Zur weiteren Differenzierung ist an dieser Stelle beispielsweise die Einleitung einer Materialliste (<i>Man braucht...</i>, <i>Ich benutze...</i>), die Einleitung einer Durchführung (<i>So wird der Versuch durchgeführt: ...</i>) und die Einleitung einer Beobachtung (<i>Ich beobachte, dass...</i>) zu kodieren.</p>		

Lexikon	In dieser Kategorie werden einzelne Wörter kodiert, die im Ermessen der kodierenden Person als bildungssprachlich gelten. Die spezifischen Unterkategorien Nomen, Verben und Adjektive mit ihren jeweiligen Zuordnungen sind einer vorangegangenen Studie zur Bildungssprache in der Vorschule entnommen (Tietze et al., 2016) und dienen der Unterbindung willkürlicher Kodierungen.		
Unterkategorien	Zuordnung	Ankerbeispiele	Abgrenzung
<i>Nomen</i>	Diese Kategorie ist bei Nomen mit hohem Abstraktionsgrad sowie differenzierenden Nomen zu kodieren. Letztere zeichnen sich durch einen exakten bildungssprachlichen Gebrauch oder einer Doppelbedeutung (Bildungssprache / Alltagssprache) aus. Bildungssprachlich genutzte Komposita sind ebenfalls zu kodieren.	<ul style="list-style-type: none"> ✓ Oberfläche ✓ Wasserspiegel 	Ist ein Nomen bereits in der Oberkategorie Sprachhandlung kodiert (Bsp.: Die <u>Begründung</u> ist...), so wird das Wort an dieser Stelle <u>nicht doppelt</u> kodiert! Begriffe aus der Versuchsanleitung (Bsp.: Messbecher) werden ebenfalls <u>nicht</u> kodiert!
<i>Verben</i>	Diese Kategorie ist bei Verben mit hohem Abstraktionsgrad, differenzierenden Verben, bildungssprachlich genutzten Verbkomposita sowie komplexen Präfixverben zu kodieren.	<ul style="list-style-type: none"> ✓ erzeugt ✓ befindet sich 	Das folgende bildungssprachlich einzuordnende Verb aus dem Wortspeicher wird <u>nicht</u> kodiert: ✗ feststellen
<i>Adjektive</i>	Diese Kategorie ist bei Adjektiven mit hohem Abstraktionsgrad und bei differenzierenden Adjektiven zu kodieren. Des Weiteren werden Partizipien in der Funktion eines Adjektivs kodiert.	<ul style="list-style-type: none"> ✓ ungefähr ✓ bestimmten Stoff 	
<i>Sonstiges</i>	Diese Kategorie ist zu kodieren, wenn ein einzelnes Wort mit hohem Abstraktionsgrad und / oder einer exakten Bedeutungszuschreibung (hohe Differenzierung) keinem der vorliegenden Unterkategorien zuzuordnen ist.		

Morphosyntax	Die Kategorie der Morphosyntax geht über die reine Wortebene (Lexikon) hinaus und verknüpft den Bereich der Morphologie mit dem Bereich der Syntax. Es geht dementsprechend um Kodierungen auf Satz- und/oder Textebene. Im Gegensatz zur Kategorie Lexikon, werden in dieser Kategorie deshalb Phrasen kodiert, welche der Bildungssprache zugeordnet werden. Darunter fallen vor allem Phrasen, in denen eine Kohäsion erkennbar ist, Sätze mit komplexem Satzgefüge sowie unpersönliche Konstruktionen. Ebenso wie die zwei vorhergehenden Oberkategorien, ist diese Kategorie mitsamt ihrer Unterkategorien entnommen aus Tietze et al. (2016). Komplexe Verbgefüge als weitere Unterkategorie werden aufgrund doppelter Kodierungen nicht mit aufgenommen. Die Zuordnungen wurden entsprechend des vorliegenden Untersuchungsinteresses angepasst.		
Unterkategorien	Zuordnung	Ankerbeispiele	Abgrenzung
<i>Kohäsion</i>	<p>„Das Kind äußert sich zusammenhängend und schlüssig [...]“ (Tietze et al., 2016, S. Anhang). Diese Kategorie ist zu kodieren, wenn Sätze oder Satzteile durch Sprache besonders miteinander verknüpft sind, z.B. durch die Nutzung von Temporaladverbien.</p> <p>Die Verknüpfung ist im Kontext sinnvoll und wurde erfolgreich gebildet. Kleinere Fehler im Satz sind ausgenommen, solange der Gesamtkontext klar erkennbar ist.</p> <p><u>Ausnahmen:</u> Kohäsion ist ein sehr weit gefasster Begriff. Die Verwendung von Rekurrenz, Textdeixis oder die Herstellung von Kohäsion durch passende</p>	<ul style="list-style-type: none"> ✓ Am Anfang... ✓ und als letztes... ✓ zur Hälfte im Wasser und zur anderen Hälfte an der Luft... <p>Beispielhafte Temporaladverbien sind:</p> <ul style="list-style-type: none"> ✓ Zuerst, erst, dann, nun, vorher, nachher, zuletzt, irgendwann... 	<p>Die folgenden Formulierungen aus dem Wortspeicher werden <u>nicht</u> kodiert:</p> <ul style="list-style-type: none"> ✗ Zunächst... / Danach... / Daraufhin... / Anschließend... / Schließlich... ✗ Das liegt daran, dass... ✗ Dadurch, dass... ✗ ..., deshalb ... <p>Beispiel für eine sprachliche Verknüpfung, die bereits in der Versuchsanleitung zum Ausdruck kommt und deshalb nicht kodiert wird:</p> <ul style="list-style-type: none"> ✗ Markiere <i>einen Messbecher</i> mit 1 und <i>den anderen Messbecher</i> mit 2.

	<p>Tempusformen und/oder Pronomen sind aufgrund ihres hohen Auftretens für die Kodierung obsolet. Da Konjunktionalsätze und Nebensätze mit Relativpronomen bereits in der Unterkategorie <i>Satzgefüge</i> aufgenommen sind, werden auch diese - trotz zusammenhängender Textstruktur - an dieser Stelle nicht kodiert.</p>		
<i>Satzgefüge</i>	<p>Diese Kategorie ist zu kodieren bei grammatikalisch anspruchsvolleren Satzstrukturen, zu denen Konjunktionalsätze und Relativsätze zählen.</p> <p>Die Satzstruktur ist im Kontext sinnvoll und wurde erfolgreich gebildet. Kleinere Fehler im Satz sind ausgenommen, solange der Gesamtkontext klar erkennbar ist.</p> <p><u>Ausnahmen:</u> Erweiterte Infinitive (zu + Infinitiv), sowie Konjunktionalsätze mit <i>und</i>, werden aufgrund fehlender Komplexität nicht kodiert. Weiterhin werden keine Satzteile kodiert, die bereits unter Kohäsion oder unter Sprachhandlung aufgenommen sind. Ist eine Sprachhandlung unter <i>Sonstiges</i></p>	<ul style="list-style-type: none"> ✓ Die Lampe leuchtet nicht, wenn... ✓ Die Blätter, die nicht im Wasser sind, kleben nicht... ✓ Wenn...,... <p>Beispielhafte Konjunktionen sind:</p> <ul style="list-style-type: none"> ✓ aber, allein, denn, doch, nur, oder, sondern, sowohl, als, desto, umso, wie, wenn, zu, (an)statt, ohne... 	<p>Die Formulierungen aus dem Wortspeicher werden <u>nicht</u> kodiert:</p> <ul style="list-style-type: none"> ✗ Meine Vermutung ist, dass... ✗ Ich vermute, dass... ✗ Ich denke, dass... ✗ Man kann beobachten, dass... ✗ Es lässt sich feststellen, dass... ✗ Wenn..., dann... ✗ Das liegt daran, dass... ✗ Dadurch, dass... <p>Beispiel für umgangssprachliche Satzstrukturen:</p> <ul style="list-style-type: none"> ✗ Das, was mir gegeben wurde. ✗ Beobachte, was passiert. <p>Beispiel für eine nicht korrekte Bildung:</p> <ul style="list-style-type: none"> ✗ In dem [Messbecher], der mit dem Salz war, ...

	erfasst, darf zusätzlich das Satzgefüge ermittelt werden. Dies ist damit zu begründen, dass die Kategorie <i>Sonstiges</i> bei Sprachhandlungen nicht in der Hauptauswertung integriert ist und lediglich als zusätzliche Darstellung dient.		Beispiel für Satzgefüge, die bereits in der Versuchsanleitung stehen und somit nicht kodiert werden: <ul style="list-style-type: none"> ✗ ..., bis du auf ... ml kommst.
<i>Unpersönliche Konstruktion</i>	Diese Kategorie ist zu kodieren bei der Verwendung des Vorgangspassivs oder der Verwendung des Zustandspassivs. Die kodierte Konstruktion ist im Kontext sinnvoll und wurde erfolgreich gebildet. Kleinere Fehler im Satz sind ausgenommen, solange der Gesamtkontext klar erkennbar ist.	<ul style="list-style-type: none"> ✓ ...wird mit einem Nagel verbunden... ✓ wird aufgelöst ✓ ist aufgelöst 	<ul style="list-style-type: none"> ✗ wird sich auflösen (Futur I) ✗ hat sich aufgelöst (Perfekt) ✗ ist dunkel geworden (Perfekt) ✗ ist gestiegen (Perfekt) ✗ ist verschwunden (Perfekt) ✗ ist hochgekommen (Perfekt) ✗ ist geblieben (Perfekt)
<i>Sonstiges</i>	Diese Kategorie ist zu kodieren, wenn eine komplexe Satzstruktur keinem der vorliegenden Unterkategorien zuzuordnen ist.		

8.2.2 Anhang B.2: Kategoriensystem – Inhalt

Inhalt	Der Inhalt dient der evaluativen Ermittlung, inwiefern die Kinder <i>korrekte</i> , <i>unzureichende</i> oder <i>fehlerhafte</i> Antworten in ihren Versuchsprotokollen geben. Die Beschränkung auf die genannten drei Unterkategorien ist darauf zurückzuführen, dass ein einfach gehaltenes Beurteilungssystem eine gute Vergleichbarkeit der Versuchsprotokolle ermöglicht.	
Unterkategorien	Zuordnung	Ankerbeispiele
<i>Anzahl der Materialien</i>	<p>Zur inhaltlichen Beurteilung der Materialangabe, wird die Anzahl der von den Kindern genannten Materialien kodiert. Zu beachten sind folgende Leitlinien:</p> <ol style="list-style-type: none"> 1. Die Anzahl eines Materials (z.B. 2 Messbecher) ist für die kodierte Anzahl der Materialien irrelevant (hier: Messbecher = 1). 2. Flasche & Wasser werden wie <i>Wasserflasche</i> als ein einzelnes Material gewertet. 3. Süßwasser & Salzwasser werden wie <i>Wasser</i> oder <i>Wasserflasche</i> als ein einzelnes Material gewertet. 4. Abkürzungen (z.B. TL) werden gewertet. 	<p><u>Material:</u></p> <ul style="list-style-type: none"> ✓ Ei, Flasche, Wasser, Salz → 3 ✓ Ei, TL, Süßwasser, Salzwasser, Salz, 2 Messbecher → 5 ✓ Ei, TL, Becher voller Salz → 3

<i>korrekt</i>	Vermutung	
	<p>Diese Kategorie ist bei einer Vermutung mit Aufgabenbezug zu kodieren.</p> <p>Wird das Personalpronomen <i>es</i> stellvertretend für das in der Aufgabenstellung genannte Hauptmaterial verwendet und die Vermutung ansonsten aufgabenbezogen aufgeführt, so wird diese Vermutung als korrekt kodiert.</p>	<ul style="list-style-type: none"> ✓ Das Ei wird in MB1 sinken. ✓ Es wird sinken. ✓ Das Salz wird unsichtbar und das Blatt wird weich.
	Durchführung	
	<p>Die Beschreibung der Versuchsdurchführung ist vollständig und grenzt sich überwiegend im Wortlaut von der Versuchsanleitung ab.</p>	<ul style="list-style-type: none"> ✓ Ich nehme einen Messbecher und gieße 100 ml Wasser rein. Dann nehme ich einen Teelöffel und probiere erst einen Teelöffel Salz und dann drei Teelöffel Salz und dann sechs Teelöffel Salz.
	Beobachtung	
	<p>Diese Kategorie ist bei einer Beobachtung mit Aufgabenbezug zu kodieren. Sie ist korrekt und passend zum Phänomen. Bezüglich des Prä-Tests ist die Beobachtung auch dann korrekt, wenn der veränderte Wasserspiegel aufgegriffen wird. Es geht in dieser Kategorie entsprechend um eine allgemeine Beobachtung, die später in der Erklärung hinsichtlich der Fragestellung differenzierter dargestellt wird.</p>	<ul style="list-style-type: none"> ✓ Das Ei befindet sich in mittlerer Höhe. ✓ Das Blatt in Mb2 fühlt sich weich an.

Erklärung	
<p>Die Erklärung passt zu der Beobachtung. Richtige fachliche Zusammenhänge müssen der Erklärung zugrunde liegen.</p> <p>Es genügt ein einfacher, aber erklärender Zusammenhang.</p> <p>Bezogen auf den Prä-Test ist das Auflösen des Salzes bereits eine passende Erklärung.</p> <p>Bezogen auf den Post-Test ist die Aussage, dass das Ei durch das Salz im Wasser getragen wird, bereits als korrekt zu kodieren.</p>	<ul style="list-style-type: none"> ✓ Das Salz löst sich bei wenig Salz auf und später nicht. ✓ Das Ei geht im Salzwasser nicht unter, weil das Salz das Ei trägt. Deswegen geht es nicht unter, sondern es schwimmt.

<i>unzureichend</i>	Vermutung	
	Diese Kategorie ist bei einer Vermutung ohne Aufgabenbezug zu kodieren. Ebenso werden Vermutungen kodiert, die teils eine Vermutung und teils eine andere Kategorie des Versuchsprotokolls ausdrücken. Sind Wörter nicht lesbar, aber eine Vermutung ist erkennbar, so zählt dies ebenso in diese Kategorie.	<ul style="list-style-type: none"> ✓ Das Ei wird platzen. ✓ Explodieren.
	Durchführung	
	Die Beschreibung der Versuchsdurchführung ist nicht vollständig und grenzt sich nur teilweise im Wortlaut von der Versuchsanleitung ab.	<ul style="list-style-type: none"> ✓ Erst malt man 1 und 2 auf Messbecher. Dann macht man 100 ml Wasser in die Messbecher. Dann 1 oder 2 Teelöffel Salz in die Messbecher.
	Beobachtung	
Diese Kategorie ist bei einer plausiblen Beobachtung mit Aufgabenbezug zu kodieren, die jedoch nicht vollständig ist oder den Bezug zum Gesehenen nicht eindeutig aufzeigt (z.B. durch die Verwendung von Passepartout-Wörtern). Werden falsche und richtige Antworten kombiniert, so ist die Antwort insgesamt als <i>unzureichend</i> zu kodieren.	<ul style="list-style-type: none"> ✓ Eins ist härter als das andere. ✓ In MB2 schwimmt das Ei, in Messbecher 1 aber nicht. 	

Erklärung	
<p>Die Erklärung ist plausibel und passt zur Beobachtung. Sie erklärt die Beobachtung jedoch nur unzureichend.</p> <p>Im Gegensatz zur korrekten Erklärung, wird hier ein Zusammenhang zwar festgestellt, aber nicht näher ausgeführt. Wird eine korrekte Erklärung mit einer inkorrekten Erklärung kombiniert, ist die gesamte Erklärung ebenfalls als unzureichend zu bewerten.</p> <p>Bezogen auf den Prä-Test ist das Auflösen des Salzes bereits eine passende Erklärung. Wird jedoch nicht ausgeführt, in welchem Fall sich das Salz auflöst (bei einer hohen Menge Salz löst sich kaum weiteres Salz in Wasser → das Wasser ist gesättigt), so ist die Erklärung als unzureichend zu bewerten.</p> <p>Bezogen auf den Follow-Up-Test kann von Kindern der 3. Klasse nicht das Verständnis für Osmose aus der Beobachtung der Salatblätter erwartet werden. Daher gilt an dieser Stelle eine hypothetische Erklärung als unzureichend, auch wenn diese fachlich falsch ist.</p>	<ul style="list-style-type: none"> ✓ Das Salz löst sich auf. (Der Satz wurde nicht weiter ausgeführt.) ✓ Alles schwimmt auf Salzwasser. ✓ Ich glaube, dass der Salat an den Wänden klebt, weil das Wasser ins Salat kommt. Der Salat wird dadurch weich und klebt an den Wänden.

<i>fehlerhaft / keine Antwort</i>	Vermutung	
	Das Ausbleiben einer Vermutung oder der Ausdruck von Unkenntnis wird an dieser Stelle kodiert.	<ul style="list-style-type: none"> ✓ - ✓ Keine Ahnung
	Durchführung	
	Die Beschreibung der Versuchsdurchführung nimmt den genauen Wortlaut der Versuchsanleitung an. Die Kategorie wird auch bei einer grob fehlerhaft beschriebenen Durchführung oder bei einer fehlenden Antwort kodiert. Eine falsch genannte Mengenangabe wird nicht als grober Fehler gewertet.	<ul style="list-style-type: none"> ✓ Wir schütten 100 Milliliter Salz in einen Messbecher. Dann nehmen wir einen Löffel und tun Salz auf den Löffel und tun Salz in den Messbecher und beobachten, was passiert.
	Beobachtung	
<p>Die Beobachtung ist nicht schriftlich festgehalten.</p> <p>Die Beobachtung lässt keinen Aufgabenbezug erkennen oder sie passt nicht zum erwarteten Phänomen. Beobachtungen, die aufgrund einer falschen Vorgehensweise bei der Versuchsdurchführung von der erwarteten Beobachtung abweichen, werden ebenfalls als <i>fehlerhaft / keine Antwort</i> kodiert.</p>	<ul style="list-style-type: none"> ✓ Das Ei ist kaputtgegangen. ✓ Wir haben die Schritte beobachtet. 	

	Sind für das Verständnis wichtige Wörter nicht lesbar, so wird die Beobachtung auch als <i>fehlerhaft</i> gewertet.	
	Erklärung	
	<p>Es gibt keinen verschriftlichten Erklärungsansatz.</p> <p>Die Erklärungen lassen keinen Aufgabenbezug erkennen oder sind fachlich inkorrekt.</p> <p>Ausnahme: Bezogen auf den Follow-Up-Test ist auch eine fachlich inkorrekte Erklärung als <i>unzureichend</i> zu bewerten.</p>	<ul style="list-style-type: none"> ✓ Es passiert gar nichts. ✓ Das Salz hat den Salat lätschig gemacht.
<i>Sonstiges</i>	Diese Kategorie ist zu kodieren, wenn der Inhalt keinem der vorliegenden Unterkategorien zuzuordnen ist.	

8.3 Anhang C: Ergebnisse der statistischen Berechnungen

8.3.1 Anhang C.1: Statistische Ergebnisse zum C-Test

Tab. 25: Deskriptive Statistik bezogen auf die Punktzahl im C-Test (AV)

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% Konfidenzintervall des Mittelwerts		<i>min</i>	<i>max</i>
					Unter- grenze	Ober- grenze		
KG	20	46.10	16.921	3.784	38.18	54.02	19	87
TG1	12	57.33	24.062	6.946	42.05	72.62	9	87
TG2	14	60.36	18.670	4.990	49.58	71.14	32	88

AV = Abhängige Variable; *n* = Stichprobengröße; *M* = Mittelwert; *SD* = Standardabweichung; *SEM* = Standardfehler; *min* = Minimum; *max* = Maximum; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 26: Test auf Normalverteilung bezogen auf die Punktzahl im C-Test (AV)

	<i>n</i>	K-S	Sig.	S-W	Sig.
KG	20	0.170	.131	0.950	.370
TG1	12	0.195	.200	0.930	.383
TG2	14	0.109	.200	0.948	.525

AV = Abhängige Variable; *n* = Stichprobengröße; K-S = Kolmogorow-Smirnov-Prüfwert mit Signifikanzkorrektur nach Lilliefors; S-W = Shapiro-Wilk-Prüfwert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 27: Test auf Varianzhomogenität bezogen auf die Punktzahl im C-Test (AV)

		Levene- Statistik	<i>df1</i>	<i>df2</i>	Sig.
C-Test	Basiert auf dem Mittelwert	1.598	2	43	.214
	Basiert auf dem Median	0.645	2	43	.530

AV = Abhängige Variable; *df1* = Zählerfreiheitsgrade; *df2* = Nennerfreiheitsgrade; Sig. = Signifikanz

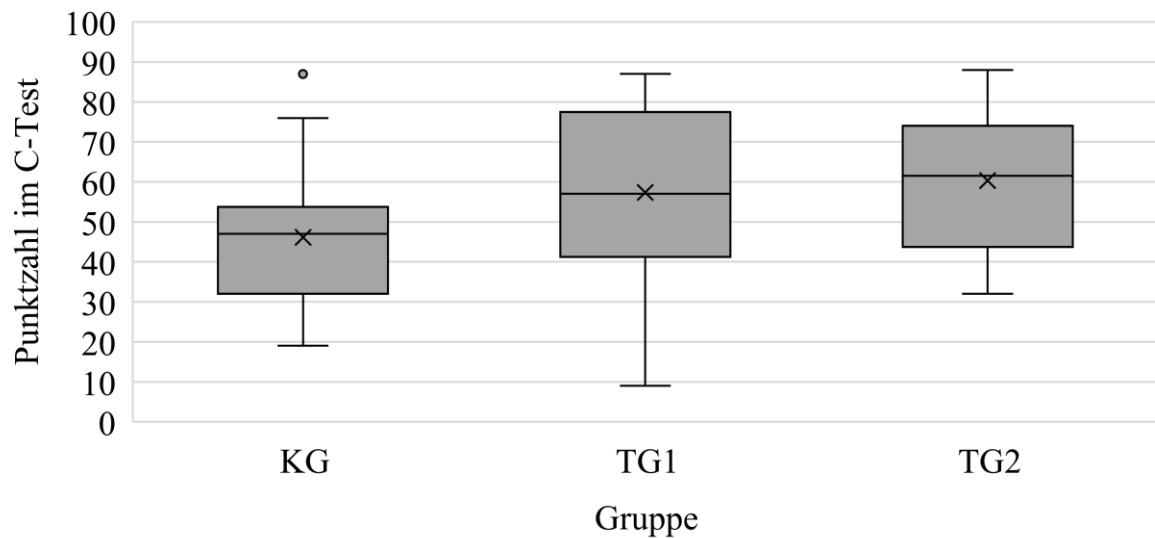


Abb. 39: Boxplot bezogen auf die Punktzahl im C-Test (AV)

AV = Abhängige Variable; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 28: 3x1 ANOVA bezogen auf die Punktzahl im C-Test (AV)

		Quadrat- summe	<i>df</i>	Mittel der Quadrate	<i>F</i>	Sig.
C-Test	Zwischen den Gruppen	1929.036	2	964.518	2.538	.091
	Innerhalb der Gruppen	16339.681	43	379.993		
	Gesamt	18268.717	45			

AV = Abhängige Variable; *df* = Freiheitsgrad; *F* = F-Wert; Sig. = Signifikanz

8.3.2 Anhang C.2: Statistische Ergebnisse zur ersten Teilstudie (F1)

Tab. 29: Deskriptive Statistik bezogen auf die Verwendung der Wortspeicher (AV)

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% Konfidenzintervall des Mittelwerts		<i>min</i>	<i>max</i>
					Unter- grenze	Ober- grenze		
PRÄ-TEST								
KG	20	.25	.550	.123	-.01	.51	0	2
TG1	12	.25	.452	.131	-.04	.54	0	1
TG2	14	.43	.852	.228	-.06	.92	0	3
Gesamt	46	.30	.628	.093	.12	.49	0	3
POST-TEST								
KG	20	.30	.571	.128	.03	.57	0	2
TG1	12	1.33	1.875	.541	.14	2.52	0	6
TG2	14	2.00	2.038	.545	.82	3.18	0	6
Gesamt	46	1.09	1.658	.244	.59	1.58	0	6
FOLLOW-UP-TEST								
KG	20	.40	.940	.210	-.04	.84	0	3
TG1	12	.58	1.165	.336	-.16	1.32	0	4
TG2	14	1.00	1.414	.378	.18	1.82	0	5
Gesamt	46	.63	1.162	.171	.29	.98	0	5

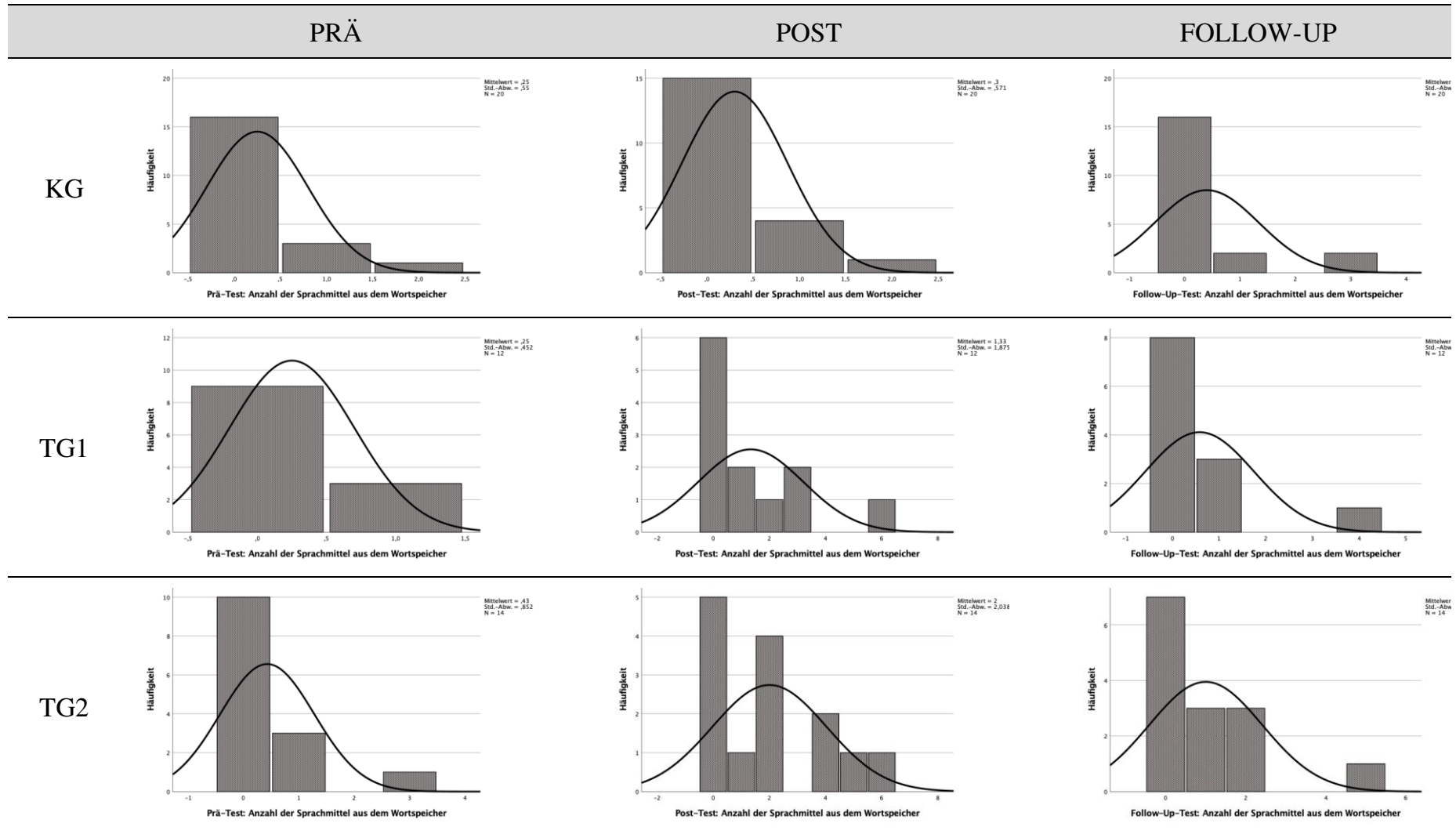
AV = Abhängige Variable; *n* = Stichprobengröße; *M* = Mittelwert; *SD* = Standardabweichung; *SEM* = Standardfehler; *min* = Minimum; *max* = Maximum; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 30: Test auf Normalverteilung bezogen auf die Verwendung der Wortspeicher (AV)

Gruppe	<i>n</i>	K-S	Sig.	S-W	Sig.
PRÄ-TEST					
KG	20	0.475	< .001	0.522	< .001
TG1	12	0.460	< .001	0.552	< .001
TG2	14	0.407	< .001	0.577	< .001
POST-TEST					
KG	20	0.450	< .001	0.583	< .001
TG1	12	0.262	.023	0.762	.004
TG2	14	0.214	.081	0.865	.036
FOLLOW-UP-TEST					
KG	20	0.465	< .001	0.483	< .001
TG1	12	0.358	< .001	0.570	< .001
TG2	14	0.260	.011	0.732	< .001

Abweichender Wert (hier: Annahme der Normalverteilung) ist grau hervorgehoben; AV = Abhängige Variable; *n* = Stichprobengröße; K-S = Kolmogorow-Smirnov-Prüfwert mit Signifikanzkorrektur nach Lilliefors; S-W = Shapiro-Wilk-Prüfwert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 31: Histogramme bezogen auf die Verwendung der Wortspeicher (AV)



AV = Abhängige Variable; Verteilungsdarstellung der Anzahl der verwendeten Sprachmittel aus dem Wortspeicher in den Versuchsprotokollen; bezogen auf die Gruppen (KG, TG1, TG2) und den Zeitpunkt (PRÄ, POST, FOLLOW-UP)

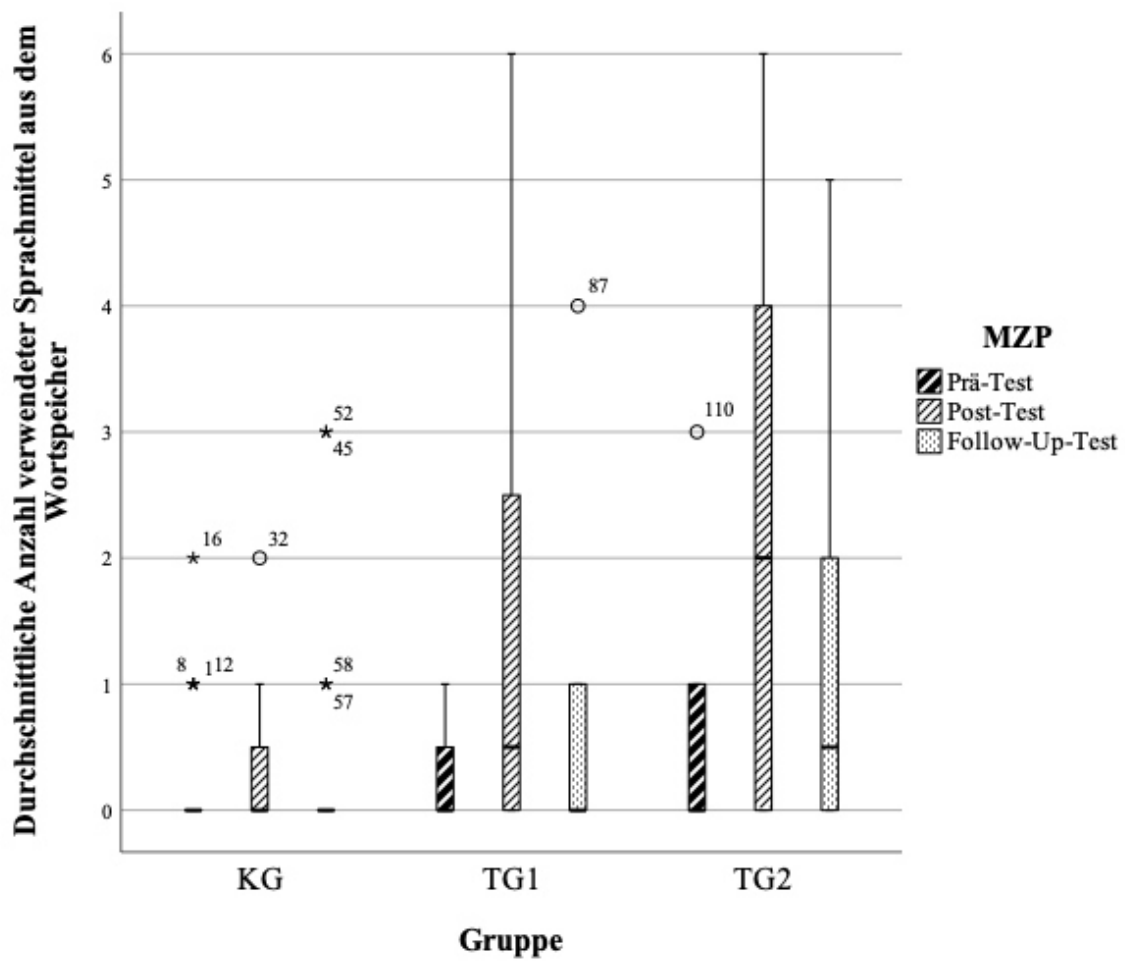


Abb. 40: Boxplot zur Analyse von extremen Ausreißern in SPSS bezogen auf die Verwendung der Wortspeicher (AV)

AV = Abhängige Variable; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 32: Test auf Varianzhomogenität bezogen auf die Verwendung der Wortspeicher (AV)

	Levene- Statistik	<i>df1</i>	<i>df2</i>	Sig.
PRÄ-TEST				
Basiert auf dem Mittelwert	1.424	2	43	.252
Basiert auf dem Median	0.383	2	43	.684
POST-TEST				
Basiert auf dem Mittelwert	7.861	2	43	.001
Basiert auf dem Median	6.546	2	43	.003
FOLLOW-UP-TEST				
Basiert auf dem Mittelwert	0.813	2	43	.450
Basiert auf dem Median	1.366	2	43	.266

Signifikante Werte sind grau hervorgehoben, AV = Abhängige Variable; *df1* = Zählerfreiheitsgrade; *df2* = Nennerfreiheitsgrade; Sig. = Signifikanz

Tab. 33: Test auf Sphärizität bezogen auf die Verwendung der Wortspeicher (AV)

Gruppe	Mauchly-W	<i>df</i>	Sig.
KG	.675	2	.029
TG1	.187	2	< .001
TG2	.560	2	< .031

Signifikante Werte sind grau hervorgehoben, AV = Abhängige Variable; Mauchly-W = Mauchlys Sphärizitätstest; *df* = Freiheitsgrade; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 34: Ergebnis der 3x1 ANOVA bezogen auf die Verwendung der Wortspeicher (AV)

		Quadrat- summe	<i>df</i>	Mittel der Quadrate	<i>F</i>	Sig.
PRÄ	Zwischen den Gruppen	0.311	2	0.155	0.383	.684
	Innerhalb der Gruppen	17.429	43	0.405		
	Gesamt	17.739	45			
POST	Zwischen den Gruppen	24.786	2	12.393	5.390	.008
	Innerhalb der Gruppen	98.867	43	2.299		
	Gesamt	123.652	45			
FOLLOW- UP	Zwischen den Gruppen	3.001	2	1.500	1.118	.336
	Innerhalb der Gruppen	57.717	43	1.342		
	Gesamt	60.717	45			

Signifikanter Wert ist grau hervorgehoben; *df* = Freiheitsgrade; *F* = F-Wert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 35: 3x1 ANOVA-Effektgröße^{a,b} bezogen auf die Verwendung der Wortspeicher (AV)

		Punktschätzung	95% Konfidenzintervall	
			Unterer	Oberer
Post-Test	Eta-Quadrat	.200	.017	.373

AV = Abhängige Variable; a = Eta-Quadrat wird basierend auf dem Modell mit festen Effekten geschätzt; b = Negative, aber weniger verzerrte Schätzungen werden beibehalten, nicht auf Null aufgerundet; AV = Abhängige Variable; *df* = Freiheitsgrad; *F* = F-Wert; Sig. = Signifikanz

Tab. 36: Ergebnis der 1x3 ANOVA bezogen auf die Verwendung der Wortspeicher (AV)

Gruppe	Quelle		ε	F	Sig.
KG	MZP	Sphärizität angenommen		0.322	.727
		Greenhouse-Geisser	.755	0.322	.666
		Huynh-Feldt (HF)	.806	0.322	.680
TG1	MZP	Sphärizität angenommen		3.215	.060
		Greenhouse-Geisser	.551	3.215	.095
		Huynh-Feldt (HF)	.568	3.215	.094
TG2	MZP	Sphärizität angenommen		5.123	.013
		Greenhouse-Geisser	.694	5.123	.027
		Huynh-Feldt (HF)	.751	5.123	.023
		Untergrenze		5.123	.041

Signifikante Werte sind grau hervorgehoben; Werte zur Interpretation (abh. von ε) sind fett markiert; ε = Epsilon als Maß für die Abweichung von Sphärizität (1 = Sphärizität ist in vollem Ausmaß gegeben); F = F-Wert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 37: 1x3 ANOVA-Effektgröße bezogen auf die Verwendung der Wortspeicher (AV)

Gruppe	Partielles Eta-Quadrat
TG2	.283

AV = Abhängige Variable; TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

8.3.3 Anhang C.3: Statistische Ergebnisse zur zweiten Teilstudie (F2)

Tab. 38: Deskriptive Statistik bezogen auf die Verwendung von Bildungssprache (AV)

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% Konfidenzintervall des Mittelwerts		<i>min</i>	<i>max</i>
					Unter- grenze	Obergrenze		
PRÄ-TEST								
Mehrfachkodierungen								
KG	20	3.25	2.881	.644	1.90	4.60	0	10
TG1	12	5.08	3.343	.965	2.96	7.21	1	12
TG2	14	2.93	2.615	.699	1.42	4.44	0	8
Gesamt	46	3.63	2.999	.442	2.74	4.52	0	12
Einfachkodierungen								
KG	20	2.40	2.010	.450	1.46	3.34	0	7
TG1	12	3.42	1.929	.557	2.19	4.64	1	6
TG2	14	2.07	2.018	.539	.91	3.24	0	7
Gesamt	46	2.57	2.018	.298	1.97	3.16	0	7
POST-TEST								
Mehrfachkodierungen								
KG	20	2.00	2.128	0.476	1.00	3.00	0	8
TG1	12	4.00	3.742	1.080	1.62	6.38	0	14
TG2	14	5.14	3.900	1.042	2.89	7.39	0	15
Gesamt	46	3.48	3.411	0.503	2.47	4.49	0	15
Einfachkodierungen								
KG	20	1.65	1.785	.399	.81	2.49	0	7
TG1	12	2.50	2.355	.680	1.00	4.00	0	9
TG2	14	3.79	2.577	.689	2.30	5.27	0	10
Gesamt	46	2.52	2.336	.344	1.83	3.22	0	10
FOLLOW-UP-TEST								
Mehrfachkodierungen								
KG	20	2.55	2.438	.545	1.41	3.69	0	8

TG1	12	4.83	4.174	1.205	2.18	7.49	0	13
TG2	14	2.64	2.790	.746	1.03	4.25	0	10
Gesamt	46	3.17	3.164	.467	2.23	4.11	0	13

Einfachkodierungen

KG	20	2.00	1.947	.435	1.09	2.91	0	7
TG1	12	3.33	3.055	.882	1.39	5.27	0	11
TG2	14	2.21	2.190	.585	.95	3.48	0	7
Gesamt	46	2.41	2.362	.348	1.71	3.11	0	11

AV = Abhängige Variable; n = Stichprobengröße; M = Mittelwert; SD = Standardabweichung; SEM = Standardfehler; min = Minimum; max = Maximum; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 39: Test auf Normalverteilung bezogen auf die Verwendung von Bildungssprache (AV)

Gruppe	<i>n</i>	K-S	Sig.	S-W	Sig.
PRÄ-TEST					
Mehrfachkodierungen					
KG	20	.168	.142	.914	.076
TG1	12	.177	.200*	.927	.349
TG2	14	.198	.142	.897	.103
Einfachkodierungen					
KG	20	.179	.093	.919	.096
TG1	12	.202	.189	.880	.087
TG2	14	.202	.125	.868	.039
POST-TEST					
Mehrfachkodierungen					
KG	20	.231	.007	.849	.005
TG1	12	.250	.037	.810	.012
TG2	14	.199	.139	.872	.045
Einfachkodierungen					
KG	20	.222	.011	.825	.002
TG1	12	.249	.038	.780	.006
TG2	14	.181	.200*	.929	.300
FOLLOW-UP-TEST					
Mehrfachkodierungen					
KG	20	.152	.200*	.888	.025
TG1	12	.170	.200*	.920	.285
TG2	14	.222	.060	.840	.016
Einfachkodierungen					
KG	20	.200	.035	.868	.011
TG1	12	.169	.200*	.877	.080
TG2	14	.282	.004	.855	.026

Abweichende Werte (hier: Annahme keine Normalverteilung) sind grau hervorgehoben; AV = Abhängige Variable; * = dies ist eine untere Grenze der echten Signifikanz; *n* = Stichprobengröße; K-S = Kolmogorow-Smirnov-Prüfwert mit Signifikanzkorrektur nach Lilliefors; S-W = Shapiro-Wilk-Prüfwert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

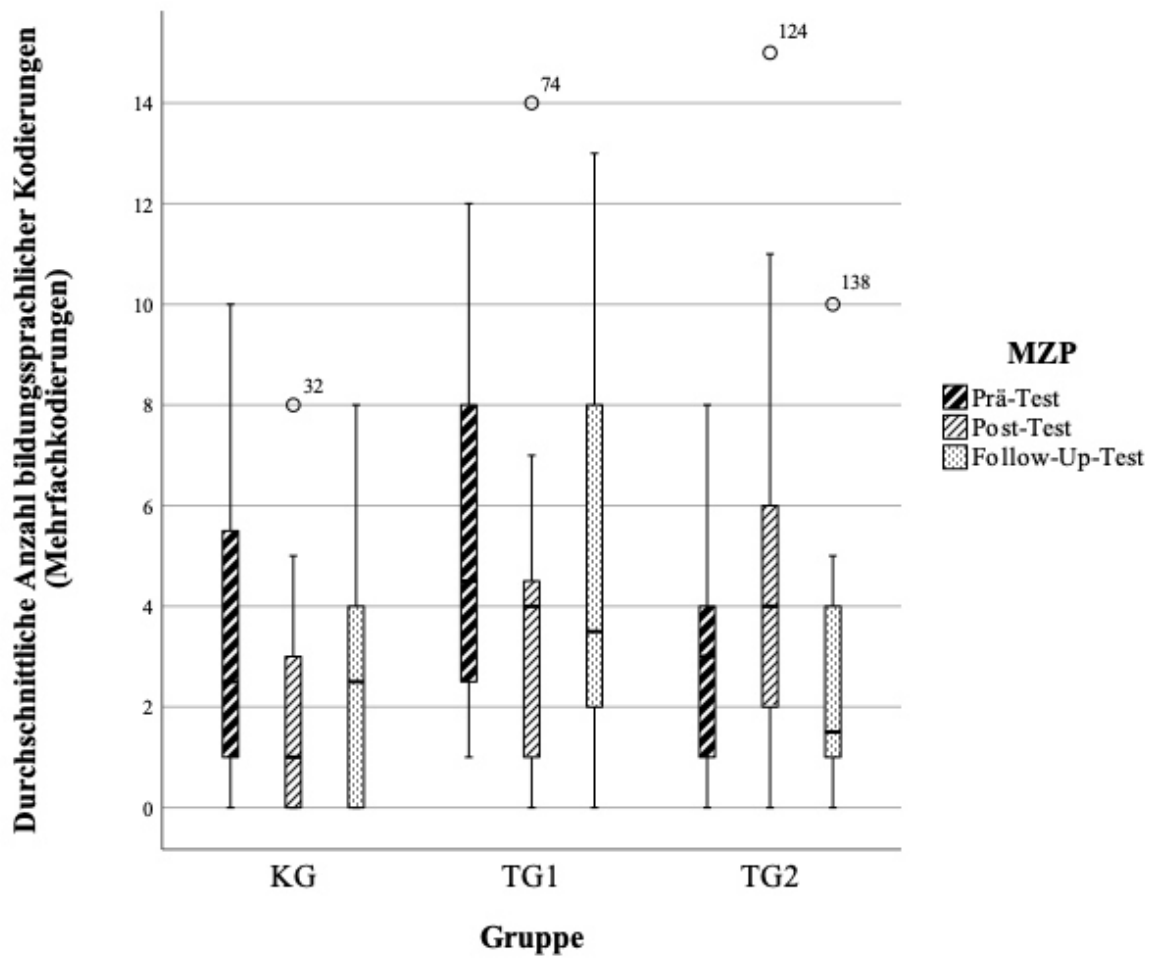


Abb. 41: Boxplot bezogen auf die Verwendung von Bildungssprache bei Mehrfachkodierungen (AV)

AV = Abhängige Variable; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

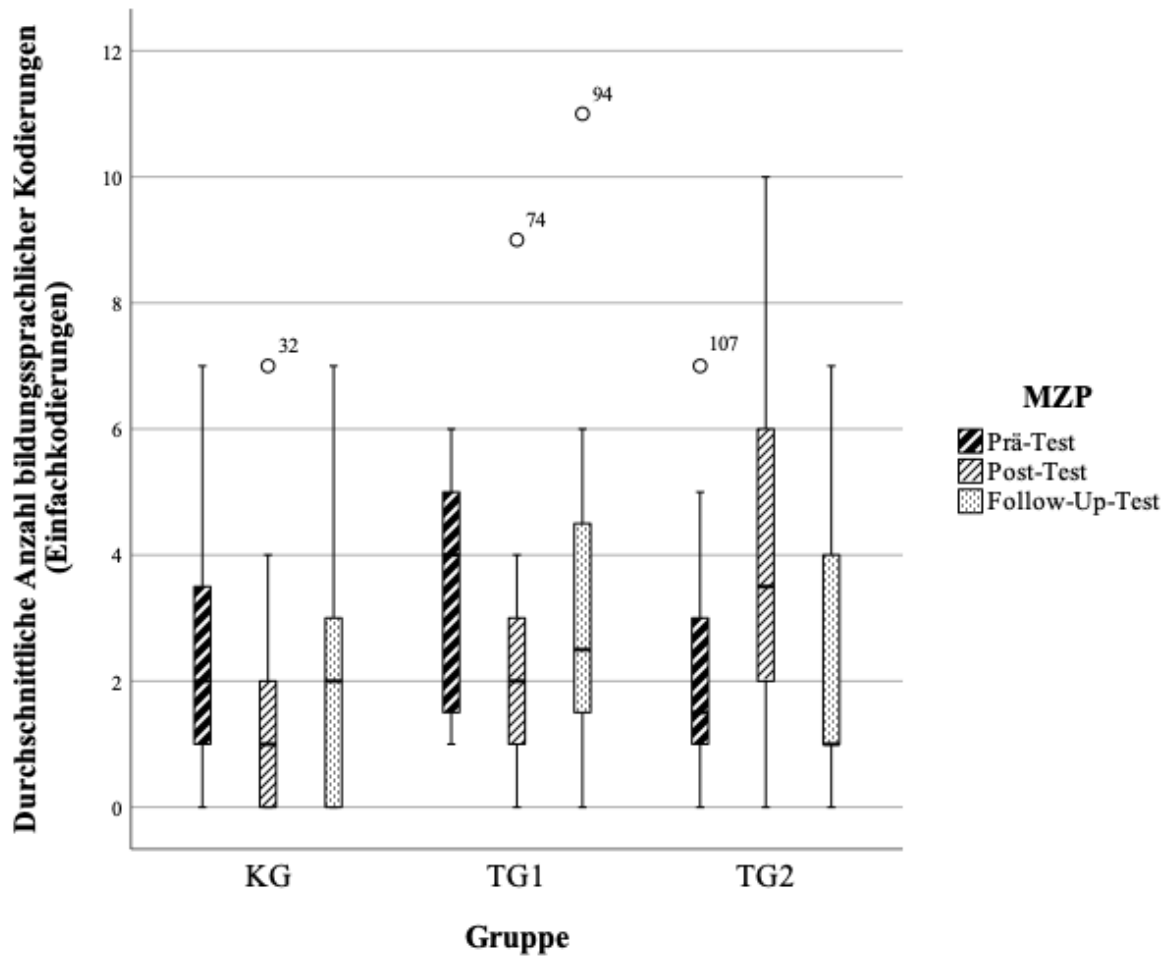


Abb. 42: Boxplot bezogen auf die Verwendung von Bildungssprache bei Einfachkodierungen (AV)

AV = Abhängige Variable; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 40: Test auf Varianzhomogenität bezogen auf die Verwendung von Bildungssprache (AV)

	Levene- Statistik	<i>df1</i>	<i>df2</i>	Sig.
PRÄ-TEST				
Mehrfachkodierungen				
Basiert auf dem Mittelwert	.210	2	43	.811
Basiert auf dem Median	.153	2	43	.859
Einfachkodierungen				
Basiert auf dem Mittelwert	.066	2	43	.936
Basiert auf dem Median	.017	2	43	.983
POST-TEST				
Mehrfachkodierungen				
Basiert auf dem Mittelwert	.982	2	43	.383
Basiert auf dem Median	.705	2	43	.500
Einfachkodierungen				
Basiert auf dem Mittelwert	.732	2	43	.487
Basiert auf dem Median	.741	2	43	.482
FOLLOW-UP-TEST				
Mehrfachkodierungen				
Basiert auf dem Mittelwert	3.348	2	43	.045
Basiert auf dem Median	1.941	2	43	.156
Einfachkodierungen				
Basiert auf dem Mittelwert	1.245	2	43	.298
Basiert auf dem Median	.717	2	43	.494

Signifikanter Wert ist grau hervorgehoben; AV = Abhängige Variable; *df1* = Zählerfreiheitsgrade; *df2* = Nennerfreiheitsgrade; Sig. = Signifikanz

Tab. 41: Ergebnis der 3x1 ANOVA bezogen auf die Verwendung von Bildungssprache (AV)

	Quadrat- summe	<i>df</i>	Mittel der Quadrate	<i>F</i>	Sig.
PRÄ-TEST					
Mehrfachkodierungen					
Zwischen den Gruppen	35.122	2	17.561	2.043	.142
Innerhalb der Gruppen	369.595	43	8.595		
Gesamt	404.717	45			
Einfachkodierungen					
Zwischen den Gruppen	12.659	2	6.330	1.595	.215
Innerhalb der Gruppen	170.645	43	3.968		
Gesamt	183.304	45			
POST-TEST					
Mehrfachkodierungen					
Zwischen den Gruppen	85.764	2	42.882	4.213	.021
Innerhalb der Gruppen	437.714	43	10.179		
Gesamt	523.478	45			
Einfachkodierungen					
Zwischen den Gruppen	37.571	2	18.786	3.885	.028
Innerhalb der Gruppen	207.907	43	4.835		
Gesamt	245.478	45			
FOLLOW-UP-TEST					
Mehrfachkodierungen					
Zwischen den Gruppen	44.778	2	22.389	2.372	.105
Innerhalb der Gruppen	405.831	43	9.438		
Gesamt	450.609	45			
Einfachkodierungen					
Zwischen den Gruppen	14.128	2	7.064	1.282	.288
Innerhalb der Gruppen	237.024	43	5.512		
Gesamt	251.152	45			

Signifikante Werte sind grau hervorgehoben; AV = Abhängige Variable; *df* = Anzahl der Freiheitsgrade; *F* = F-Wert; Sig. = Signifikanz

Tab. 42: 3x1 ANOVA-Effektgröße^{a,b} bezogen auf die Verwendung von Bildungssprache (AV)

		Punktschätzung	95% Konfidenzintervall	
			Unterer	Oberer
Post-Test, Mk	Eta-Quadrat	.164	.002	.335
Post-Test, Ek	Eta-Quadrat	.153	.000	.323

AV = Abhängige Variable; Mk = bei Mehrfachkodierungen, Ek = bei Einfachkodierungen; a = Eta-Quadrat wird basierend auf dem Modell mit festen Effekten geschätzt; b = Negative, aber weniger verzerrte Schätzungen werden beibehalten, nicht auf Null aufgerundet; *df* = Freiheitsgrad; *F* = F-Wert; Sig. = Signifikanz

Tab. 43: Test auf Sphärizität bezogen auf die Verwendung von Bildungssprache (AV)

Gruppe	Mauchly-W	<i>df</i>	Sig.
Mehrfachkodierungen			
KG	.985	2	.875
TG1	.986	2	.931
TG2	.428	2	.006
Einfachkodierungen			
KG	.961	2	.696
TG1	.786	2	.299
TG2	.751	2	.179

Signifikanter Wert ist grau hervorgehoben; AV = Abhängige Variable; Mauchly-W = Mauchlys Sphärizitätstest; *df* = Freiheitsgrade; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 44: Ergebnis der 1x3 ANOVA bezogen auf die Verwendung von Bildungssprache (AV)

Gruppe	Quelle		ϵ	F	Sig.
Mehrfachkodierungen					
KG	MZP	Sphärizität angenommen		2.039	.144
		Greenhouse-Geisser	.986	2.039	.145
		Huynh-Feldt (HF)	1.000	2.039	.144
		Untergrenze	.500	2.039	.170
TG1	MZP	Sphärizität angenommen		.434	.653
		Greenhouse-Geisser	.986	.434	.651
		Huynh-Feldt (HF)	1.000	.434	.653
		Untergrenze	.500	.434	.523
TG2	MZP	Sphärizität angenommen		5.428	.011
		Greenhouse-Geisser	.636	5.428	.026
		Huynh-Feldt (HF)	.674	5.428	.024
		Untergrenze	.500	5.428	.037
Einfachkodierungen					
KG	MZP	Sphärizität angenommen		1.159	.325
		Greenhouse-Geisser	.962	1.159	.324
		Huynh-Feldt (HF)	1.000	1.159	.325
		Untergrenze	.500	1.159	.295
TG1	MZP	Sphärizität angenommen		.836	.447
		Greenhouse-Geisser	.824	.836	.429
		Huynh-Feldt (HF)	.950	.836	.442
		Untergrenze	.500	.836	.380
TG2	MZP	Sphärizität angenommen		4.016	.030
		Greenhouse-Geisser	.801	4.016	.041
		Huynh-Feldt (HF)	.896	4.016	.036
		Untergrenze	.500	4.016	.066

Signifikante Werte sind grau hervorgehoben; Werte zur Interpretation (abh. von ϵ) sind fett markiert; ϵ = Epsilon als Maß für die Abweichung von Sphärizität (1 = Sphärizität ist in vollem Ausmaß gegeben); F = F-Wert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 45: 1x3 ANOVA-Effektgröße bezogen auf die Verwendung von Bildungssprache (AV)

Gruppe	Partielles Eta-Quadrat
TG2, Mk	.295
TG2, Ek	.236

AV = Abhängige Variable; Mk = bei Mehrfachkodierungen, Ek = bei Einfachkodierungen; TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

8.3.4 Anhang C.4: Statistische Ergebnisse zur dritten Teilstudie (F3)

Tab. 46: Deskriptive Statistik bezogen auf die inhaltliche Punktzahl (AV)

	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% Konfidenzintervall des Mittelwerts		<i>min</i>	<i>max</i>
					Unter- grenze	Ober- grenze		
PRÄ-TEST								
KG	20	10.85	2.889	.646	9.50	12.20	3	15
TG1	12	9.42	2.906	.839	7.57	11.26	3	13
TG2	14	9.64	2.437	.651	8.24	11.05	4	12
Gesamt	46	10.11	2.783	.410	9.28	10.94	3	15
POST-TEST								
KG	20	11.10	2.770	0.619	9.80	12.40	6	16
TG1	12	9.42	4.078	1.177	6.83	12.01	4	16
TG2	14	11.43	3.005	0.803	9.69	13.16	4	15
Gesamt	46	10.76	3.254	0.480	9.79	11.73	4	16
FOLLOW-UP-TEST								
KG	20	7.20	2.783	0.622	5.90	8.50	3	12
TG1	12	8.50	2.195	0.634	7.11	9.89	4	12
TG2	14	8.07	2.615	0.699	6.56	9.58	4	13
Gesamt	46	7.80	2.596	0.383	7.03	8.58	3	13

AV = Abhängige Variable; *n* = Stichprobengröße; *M* = Mittelwert; *SD* = Standardabweichung; *SEM* = Standardfehler; *min* = Minimum; *max* = Maximum; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 47: Test auf Normalverteilung bezogen auf die inhaltliche Punktzahl (AV)

Gruppe	<i>n</i>	K-S	Sig.	S-W	Sig.
PRÄ-TEST					
KG	20	.184	.074	.921	.102
TG1	12	.193	.200*	.919	.282
TG2	14	.201	.130	.874	.048
POST-TEST					
KG	20	.173	.120	.937	.211
TG1	12	.216	.129	.914	.241
TG2	14	.158	.200*	.916	.191
FOLLOW-UP-TEST					
KG	20	.163	.171	.936	.199
TG1	12	.164	.200*	.948	.609
TG2	14	.297	.002	.878	.054

Abweichende Werte (hier: Annahme keine Normalverteilung) sind grau hervorgehoben; AV = Abhängige Variable; * = dies ist eine untere Grenze der echten Signifikanz; *n* = Stichprobengröße; K-S = Kolmogorow-Smirnov-Prüfwert mit Signifikanzkorrektur nach Lilliefors; S-W = Shapiro-Wilk-Prüfwert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

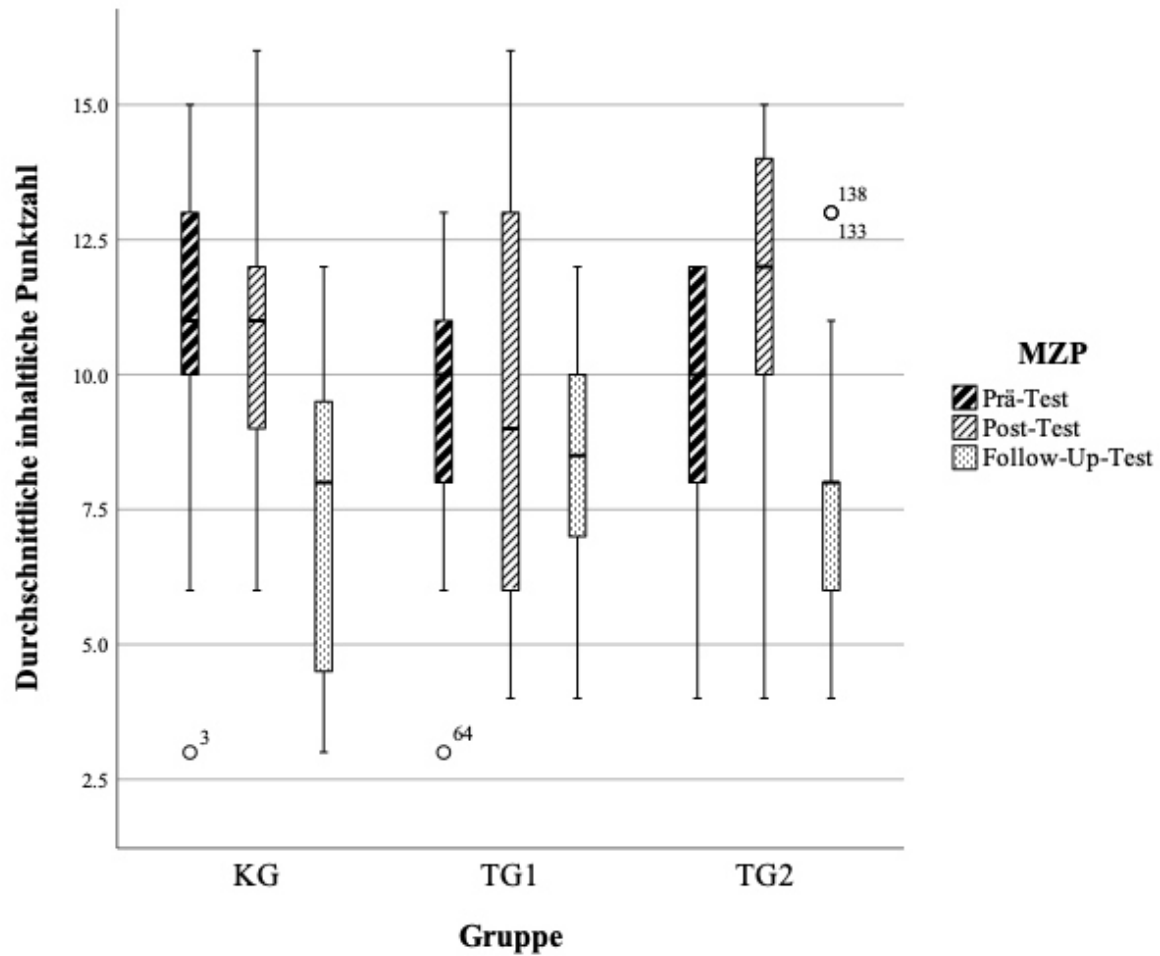


Abb. 43: Boxplot bezogen auf die inhaltliche Punktzahl (AV)

AV = Abhängige Variable; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 48: Test auf Varianzhomogenität bezogen auf die inhaltliche Punktzahl (AV)

	Levene- Statistik	<i>df1</i>	<i>df2</i>	Sig.
PRÄ-TEST				
Basiert auf dem Mittelwert	.094	2	43	.911
Basiert auf dem Median	.105	2	43	.900
POST-TEST				
Basiert auf dem Mittelwert	3.008	2	43	.060
Basiert auf dem Median	2.890	2	43	.066
FOLLOW-UP-TEST				
Basiert auf dem Mittelwert	.677	2	43	.514
Basiert auf dem Median	.436	2	43	.649

AV = Abhängige Variable; *df1* = Zählerfreiheitsgrade; *df2* = Nennerfreiheitsgrade; Sig. = Signifikanz

Tab. 49: Ergebnis der 3x1 ANOVA bezogen auf die inhaltliche Punktzahl (AV)

	Quadrat- summe	<i>df</i>	Mittel der Quadrate	<i>F</i>	Sig.
PRÄ-TEST					
Zwischen den Gruppen	19.776	2	9.888	1.294	.285
Innerhalb der Gruppen	328.681	43	7.644		
Gesamt	348.457	45			
POST-TEST					
Zwischen den Gruppen	30.224	2	15.112	1.457	.244
Innerhalb der Gruppen	446.145	43	10.375		
Gesamt	476.370	45			
FOLLOW-UP-TEST					
Zwischen den Gruppen	14.111	2	7.055	1.049	.359
Innerhalb der Gruppen	289.129	43	6.724		
Gesamt	303.239	45			

AV = Abhängige Variable; *df* = Anzahl der Freiheitsgrade; *F* = F-Wert; *Sig.* = Signifikanz

Tab. 50: Test auf Sphärizität bezogen auf die inhaltliche Punktzahl (AV)

Gruppe	Mauchly-W	df	Sig.
KG	.998	2	.979
TG1	.887	2	.551
TG2	.954	2	.753

AV = Abhängige Variable; Mauchly-W = Mauchlys Sphärizitätstest; df = Freiheitsgrade; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 51: Ergebnis der 1x3 ANOVA bezogen auf die inhaltliche Punktzahl (AV)

Gruppe	Quelle	ε	F	Sig.	
KG	MZP	Sphärizität angenommen	11.835	<.001	
		Greenhouse-Geisser	.998	11.835	<.001
		Huynh-Feldt (HF)	1.000	11.835	<.001
		Untergrenze	.500	11.835	.003
TG1	MZP	Sphärizität angenommen	.613	.551	
		Greenhouse-Geisser	.899	.613	.535
		Huynh-Feldt (HF)	1.000	.613	.551
		Untergrenze	.500	.613	.450
TG2	MZP	Sphärizität angenommen	6.893	.004	
		Greenhouse-Geisser	.956	6.893	.005
		Huynh-Feldt (HF)	1.000	6.893	.004
		Untergrenze	.500	6.893	.021

Signifikante Werte sind grau hervorgehoben; Werte zur Interpretation (abh. von ε) sind fett markiert; ε = Epsilon als Maß für die Abweichung von Sphärizität (1 = Sphärizität ist in vollem Ausmaß gegeben); F = F-Wert; Sig. = Signifikanz; KG = Kontrollgruppe, TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

Tab. 52: 1x3 ANOVA-Effektgröße bezogen auf die inhaltliche Punktzahl (AV)

Gruppe	Partielles Eta-Quadrat
KG	.384
TG2	.346

AV = Abhängige Variable; KG = Kontrollgruppe, TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform)

8.4 Anhang D: Weitere Ergebnisse

8.4.1 Anhang D.1: Häufigkeitsanalyse einfachkodierter Sprachhandlungen

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Vermuten												
KG	2	0	1	0.1	0	0.05	0	0	0	18	20	19
TG1	6	1	3	0.5	0.08	0.25	0	0	0	7	11	9
TG2	1	4	3	0.07	0.29	0.21	0	0	0	13	10	12
Begründen												
KG	2	3	1	0.1	0.15	0.05	0	0	0	18	17	19
TG1	3	1	5	0.25	0.08	0.42	0	0	0	9	11	8
TG2	0	18	6	0	1.29	0.43	0	1	0	14	6	10
Gesamt												
KG	4	3	2	0.2	0.15	0.1	0	0	0	16	17	18
TG1	9	2	8	0.75	0.17	0.67	0.5	0	0	6	10	7
TG2	1	22	9	0.07	1.57	0.64	0	1	0	13	5	9

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = absolute Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

8.4.2 Anhang D.2: Häufigkeitsanalyse einfachkodiertes Lexikon

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Nomen												
KG	11	3	2	0.55	0.15	0.1	0	0	0	13	17	18
TG1	6	7	3	0.5	0.58	0.25	0	0	0	7	7	9
TG2	5	9	2	0.36	0.64	0.14	0	1	0	9	6	12
Verben												
KG	3	3	9	0.15	0.15	0.45	0	0	0	17	17	11
TG1	5	5	4	0.42	0.42	0.33	0	0	0	8	9	9
TG2	1	2	2	0.07	0.14	0.14	0	0	0	13	12	13
Adjektive												
KG	3	0	2	0.15	0	0.1	0	0	0	17	20	18
TG1	4	0	1	0.33	0	0.08	0	0	0	8	12	11
TG2	6	1	1	0.43	0.07	0.07	0	0	0	9	13	13
Sonstiges												
KG	1	1	0	0.05	0.05	0	0	0	0	19	19	20
TG1	0	2	1	0	0.17	0.08	0	0	0	12	10	11
TG2	0	0	3	0	0	0.21	0	0	0	14	14	11
Gesamt												
KG	18	7	13	0.9	0.35	0.65	0.5	0	1	10	15	9
TG1	15	14	9	1.25	1.17	0.75	1	0.5	0	4	6	7
TG2	12	12	8	0.86	0.86	0.57	1	1	0	6	5	8

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)

8.4.3 Anhang D.3: Häufigkeitsanalyse einfachkodierte Morphosyntax

	PRÄ Anzahl	POST Anzahl	FOLLOW Anzahl	PRÄ <i>M</i>	POST <i>M</i>	FOLLOW <i>M</i>	PRÄ <i>Md</i>	POST <i>Md</i>	FOLLOW <i>Md</i>	PRÄ -	POST -	FOLLOW -
Kohäsion												
KG	17	15	14	0.85	0.75	0.7	0	1	0	12	9	11
TG1	9	7	13	0.75	0.58	1.08	0.5	1	1	6	5	3
TG2	10	10	9	0.71	0.71	0.64	0.5	1	1	7	6	6
Satzgefüge												
KG	6	7	10	0.3	0.35	0.5	0	0	0	14	14	13
TG1	7	7	8	0.58	0.58	0.67	0	0	0	7	7	7
TG2	5	6	4	0.36	0.43	0.29	0	0	0	12	9	10
Unpersönliche Konstruktion												
KG	3	1	1	0.15	0.05	0.05	0	0	0	17	19	19
TG1	1	0	2	0.08	0	0.17	0	0	0	11	12	10
TG2	1	3	1	0.07	0.21	0.07	0	0	0	13	11	13
Gesamt												
KG	26	23	25	1.3	1.15	1.25	1	1	1	9	7	9
TG1	17	14	23	1.42	1.17	1.92	1	1	1.5	4	4	2
TG2	16	19	14	1.14	1.36	1	0.5	1.5	1	7	4	5

Bedeutsame Werte sind grau hervorgehoben; Anzahl = absolute Anzahl der Kodierungen; *M* = Mittelwert; *Md* = Median; - = Anzahl der Dokumente ohne Kodierungen; KG = Kontrollgruppe ($n = 20$), TG1 = Treatmentgruppe 1 (auditiv vermittelte Wortspeicher / $n = 12$), TG2 = Treatmentgruppe 2 (Wortspeicher in Plakatform / $n = 14$)