

Katharina Fliesser

## **Verständlichkeit physikalischer Sachtexte**

Untersuchungen zum Wirkungsgefüge  
zwischen sprachlicher Textgestaltung und der  
Behaltensleistung sowie der Textwahrnehmung  
im Schulfach Physik

λογος

# Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

Diese Reihe im Logos Verlag Berlin lädt Forscherinnen und Forscher ein, ihre neuen wissenschaftlichen Studien zum Physik- und Chemielernen im Kontext einer Vielzahl von bereits erschienenen Arbeiten zu quantitativen und qualitativen empirischen Untersuchungen sowie evaluativ begleiteten Konzeptionsentwicklungen zu veröffentlichen. Die in den bisherigen Studien erfassten Themen und Inhalte spiegeln das breite Spektrum der Einflussfaktoren wider, die in den Lehr- und Lernprozessen in Schule und Hochschule wirksam sind.

Die Herausgeber hoffen, mit der Förderung von Publikationen, die sich mit dem Physik- und Chemielernen befassen, einen Beitrag zur weiteren Stabilisierung der physik- und chemiedidaktischen Forschung und zur Verbesserung eines an den Ergebnissen fachdidaktischer Forschung orientierten Unterrichts in den beiden Fächern zu leisten.

Martin Hopf und Mathias Ropohl

*Studien zum Physik- und Chemielernen*

Band 382



Katharina Flieser

# **Verständlichkeit physikalischer Sachtexte**

Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik

Logos Verlag Berlin



## *Studien zum Physik- und Chemielernen*

Martin Hopf und Mathias Ropohl [Hrsg.]

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.



Dieses Werk ist lizenziert unter der Creative Commons Lizenz CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z. B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Logos Verlag Berlin GmbH 2024

ISBN 978-3-8325-5858-1

ISSN 1614-8967

DOI 10.30819/5858

Logos Verlag Berlin GmbH  
Georg-Knorr-Str. 4, Geb. 10  
D-12681 Berlin

Tel.: +49 (0)30 / 42 85 10 90

Fax: +49 (0)30 / 42 85 10 92

<https://www.logos-verlag.de>

# Verständlichkeit physikalischer Sachtexte

Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung  
und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik



## DISSERTATION

ZUR ERLANGUNG DES AKADEMISCHEN GRADES EINER DOKTORIN  
DER DIDAKTIK DER NATURWISSENSCHAFTEN (DR. PHIL. NAT.)

DER FAKULTÄT FÜR PHYSIK  
AN DER UNIVERSITÄT REGENSBURG

vorgelegt von

**Katharina Flieser**

aus Landshut

im Jahr 2023

Promotionsgesuch eingereicht am: 20.12.2023

Die Arbeit wurde angeleitet von: Prof. Dr. Karsten Rincke

Prüfungsausschuss:

- Prof. Dr. Oliver Tepner (Vorsitzender)
- Prof. Dr. Karsten Rincke (1. Gutachter)
- Prof. Dr. Roland Berger (2. Gutachter)
- Prof. Dr. Heiko Krabbe (3. Gutachter)
- Prof. Dr. Arne Dittmer (weiterer Prüfer)

Termin Disputation: 25.04.2024

## Verständlichkeit physikalischer Sachtexte

Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik

### KURZZUSAMMENFASSUNG

Geschriebener Text ist ein fundamentales Medium der Wissensvermittlung. Das Verstehen des Lesens ist daher zentral für den Schulunterricht. Dabei bringen die spezifischen Inhalte sowie sprachlichen Charakteristika der einzelnen Fächer unterschiedliche Anforderungen mit. Die sprachlichen Merkmale, durch die sich geschriebene Texte speziell im Fach Physik auszeichnen, gelten als besonders anspruchsvoll. Um Lernende an die Bewältigung naturwissenschaftlicher Texte heranzuführen, ist adressatengerechtes, verständliches Textmaterial nötig.

Dieses Forschungsprojekt widmet sich der Frage, inwiefern bestimmte sprachliche Gestaltungsmittel die Wirkung von Texten auf Lesende beeinflussen. Es werden dabei sowohl kognitive als auch affektive Lernerfolgskriterien betrachtet. Dazu wurde ein theoretisches Fundament erarbeitet, das in einer umfangreichen Sammlung und systematischen Ordnung sprachlicher Gestaltungsmittel für Texte im Allgemeinen besteht sowie in einer gründlichen Erfassung und Strukturierung der Aspekte affektiver Textwirkungen. In der darauf gründenden empirischen Studie, in der die Teilnehmenden sich mit sprachlich unterschiedlich gestalteten Einführungstexten zum Thema *Die elektrische Spannung* auseinandersetzten, nahmen 787 Lernende an bayerischen Gymnasien und Realschulen teil.

Es zeigte sich, dass eine anspruchsgeminderte Wortwahl, anspruchsgeminderte Satzstrukturen und ein ausgeprägter Grad an Personalisierung zu wünschenswerten Textwirkungen auf die Lernenden führten. Dabei ließ sich die affektive Textwahrnehmung sensibler durch die sprachliche Gestaltung der Texte beeinflussen als die Behaltensleistung. Vor allem die Wahrnehmungsfacetten der *Involviertheit* hing stark vom Grad der Personalisierung ab. Andere Wahrnehmungsfacetten sowie das kognitive Kriterium der Behaltensleistung wurden in erster Linie durch die Wortwahl, vereinzelt auch durch die Satzstrukturen beeinflusst. Tendenziell profitierten Lernende mit schwächeren Ausgangsbedingungen (insbesondere Lesefähigkeiten und Vorwissen) stärker von Textveränderungen.

Die Studie liefert wertvolle Erkenntnisse zur Textverständlichkeit in den Naturwissenschaften und schafft eine wesentliche Grundlage für die weitere Erforschung insbesondere affektiver Textwirkungen und deren lernförderlicher Konsequenzen.



## DANKSAGUNG

Das dieser Arbeit zugrundeliegende Forschungsprojekt entstand im Rahmen des Projekts IMPULS+ als eine Maßnahme von KOLEG2<sup>1</sup>.

Die folgenden Zeilen sind all denjenigen gewidmet, die dazu beigetragen haben, dass ich diese Arbeit fertigstellen konnte. **DANKE!**

An erster Stelle richtet sich mein Dank an meinen Doktorvater Karsten Rincke. Lieber Karsten, ich danke dir für dein Vertrauen in mich. Es hat mich darin bestärkt, diese Promotion in Angriff zu nehmen und mich wieder und wieder zuversichtlich den damit verbundenen Herausforderungen zu stellen. Danke, dass du mir diese Gelegenheit geschenkt hast, selbst zu lernen und zum Lernerfolg Anderer beitragen zu können. Dank dir blicke ich auf die gemeisterten Schwierigkeiten als erfolgreich absolvierte Übungen zurück. Deine Unterstützung war in wissenschaftlicher sowie in menschlicher Hinsicht wegweisend.

Ich danke auch Herrn Prof. Roland Berger, der sich mit Offenheit, Interesse und Freundlichkeit für die Begutachtung meiner Dissertation bereit erklärt hat. Auch allen weiteren Mitgliedern des Prüfungsausschusses danke ich, dass Sie dieser Arbeit Ihr Interesse und Ihre Zeit schenkten.

Ein ganz großes Dankeschön geht an die AG der Physikdidaktik an der Uni Regensburg. An Jana, die mir immer wieder ausdrücklich *gerne* zu Hilfe kam. Liebe Jana, du und dein Wissen, dein Humor und dein Teesortiment waren immer wichtig. An Christian, der schier unermüdlich war – beim Hinweisen geben sowie auch beim gemeinsamen Laufen. Danke, dass du Mut machst, ein offenes Ohr für Probleme hast und die gute Stimmung im Blick behältst. Liebe Domi, danke für unser wertvolles „Thesiswriting-Worldwide“-Projekt. Mit deiner optimistischen Art und deinem sonnigen Gemüt warst du eine wunderbare Begleiterin nicht nur beim finalen Zusammenschreiben. Lieber Paul, liebe Patricia, liebe Laura, lieber Jonas, liebe Barbara, liebe Steffi und lieber Hans – von einer solchen Arbeitsgruppe umgeben zu sein, war ein ganz besonderes Geschenk.

Ich danke allen meinen Korrekturlektor\*innen: Andi, Marinus, Marlene, Robert, meiner Mama, Jana, Christian, Domi, Paul und Laura. Ich danke Liesel Kraus und Lena Schiestel für die statistische Beratung, meinen studentischen Hilfskräften Sheyenne, Korbinian, Nico,

---

<sup>1</sup>KOLEG2 wurde im Rahmen der gemeinsamen „Qualitätsoffensive Lehrerbildung“ von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01JA1812 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin.

Alessa und Christof für ihre Verlässlichkeit und allen IMPULS+-Beteiligten für die angenehme Projektarbeit. Besonders bedanken möchte ich mich bei Robert für die gemeinsamen *Schreibzeiten*, welche die Mittwoch-Nachmittage zur Garantie für produktives Schreiben machten.

Tiefer Dank geht an meine Familie. Danke, Mama und Papa, für eure schier endlose Hilfsbereitschaft und dafür, dass ich mich bedingungslos auf euch verlassen kann. Danke, dass ihr mir Freiräume schenkt und ich dennoch stets auf wertvollen Rat von euch zählen kann. Ich hoffe, ihr seid stolz auf mich. Vielen Dank auch an dich, Caro, für deinen Zuspruch, deinen Humor und deinen unerschütterlichen Glauben an mich. Und danke an meine liebenswerte, große Verwandtschaft. Ich habe mich zu jeder Zeit unterstützt gefühlt.

Nicht nur aber auch während meiner Promotion war der Sport für mich von besonderem Wert. Und noch mehr die Freundschaften, die daraus für mich entstanden. Ich danke euch allen. Beim Bouldern und Klettern fand ich zusammen mit Thea, Krissy, Veri und Christian immer einen Weg nach oben und dafür die nötige Trittsicherheit und den festen Griff. Beim Mountainbiken steuerten Alex, Andi, Jojo, Marinus, Kasu und ich gemeinsam durch schwierige Passagen und hielten dabei (fast) immer die Balance. Zum Kopf-frei-Kriegen wurden die Laufschuhe geschnürt. Die Runden mit Tommy, Theresa oder Eva taten immer gut. Flo, danke für die musikalischen Tipps. Tiefsten Dank auch euch, Hanna und Louisa, für euren seelischen Beistand und unser gemeinsames herzliches Lachen.

Und dir, Simon, gebührt mein aufrichtigster Dank für alles, was du für mich tust und für mich bist. Nur mit dir gemeinsam konnte ich so manche Krisen meistern, weil du mir die nötige Ruhe schenkst, tiefe Zuversicht vermittelst, lösungsorientierte Vorschläge bringst und tatkräftig an meiner Seite stehst. Es ist das Größte für mich, den Erfolg, diese Arbeit fertiggestellt zu haben, mit dir teilen zu dürfen.

# INHALTSVERZEICHNIS

<b>1. Einleitung</b>	<b>11</b>
<b>2. Physikunterricht und Sprache</b>	<b>17</b>
2.1. Funktionen der Sprache im Physikunterricht . . . . .	18
2.1.1. Denken anhand von Sprache . . . . .	18
2.1.2. Kommunizieren anhand von Sprache . . . . .	22
2.1.3. Fazit mit Blick auf den Physikunterricht . . . . .	26
2.2. Bildungs- und fachbezogene Sprachformen . . . . .	28
2.2.1. Begriffliche Einordnung von Bildungssprache, Unterrichtssprache, Fachsprache und Alltagssprache . . . . .	28
2.2.2. Die Kennzeichen von Fachsprache im Fach Physik . . . . .	32
<b>3. Wissensvermittlung anhand von Texten</b>	<b>37</b>
3.1. Verstehendes Lesen: Begrifflichkeiten . . . . .	38
3.2. Personenbezogene Abläufe und Voraussetzungen . . . . .	43
3.2.1. Teilprozesse beim Leseverstehen . . . . .	45
3.2.2. Einfluss personenbezogener Dispositionen . . . . .	47
3.2.3. Maßnahmen zur Leseförderung . . . . .	49
3.3. Verständliche Textgestaltung im Allgemeinen . . . . .	50
3.4. Stand der fachspezifischen Leseforschung . . . . .	58
<b>4. Forschungsinteresse</b>	<b>63</b>
4.1. Zentrale Begriffe und Leitgedanken der Studie zur Textverständlichkeit . . .	63
4.2. Die Forschungsfragen der Studie zur Textverständlichkeit . . . . .	68
<b>5. Rekonstruktion des Konstruktraums für Textverständlichkeit</b>	<b>73</b>
5.1. Die Ebene der sprachlichen Mittel . . . . .	76
5.2. Die Ebene der Textwirkung . . . . .	79
<b>6. Methode der empirischen Studie: Material und Durchführung</b>	<b>87</b>
6.1. Gesamtkonzeption der empirischen Studie . . . . .	87

6.2.	Texterstellung . . . . .	89
6.2.1.	Inhaltliche Konzeption der Texte . . . . .	91
6.2.2.	Erstellung der sprachlich variierten Textversionen . . . . .	93
6.3.	Testinstrumente . . . . .	101
6.3.1.	Messung der sprachlichen Fähigkeiten: der LGVT . . . . .	102
6.3.2.	Messung der Behaltensleistung: Entwicklung eines Wissenstests . . .	102
6.3.3.	Messung der Textwahrnehmung: Entwicklung eines Bewertungsfragebogens . . . . .	108
6.4.	Statistische Datenauswertung . . . . .	117
6.4.1.	Organisatorische Vorbereitungen für die Datenverarbeitung . . . . .	117
6.4.2.	Verwendete Software . . . . .	118
6.4.3.	Statistische Methoden im Einzelnen . . . . .	118
6.5.	Durchführung und Stichprobe . . . . .	121
6.5.1.	Zeitlicher und organisatorischer Ablauf der Studie in Schulen . . . . .	122
6.5.2.	Beschreibung der Stichprobe . . . . .	124
<b>7.</b>	<b>Ergebnisse und Diskussion</b>	<b>125</b>
7.1.	Ergebnisse aus dem Wissenstest . . . . .	125
7.1.1.	Deskriptivstatistische Ergebnisse aus dem Wissenstest . . . . .	126
7.1.2.	Prä-Post-Vergleiche im Wissenstest . . . . .	127
7.1.3.	Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung (ohne Moderatorvariablen) . . . . .	135
7.1.4.	Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung (mit Moderatorvariablen) . . . . .	146
7.1.5.	Zusammenfassung und summarische Diskussion der Wissenstest-Ergebnisse . . . . .	154
7.2.	Ergebnisse aus dem Bewertungsfragebogen . . . . .	157
7.2.1.	Deskriptivstatistische Ergebnisse aus dem Bewertungsfragebogen . . .	157
7.2.2.	Innere Struktur der affektiven Textwahrnehmung . . . . .	159
7.2.3.	Zusammenhang zwischen sprachlicher Gestaltung und Globalurteil . . .	172
7.2.4.	Zusammenhang zwischen sprachlicher Gestaltung und affektiven Wahrnehmungsfacetten . . . . .	178
7.2.5.	Zusammenfassung und summarische Diskussion der Bewertungsfragebogen-Ergebnisse . . . . .	187
7.3.	Zusammenhang zwischen affektiver Textwahrnehmung und Behaltensleistung	190
<b>8.</b>	<b>Resümee und Ausblick</b>	<b>195</b>
	<b>Literaturverzeichnis</b>	<b>201</b>

<b>A. Zu den Lesetexten für die empirische Studie</b>	<b>211</b>
A.1. Die Lesetexte . . . . .	211
A.2. Zusatzmaterial zu den Lesetexten . . . . .	214
A.2.1. Outputs des Textanalyseprogramms RATTE . . . . .	214
A.2.2. Liste mit allen variierten Textstellen . . . . .	214
<b>B. Zu den Testinstrumenten der empirischen Studie</b>	<b>221</b>
B.1. Zur Messung der Behaltensleistung: der Wissenstest . . . . .	221
B.2. Zur Messung der Textwahrnehmung: der Bewertungsfragebogen . . . . .	223
<b>C. Zum Konstruktraum für die sprachlichen Gestaltungsmittel</b>	<b>227</b>
<b>D. Statistischer Anhang</b>	<b>235</b>
D.1. Testkennwerte aus der Pilotierungsstudie . . . . .	235
D.2. Testkennwerte aus der Hauptstudie . . . . .	238
D.3. Robustheitsanalyse des Wissenstests . . . . .	241
D.4. Robustheitsanalyse des Bewertungsfragebogens . . . . .	246
D.5. Zusammenhangsanalysen sprachliche Gestaltung – Post-Test-Score . . . . .	258
D.6. Moderierte Zusammenhangsanalysen: sprachliche Gestaltung – Gain-Score	260
D.7. PCA für den Bewertungsfragebogen . . . . .	266
D.8. Moderierte Zusammenhangsanalysen: sprachliche Gestaltung – Globaleindruck	268
<b>E. Zum Seminar „Sprache im Physikunterricht“</b>	<b>275</b>



*Denn, was man schwarz auf weiß besitzt,  
kann man getrost nach Hause tragen.*

---

J. W. von Goethe,  
Faust: Der Tragödie Erster Teil

## 1 | EINLEITUNG

Vertrauenswürdigkeit, Beständigkeit, Verfügbarkeit – mit derartigen Werten verknüpft dieser Auszug aus der Tragödie *Faust I* (von Goethe, 1808, Vers 1966 f.) etwas Aufgeschriebenes in Form von schwarzer Tinte auf weißem Papier. Und zugleich verführt der Vers, sodann man ihn in Verbindung mit Bildung und Schule setzt, zu einem träumerischen Bild von Lernenden, die wichtige Lerninhalte in schriftlicher Form vor sich haben, die Bedeutung der Texte erfassen und in ihre alltägliche Lebenswelt mitnehmen. Im *Faust* erscheint die Bedeutsamkeit geschriebener Texte für das Lernen unhinterfragt, sind sich Mephistopheles und ein lernwilliger Schüler, mit dem er in der betreffenden Szene spricht, doch sofort einig, dass es aus Büchern zu lernen und im Unterricht mitzuschreiben gilt.

Laut Duden bedeutet die Wendung *schwarz auf weiß*: „zur Sicherheit, zur Bekräftigung schriftlich, sodass man sich darauf verlassen kann.“ („Duden“, o.J.). Ob man also der Lyrik Goethes folgt oder die Bedeutung der Wendung im alltäglichen Sprachgebrauch laut aktuellem Wörterbuch heranzieht: abgedruckte Informationen haben mit Verlässlichkeit und Glaubwürdigkeit zu tun.

Von diesem Standpunkt aus ist es nicht überraschend, dass der Umgang mit geschriebenem Text eine zentrale Rolle im Unterricht der verschiedenen Fächer einnimmt. Selbst in Zeiten des rasanten (digitalen) Wandels, der neuartige Unterrichtsmedien wie beispielshalber Videos mit sich bringt und stets verfügbar macht, bleibt der Text ein fundamentales Medium der Wissensvermittlung. Eine Auseinandersetzung mit der Frage, wie der Einsatz von Texten im Unterricht möglichst günstig verlaufen kann, ist daher von großer Bedeutsamkeit für das grundlegende Ziel, guten Fachunterricht zu ermöglichen. Eine moderne Auffassung von Unterricht gründet auf einer Denkfigur von Angebot und Nutzung (für eine kompakte Darstellung des zugehörigen Modells siehe Kleinbub, 2018; im Detail siehe die Ausführungen bei Helmke, 2015, S. 71 ff.). Damit hebt sich die heutige von der behavioristischen Unterrichtsphilosophie des frühen bis mittleren 20. Jahrhunderts deutlich ab (Klieme, 2006). Der Gegensatz besteht im Kern darin, dass nicht wie im Behaviourismus angenommen allein die Lehrperson mit ihren unterrichtlichen Handlungen und dem eingesetzten Unterrichtsmaterial als ursächlich für Wissens- und Verhaltensveränderungen bei Schüler\*innen gilt und dass Lernen nicht allein durch *Instruktion* geschieht, sondern die Interaktion von Lehrenden und

## 1. Einleitung

Lernenden in die *Konstruktion* von Wissen und den Aufbau von bildungsrelevanten Kompetenzen mündet (Unterricht als „Ko-Produktion“ der Beteiligten, siehe Klieme, 2006, S. 765). Dies entspricht der zeitgemäßen Auffassung von Konstruktion und Instruktion als komplementäre Unterrichtsformen, die weder in einem rein konkurrierenden Verhältnis zueinander stehen (Instruktion *versus* Konstruktion), noch schlichtweg im Unterricht koexistieren (Instruktion *und* Konstruktion). Am zeitgemähesten ist es demnach, im Unterricht Konstruktion *durch* Instruktion zu unterstützen (siehe dazu ausführlich Möller, 2012). Seitens der Schüler\*innen setzt dies einen aktiven Umgang mit Lerninhalten und mit Lernmaterial voraus.

Im Einvernehmen damit ist auch das Textverstehen interaktiv zu betrachten: Die vermittelnde Instanz bildet der Text. Lesende müssen den Text verstehen um daraus zu lernen, das heißt, die gebotenen Informationen in einen intratextuellen Zusammenhang bringen und sie mit über den Text hinausreichenden Wissensgebieten in Beziehung setzen. Das verstehende Lesen besitzt einen hochgradig aktiven Charakter (Lenhard, 2013, S. 27).

Es wird deutlich, dass ein günstiger Einsatz von Texten im Unterricht bedeutet, eine möglichst günstige Interaktion zwischen Lesenden und Texten zu ermöglichen. Für das Grundschulalter sieht der Lehrplan des Bundeslands Bayern hierfür einen expliziten Rahmen vor (Staatsinstitut für Schulqualität und Bildungsforschung, ISB, 2022a). Ein Blick auf die Lehrplaninhalte sowie Beobachtungen aus der Unterrichtspraxis (z. B. bei Holliday und Cain, 2012, oder Schilcher und Wild, 2018) legen jedoch die Feststellung nahe, dass Schülerinnen und Schüler zwar in der Primarstufe systematisch lernen zu lesen, Lernende der Sekundarstufe aber bereits lesen um zu lernen (siehe auch Fang, 2006, zum amerikanischen Schulsystem). Lernende der Sekundarstufe müssen weitestgehend selbstständig in der Lage sein, Informationen aus Texten zu entnehmen und zu verarbeiten. Der zentrale Begriff, aus dem das Ziel erwächst, Schülerinnen und Schüler unter anderem in der Sekundarstufe dabei zu unterstützen, aktive Leser\*innen zu werden und sich erfolgreich mit den Texten der einzelnen Schulfächer auseinander zu setzen, ist die Textverständlichkeit.

### **Der Blick auf die Lesenden**

*Textverständlichkeit* ist ein Kompositum, das die *Verständlichkeit eines Texts* bezeichnet. Damit legt der Begriff nahe, die Verständlichkeit auf ein Merkmal zu reduzieren, das ein Text objektiv hat oder nicht hat. Eine solche Auffassung ist jedoch fehlerhaft, da sich die Verständlichkeit eines Texts erst in der Interaktion einer lesenden Person mit dem Text manifestiert. Die Textverständlichkeit ist eine resultierende Wirkung aus dem Zusammenspiel von text- und leserseitigen Einflussfaktoren. Es ist ein zentrales Anliegen dieser Arbeit, klar zwischen objektiven Textmerkmalen, sogenannten *sprachlichen Mitteln* einerseits und Textwirkungen auf Lesende andererseits zu trennen. Die Wirkungen wiederum werden in dieser Arbeit in eine kognitive, auf die Behaltensleistung beim Lesen bezogene, und eine affektive, auf den subjektiven Eindruck beim Lesen bezogene Komponente differenziert. Texte verständlich zu machen heißt folglich, beide Seiten zu bedenken – Texte wie Lesende. Nur so rückt ins Zentrum der Zielsetzung, was Lesende mit dem Text *tun*. Für Lesende heißt das, sie müssen

günstige Voraussetzungen mitbringen und günstige Aktivitäten an den Tag legen, damit sie einen Text verstehend lesen können. Das wechselseitige Zusammenspiel dessen, was Autor\*innen äußern, mit dem, was Lesende daraus rekonstruieren, ist zunächst eine theoretische Implikation verschiedener Kommunikations- und Lesekompetenzmodelle. Eine weit verbreitete, sprachwissenschaftlich etablierte Grundlage für die Betrachtung von Kommunikationssituationen stellt beispielshalber das *Semiotische Dreieck* nach Ogden und Richards dar (Ogden und Richards, 1930, siehe dazu Abschnitt 2.1.2 in dieser Arbeit). Darüber hinaus ist es Bestandteil empirischer Forschung zum Textverstehen, sich mit dem Zusammenspiel aus geäußertem Text und Rekonstruktion der Lesenden zu befassen. Die Kapazitäten der Lesenden hinsichtlich ihrer Fähigkeiten zum Textumgang sind also mitentscheidend für das Leseverstehen (siehe dazu Unterkapitel 3.2).

Der Blick auf die Seite der Lesenden eröffnet zahlreiche Möglichkeiten, einen erfolgreichen Umgang mit Texten im Unterricht zu fördern und zu unterstützen (siehe dazu ausführlich Abschnitt 3.2.3 in dieser Arbeit). Zusammenfassend lässt sich sagen, dass sich unterrichtliche Maßnahmen zur Steigerung der Lesemotivation oder zur Verbesserung der Metakognition und Selbstregulation beim Lesen vor allem auf die Expertise eines (Lese-)Vorbilds (Lernen am Modell, reziprokes Lehren und Lernen) stützen und auf kooperative Bewertungs- und Erfahrungsdiskussionen innerhalb der Lernendengruppe setzen. So unterschiedlich die konkreten Zielsetzungen und die realisierten Praktiken der einzelnen Fördermaßnahmen auch sein mögen, ihnen allen ist gemein, dass sie geeignetes Textmaterial benötigen. Diesbezüglich betonen Artelt et al. (2005, S. 67 f.) den Bedarf nach Texten, die das (insbesondere alters- und geschlechterspezifische) Interesse der Lesenden wecken und bündeln.

### **Der Blick auf die Texte**

Einen Text verstehend zu lesen bedeutet, ausgehend von der augenscheinlich wahrnehmbaren Oberfläche des Texts ein geistiges Abbild seines Inhalts zu bilden (Lenhard, 2013, S. 14). In einem Text erscheinen Informationen in kodierter Form. Das entspricht gewissermaßen einer Anschauung des Textinhalts als zu entdeckende, zu entschlüsselnde Entität, die von den sichtbaren Zeichen und Symbolen *be-schrieben* (sic) wird und einhergehend mit persönlichen Inferenzen und Vorstellungen von Lesenden eine individuelle Bedeutung entfaltet. Inhalt und Oberfläche sind es demzufolge, die von Seiten des Texts Einfluss darauf nehmen können, wie ein Text auf Lesende wirkt (Bamberger & Vanacek, 1984, S.19). Oder andersherum: Ob ein Text ansprechend wirkt und zum Lernen anregt, hängt von inhaltlichen sowie von Aspekten der Textoberfläche ab. Ein Modell von Susanne Göpferich für die Textproduktion und -bewertung (Göpferich, 2002, S. 155 ff., siehe dazu ausführlich Unterkapitel 3.3) macht deutlich, dass beim Verfassen eines Texts die *inhaltlichen* Entscheidungen der Festlegung auf einen sprachlichen Code vorausgehen. Es lässt sich in gewissem Maß eigenständig und losgelöst vom Inhalt festlegen, welche Buchstaben, Wörter, Phrasen und Sätze letztendlich in welcher Reihenfolge niedergeschrieben sind, wie das Layout gestaltet wird und ob und in welchem Umfang Bilder verwendet werden. Anhand von Sprache werden die Wissensstruk-

## 1. Einleitung

turen des\*der Textautor\*in für Lesende zugänglich, sodass komplexe Inhalte in die eigenen Wissensstrukturen aufgenommen und eingegliedert werden können. Dabei muss die Sprache die zu vermittelnden Bedeutungseinheiten sach- und adressatengerecht darstellen. Das bedeutet, sie muss einerseits fachlichen Ansprüchen und Konventionen genügen und andererseits die Aufmerksamkeit der Lesenden gewinnen und bündeln können (Kulgemeyer & Schecker, 2009). Insbesondere ist zu vermeiden, dass aufgrund eines ungünstigen Sprachgebrauchs von sich aus komplexe Inhalte noch kryptischer und komplizierter für Schüler\*innen erscheinen. Sinnbildlich gesprochen soll Sprache als Brücke dienen und keine Barriere darstellen. Aus diesem Grund bietet die Untersuchung der Wirkungen, die sich durch unterschiedliche sprachliche Textgestaltung ergeben, eine Möglichkeit, dem Wunsch nach möglichst günstigen Texten für den Fachunterricht nachzukommen.

### **Textverständlichkeit unter physikdidaktischer Perspektive**

Im Kontext des Fachunterrichts ist es dabei nötig, sich der spezifischen Anforderungen verstehenden Lesens im Fach bewusst zu werden. Speziell in den Naturwissenschaften stellt die Auseinandersetzung mit Fachtexten einen wesentlichen Bestandteil des fachspezifischen Arbeitens dar und zählt daher explizit zu den prozessbezogenen Kompetenzen im Unterrichtsfach Physik. Zum Kompetenzbereich *Kommunizieren* zählt gemäß dem LehrplanPLUS für das Gymnasium in Bayern - Fachprofil Physik (ISB, 2022b) „das Verständnis und die sichere Verwendung der Fachsprache“ und, dass die Schülerinnen und Schüler „in der Lage [sind], sich physikalische Kenntnisse aus Fachtexten zu erschließen, Informationen zu recherchieren und Arbeitsergebnisse z. B. in Dokumentationen und Präsentationen sach- und adressatengerecht aufzubereiten“. Es ist unbestritten, dass naturwissenschaftliche Informationstexte spezielle und mitunter als besonders herausfordernd empfundene Ansprüche an Schülerinnen und Schüler stellen (siehe auch Unterkapitel 3.4. Eine ausführliche Abhandlung über die Herausforderungen rund um das *Science Reading* bietet Fang, 2006. Weitere Hinweise auf spezifische Anforderungen finden sich bei Starauschek, 2006, S. 128; Schmellentin, Dittmar, Gilg und Schneider, 2017, S. 73 ff.; Härtig, Bernholt, Prechtel und Retelsdorf, 2015, S. 58 ff. und Patterson, Roman, Friend, Osborne und Donovan, 2018, S. 292 und S. 296). Während der Lehrplan zwar die Beherrschung der Fachsprache fordert, lässt er unbestimmt, was ebendiese konkret ausmacht und von anderen Sprachformen unterscheidet. Für eine inwendige Definition lassen sich deskriptive Merkmalslisten z. B. aus korpuslinguistischen Analysen heranziehen. Für eine Abgrenzung nach außen – etwa gegenüber Alltagssprache, Unterrichtssprache und Bildungssprache – bietet sich die Einordnung in sprachtheoretische Rahmenmodelle an. In Unterkapitel 2.2 wird die Fachsprache daher sowohl anhand eines Rastermodells neben anderen Sprachregistern verortet als auch durch charakteristische Merkmale beschrieben.

Nicht zuletzt um Schüler\*innen dabei zu unterstützen, Fachsprache zu beherrschen und zu verstehen, müssen sich Lehrkräfte in den Naturwissenschaften sprachlicher Herausforderungen im Unterricht bewusst sein. Beispielshalber bei der Beurteilung und Auswahl beziehungsweise beim Verfassen passender Unterrichtstexte sind nicht nur fachlich-inhaltliche

Gesichtspunkte, sondern auch ein ausgeprägtes Bewusstsein für Sprache und insbesondere für fachsprachliche Spezifika relevant. Vor dem Hintergrund der Bildungsstandards, die die Leseförderung zum übergreifenden Ziel der einzelnen Fächer erklären (Kultusministerkonferenz, 2019), zeigt sich die Bedeutsamkeit und Aktualität der Thematik für die Fachdidaktik. So verfolgt zum Beispiel das Staatsinstitut für Schulqualität und Bildungsforschung in Bayern mit der Initiative *#lesen.bayern* das Ziel, Lehrkräfte, Schulleitungen und Eltern in der lesebezogenen Lehr- und Unterrichtspraxis zu unterstützen ([www.lesen.bayern.de](http://www.lesen.bayern.de)). Speziell in den Naturwissenschaften wurden im Zusammenhang mit dem Lesen im Fach aufbauend auf allgemeinsprachlichen, fach-unspezifischen Arbeiten zur Textverständlichkeit (siehe hierzu Unterkapitel 3.3) bereits diverse Untersuchungen zu den Auswirkungen sprachlich variierten Texte oder Textaufgaben auf den Wissenszuwachs beziehungsweise die Lösewahrscheinlichkeit durchgeführt. Die Befundlage ist jedoch oftmals erwartungswidrig oder uneindeutig und lässt Fragen offen (siehe Unterkapitel 3.4).

Diese Promotionsarbeit knüpft aus diesem Grund hier an. Sie legt insbesondere ein zentrales Interesse auf eine bisher kaum untersuchte Variable im Textverstehensprozess, nämlich auf die affektive Wahrnehmung, die Lesende von einem Text haben. Die Arbeit geht somit der Frage nach den Wirkungszusammenhängen zwischen sprachlicher Textbeschaffenheit, Behaltensleistung und einzelnen Facetten der affektiven Textwahrnehmung auf Seiten der Lesenden im Unterrichtsfach Physik nach. Gemäß der Prämisse, dass die affektive Wahrnehmung mediatorisch Einfluss darauf nimmt, inwiefern Lesende geneigt sind, sich engagiert mit einem Text zu beschäftigen, fokussiert diese Arbeit damit neben den kognitiven Lernerfolgskriterien vor allem auch den individuellen Eindruck vom Text, den Schüler\*innen beim Lesen gewinnen. Es sollen grundlegende Erkenntnisse darüber gewonnen werden, inwiefern die sprachliche Gestaltung eines Texts dazu beitragen kann, dass sich Lesende einen günstigen, lernförderlichen Eindruck vom Text erhalten. Welche sprachlichen Gestaltungsmittel bieten sich also an, damit physikalische Sachtexte für Schüler\*innen verständlich wirken und das Wissen hinter der Textoberfläche für Lesende zugänglich wird?

Die Arbeit ist in das *Impuls+*-Projekt der Universität Regensburg eingebunden, das im Rahmen der *Qualitätsoffensive Lehrerbildung* des Bundesministeriums für Bildung und Forschung gefördert wird. Das Ziel des Projekts ist es, aus der Richtung verschiedener Fächer den Umgang mit heterogenen Lernvoraussetzungen im Unterricht in den Blick zu nehmen. Diese Arbeit spezialisiert sich auf sprachliche Aspekte und möchte herausarbeiten, welche Lernvoraussetzungen sich in Hinsicht auf die Auseinandersetzung mit schriftlichem Textmaterial im Fach Physik in Abhängigkeit vom sprachlichen Inventar der Informationstexte ergeben.

Im Folgenden führt die vorliegende Dissertationsschrift zunächst aus, auf welche theoretischen Grundlagen sich das Forschungsinteresse der Arbeit stützt. Den Ausgangspunkt bilden dabei zunächst Ausführungen über die allgemeine Bedeutung von Sprache für den

## *1. Einleitung*

Physikunterricht (Kapitel 2). Eine spezielle Darstellungsform von Sprache, der im Unterricht große Relevanz zukommt, ist geschriebener Text. Die Wissensvermittlung anhand von Texten ist Gegenstand des Kapitels 3. Es werden darauf aufbauend die Forschungsfragen herausgearbeitet (Kapitel 4). In Kapitel 5 wird die Rekonstruktion des Konstruktraums beschrieben, in dem sich die untersuchten Variablen der Textverständlichkeit verorten lassen. Die Konzeption und Umsetzung einer empirischen Studie zur Textverständlichkeit im Fach Physik, die im Rahmen dieser Promotion durchgeführt wurde, wird in Kapitel 6 beschrieben. Abschließend werden zunächst die Ergebnisse präsentiert und diskutiert und in Beziehung zum aktuellen Forschungsstand gesetzt (Kapitel 7), bevor im Ausblick die Bedeutsamkeit der Ergebnisse und mögliche Anschlussfragen aufgezeigt werden (Kapitel 8).

## 2 | PHYSIKUNTERRICHT UND SPRACHE

Sprache und Denken sind ein unzertrennliches Paar. Das Zitat des österreichischen Philosophen Ludwig Wittgenstein, das diesem Kapitel voransteht, stellt diesbezüglich einen interessanten Standpunkt dar. Es besagt, dass die Sprache eines Menschen ebenso weit reiche wie das eigene Wissen, die Kenntnisse, das Bewusstsein, die eigene *Welt*. Mit anderen Worten: Bis zum eigenen Horizont sei alles Wahrnehmbare mit Sprache fassbar und genauso habe alles, was mit Sprache fassbar ist, einen Platz in der eigenen Welt. Und tatsächlich betrachtet die Kognitionspsychologie die Verschränktheit von Wahrnehmung und Sprache als einen von vielen wichtigen Aspekten des Zusammenhangs zwischen Sprache und Denken (Solso, 2005, S. 294 ff.).

Das folgende Kapitel wird die einleitende Behauptung der Unzertrennlichkeit von Sprache und Denken einerseits stützen und andererseits in einen unterrichtlichen Kontext bringen. Es wird deutlich werden, inwiefern das Lernen im Unterricht ein Lernen von und mit Sprache ist. Denn erwiesenermaßen hängt der Lernerfolg in der Schule von sprachlichen Fähigkeiten ab. Mangelnde sprachliche Fähigkeiten behindern das Lösen schulischer Aufgaben und den selbstständigen Wissenserwerb (Lenhard, 2013, S. 11 f.) und schmälern allgemein die Chancen auf Bildungserfolg (siehe z. B. Köller et al., 2022, S. 19; Gogolin & Duarte, 2016). Die „Grenzen der eigenen Welt“ werden enger.

Wenn es um Sprache im Unterricht geht, geht es jedoch nicht nur um die Verbindung zur Kognition (epistemische Funktion der Sprache), sondern auch darum, Fachinhalte für Schüler\*innen zugänglich zu machen und sich gemeinsam über Fachliches zu unterhalten (mediale Funktion der Sprache). In diesem Zusammenhang ist oftmals von Unterrichtssprache (z. B. Feilke, 2013) oder Bildungssprache (z. B. Gogolin und Duarte, 2016, oder Morek und Heller, 2012) die Rede. Für die einzelnen Schulfächer ist außerdem die jeweilige Fachsprache relevant. Diese bringt neben der epistemischen und der medialen Funktion noch den Aspekt in den Unterricht, dass die Sprache selbst ein Lerngegenstand ist. Die Sprache tritt also nicht nur in verschiedenen *Funktionen* auf, es sind auch verschiedene *Sprachformen* Teil der Unterrichtspraxis.

Die folgenden Ausführungen in Unterkapitel 2.1 beleuchten zunächst die epistemische und die mediale Funktion von Sprache im Unterricht. Im Anschluss wird die physikalische

(beziehungsweise treffender: naturwissenschaftliche) Fachsprache neben weiteren unterrichtsrelevanten Sprachformen verortet und anhand typischer Merkmale charakterisiert (Unterkapitel 2.2).

## **2.1 Funktionen der Sprache im Physikunterricht**

In den einleitenden Zeilen zum aktuellen Kapitel 2 wurde der Gedanke geäußert, dass das Lernen im Unterricht ein „Lernen von und mit Sprache“ sei (siehe Seite 17). Dabei ist das „Lernen mit Sprache“ in zweierlei Weise zu deuten: Bezogen auf die kognitiven Lernprozesse von Schüler\*innen ist damit gemeint, dass die Sprache ein Mittel oder „Werkzeug“ des Denkens ist. Sie erfüllt damit eine epistemische Funktion. In Hinblick auf den kommunikativen Charakter von Unterricht bedeutet der Ausdruck, dass die Sprache ein Mittel des Wissenstransfers ist und gewissermaßen den „Baustoff“ für die übermittelten Botschaften darstellt (Leisen, 2005, stellt hier den Begriff des „Konstruktionsmittels“ kontrastierend dem Begriff des „Transportmittels“ gegenüber, um zu betonen, dass aktive Sprachverarbeitungsprozesse bei den rezipierenden Personen nötig sind). Sie erfüllt also auch eine mediale Funktion. Die folgenden beiden Abschnitte 2.1.1 und 2.1.2 wenden sich der Reihe nach diesen Funktionen von Sprache im Unterricht zu. Der abschließende Abschnitt 2.1.3 zieht ein Fazit hinsichtlich der Relevanz von Sprache im (Physik-)Unterricht und geht dabei insbesondere auf den Aspekt der Sprache als Lerngegenstand („Lernen von Sprache“) ein.

### **2.1.1 Denken anhand von Sprache**

Die Zusammenhänge von Denken und Sprache waren schon immer und sind bis heute ein zentrales Interessensgebiet der Kognitionspsychologie. Schon seit den Anfängen der modernen Psychologie (etwa Ende des 19. Jahrhunderts) stellt die Frage, inwieweit Sprache und Denken verknüpft sind, ob das Eine ohne das jeweils Andere überhaupt möglich sei, ein intensiv und kontrovers diskutiertes Thema dar. Das Verhältnis von Sprache und Denken hier in wenigen Zeilen ausschöpfend zu beschreiben sprengt dementsprechend bei Weitem den Rahmen des Möglichen. Dennoch können einige zentrale Aspekte aufgezeigt werden, anhand derer die Relevanz von Sprache für das Denken und damit für das Lernen deutlich wird.

Solso (2005, S. 294 ff.) differenziert hinsichtlich des Zusammenhangs zwischen Sprache und Denken verschiedene Aspekte. Sie werden hier bündig zusammengefasst; für vertiefte Erläuterungen sei auf die Originalquelle verwiesen. Besonders grundlegend ist der erste Aspekt, der auch schon in der obenstehenden Einleitung dieses Kapitels 2 Erwähnung fand – nämlich die Abhängigkeit und gegenseitige Beeinflussung von Sprache und Wahrnehmung. Wie ein Mensch über eine Sache spricht, verändert die Art, wie er die Sache wahrnimmt. Umgekehrt entwickelt sich die eigene Sprache maßgeblich auf der Grundlage wahrgenommener, äußerer

Einflüsse<sup>1</sup>. Ein zweiter Aspekt des Zusammenhangs von Sprache und Denken liegt in der Nutzung der Sprache als Mittel des Informationsaustauschs (dass die Sprache somit auch als *Kommunikationsmittel* dient, ist Thema des Abschnitts 2.1.2). Die Auf- und Übernahme von Information ins Gehirn ist auf Prozesse der Sprachverarbeitung angewiesen, indem beispielshalber die grammatikalische Ordnung von Sätzen aufgeschlüsselt und begriffliche Assoziationen aufgesucht, aktiviert und verknüpft werden müssen. Der Zusammenhang zwischen Sprache und Denken wird drittens deutlich, wenn die Sprache als *Werkzeug des Denkens* verwendet wird: Das Lösen von Problemen findet oftmals innerlich verbalisiert statt. Zwar gibt es Problemstellungen, die auf ganz konkrete oder bildliche Sachverhalte/Objekte bezogen sind, sodass die Denkprozesse beim Lösen solcher Aufgaben unter Umständen nicht auf sprachlich-abstrahierende Repräsentationsformen im Gedächtnis angewiesen sind (siehe dazu auch Beyer & Gerlach, 2018, S. 2 f.). Im Allgemeinen läuft das Problemlösen jedoch losgelöst von äußeren Reizen und stattdessen in Verbindung mit innerem oder auch lautem Sprechen ab. An diesen dritten Aspekt des verbalisierten Problemlösens knüpft direkt die Möglichkeit der mentalen Abstraktion anhand von Sprache an. Mit der Sprachentwicklung beim Menschen geht eine abstrahierende Begriffsbildung einher: Indem Menschen Sprache benutzen, müssen sie sich nicht ausschließlich auf konkrete Objekte beziehen, sondern sind in der Lage, Objekte anhand von Begriffen zu klassifizieren und somit Bezüge auf abstrakter Ebene herzustellen. Beyer und Gerlach führen dazu aus, dass die produktive und rezeptive Nutzung von Zeichensystemen, die Herausbildung und Ausschärfung von Begriffen auslöst. Erfundene Wörter dienen dazu, dass zugehörige Begriffe gefunden werden. Anhand von inneren, kognitiven Prozessen werden also Begriffe generiert, welche den Wörtern Bedeutung geben und insbesondere eine Klassifizierung der bezeichneten Bezugsobjekte vornehmen (vergleiche Beyer & Gerlach, 2018, S. 23 f.).

Aus psychologischer Sicht kann man sich die Wissensrepräsentation<sup>2</sup> im menschlichen Gedächtnis als ein Netz aus solchen Begriffen und aus Relationen zwischen den Begriffen vorstellen. Dieses sogenannte *semantische* Gedächtnis stellt jenes Gedächtnissystem dar, welches deklaratives und konzeptuelles Wissen enthält<sup>3</sup>. Deklaratives Wissen beschreiben Anderson und Krathwohl (2001, S. 29) als Basiselemente einer Disziplin, die es zu kennen gilt, um mit der Disziplin vertraut zu sein (im Original: „*to be acquainted with a discipline*“) und Probleme innerhalb der Disziplin lösen zu können. Unter konzeptuellem Wissen verstehen die Autoren

---

<sup>1</sup>Es sind verschiedene Annahmen darüber verbreitet, in welchem Ausmaß sich Sprachentwicklung tatsächlich durch Umwelteinflüsse beeinflussen lässt. Der behaviouristischen Theorie nach B. F. Skinner zufolge erfolgt die Sprachentwicklung durch Imitationsverhalten und äußere Verstärkung. N. Chomsky hingegen geht davon aus, dass dem Menschen eine Neigung zu Sprache angeboren sei. Diese Neigung wiederum sei die Begründung für eine *Universalgrammatik*, welche allen Sprachen gemein ist (siehe dazu Solso, 2005, S. 303 ff.).

<sup>2</sup>Als Wissen sei hier in Anlehnung an Solso (2005, S. 242) das Ergebnis einer *Speicherung, Integration und Organisation von Information im Gedächtnis* verstanden.

<sup>3</sup>Es existieren noch weitere Gedächtnissysteme für z. B. prozedurales oder episodisches Wissen (siehe Urhahne, Dresel & Fischer, 2019, S. 25 ff.). Auch werden noch weitere Wissensarten unterschieden (Anderson & Krathwohl, 2001, S. 27 ff.). Auf diese wird jedoch jeweils hier nicht näher eingegangen.

(funktionelle) Abhängigkeiten zwischen diesen Basiselementen innerhalb einer größeren Struktur. Zum deklarativen und konzeptuellen Wissen zählen beispielshalber allgemeine Fakten (Wie viele Beine hat ein Hund?), Kausalitäten sowie wechselseitige Beziehungen (UV-Strahlung ist schädlich für die Haut.) und Wortbedeutungen (Was bedeutet „Emmergenz“?) (Urhahne, Dresel & Fischer, 2019, S. 30 f.). Für den *Aufbau* solcher Wissensstrukturen spielt die Sprache eine entscheidende Rolle. Man spricht von der *epistemischen* Funktion der Sprache (Morek & Heller, 2012), wenn die Sprache in diesem Sinne als *Werkzeug des Denkens* aufgefasst wird (Vollmer & Thürmann, 2010).

Speziell in den Naturwissenschaften gelten das sprachliche und das fachliche Lernen als eng verknüpft (Busch & Ralle, 2013; Rincke, 2010; Leisen, 2003). Nach Halliday (2004) sei die Aneignung naturwissenschaftlicher Sprache (im Sinne fachsprachlicher Wörter und Wendungen) direkt an ein fachliches Begriffs-, Konzept- und Zusammenhangslernen geknüpft. Dieser Annahme folgend bedeutet korrekt mit Fachsprache umzugehen (rezeptiv und produktiv), die Bedeutung des Fachvokabulars zu erfassen und zugrundeliegende Konzepte zu verstehen und zu verknüpfen. Und auch umgekehrt bewirkt fachbezogene Kommunikation ein fachliches Lernen. Bei Patterson, Roman, Friend, Osborne und Donovan (2018) heißt es hierzu in Bezug auf die Vermittlung naturwissenschaftlich etablierter Vorstellungen („Ideen“) an Schüler\*innen: „[S]cience is fundamentally a *set of ideas* about the natural world – ideas that are not readily accessible only from empirical observation, but rather, must also be learnt through reading and writing.“ (Patterson, Roman, Friend, Osborne & Donovan, 2018, S. 292, Hervorh. im Original). Die Naturwissenschaft als *set of ideas* aufzufassen leuchtet ein, wenn man sich vor Augen führt, dass sie sich häufig mit abstrakten Inhalten befasst und dementsprechend auch ihre Begriffe abstrakten Charakter tragen (Fang, 2006). Naturwissenschaftliche Begriffe beziehen sich oft auf etwas Unsichtbares, das nicht direkt erfahrbar ist (z. B. Energie, elektrische Spannung, elektromagnetische Schwingung) oder sie bergen Generalisierungen (z. B. beschleunigte beziehungsweise gleichförmige Bewegung, Potential(feld), Erhaltungs-/Prozess-/Zustandsgröße). Wenn also die Begriffe der Naturwissenschaften – und der Physik im Besonderen – Bezug auf *Ideen über die Natur* nehmen, die über konkret Beobachtbares hinausgehen, liegt es nahe, dass der Aufbau von Verständnis sowie die Ansichten darüber, was als fachlich korrekt gilt, vor allem in den Naturwissenschaften stark dadurch beeinflusst sind, wie über eine Sache gedacht und gesprochen wird (Leisen, 2005)<sup>4</sup>.

Dennoch darf der Aufbau von Verständnis auf der Grundlage von Sprechweisen nicht

---

<sup>4</sup>Folgendes Beispiel soll diese Sichtweise illustrieren:

Das Standardmodell der Teilchenphysik ist ein Modell, in dem der aktuelle Erkenntnisstand in der Physik zum Aufbau der sichtbaren Materie enthalten ist. Gemäß dem Standardmodell setzen sich die *hadronischen* Materieteilchen (z. B. Protonen und Neutronen) aus sogenannten *Quarks* zusammen. Die Vorstellung von Quarks ermöglichte die Klassifizierung des *Teilchenzoos*, einer großen Menge beobachteter Hadronen. Die Vorstellung entstand somit, ohne jemals zuvor Quarks beobachtet zu haben. Die letztendliche Benennung von drei Quarkfamilien war also eine *Idee*, die eine passende Systematik für die tatsächlich beobachteten Hadronen und ihre Eigenschaften mitbrachte. Sie bedeutete damit einen hervorragenden Erklärungsrahmen mit starker Vorhersagekraft und wurde gewissermaßen zur physikalisch akzeptierten, etablierten Vorstellung über den Aufbau von Hadronen.

damit gleichgesetzt werden, zu lernen, wie Dinge zu *benennen* sind. Der Physiker und Nobelpreisträger Richard Feynman verdeutlicht den Unterschied zwischen *knowing something* und *knowing the name of something* in einem weitbekannten Interview mit den folgenden Worten in eingängiger Weise<sup>5</sup>: Über eine Sache zu lernen bedeute mehr, als zu wissen, wie sie (z. B. in verschiedenen Landessprachen) genannt wird. Den Namen einer Sache zu kennen bedeute vielmehr, etwas über *Personen* zu wissen: nämlich, wie sie die Sache nennen<sup>6</sup>. Eine korrekte Benennung macht also lediglich einen Teil des Wissens über eine Sache aus.

In naturwissenschaftlichen Lehr-Lern-Prozessen kann es bisweilen bewusst geschehen und für Lehrkräfte sinnvoll sein, eine Sache zu benennen, ohne ein fundamentales Verständnis bei den Lernenden aufbauen zu wollen. Eine schlichte *Benennung* kann dazu dienen, wissentlich den Deckmantel des Schweigens über tieferliegende und komplexere (kausale) Zusammenhänge zu legen und damit Ablenkung zu vermeiden. Beispielshalber könnte ein Ziel zu Beginn des Unterrichts über Elektrizitätslehre sein, dass Schüler\*innen einen einfachen Stromkreis aus Netzteil, Draht und Glühlämpchen korrekt beschreiben, indem sie das Zusammenwirken von elektrischer Spannung, elektrischem Strom und elektrischer Ladung erläutern. Dabei genügt es wohl, die Bauteile wie z. B. das Netzteil zutreffend zu *benennen* ohne detailliert deren Aufbau und Funktionsweise zu kennen. Jedoch kann eine einfache *Benennung* im Sinne einer „leeren Phrase“ auch die unerwünschte Verschleierung von Wissenslücken bedeuten. Beschreibt beispielshalber eine Physikstudentin den Phasenübergang von Wasser aus der flüssigen in die gasförmige Phase, indem sie erläutert: „Aus *energetischen Gründen* verdampft Wasser bei 100 °C.“, so bleibt fraglich, inwiefern sie den Ausdruck „energetische Gründe“ nutzt, um die Vorgänge auf mikroskopischer Ebene (Steigerung der mittleren kinetische Energie, Lösen von Bindungen) treffend zusammenzufassen oder ob der Ausdruck für sie inhaltsleer ist und sie ihn lediglich nachspricht, weil sie ihn oft genug in diesem oder ähnlichen Zusammenhängen gehört hat.

Zuletzt sei ein Problem erwähnt, das sich ergibt, wenn übliche Sprech- und fachlich korrekte Denkweisen nicht zur Deckung kommen. Oftmals stehen typische Fehl- oder Alltagsvorstellungen in Verbindung mit gängigen, aber unpräzisen oder irreführenden Ausdrücken. So suggeriert der Terminus *Wärmekapazität* fälschlicherweise, dass die Wärme eine speicherbare Größe sei; der (*elektrische*) *Verbraucher* scheint elektrischen Strom zu verbrauchen; und in einem *Kraftwerk* müsste dem Wortsinn nach Kraft erzeugt werden. Durch solche Ausdrucksweisen liegen also Vorstellungen nahe, die quer zur fachlich zutreffenden Denkweise liegen. Der Umgang mit Fehl- und Alltagsvorstellungen und damit verbundene passende Sprechweisen sind daher zentral für das Lernen und Lehren von Physik und somit für die Gestaltung von Physikunterricht (Schecker & Duit, 2018; Wilhelm & Schecker, 2018).

---

<sup>5</sup>Das Interview mit dem oft zitierten Ausspruch „I learned very early the difference between knowing the name of something and knowing something.“ findet sich im Internet z. B. unter [https://youtu.be/px\\_4TxC2mXU](https://youtu.be/px_4TxC2mXU)

<sup>6</sup>Hier schwingt der Gedanke mit, dass Sprache eine *soziale* Funktion erfüllen kann, indem sich eine bestimmte Personengruppe durch ihren Sprachgebrauch auszeichnet. Denn sich zur Gemeinschaft der Naturwissenschaftler\*innen zugehörig zu fühlen ist unter anderem damit verbunden, die naturwissenschaftliche Fachsprache zu beherrschen. Siehe dazu auch Seite 31 in Abschnitt 2.2.1

Ausgehend von der allgemeinen Verbindung zwischen Sprache und Denken zeigt sich also, dass der Sprache eine zentrale Funktion für das Lehren und Lernen speziell in den Naturwissenschaften zukommt.

### 2.1.2 Kommunizieren anhand von Sprache

Neben der epistemischen Funktion kommt der Sprache eine weitere, besonders offensichtliche Funktion zu: Sprache dient den Menschen als Kommunikationsmittel. Es liegt auf der Hand, dass auch diese mediale Funktion der Sprache von großer Relevanz für den Unterricht aller Schulfächer ist. Auch in Bezug auf Kommunikation gilt jedoch, dass im Rahmen dieser Arbeit kein umfassender Überblick über den Stand der Forschung gegeben werden kann. Die Kommunikation wird aus Sicht vielfältiger Wissenschaftsdisziplinen wie der Sozialpsychologie, der Semiotik, der Neurobiologie oder der Medienwissenschaft untersucht, wobei jede Disziplin ihre eigenen Theorien und Modelle für Kommunikation mitbringt. Für diese Arbeit, in der es um die Wissensvermittlung anhand geschriebenen Texts geht, bietet sich der Blickwinkel der Semiotik (Zeichenlehre) an, um Funktion der Sprache für die Kommunikation insbesondere im Unterricht zu beleuchten.

Kommunikation lässt sich als Mitteilung von Informationen verstehen, die eine sprachproduzierende Person einer rezipierenden Person vermittelt<sup>7</sup>. Dazu werden Zeichen (Buchstaben, Symbole, Lautfolgen, Wörter, ...) genutzt, die für einen Referenten (Bezugsobjekt, das heißt ein Gegenstand oder Sachverhalt) stehen. Für die Beziehung zwischen Zeichen und Referent ist das *Semiotische Dreieck* nach Ogden und Richards (1930) eine in der Semiotik weit verbreitete Anschauung (S. 18 ff.). Das Modell ist in Abbildung 2.1 dargestellt. Ogden und Richards postulieren, dass die Verbindung zwischen Zeichen und Referent keine echte, eigenständige Verbindung darstelle, sondern lediglich indirekt über das Mentale aufgebaut (im Original: *imputed*) werde. In der Abbildung wird dies anhand der gestrichelten Grundlinie des Dreiecks veranschaulicht. Zeichen und Referent stehen aber jeweils in direkter Beziehung zu einer dritten Instanz, dem Gedanken beziehungsweise Bezug. Das, was die Zeichen symbolisieren beziehungsweise kausal hervorrufen, liege zunächst im Mentalen der sprachproduzierenden beziehungsweise rezipierenden Person. Mit dieser Beziehung greift das Semiotische Dreieck die Anschauungen de Saussures<sup>8</sup> auf. In dessen Zeichenlehre besteht ein Zeichen nicht in etwas Substantiellem wie einer Lautfolge oder geschriebenen Symbolen, sondern ein Zeichen ist relational zu betrachten. Es besteht in ebendieser Beziehung zwischen dem Bezeichnenden („*signifiant*“, psychischer Reflex eines Lautbilds) und dem Bezeichneten („*signifié*“, mental aktivierte Vorstellung). Ein Zeichen ist damit eine dyadische Relation, die rein im Mentalen besteht – und zwar zwischen dem Bezeichnenden auf der Ausdrucksseite

---

<sup>7</sup>Vergleiche dazu das Sender-Empfänger-Modell von Shannon und Weaver (1964): Eine aus einer *Quelle* stammende Information oder Nachricht wird von einer *sendenden* Person als Signal durch einen *Kanal* an eine *empfangende* Person, die das *Bestimmungsziel* der Nachricht ist, übermittelt.

<sup>8</sup>Ferdinand de Saussure ist ein bedeutender Sprachwissenschaftler, dessen Vorlesungsreihe *Cours de Linguistique Général* zu den Standardwerken der modernen Linguistik zählt: de Saussure, 1967 (dt. Ausgabe)

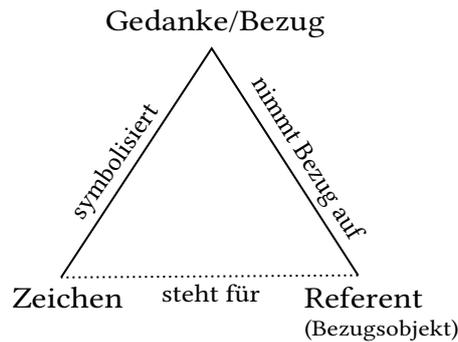


Abbildung 2.1.: Semiotisches Dreieck nach Ogden und Richards. Die durchgezogenen Linien repräsentieren direkte Verbindungen, während die gestrichelte Linie verdeutlicht, dass zwischen Zeichen (*symbol* im Original) und außersprachlichem Bezugsobjekt (Referent; *referent* im Original) nur eine mittelbare Verbindung besteht. Diese wird über eine mentale Vorstellung (Gedanke/Bezug; *thought or reference* im Original) aufgebaut. *Anmerkung: Die Elemente an den Eckpunkten dieser dreiteiligen semiotischen Relation werden in verschiedenen Quellen verschieden genannt. Diese Abbildung hier richtet sich nach der Darstellung in Ogden und Richards (1930, S. 10) mit eigenen Übersetzungen vom Englischen ins Deutsche.*

und dem Bezeichneten auf der Inhaltsseite (Köller, 1977, S. 22 f.). Im Semiotischen Dreieck wird jene Ausdrucksseite als *Gedanke* gefasst und in unmittelbarer kausaler Beziehung zum Referenten betrachtet, das heißt zu einem außersprachlichen Bezugsobjekt, über welches kommuniziert wird. Der Referent kommt bei de Saussure nicht explizit vor; die indirekte Verbindung zwischen Referent und Zeichen ist ein wesentlicher Baustein der semiotischen Theorie nach Ogden und Richards.

Dass Zeichen und Referent als indirekt verbunden gelten können, hat insbesondere Folgen für die rezeptive Seite einer Kommunikation. Anderson, Reynolds, Schallert und Goetz (1977) plausibilisieren theoretisch und experimentell ihre Behauptung: „The meanings of the individual words in a sentence clearly depend upon the interaction of world knowledge and context“ (S. 368). Die Verbindung zwischen Zeichen und Referent kann also nicht nur als indirekt (siehe oben; nach Ogden und Richards, 1930), sondern auch als uneindeutig und fragil betrachtet werden. Dies leuchtet anhand der folgenden Überlegungen ein:

- Mit Veränderung des Code-Systems (etwa: deutsche/englische/chinesische/... Sprache) können sich Wörter ändern, während die realen Referenten, die beschrieben werden, dieselben bleiben.
- Die Bedeutung eines Worts ist abhängig vom Kontext, in dem es auftaucht. Besonders offensichtlich wird das anhand von Homonymen (mehrdeutigen Wörtern): Wenn von „der Umlaufbahn der *Erde* um die Sonne“ die Rede ist, wird sicherlich ein ganz anderes Bild von der „*Erde*“ vor dem inneren Auge entstehen, als wenn jemand von „*Erde* in einem Blumenbeet“ spricht.
- Das Phänomen, dass die Beziehung zwischen einem Wort und dem damit verbundenen Referenten kontextabhängig ist, ist nicht auf Homonyme beschränkt (vergleiche hierzu

Anderson, Reynolds, Schallert & Goetz, 1977, S. 368). Sowohl der mentale Bezug als auch der Referent, die zu dem Wort „Schuh“ gehören, sind in den folgenden beiden Sätzen unterschiedlich, obwohl das Wort „Schuh“ nicht im eigentlichen Sinne mehrdeutig gebraucht wird: „Dank eines verlorenen *Schuhs* fand der Prinz sein Aschenputtel wieder.“ und „Auf steinigen Wanderwegen sind geeignete *Schuhe* besonders wichtig.“

Inwiefern demzufolge das Vorwissen der Adressat\*innen einer Sprachhandlung und somit insbesondere die Aktivierung dieses Vorwissens relevant für den Verstehensprozess beim Lesen von Lehrtexten ist, wird auf Seite 48 in Abschnitt 3.2.2 erläutert.

Eine Folge der indirekten Verbindung zwischen Zeichen und Referent für die produktive Seite der Kommunikation ist, dass die gemeinte Bedeutung und die intendierte Wirkung einer Sprachhandlung vom resultierenden mentalen Bezug und der resultierenden Wirkung der Rezipient\*innen zu differenzieren sind: „When we hear what is said, the symbols both cause us to perform an act of reference and to assume an attitude which will, according to circumstances, be more or less similar to the act and the attitude of the speaker“ (Ogden & Richards, 1930, S. 11). Für eine gelingende Kommunikation müssen Intention und Resultat zur Deckung kommen. Dabei stellt sich auf der Seite der sprachproduzierenden Person die Frage, ob sie „objektiv [sagt], was sie subjektiv meint zu sagen“ (Salzborn, 2016, S. 269, über John Searles *Speech Acts*<sup>9</sup>). Denn Gesagtes und Gemeintes sind im Allgemeinen verschieden (z. B. gesagt: „Es zieht!“ und gemeint: „Bitte schließe das Fenster.“<sup>10</sup>). Maßgeblich für das „Gesagte“ sind bekannte und eingeübte Formulierungskonventionen sowie situative Umstände (vergleiche die bündige Darstellung der pragmatischen Komponente sprachlicher Handlungen bei Beyer & Gerlach, 2018, S. 25 ff.). Ogden und Richards (1930, S. 10 f.) erläutern, dass die Wahl des Zeichensystems beim Sprechen durch den eigenen mentalen Bezug (*reference*), den Zweck der Sprachhandlung (*purpose*), die beabsichtigte Wirkung (*proposed effect*) der genutzten Zeichen und durch die eigene Haltung (*own attitude*) bestimmt sei.

Die Auswahl des Zeichensystems (im Sinne von konkreten Formulierungen), die sich nach Referent und nach intendierter Wirkung auf die empfangende Person richtet, findet sich auch in einem Modell für Kommunikationskompetenz im Fach Physik wieder. Kulgemeyer und Schecker (2009) differenzieren die *sachgerechte* und die *adressat\*innengemäße* Perspektive bei der Konstruktion eines Kommunikats. Als „Kommunikat“ (S. 135) bezeichnen die Autoren das Angebot, das rezipierende von produzierenden Kommunikationspartner\*innen erhalten; es sei „mit Kommunikateigenschaften (thematische, referentielle und stilistische Eigenschaften; dies entspricht dem Inhalt).“ ausgestattet und anhand von „Kommunikatbasen (Bilder, Schrift, gesprochene Sprache)“ produziert. Hinsichtlich beider Perspektiven – sachgerecht und adressat\*innengemäß – seien die vier Aspekte Kontext, Code, Darstellungsform und Sachaspekt maßgeblich für den Erfolg eines Kommunikationsvorgangs. Die Sachgerechtigkeit

---

<sup>9</sup>John R. Searle (geboren 1932 in Denver, Colorado) ist Autor zahlreicher bedeutender Werke der Sprachphilosophie. Dazu zählt das Buch *Speech Acts: An Essay in the Philosophy of Language*.

<sup>10</sup>In diesem Beispiel wird ein *illokutionärer Akt* (Sammelbegriff für Behaupten, Fragen, Befehlen, Versprechen) bei der Sprachhandlung vollzogen; siehe dazu Salzborn (2016)

sei dadurch bestimmt, inwiefern ein Kommunikat den Ansprüchen entspricht, die der Sachinhalt mit sich bringt. Kulgemeyer und Schecker (2009) zufolge zählen hierzu die fachliche Adäquanz, Korrektheit und Konsistenz und die Beachtung fachlicher Konventionen. Jedoch hängen Formulierungsentscheidungen auch unmittelbar mit der Vermittlungsabsicht, dem Zweck und der intendierten Wirkung einer Botschaft zusammen (adressat\*innengemäße Kommunikation). Bezogen auf ein mündliches oder schriftliches Gespräch heißt das, dass der Inhalt allein nicht die passende Formulierung festlegt<sup>11</sup>.

Der Inhalt lässt also Freiräume für die sprachproduzierende Seite offen, wenn es um konkrete Formulierungsentscheidungen geht. Insbesondere im Zusammenhang mit Lehrtexten gibt die Sachstruktur nicht zwingend und endgültig vor, wie der zugehörige Lehrtext optimalerweise zu strukturieren sei (Rincke, 2010). Um Lesende dabei zu unterstützen, die Sache kognitiv zu durchdringen, müssen Textautor\*innen den verbleibenden Gestaltungsspielraum passend nutzen. Die Formulierung der zu vermittelnden Botschaft muss an einer adäquaten didaktischen Rekonstruktion der Sache<sup>12</sup> und an den Fähigkeiten der Adressat\*innengruppe orientiert sein. Entsprechend bringt Rincke (2010) mit dem Vergleich des Sprachdukus einerseits in einem Physikschulbuch und andererseits in einem Lehrbuch für die Hochschulphysik ein Beispiel dafür, dass sich Sprachformen mit der sozialen Gemeinschaft, in der gesprochen wird, ändern<sup>13</sup>. Schneider, Gilg, Dittmar und Schmellentin (2019) arbeiten hinsichtlich des Lehr-Lern-Zwecks schulischer Texte heraus, inwiefern Schulbuchtexte zwar durchaus Züge wissenschaftlicher Texte aufweisen, sich aber durch ihre didaktische Aufbereitung auszeichnen, da sie an (junge) Lernende gerichtet seien. Dies äußere sich in einer besonders anschaulichen Darstellung, in einer an Lernbarkeit orientierten Anordnung und in einer Anpassung der fachsprachlichen Oberfläche<sup>14</sup>.

Seitens der Lernenden spricht man bei den speziellen Fähigkeiten, die der Verarbeitung von bildungsbezogener Sprache dienen von *CALP (Cognitive Academic Language Proficiency)*. Diesen Fähigkeiten steht ein basalerer Fähigkeitskomplex (*BICS – Basic Interpersonal*

---

<sup>11</sup>Ein etwas überspitztes, aber sehr treffendes Beispiel findet sich im Roman „Die Vermessung der Welt“ von Daniel Kehlmann (2021): Der Forschungsreisende Alexander von Humboldt möchte den Bewohner\*innen des Rio-Negro-Gebiets in Südamerika eine Kostprobe höchster deutscher Literaturkunst präsentieren, indem er das „schönste deutsche Gedicht“ vorträgt. Humboldt liefert – ganz seiner sachlich-nüchternen Denkweise entsprechend – eine inhaltlich treffende Darbietung des Gedichts von *Über allen Gipfeln* von Johann Wolfgang von Goethe: „Oberhalb aller Bergspitzen sei es still, in den Bäumen kein Wind zu fühlen, auch die Vögel seien ruhig, und bald werde man tot sein“. Humboldt beschließt dann die Rezitation nach einer kurzen Pause: „Fertig“. Durch diese prägnante Zusammenfassung des lyrischen Meisterwerks, die auf der Sprachoberfläche wenig mit Goethe zu tun hat, verliert das schönste deutsche Gedicht natürlich vollkommen seine Wirkung. Die Reaktionen der Zuhörer\*innen zeugen entsprechend von etwas peinlich berührter Irritation: „Ja wie“ antwortet Bonpland und „Entschuldigung [...] Das könne doch nicht alles gewesen sein“ erwidert Julio.

<sup>12</sup>Das Modell der Didaktischen Rekonstruktion nach Kattmann, Duit, Gropengießer und Komorek (1997) umfasst drei jeweils miteinander wechselwirkende Komponenten, nämlich die *fachliche Klärung*, das *Erfassen von Vorstellungen der Lernenden* und die *didaktische Strukturierung*.

<sup>13</sup>Für die damit verbundene Auffassung von Sprachformen als *Soziolekte* sei auf die kurze Erläuterung in Abschnitt 2.2.1 auf Seite 31 verwiesen.

<sup>14</sup>In Zusammenhang mit einer „angepassten fachsprachlichen Oberfläche“ sei hier auf die Ausführungen zu den unterschiedlichen, im Unterricht relevanten Sprachformen (insbesondere Unterrichts- und Bildungssprache) in Abschnitt 2.2.1 verwiesen.

*Communication Skills*) gegenüber. Die beiden Bezeichnungen gehen auf Cummins zurück, der mit diesen beiden Fähigkeitskomplexen zwischen den Anforderungen bei bildungsbezogenem Wissenserwerb und den Anforderungen in alltäglichen Kommunikationssituationen unterscheidet (Cummins, 1979). In Bezug auf die Aneignung fachlichen, mitunter komplexen Wissens, ist vor allem *CALP* gefragt. Denn *CALP* besteht in solchen Fähigkeiten, welche die Verarbeitung abstrakter, informationsdichter, schriftlich realisierter Sprachformen betreffen. *CALP* wird daher mitunter als Voraussetzung für fachliches Lernen anhand von Sprache betrachtet (Gogolin & Duarte, 2016).

Eine weitere Voraussetzung für fachliches Lernen, die allerdings nicht direkt die Dispositionen der Rezipient\*innen betrifft, sondern in erster Linie Ansprüche an die Gestaltung des Lernangebots stellt, liegt in der Attraktivität des Angebots für Lernende. Kulgemeyer und Schecker (2009) weisen darauf hin, dass die Attraktivität eines Kommunikationsangebots entscheidend dafür sei, ob die empfangende Person sich konstruktiv und verstehend dem Angebot zuwendet (vergleiche dazu auch die interaktive Sichtweise auf das Leseverstehen, die in Unterkapitel 3.1 auf Seite 43 beschrieben wird). Auch das Modell für die Grunddimensionen der Unterrichtsqualität von Klieme (2006) beinhaltet explizit, dass Unterrichtsqualität unter anderem daran festzumachen sei, wie Schüler\*innen ein unterrichtliches Angebot *nutzen* und *wahrnehmen*. Genauer: die Verarbeitungstiefe, das Erfahren von Selbstbestimmung und die Zeitspanne aktiver Auseinandersetzung mit einem Material entscheiden über die erwünschten Wirkungen, nämlich systematischen Wissensaufbau und Motivation. Jene beiden Wirkungskomponenten bedingen sich gegenseitig und tragen gemeinsam zum Unterrichtserfolg bei. Im Kontext der Schullehrbücher impliziert dies, dass Lehrtexte ein Gefühl von Involviertheit erzeugen sollten, damit Lesende geneigt sind, sich aktiv mit dem Inhalt zu beschäftigen. Welche Rolle insbesondere die sprachliche Oberfläche für die Wahrnehmung und Nutzung von Unterrichtsangeboten wie z. B. Lehrtexten spielt beziehungsweise wie sie dahingehend optimal einzusetzen ist, ist jedoch nicht geklärt. Die affektive Textwahrnehmung gezielt als Wirkung von sprachlicher Gestaltung zu betrachten, ist daher ein neuartiger Forschungsansatz, den diese Arbeit verfolgt (siehe Kapitel 4).

### 2.1.3 Fazit mit Blick auf den Physikunterricht

Es wurde bis hierhin deutlich, dass die Sprache den Menschen einerseits als einzigartiges Werkzeug des Denkens dient (Abschnitt 2.1.1) und andererseits als zentrales Kommunikationsmittel (Abschnitt 2.1.2). Damit liegen zwei bestechende Argumente für eine intensive Beschäftigung und einen sorgfältigen Umgang mit Sprache beim Lehren und Lernen auf der Hand.

Unter epistemischer Perspektive (Denken anhand von Sprache) ist die Sprache unter anderem relevant für die Wissensrepräsentation im menschlichen Gedächtnis. Diese kann als ein Netz aus Begriffen und Relationen zwischen diesen Begriffen betrachtet werden. Begriffe versehen erstens Wörter mit Bedeutung und zweitens klassifizieren sie reale Objekte und

Sachverhalte. *Lernen* im Sinne des Wissenserwerbs<sup>15</sup> bedeutet den „Aufbau und die [...] Modifikation von Wissensrepräsentationen“ (Urhahne, Dresel & Fischer, 2019, S. 42) und entsprechend die Erweiterung und den Umbau des Netzwerks. Für die Herausbildung und Ausschärfung der Begriffe ist die Sprache mit ihrem Potenzial für die mentale Abstraktion fundamental.

Lenkt man nun die Aufmerksamkeit auf den Bildungskontext und die *Vermittlung* von Wissen, erscheinen spezielle sprachliche Fähigkeiten (*CALP*) für Lernende notwendig. Die Sprache wird zur Voraussetzung für das Lernen. Im Unterricht ist es daher wichtig, bei der Vermittlung von Wissen die Adressat\*innen zu bedenken. Abschnitt 2.1.2 zeigt auf, dass der außersprachliche Bezug (*Referent*) einer Sprachhandlung Gestaltungsspielräume für die konkrete Formulierung offenlässt. Insbesondere das Modell für physikspezifische Kommunikationskompetenz von Kulgemeyer und Schecker (2009) beinhaltet erstens die Perspektive *adressat\*innengemäße* Kommunikation (neben der *Sachgerechtigkeit*) und zweitens als einen eigenständigen Aspekt von Kommunikationskompetenz den *sprachlichen Code* (neben der Darstellungsform, dem Kontext und dem Sachaspekt). Es entscheidet sich also (mitunter) auf der sprachlichen Oberfläche, ob der Inhalt nicht nur sach-, sondern auch adressat\*innengemäß vermittelt wird und die Kommunikation somit erfolgreich ist (siehe auch die Erläuterungen zum Verstehen, Missverstehen und Nicht-Verstehen in Unterkapitel 3.1 auf Seite 40).

Hinsichtlich dieser sprachlichen Oberflächengestaltung spielen im Fachunterricht verschiedene Sprachformen eine Rolle. Auf diese wird im folgenden Unterkapitel 2.2 im Einzelnen eingegangen, es sei hier jedoch bereits ein besonderer Aspekt der naturwissenschaftlichen *Fachsprache* erwähnt: Die Fachsprache ist aus naturwissenschaftsdidaktischer Perspektive als eigenständiger Lerngegenstand des Unterrichts zu betrachten. Dies lässt sich dadurch begründen, dass sich naturwissenschaftliche Begriffe durch einen besonders abstrakten Charakter auszeichnen und dass die Art und Weise, wie über eine Sache gesprochen wird, maßgeblich dafür ist, was als fachlich korrekt gilt (siehe Seite 20 im Abschnitt 2.1.1). Bezüglich der naturwissenschaftlichen Begriffe, die einerseits im Fachvokabular und in fachsprachlichen Wendungen zum Ausdruck kommen und andererseits als mentale Repräsentation ebendieser fachsprachlichen Formulierungen zu verstehen sind, sind Inhalt und Sprachoberfläche kaum trennbar. Rincke (2010) arbeitet aus Wygotski (1979) den Gedanken heraus, dass „[d]ie Anschauung [...] im Fall der Fachsprache das Ziel des Entwicklungsprozesses [der Fachsprache]“ sei (S. 246). Dabei weist „das Erlernen eines angemessenen Verständnisses und Gebrauchs der Fachsprache Ähnlichkeiten mit Sprachlernprozessen“ (im Sinne einer Fremdsprache) auf (S. 246). Lernende im naturwissenschaftlichen Unterricht finden sich bisweilen in einer Situation wieder, in der nicht nur Inhalte neu sind. Auch die Vokabeln, mithilfe derer die Inhalte vermittelt werden, sind zunächst unvertraut, sollen aber letztendlich in den eigenen Sprachgebrauch Einzug finden. In den Naturwissenschaften ist daher der rezeptive sowie

---

<sup>15</sup>Unter *Lernen* werden im Allgemeinen auch verhaltensbezogene Veränderungen gefasst (vergleiche Urhahne, Dresel & Fischer, 2019, S. 4 und S. 42).

produktive Umgang mit fachsprachlichen Wendungen als eigenständiger Lerngegenstand zu betrachten (Höttecke, Ehmke, Krieger & Kulik, 2017). Entsprechend zählt „die Fähigkeit, diese [Fachsprache] zu nutzen, um fachbezogene Informationen zu erschließen, adressat[\*inn]en- und situationsgerecht darzustellen und auszutauschen“ explizit zu den Bildungsstandards für das Fach Physik (Kultusministerkonferenz, 2020).

## 2.2 Bildungs- und fachbezogene Sprachformen

Im Diskurs über die Rolle von Sprache im Unterricht und speziell im Unterricht der verschiedenen Schulfächer ist von unterschiedlichen Sprachformen die Rede (Alltagssprache, Unterrichtssprache, Fachsprache, Umgangssprache, Muttersprache, Schulsprache, Wissenschaftssprache, etc.). Es fehlt hier an einheitlichen Definitionen, anhand derer diese Sprachformen klar gefasst und gegeneinander abgegrenzt werden könnten. An einem Beispiel sei diese mangelnde Einheitlichkeit demonstriert: Härtig, Bernholt, Prechtel und Retelsdorf (2015) fassen die Fachsprache (auf schulischem Niveau) als „Erweiterung der Unterrichtssprache“ um bestimmte, fachsprachliche Mittel auf. Demgegenüber beschreibt Rincke (2010) die Fachsprache im Rahmen einer Arbeitsdefinition losgelöst von einer Unterrichtssituation: Sie sei gekennzeichnet durch ihr „sprachliches Inventar“ (Vokabeln und verbindende Elemente zwischen Vokabeln) und durch ihre „Rücksichtnahme auf die vorliegende Kommunikationssituation“.

Im Folgenden werden die Alltagssprache, die Fachsprache, die Unterrichtssprache und die Bildungssprache vergleichend umrissen. Denn dass die Fachsprache für die Wissenseignung und -vermittlung im naturwissenschaftlichen Unterricht (und damit insbesondere im Kontext dieser Arbeit) besonders bedeutsam ist, wurde im Abschnitt 2.1.1 (epistemische Funktion: Begriffslernen und Sprache als Lerngegenstand) und im Abschnitt 2.1.3 (Bezug zum Physikunterricht) herausgearbeitet. Sie lässt sich erstens anhand einer (funktionalen) Abgrenzung von der Alltagssprache (siehe den folgenden Abschnitt 2.2.1) und zweitens anhand einer deskriptiven Charakterisierung ihrer sprachoberfläche gut fassen (siehe dazu Abschnitt 2.2.2). Eine weitere Sprachform, die in Bezug auf wissensvermittelnde Kommunikation und auf Wissenseignung von Schüler\*innen relevant ist, ist die Bildungssprache (Morek & Heller, 2012). Inwiefern sie besonders im Kontext dieser Arbeit, das heißt im Kontext der Wissensvermittlung für Schüler\*innen anhand von Texten, bedeutsam ist, wird durch ihre Verortung neben den verwandten, unterrichtsrelevanten Sprachformen Alltags-, Fach- und Unterrichtssprache gezeigt.

### 2.2.1 Begriffliche Einordnung von Bildungssprache, Unterrichtssprache, Fachsprache und Alltagssprache

Bei der Unterscheidung von *Alltagssprache*, *Fachsprache*, *Unterrichtssprache* und *Bildungssprache* sind in der Literatur die Begriffe der *Sprachregister* (z. B. Krabbe, Rincke & Aleksov,

2021) und der *sprachlichen Varietäten* (z. B. Roelcke, 2020, S. 15) gebräuchlich. Unter einem *Sprachregister* versteht man ein verbales Repertoire, welches erstens in inhaltlich-spezifischen Situationen, zweitens in spezifischen sozialen Gruppen und drittens je nach Kommunikationsabsicht bewusst ausgewählt und verwendet wird (Glück & Rödel, 2016, S. 559 f.). Eine *sprachliche Varietät* ist nach Roelcke (2020, S. 15) ein „sprachliches Subsystem“, welches nicht eigenständig, sondern „einer Einzelsprache untergeordnet“ ist und anhand von „inersprachlichen Merkmalen“ (z. B. Vokabular und Satzstrukturen sowie Grammatik und Semantik) und „außersprachlichen Merkmalen“ (Personengruppen oder Tätigkeitsbereiche, in denen die Varietät typischerweise genutzt wird) von weiteren Sprachvarietäten abgegrenzt wird. Für den hier vorliegenden Zweck einer funktionalen Unterscheidung von Alltags-, Fach-, Unterrichts- und Bildungssprache genügt die einheitliche Verwendung des Begriffs der *Sprachformen*. Unter *Sprachformen* sind sowohl Register als auch Varietäten gefasst. Der Begriff bringt allgemein zum Ausdruck, dass von unterscheidbaren – wenn auch nicht vollkommen trennscharfen – Erscheinungsformen einer Einzelsprache (hier: des Deutschen) die Rede ist.

Für eine vergleichende Betrachtung der vier Sprachformen können kategorisierende Modelle hilfreich sein. Anhand solcher Modelle werden Unterschiede und Gemeinsamkeiten der einzelnen Sprachformen deutlich. Dabei scheint auf den ersten Blick eine Einteilung in *geschriebene* und in *gesprochene* Sprache je nach Realisierungsform – das heißt schriftlich fixiert oder mündlich geäußert – naheliegend zu sein (vergleiche Dürscheid, 2012, S. 24 ff.). Eine treffendere und differenziertere Einordnung von Sprachformen ist jedoch mithilfe von Hallidays *Registertheorie* möglich, das heißt anhand der Dimensionen *Tenor* (Art der Interaktion der Gesprächspartner\*innen, Diskursstil), *Mode* (Modalität; das heißt mediale Äußerungsform und Funktion des Texts) und *Field* (Themen-/Handlungsfeld) (Halliday, 1966). Die Dimensionen werden anhand der jeweiligen Pole, zwischen denen sie sich aufspannen, greifbarer (vergleiche hierzu Gregory, 1967): Der Diskursstil (*Tenor*) eines Sprachregisters liegt zwischen einer *persönlichen, informellen Beziehung der Sprachteilnehmenden auf Augenhöhe* und einer *nüchternen, formellen, Expert\*innen-Noviz\*innen-Beziehung*. Das Modalitätskontinuum (*Mode*) erstreckt sich zwischen *konzeptionell mündlich / spontan / nächersprachlich* und *konzeptionell schriftlich / geplant / distanzsprachlich* (mit dieser Unterscheidung haben sich ausführlich auch Koch und Oesterreicher, 1985, auseinandergesetzt). Das Themen-/Handlungsfeld (*Field*) schließlich hat die Ausprägungsformen *kontextübergreifend/allgemein* und *kontextgebunden/fachspezifisch*.

Eine daran anknüpfende Möglichkeit der Verortung unterschiedlicher Sprachformen speziell im Kontext Schule und naturwissenschaftliche Bildung findet sich bei Krabbe, Rincke und Aleksov (2021) (siehe Abbildung 2.2). Dort werden die Alltagssprache (*everyday language*, auch *colloquial language*), die Bildungssprache (*academic language*, auch *language of schooling*), die Unterrichtssprache (*language of teaching and learning*<sup>16</sup>) und die Fachsprache (*scientific*

<sup>16</sup>In der Originalquelle ist der Naturwissenschafts-Jargon (*scientific jargon*) an der gleichen Stelle wie die Unterrichtssprache verortet (Krabbe, Rincke & Aleksov, 2021, S. 367 f.). Dennoch sind beides unterschiedliche Sprachformen und im hier vorliegenden Kontext ist lediglich die Unterrichtssprache von Belang.

## 2. Physikunterricht und Sprache

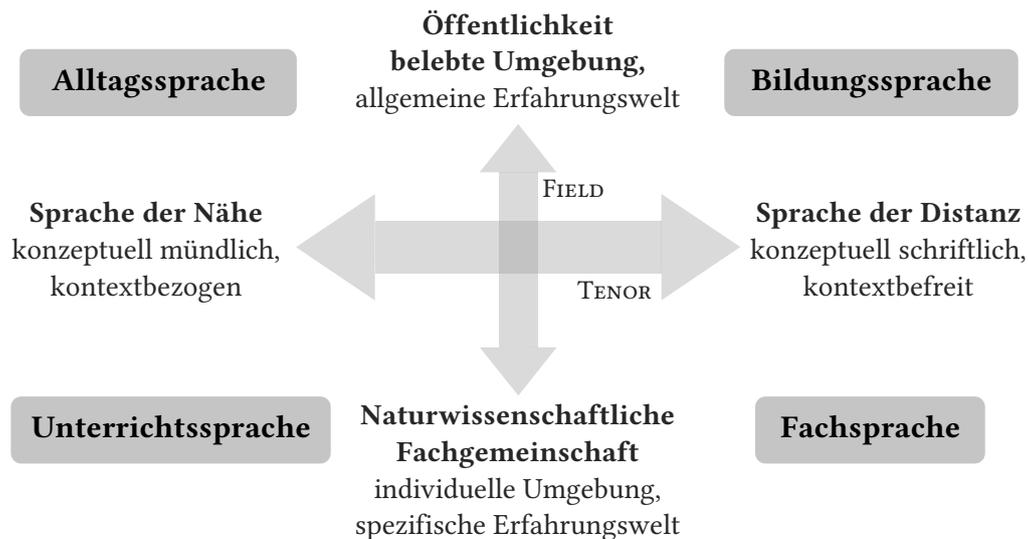


Abbildung 2.2.: Verortung der für den Physikunterricht relevanten Sprachformen *Alltagssprache*, *Bildungssprache*, (physikalische) *Unterrichtssprache* und (physikalische) *Fachsprache* entlang der beiden Dimensionen *Tenor* (Diskursstil) und *Field* (Themen-/Handlungsfeld). Dies ist eine eigene, verkürzte und ins Deutsche übersetzte Darstellung der entsprechenden Abbildung in Krabbe, Rincke und Aleksov, 2021, S. 367.

*language*, auch *technical language*) entlang zweier Dimensionen verortet. Die erste Dimension entspricht Hallidays *Tenor*-Kategorie und sie spannt sich zwischen den beiden Polen *Sprache der Nähe* und *Sprache der Distanz* auf (für diese Unterscheidung sei erneut auch auf Koch und Oesterreicher, 1985, hingewiesen). Zwischen der *Adressierung der breiten Öffentlichkeit und allgemeinen Erfahrungswelt* und der *Adressierung ausgewählter Personenkreise* (scientific community) liegt die zweite Dimension, die das Handlungs-/Themenfeld betrifft (*Field* im Original bei Halliday, 1966).

Zunächst sollen die Alltagssprache und die Fachsprache voneinander abgegrenzt werden. Dabei bietet sich vor allem der Blick auf die *Funktion* der Fachsprache an (Roelcke, 2020, S. 15). Das bedeutet, die Alltags- und die Fachsprache als zwei unterschiedliche *Funktiolekte* aufzufassen, sie also dahingehend zu unterscheiden, zu welchem Zweck sie gebraucht werden<sup>17</sup>. Die Alltagssprache dient in erster Linie der gesprochenen Kommunikation im Privaten, das heißt sie wird für „gewöhnliche, soziale Interaktionen“ gebraucht (Glück & Rödel, 2016, S. 30). Somit folgt sie mündlichen Sprachgebrauchsmustern und trägt oftmals eine implizite Bedeutungsebene (Prediger, 2013). Das heißt, sie funktioniert auf spezifische Kontexte bezogen. Sie ist damit tendenziell eine Sprache der Nähe. Die Alltagssprache wird nicht in exklusiven, speziellen sozialen Gemeinschaften genutzt, sondern sie dient ganz generell der Kommunikation über Lebenswelt und Erfahrungen und ist damit jedem zugänglich (vergleiche Abbildung 2.2). Die Fachsprache hingegen dient einer effizienten Kommunikation (im Sinne von Wissenstransfer,

<sup>17</sup>Einen anderen Blickwinkel zur begrifflichen Klärung von Fach- und Alltagssprache bieten unter anderem *deskriptive* Unterscheidungsmerkmale. Einer solchen sprachoberflächlichen Charakterisierung der Fachsprache wendet sich weiter unten Abschnitt 2.2.2 zu. Die Auffassung von Fachsprache als die Sprache einer bestimmten sozialen Gemeinschaft wird später in diesem Absatz kurz erläutert.

Abstraktion, Konkretion, Kategorisierung, Assoziation, Dissoziation, etc.; siehe Leisen, 2022, S. 38) innerhalb eines Fachgebiets beziehungsweise einer Fach-Community. Im Sprachlexikon von Glück und Rödel (2016, S. 195) ist die Fachsprache als „Gesamtheit der sprachlichen Mittel, die in einem Fachgebiet verwendet werden“ definiert. Unter einem Fachgebiet sei ein Berufsfeld oder Hobby zu verstehen. Mit Blick auf Abbildung 2.2 ist die Fachsprache als tendenziell distanzsprachlich einzuordnen. Fachsprachliche Begriffe sind allgemeingültig und präzise definiert, sodass sie auch mit reduziertem Kontextbezug eindeutig sind. Allerdings erfordert die fachsprachliche Kommunikation einen zwischen den Gesprächsteilnehmenden etablierten Konsens über die Bedeutung ebensolcher Begriffe und einen geübten Umgang mit weiteren fachsprachlichen Mustern (spezielle Satzkonstruktionen, Abkürzungen, etc.; siehe Abschnitt 2.2.2). Die Fachsprache wird daher innerhalb ausgewählter Personenkreise genutzt und verstanden, sie adressiert nicht die breite Öffentlichkeit. Im Zusammenhang mit der exklusiven Nutzung der Fachsprache in bestimmten Personenkreisen steht eine weitere Funktion von Sprache, die hier jedoch nur am Rande erwähnt wird: fachsprachliche Praktiken können Rollen definieren, Identität stiften und somit einen sozial-symbolischen Charakter tragen (Morek & Heller, 2012; Gogolin & Duarte, 2016). In diesem Sinne können verschiedene Sprachformen auch als *Soziolekte* voneinander unterschieden werden (Glück & Rödel, 2016, S. 630).

In dieser funktionalen Gegenüberstellung lässt sich erkennen, dass beide Sprachformen für den Unterricht relevant sind. Die Alltagssprache stellt gleichsam *bekanntes Terrain* dar, welches der neuen, zu erlernenden, unbekanntem fachsprachlichen Ausdrucksweise konfrontierend begegnet (diesen Standpunkt vertritt insbesondere Muckenfuß; siehe z. B. Muckenfuß, 1995, S. 259; dazu und zum Verhältnis von Fach- und Alltagssprache im Allgemeinen bieten Rincke und Leisen, 2015, weitere Ausführungen).

Gewissermaßen als vermittelnde Instanz, die zwischen Alltags- und Fachsprache liegt, kann die *Unterrichtssprache* verstanden werden. Leisen (2005) spricht in diesem Zusammenhang von Unterrichtssprache als einer „Noch-Nicht-Fachsprache“. Die Unterrichtssprache ist diejenige Sprachform, anhand derer sich Schüler\*innen Fachinhalte im Unterricht aneignen. Sie hat daher schulfachspezifische Ausprägungsformen (siehe Fang, 2006, wo im Englischen von *expository language* die Rede ist). Die Unterrichtssprache trägt überwiegend Merkmale einer Sprache der Nähe, die nicht-öffentlich adressiert ist, da sie sich im mündlichen oder schriftlichen, persönlichen Diskurs zwischen Lehrkräften und Schüler\*innen abspielt. Sie ist in Abbildung 2.2 entsprechend verortet.

Weniger auf die Institution Schule bezogen, aber eng mit dem Begriff der Unterrichtssprache verwandt, ist die *Bildungssprache*. Unter Rückbezug auf Habermas (1977) beschreiben Gogolin und Duarte (2016) sie als diejenige Sprachform, anhand derer „man sich in der Schulbildung Wissen verschaffen kann“ (Gogolin & Duarte, 2016, S. 483). Sie sei damit einerseits Transportmittel für curriculare Inhalte und gleichzeitig sei die Schule in der Pflicht, den Lernenden Zugang zur Bildungssprache zu verschaffen. Die Sprache, die zur Vermittlung der Inhalte genutzt wird, ist also als eigenständiger Lerngegenstand zu betrachten (verglei-

che die Ausführungen zur Fachsprache als Lerngegenstand in Abschnitt 2.1.3). Weiter den Ausführungen von Gogolin und Duarte (2016, S. 485) folgend gilt für die Bildungssprache hinsichtlich der Registerdimensionen (siehe Seite 29), dass ihr *Tenor* sich durch an die breite Öffentlichkeit gerichtete Mitteilungsabsichten auszeichne. Die Diskursteilnehmenden seien also eher distanziert und würden hierarchisch getrennte Rollen einnehmen (anders als die Unterrichtssprache, die im persönlichen Diskurs zur Anwendung kommt). Der *Mode* der Bildungssprache tendiere zu eher schriftförmigen, monologischen Ausdrucksformen. Die schulische Bildung stelle das *Field* der Bildungssprache dar, sodass als Einsatzzweck der Bildungssprache die Vermittlung beziehungsweise Aneignung von Wissen zu nennen sei. Entsprechend der Verortung in Abbildung 2.2 ist dabei die Bildungssprache als eine schulfachübergreifende Sprachform zu betrachten, die nicht an fachspezifische Personengruppen gerichtet ist.

Auf den vorangehenden Seiten wurden die für den Physikunterricht relevanten Sprachformen Alltags-, Fach-, Unterrichts und Bildungssprache kategorisierend gegenübergestellt. Dabei wurde deutlich, dass die verschiedenen Sprachformen einem jeweils eigenen Zweck dienen. Leisen (2003, S. 6) formuliert bezüglich dieser Sprachformen (und weiterer Formen wie z. B. nonverbaler Sprache) einen „didaktische[n] Imperativ: Der Lehrer[\*Die Lehrerin] nutze jeweils die Sprache im Fachunterricht, in der das fachliche Handeln, Erleben und Verstehen sinnvoll wird“. Die hier betrachteten Sprachformen versteht Leisen (2003) als *Verbalsprachen* (S. 3), deren Sinnhaftigkeit für Unterricht sich also insbesondere anhand von spezifischen verbalen Merkmalen bestimmt. Ohne Sprache allein auf ihre Oberfläche reduzieren zu wollen, macht dennoch das charakteristische verbale Repertoire auf Wort-, Satz- und Textebene der jeweiligen Sprachform einen Teil der Differenzierung aus (siehe dementsprechend die Merkmale der naturwissenschaftlichen Fachsprache in Abschnitt 2.2.2). Der didaktische Imperativ impliziert somit, dass unterschiedliche Merkmale auf verbaler Ebene in unterschiedlichen Situationen im Fachunterricht nützlich sind. Ein besonderes Augenmerk ist im Unterricht auf solche Merkmale zu richten, die eine mögliche Quelle für Schwierigkeiten beim Lernen darstellen. Solche Merkmale bringt die Fachsprache in zahlreicher Weise mit (Fang, 2006; Leisen, 2003, S. 5 f.). Im folgenden Abschnitt 2.2.2 wird daher die funktionale um eine deskriptive Betrachtung von Fachsprache erweitert, indem charakteristische Merkmale naturwissenschaftlicher Texte vorgestellt werden.

### 2.2.2 Die Kennzeichen von Fachsprache im Fach Physik

Die deskriptive Charakterisierung der Fachsprache im Fach Physik bezieht sich auf das typische verbale Inventar, auf typische Phrasen- und Satzstrukturen sowie auf typische Muster im inhaltlichen und strukturellen Aufbau eines Texts. Die einschlägige Literatur zeichnet hinsichtlich entsprechender Merkmalslisten für physikalische beziehungsweise naturwissen-

schaftliche Textcharakteristika ein recht einheitliches Bild<sup>18</sup>. Die folgenden Ausführungen basieren auf den Publikationen von Leisen (2003), Rincke (2010), Härtig, Bernholt, Prechtel und Retelsdorf (2015), Gogolin und Duarte (2016), Fang (2006), Patterson, Roman, Friend, Osborne und Donovan (2018) und Heine et al. (2018). Anhand der Beispiele in kursiver Schrift werden die sprachlichen Charakteristika jeweils konkretisiert.

### Wortebene

- Auf Wortebene sind zunächst solche Vokabeln typisch für Fachsprache, die exklusiv in der Fachsprache existieren.

*Wirkungsquerschnitt, Entropie, absorbieren, Verdampfungswärme.*

- Häufig sind solche Wörter Fremdwörter. Das heißt, sie sind aus anderen Sprachen wie dem Griechischen oder dem Lateinischen entnommen.

Griechisch: *Atom, nano, Thermodynamik.*

Latein: *konzentrisch, Impuls, oszillieren.*

- Zum fachsprachlichen Vokabular zählen solche Wörter, die im Allgemeinen verhältnismäßig selten verwendet werden.

*koppeln, unter-/überschreiten, unzulänglich.*

- Andererseits treten zahlreiche Wörter auf, die neben der Fachsprache auch in anderen Sprachformen wie der Alltagssprache existieren, dort aber mitunter eine andere Bedeutung tragen. In diesem Fall ist oftmals das bezeichnende Wort schon vorhanden, es ändert sich aber das Bezeichnete (vergleiche *signe* und *signifié* nach de Saussure, Seite 22 in Abschnitt 2.1.2).

*Masse, Spannung, strahlen.*<sup>19</sup>

- Hinsichtlich fachsprachlicher Wortwahl sind auch solche Ausdrücke gängig, die eine differenzierende oder abstrahierende Bedeutung haben.

*hochverdichtete* Materie (differenzierend), *Massepunkt* (abstrahierend)

- Hinsichtlich der Morphologie (etwa: Wortgestalt) lassen sich ebenfalls typische Auffälligkeiten nennen. Hierzu zählen Adjektive, die durch Suffix-Konstruktionen entstehen, die also von anderen Wortarten durch Anfügen spezifischer Endungen abgeleitet werden (sogenannte Derivata).

*masselos, reibungsarm, antistatisch, hooksche* Feder, *newtonsche* Mechanik.

- Genauso zählen Präfix-Komposita zur Fachsprache, das heißt Adjektive, Verben und

<sup>18</sup> Gerade im englischsprachigen Raum ist es üblich, von *science*, also *Naturwissenschaften*, anstelle von Einzeldisziplinen wie Physik, Chemie oder Biologie zu sprechen. Und auch im Deutschen wird im Diskurs über die Fachsprache in den naturwissenschaftlichen Fachbereichen in der Regel nicht unterschieden, sondern gemeinhin von *naturwissenschaftlicher* Fachsprache gesprochen. Speziell in Hinblick auf *Fachbegriffe* weist zwar jede Disziplin ihr eigenes Vokabular auf, bezüglich der Merkmale auf Satz- und Textebene jedoch, ist eine zusammenfassende Beschreibung *naturwissenschaftlicher* Fachsprache passend.

<sup>19</sup> Um die Beispiele genauer zu erläutern: *Masse* wird alltagssprachlich als unförmiger Stoff oder auch als große Anzahl verstanden; physikalisch ist eine Eigenschaft von Materie gemeint. *Spannung* kann alltagssprachlich einen erwartungsvollen Gefühlszustand oder auch eine latente Unstimmigkeit in einem sozialen Verhältnis bezeichnen, sie ist physikalisch aber als elektrische Potenzialdifferenz definiert. Wenn eine Person *strahlt*, dann trägt sie einen freundlichen Gesichtsausdruck – physikalische Objekte, die *strahlen*, emittieren z. B. Photonen oder Elektronen.

## 2. Physikunterricht und Sprache

Substantive mit einer Vorsilbe und einem Hauptwort.

*antistatisch, entartet, entladen, Paramagnet, Dipol.*

- Allgemein werden häufig Einzelwörter zu mehrgliedrigen Komposita zusammengesetzt. Insbesondere kommen dabei Abkürzungen oder Eigennamen zur Verwendung.

*Elektromotor, Konkavlinse, Hochspannung, hochohmig, E-Feld.*

- Eine weitere morphologische Auffälligkeit der Fachsprache sind Partizipien.
- Auch Nominalisierungen von Verben und Adjektiven sind ein charakteristisches Merkmal der Fachsprache in Hinblick auf die Morphologie.

*das Erhitzen, Verschiebung ins Rote, das soeben Gesehene*

- Schließlich ist die Wortwahl in der Fachsprache durch eine Häufung von Konnektiva geprägt. Anhand von Konnektiva werden semantische Relationen zwischen Einzelaussagen explizit zum Ausdruck gebracht. Das bedeutet, dass Erklärungen, Argumente und Beschreibungen anhand von kausalen, konsekutiven, finalen, konzessiven, adversativen, explikativen, konditionalen und weiteren Verbindungswörtern innerlich konsistent und im Aufbau transparent gemacht werden (sollen).

*da, somit, um...zu, zwar, während, das heißt, falls*

### Phrasen- und Satzebene

- Auf der Ebene von Phrasen und Sätzen ist ein ausgeprägter Nominalstil typisch für Fachsprache. Dabei werden Verben und Adjektive in substantivischer Form gebraucht und oftmals zu ganzen Gruppen aus Nominalisierungen zusammengefügt, sodass Mehrwortkomplexe entstehen. Auffällig ist in diesem Zusammenhang auch die häufige Verwendung des Genitiv.

*unter Vernachlässigung des Luftwiderstands,*

*die Bewegung mit Lichtgeschwindigkeit im Vakuum.*

- Neben den Nominalisierungen gelten auch mehrgliedrige Partizipialkonstruktionen als typisch fachsprachliche Mehrwortkomplexe. Solche komplexen Satzglieder fungieren häufig als Attribute. Sie treten an die Stelle von (attributiven) Nebensätzen oder einzelnen, kürzeren Hauptsätzen.

*die kontinuierlich steigende Temperatur,*

*die gemäß der Lenzschen Regel entgegengerichtete Spannung.*

- Zweitens ist es üblich, Sätze mit passiven Verbformen zu formulieren.

*Teilchen werden beschleunigt, Licht wird emittiert*

- Neben dem Nominalstil und dem Passiv werden Verben auch häufig in Reflexivkonstruktionen sowie in Formulierungen in der unpersönlichen *man*-Form verwendet.

*man unterscheidet... , Nukleonen lassen sich einteilen in...*

- Dennoch gelten auch komplexe Nebensatzkonstruktionen als charakteristisch für Fachsprache. Oftmals werden diese mit logischen Konnektiva (siehe den entsprechenden Punkt unter *Wortebene* oben) verknüpft.

*Es folgt, dass xy gilt, weil yz zusammenhängt, aber xz unabhängig ist.*

- Eine Alternative für attributive Zusatzinformationen sind Einschübe innerhalb eines Satzes.

Das Halbleiter-Bauteil – *wie es in LEDs verwendet wird* – setzt kaum Energie um.

- Einzelne Quellen (Fang, 2006, Heine et al., 2018) zählen auch Ellipsen zu einem (herausfordernden) Merkmal von Fachsprache. Fang (2006, S. 497) führt hierzu aus, dass dies generell ein Merkmal konzeptuell schriftlicher Texte sei und damit insbesondere für naturwissenschaftliche Fachtexte relevant.

*Protonen befinden sich im Atomkern, Elektronen in der Hülle.*

anstelle von: *Elektronen befinden sich in der Atomhülle.*

- Ein typisches Merkmal für fachsprachliche Phrasen sind außerdem Funktionsverbgefüge. Funktionsverben tragen eine ganz spezifische Bedeutung innerhalb eines feststehenden Ausdrucks oder existieren sogar nur in Verbindung mit ganz speziellen Ausdrücken.

*Arbeit verrichten, Daten erheben*

- Und schließlich sind in fachsprachlichen Texten oftmals Proformen, z. B. Pronomen und (abstrahierende) Rekurrenzen, zu finden.

*diese/dieser/dieses, ein solcher Vorgang, ein derartiger Aufbau*

### Textebene

- Zuletzt werden charakteristische Muster auf Textebene aufgezählt. Ganz generell zeichnen sich Fachtexte oftmals durch logisch aufgebaute, schlüssige Argumentations- oder Erklärungsketten aus und unterscheiden sich dadurch beispielshalber von Erzählsträngen mit zeitlicher Ordnung. Die ausgeprägte Logik von Fachtexten wird beispielshalber anhand von Vor- und Rückverweisen oder Frage-Antwort-Konstruktionen zum Ausdruck gebracht.

*wie zu Beginn gesehen, das eben erwähnte Teilchen, nachfolgend wird... erläutert.*

*Worin unterscheiden sich Kometen und Asteroiden?*

- Für längere Zusatzinformationen sind Fußnoten gebräuchlich.<sup>20</sup>
- Um gewisse Ausdrucksweisen explizit zu machen oder Fachausdrücke einzuführen, werden metasprachliche Kommentare eingebracht (Sprechen über die Sprache).

Bei dieser Schwingung *spricht man von* einem harmonischen Oszillator.

*Der Ausdruck* „A übt eine Kraft auf B aus“ *bedeutet...*

- Ein entscheidendes Merkmal für naturwissenschaftliche Fachtexte ist ihre Multimodalität. Damit ist der Wechsel zwischen verschiedenen Darstellungsformen (Text, Diagramm, Abbildung, Tabelle, Formel, etc.) gemeint. Der Darstellungswechsel zählt als besonders herausfordernd für Lernende in den Naturwissenschaften und stellt daher ein eigenes Forschungsgebiet für die Verständlichkeit von Texten beziehungsweise Erklärungen dar.<sup>21</sup>

<sup>20</sup>Ein Beispiel in Form eines längeren Textabschnitts mit einer passenden Stelle für eine Fußnote wurde hier nicht eingefügt. *Stattdessen soll diese Fußnote als Beispiel dienen.*

<sup>21</sup>Auch für dieses Merkmal wird an dieser Stelle auf die explizite Darstellung eines Beispiels verzichtet. Vertiefte Einblicke bieten z. B. Rau (2017) oder Berger und Starauschek (2009).

## 2. Physikunterricht und Sprache

Die oben genannten Merkmale beschreiben die Oberfläche typisch fachsprachlicher Texte. Aus den Merkmalen ergeben sich charakteristische *Eigenschaften* von naturwissenschaftlichen Texten. Es sei hier betont, dass Eigenschaften im Sinne einer *Folge* der Gestaltung der Textoberfläche zu verstehen sind und somit eine *Wirkung* auf Lesende beschreiben. Eine solche Folge einer fachsprachlichen Textoberflächengestaltung, die als oft genanntes Charakteristikum naturwissenschaftlicher Texte gilt, ist die *Deagentivierung* der Formulierungen (auch: Anonymisierung/Objektivierung). Das bedeutet, dass die Handlungstragenden einer Tätigkeit oder eines Vorgangs in der Regel nicht explizit genannt werden. Die Deagentivierung entsteht z. B. durch Passiv- und Reflexivkonstruktionen, durch Nominalisierungen und Partizipialkonstruktionen. Als weitere charakteristische Eigenschaft zählt die *hohe Informationsdichte* naturwissenschaftlicher Texte. Indem Fachbegriffe, attributive Satzglieder, abstrahierende und differenzierende Rekurrenzen, logische Konnektiva, Ellipsen und Einschübe gebraucht werden, wird mit möglichst wenig Zeichenaufwand möglichst viel Information transportiert. Dementsprechend liegt es nahe, dass die naturwissenschaftliche Fachsprache *herausfordernd* auf Schüler\*innen wirken kann (Fang, 2006). Nicht zuletzt daher ist der Einsatz von Sprache besonders im naturwissenschaftlichen Unterricht umsichtig und überlegt zu gestalten.

Ein zentrales Unterrichtsmedium, welches maßgeblich durch seine sprachliche Gestaltung ausgezeichnet ist, sind geschriebene Texte. Ihre Rolle als Medium der Wissensvermittlung wird nun im nächsten großen Kapitel eingehend erläutert.

### 3 | WISSENSVERMITTLUNG ANHAND VON TEXTEN

Mit seinem berühmten Ausspruch über das „Buch der Natur“ verleiht der Naturforscher Galileo Galilei der engen Verbindung zwischen Mathematik und Natur(wissenschaft) bildhaft Ausdruck. Das Zitat legt nahe, dass die Mathematik geeignet oder sogar nötig sei, um die Natur beschreiben zu können. Und tatsächlich ist es ein Wesenszug der Naturwissenschaften – der Physik im Speziellen – in der vielfältigen Natur nach Regelmäßigkeiten zu suchen. So werden die Phänomene der Natur anhand von allgemeingültigen physikalischen Gesetzen und Prinzipien zu erklären versucht beziehungsweise mithilfe logischer Ursache-Wirkungszusammenhänge vorhergesagt. Klafki beschreibt es als Wesenszug der Physik, dass sich im Besonderen und im Einzelnen das Allgemeine erkennen lasse (Klafki, 1963, S. 443). Es kann durchaus *Faszination* bedeuten, sich der erstaunlichen Verbindung zwischen einer abstrakten, formalisierten, kompakten Darstellungsform (*Mathematik*) und der lebhaften, realen, komplexen Erlebniswelt (*Natur*) bewusst zu werden und sich ihrer zu bedienen um Naturwissenschaft zu betreiben. Nicht zuletzt daher mag das Galileo-Zitat so bekannt sein.

Es wäre jedoch zu kurz gegriffen, die Sprache der Physik in Anknüpfung an Galilei allein auf die Mathematik und insbesondere die damit verbundene Darstellungsform zu beschränken. Denn die Sprache der Physik zu verwenden bedeutet auch, physikalische Begriffe (im wörtlichen und im mentalen Sinn; siehe Abschnitt 2.1.1) aufzubauen – vor allem in Bezug auf das Lernen und Lehren von Physik. Dafür sind Worte notwendig, Analogien hilfreich, reale, fiktive und mentale Bilder dienlich und Zahlen und Diagramme grundlegend. Mathematik ist also *eine von vielen* (und eine besonders formale) Darstellungsform um über Physik zu sprechen. Wagenschein betont entsprechend, wie wertvoll die lebendige, bildhafte, unfachliche Sprache der Kinder für das Lernen sei (Wagenschein, 1970, S. 376 und S. 464). Er zitiert außerdem Einstein, welcher es als „Glück“ bezeichnete, „wesentliche Ergebnisse und Methoden der gesamten Naturwissenschaften in einer vortrefflichen, populären, fast durchweg aufs Qualitative beschränkten Darstellung kennen zu lernen“<sup>1</sup>.

Die „Methoden und Erkenntnisse der Naturwissenschaften“ den Lernenden näher zu bringen ist ohne Frage eine Aufgabe des Physikunterrichts. Damit ist sie an Vermittlungskanäle

---

<sup>1</sup>Laut Wagenschein stamme dieses Zitat aus „*Albert Einstein als Philosoph und Naturforscher* (Stuttgart 1951, S. 5 f.)“ und beziehe sich wohl auf „A. Bernsteins *Naturwissenschaftliche Volksbücher*. 5 Bde. Berlin 1891“.

und Darstellungsformen gebunden. Die Vermittlung der Lehr-Lern-Inhalte kann dabei zum Einen mündlich geschehen – durch die Lehrkraft, anhand von Mitschüler\*innenbeiträgen oder über externe Quellen wie Erklärvideos und Expert\*innenbeiträge. Zum Anderen können Inhalte in schriftlicher Form für die Lernenden bereitgestellt werden. Wichtig ist in jedem Fall, dass mit der Bereitstellung auch Zugänglichkeit verbunden ist. An diesem Punkt rückt vor allem die Darstellung der Inhalte in den Mittelpunkt und damit die Frage, auf welche Art Lehrkräfte beziehungsweise Textautor\*innen sprachliche Hürden möglichst gering halten oder abbauen können (vergleiche Gogolin & Duarte, 2016). Schleppegrell (2019) betont, dass aus didaktischer Sicht eine intensive, fachspezifische Auseinandersetzung mit der Sprache in ihrer darstellenden Funktion für fachliches Wissen einen wichtigen Ansatzpunkt dafür darstellt, den vielfältigen Lernvoraussetzungen im Klassenzimmer zu begegnen.

Eine solche intensive Auseinandersetzung in Hinblick auf geschriebene Texte leisten die Forschungsfelder rund um das verstehende Lesen und die Textverständlichkeit. Es beschäftigen sich verschiedene Disziplinen mit diesen Forschungsfeldern, die jeweils ihre eigenen Blickwinkel und Schwerpunkte mitbringen: Die (Kognitions-)Psychologie beispielshalber beschäftigt sich mit den mentalen Prozessen, die bei lesenden Personen ablaufen. Die Linguistik setzt sich mit Texteigenschaften auseinander. Die (Fach-)Didaktik wiederum begründet die Ziele und den Zweck eines Leseakts (mit unterrichtsfachspezifischer Schwerpunktsetzung) und rückt den interaktiven Charakter des Lesens in den Fokus. Sie legt Kriterien für gelingendes Lernen, das heißt Kompetenzerwartungen und Lernziele, fest und erforscht die Bedingungen dafür (vergleiche zu diesen drei Perspektiven auch Scheller, 2010, S. 10). In den folgenden Unterkapiteln werden der Reihe nach diese Blickwinkel aufgegriffen: Nach einer Begriffsklärung in Unterkapitel 3.1 folgen eine Erläuterung der personenbezogenen Prozesse beim Lesen (kognitionspsychologischer Blickwinkel; Unterkapitel 3.2), ein Überblick über etablierte Kriterien und Modelle zur Textverständlichkeit im Allgemeinen (linguistischer Blickwinkel, Unterkapitel 3.3) und eine Zusammenstellung des empirischen Forschungsstands zur Textverständlichkeit speziell hinsichtlich des naturwissenschaftlichen Unterrichts (didaktischer Blickwinkel, Unterkapitel 3.4).

## 3.1 Verstehendes Lesen: Begrifflichkeiten

Zum begrifflichen Inventar zum Thema Leseverstehen zählen zahlreiche ähnlich klingende Vokabeln, die Verwandtes, aber nicht Identisches meinen. Eine trennscharfe Verwendung der Begriffe auf Basis einer präzisen Abgrenzung ist daher besonders wichtig, zumal die Begriffe in der Literatur aufgrund verschiedener Schwerpunktsetzungen einzelner Theorie-traditionen eine gewisse Unschärfe in ihrem Gebrauch erfahren (Lenhard, 2013, S. 43). Die folgenden Beschreibungen zum Lesevorgang stützen sich in erster Linie auf ein Modell für die Teilprozesse beim Leseverstehen nach Lenhard (2013, S. 14 ff.); die auf die Textproduktion bezogenen Ausführungen basieren vor allem auf dem „kommunikationsorientierten,

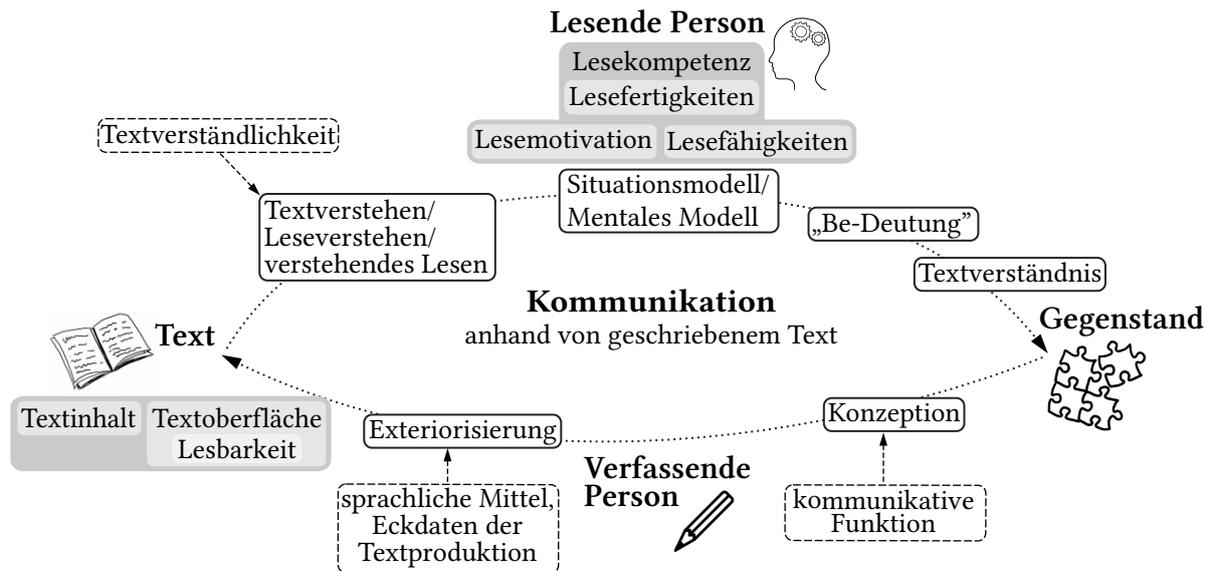


Abbildung 3.1.: Einordnung der Begriffe zum Thema Leseverstehen anhand einer Kommunikationssituation, in der Sender\*in und Empfänger\*in via einen geschriebenen Text kommunizieren. In der unteren Bildhälfte sind entlang des gebogenen Pfeils die Aktivitäten der Person verzeichnet, die den Text verfasst. Der obere Bildteil bezieht sich darauf, wie Lesende einen Text verstehend lesen, eine mentale Repräsentation des Text aufbauen und schließlich in Form des Textverständnisses Wissen über den Gegenstand aufbauen. Dies ist eine eigene Darstellung in Anlehnung an das Modell zur Wissensvermittlung bei Schnotz, 1994, S. 13.

integrativen Bezugsrahmen“ für Textbewertungen von Göpferich (2002, S. 155 ff.).

Die Abbildung 3.1 stellt die zentralen Begriffe kompakt zusammen und setzt sie miteinander in Verbindung. Der Abbildung liegt in Anlehnung an Schnotz (1994, S. 13) eine symbolhaft dargestellte Kommunikation zwischen einer textverfassenden Person und einer lesenden Person zugrunde, wobei sich der Text auf einen unbestimmten – real oder gedanklichen – Gegenstand bezieht. In der Kommunikationssituation stehen verfassende Person und lesende Person einerseits über den Text (linke Hälfte in der Darstellung) und andererseits über den Gegenstand (rechte Seite) miteinander in Verbindung. Die Verbindung anhand des Texts ist in der greifbaren Wirklichkeit realisiert, diejenige über den Gegenstand besteht in der Gedankenwelt. Das verstehende Lesen schlägt nun die Brücke von einem vorhandenen Text zum mentalen Abbild.

Die rechte Modellhälfte bezieht sich also auf ein implizites, gedankliches Verhältnis: Dieses Verhältnis besteht konkret zwischen dem *Textverständnis* auf Seiten der lesenden Person und dem *textbezogenen Gegenstandskonzept* der verfassenden Person. Das Textverständnis der lesenden Person ist das auf den Textgegenstand bezogene Ergebnis des Verstehensprozesses beim Lesen. Es bezeichnet die Wissensstrukturen, die sich die lesende Person aus dem Text rekonstruiert hat. Das Textverständnis kann als Pendant zum textbezogenen Gegenstandskonzept der verfassenden Person betrachtet werden. Letzteres meint die Wissensstrukturen über den Gegenstand, welche die verfassende Person in den Text einfließen lässt. Im Zuge

### 3. Wissensvermittlung anhand von Texten

der *Textkonzeption* berücksichtigt die verfassende Person, welche *kommunikative Funktion* der Text erfüllen soll (Lesezweck, Verhältnis der verfassenden Person zum Textgegenstand, Dispositionen der Adressat\*innen, siehe dazu im Einzelnen Unterkapitel 3.3). Die verfassende Person modelliert dann passend zur kommunikativen Funktion eine textliche Repräsentation des eigenen Gegenstandskonzepts. Im Vergleich zwischen Gegenstandskonzept und Textverständnis entscheidet sich, ob eine lesende Person den Text verstanden, nicht verstanden oder missverstanden hat. Kommen das Verständnis der lesenden Person und das Konzept der verfassenden Person zur Deckung, so wurde der Text *verstanden*. Ein lückenhaftes Verständnis bei der lesenden Person, das solche Fragen und gedankliche Widersprüche aufwirft, die nicht von der verfassenden Person vorgesehen waren, bedeutet ein *Nicht-Verstehen*. Vom *Missverstehen* ist die Rede, wenn zwar keine Widersprüche für die lesende Person offenkundig sind, sich jedoch das Verständnis vom textbezogenen Gegenstandskonzept der verfassenden Person (zu stark) unterscheidet. Eine gewisse individuelle Note wird ein zutreffendes Verständnis jedoch stets tragen, da der Internalisierungsprozess maßgeblich von Personenvariablen abhängig ist (Schnotz gebraucht den Ausdruck „*Be-deutung* (sic) eines Texts“, er betont also den transitiven Charakter des Verbs *deuten*, um deutlich zu machen, dass eine lesende Person aktiv aus dem Text Bedeutung rekonstruieren muss; siehe Schnotz, 1994, S. 14). Zusätzlich zu dem bis hierhin beschriebenen *kognitiven Gelingenskriterium* des verstehenden Lesens, das die aufgebauten Wissensstrukturen betrifft, ist im Kontext der Schule und damit der Leseförderung als Bildungsziel auch der *motivatorischen Komponente* Beachtung beizumessen (eine prägnante Zusammenfassung der kognitiven und motivationalen Gelingenskriterien verständlichen Lesens findet sich bei Groeben, 1978, S. 80 f.).

Auf der linken Seite der Abbildung 3.1 ist dargestellt, inwiefern die verfassende und die lesende Person durch den Text miteinander in Verbindung stehen. Der Text entsteht, indem die verfassende Person ihr Gegenstandskonzept *exteriorisiert*, das heißt für andere wahrnehmbar zum Ausdruck bringt. Im Falle eines geschriebenen Texts, der eine Exteriorisierung anhand von Schriftzeichen bedeutet, ist das *Wahrnehmbare*, das die Lesenden rein über den Sehvorgang aufnehmen, die *Textoberfläche* (auch: *Code*). Sie ist das Ergebnis der *Kodierung*, welche im hier diskutierten Kontext von Kommunikation die Festlegung der verfassenden Person auf die typographische Textgestaltung und auf bestimmte Worte und Formulierungen bezeichnet. Das sprachliche Inventar, das in diesem schöpferischen Prozess als Baustoff zur Verfügung steht, sind die *sprachlichen Mittel*. Sie bezeichnen die Werkzeuge beim Zusammenstellen eines Texts aus Wörtern, Phrasen und Sätzen<sup>2</sup>. Beispiele für sprachliche Mittel sind Synonyme, Fremdwörter, Funktionalverben, Ellipsen, Metaphern, Relativsatz-Konstruktionen, Nominalphrasen, etc. Sie werden so eingesetzt, dass das Konzept sach- und adressat\*innengerecht dargestellt wird (für Näheres dazu siehe Abschnitt 2.1.2) und der Text seiner kommunikativen

---

<sup>2</sup>Ein pointierter, anschaulicher, allerdings wenig wissenschaftlicher Begriff, der all diejenigen Maßnahmen zusammenfasst, welche bei der Gestaltung der sprachlichen Textoberfläche angewendet werden können ohne dabei die Inhaltsebene zu beeinträchtigen, ist die „Textkosmetik“. Den Begriff nutzen z. B. Spitta (2010) und Heiser (2019).

Funktion (siehe Seite 40) und weiteren Rahmenbedingungen entspricht. Letztere Bedingungen sind *Eckdaten*, die sich aus redaktionellen, konventionellen oder inhaltlichen Vorgaben ergeben (siehe dazu im Detail die Ausführungen zum Karlsruher Verständlichkeitskonzept auf Seite 55 in Unterkapitel 3.3). Die lesende Person nimmt den rezipierenden Part ein, wobei die Rezeption des Texts keineswegs aus einer passiven Position heraus geschieht, sondern einen aktiven *Verstehensprozess* benötigt. Das *verstehende Lesen* (synonym: *Leseverstehen* oder *Textverstehen*) meint die Genese von Bedeutung aus geschriebenem Text (z. B. Schnotz, 1994, S. 13 f.) und ist somit die Gesamtheit an Tätigkeiten, selbstregulatorischen Maßnahmen und kognitiven Teilprozessen beim Rekonstruieren von Wissensstrukturen aus dem Text (zusammenfassend nach Lenhard, 2013, S. 27; einen tiefgehenden Einblick in die Selbstregulation beim Lesen bieten Philipp und Schilcher, 2012; die Teilprozesse beim Leseverstehen werden in Abschnitt 3.2.1 genauer erläutert). Das Ergebnis verstehenden Lesens ist ein Textverständnis, welches die Informationen für den Aufbau einer mentalen Repräsentation des Textinhalts gibt, des sogenannten *Situationsmodells* oder *mental Modells* (siehe dazu auch Abschnitt 3.2.1). Der Begriff „situation model“ wurde in der englischen Sprache maßgeblich durch die Arbeiten von Walter Kintsch und Teun van Dijk geprägt (siehe z. B. Van Dijk, 1995) und so ins Deutsche übernommen (z.B. bei Holle, 2010; Lenhard, 2013; Artelt et al., 2005; Härtig, Fraser, Bernholt & Retelsdorf, 2019). Ein in der Literatur häufig verwendetes Synonym lautet *mentales Modell* (z.B. bei Staraschek, 2006; Härtig, Bernholt, Prechtel & Retelsdorf, 2015; Schnotz, 1994; englisch „mental model“ bei Johnson-Laird, 1981). Für die Teilprozesse beim Aufbau eines Situationsmodells sind individuelle Dispositionen der Lesenden fundamental. Als entscheidende Einflussgrößen für die Textverarbeitung gelten allgemein-kognitive Ausgangsbedingungen wie inhaltliches und lexikalisches Wissen, Vertrautheit mit dem Textformat, schlussfolgerndes Denken sowie selbstregulatorische Maßnahmen sowie motivationale Voraussetzungen. Denn das Situationsmodell einer lesenden Person entsteht unter anderem unter Rückbezug auf inhaltliche Deduktionen und Induktionen, Gedankenverbindungen, vermutete Ähnlichkeitsbeziehungen zu themenverwandten Vorwissensbeständen, etc. (Unterkapitel 3.2 geht darauf im Einzelnen ein). Schlussendlich reicht das Situationsmodell also über die Textbasis hinaus. Es erhält eine reale Bedeutung, indem die lesende Person die Buchstaben und Symbole im Text hinsichtlich des Textgegenstands *deutet*.

Das Situationsmodell als Ergebnis des Leseverstehens ist also abhängig von individuellen Eigenschaften, Erfahrungen und Aktionen der Lesenden. Neben den textbezogenen Faktoren bestimmen also insbesondere die Dispositionen der Lesenden die *Gelingsbedingungen* für verstehendes Lesen. Den einzelnen Personenfaktoren wendet sich Abschnitt 3.2.2 im Detail zu. Die dabei zentralen Begriffe werden hier erläutert: Um Texte erfolgreich verstehen zu können, bedarf es basaler *Lesefähigkeiten*, die das Re- und Dekodieren von Buchstaben und Wörtern und die Satzerkennung meinen, sowie komplexer kognitiver Fähigkeiten, die sich auf Selbstregulation und auf logisches Schlussfolgern und Verknüpfen beziehen. Die Kombination aus den genannten Fähigkeiten und der *Motivation*, sich mit einem Text auseinanderzusetzen, macht die *Lesefertigkeit* einer lesenden Person aus. Schilcher und Wild, 2018, betonen, inwiefern

### 3. Wissensvermittlung anhand von Texten

Leserfolg von beiden Komponenten – Fähigkeiten und Motivation – abhängen und inwiefern verschiedene Schul- und Leseleistungsstudien dies erfassen. Die Lesemotivation lässt sich in Anlehnung an Krapps Ausführungen zur Lernmotivation als Anstrengungsbereitschaft mit intentionalem Charakter auffassen, also als die bewusste und zielgerichtete Bereitschaft, sich aktiv und wirkungsvoll mit einem Text zu beschäftigen (Krapp, 1999 und 1993). Für diese Arbeit besonders relevant ist der von Krapp (1993, S. 189) eingeführte Begriff der *aktualisierten Motivation*, welche die „in einer bestimmten Situation tatsächlich auftretende [...] Motivation“ meint. Sie ist eng an das situationale Interesse gekoppelt, also an die kurzfristige, von einer konkreten Wahrnehmung ausgelöste Neigung der Lesenden, sich aktiv mit dem Text beziehungsweise mit seinem Inhalt auseinanderzusetzen, und an eine damit verbundene spontan empfundene Lesefreude (vergleiche Lexikon der Psychologie, 2021; siehe zur Relevanz des situationalen Motivationszustands auch Groeben, 1978, S. 12, S. 38 f. und S. 48). Von dieser Auffassung von Motivation lässt sich die habituelle Motivation abgrenzen, die sich auf Lesegewohnheiten bezieht<sup>3</sup>. Noch eine Stufe umfassender als die Lesefertigkeit wird der Begriff der *Lesekompetenz* aufgefasst. Mit der inflationären Verwendung des Kompetenzbegriffs in allen Bildungswissenschaften geht ein gewisser Verlust an Präzision einher. Es besteht jedoch weitestgehend Konsens darüber, dass Kompetenzen erfassen, inwiefern eine Leistung erbracht wird und mit-meinen (sollten), was mit einer erbrachten Leistung anzufangen ist (Artelt et al., 2005, S. 11; Schilcher & Wild, 2018). Es geht also bei der Lesekompetenz um die grundsätzliche Fähigkeit einen Text zu lesen, die Umsetzung dieser Fähigkeit und um die weiterführende Auseinandersetzung mit dem Text – insbesondere in einem sozialen Kontext (Lenhard, 2013, S. 47 f.). Der Lesekompetenzbegriff, wie er in der PISA-Studie verwendet wird, fügt sich in diese Auffassung: „Lesekompetenz (reading literacy) heißt, geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potenzial weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (Deutsches PISA-Konsortium, 2013, S. 58). Und: „Sie [die Lesekompetenz] enthält nicht nur die kognitive Leistungskomponente, sondern auch Haltungen, Einstellungen und förderliche Strategien und Routinen, die einen aktiven Leser auszeichnen. [...] Lesekompetenz ist also die Verfügbarkeit und das effektive Zusammenspiel der am Lesen beteiligten Komponenten“ (Deutsches PISA-Konsortium, 2013, S. 57). Trotzdem ist nicht unumstritten, ob die in der PISA-Studie umgesetzte Operationalisierung und damit die Erfassung der Lesekompetenz diesem grundbildungsorientierten Verständnis Rechnung zu tragen vermag. Schilcher und Wild (2018) merken an, dass die Studie ein stark instrumentalisiertes Konstrukt messe, das die *Informationsentnahme aus Texten* fokussiert und damit interaktive, motivationale und interessenbezogene Facetten, die Relevanz des Lesezwecks und des soziokulturellen Hintergrunds übersieht. Eine umfassendere Auffassung von Lesekompetenz wird bei Philipp und Schilcher (2012, S. 38 ff.) beschrieben. Ein solches Verständnis von Lesekompetenz, welches basale

---

<sup>3</sup>Solche Lesegewohnheiten und – sehr allgemein – kommunikative Praktiken in einer soziokulturellen Gemeinschaft greift der Begriff der *Literalität* auf (Lenhard, 2013, S. 48). Der Begriff ist jedoch für diese Arbeit weniger bedeutsam.

und komplexere leserbezogene Lesefähigkeiten, Lesemotivation, Lesestrategiewissen, die Fähigkeit zur Selbstregulation beim Lesen sowie die kritische und schlussfolgernde Auseinandersetzung mit dem Text umfasst, liegt der Studie, die im Rahmen dieser Arbeit entwickelt, durchgeführt und ausgewertet wurde, zugrunde. Die Lesekompetenz ist damit eine ausschlaggebende Größe für die subjektive Haltung der Lesenden gegenüber dem Text als auch für das kognitive Leseergebnis, also den Wissenszuwachs auf Grundlage des Textverstehens.

Den personenbezogenen Gelingenbedingungen für verstehendes Lesen stehen die textbezogenen Faktoren gegenüber. Beim verstehenden Lesen treten Lesende in *Interaktion* mit dem Text. Die lesende Person bringt individuelle Voraussetzungen mit, welche die Prozesse zur Verarbeitung des Texts laufend beeinflussen und somit auch das resultierende Textverständnis mitbestimmen (Lenhard, 2013, S. 27 f.; Artelt et al., 2005, S. 12 f.). Der Text seinerseits wirkt ebenfalls auf die Interaktion ein: es lassen sich Merkmale auf Wort-, Satz- und Textebene identifizieren, die maßgeblich für die Textverständlichkeit sind (Patterson, Roman, Friend, Osborne und Donovan, 2018, S. 296; Frickel, 2018, S. 190; Lenhard, 2013, S. 27; siehe Unterkapitel 3.3).

Als *Textverständlichkeit* kann die kognitive und affektive Wirkung zusammengefasst werden, die der Text insgesamt auf Grundlage des Zusammenspiels aus personen- und textbezogenen Faktoren beim verstehenden Lesen entfaltet. Der vorhergehende Absatz macht deutlich, dass die Textverständlichkeit keine objektive Größe sein kann, sondern sich erst in der Interaktion zwischen Text und lesender Person manifestiert. Besonders deutlich zum Ausdruck bringt dies der Begriff der *Textpassung*, welcher ebenfalls eine resultierende Textwirkung bezeichnet und dabei den Fokus darauf legt, inwiefern die Ausprägung der Faktoren geeignet für die Zielgruppe ist. Ein verwandter Begriff, der zwar ebenso mit Textverständlichkeit verbunden ist, dabei aber die individuellen Lesenden außer Acht lässt, ist die *Lesbarkeit* (selten auch: *Leserlichkeit*, z.B. bei Göpferich, 2002) eines Texts. Sie setzt sich aus rein objektiven und quantitativen Merkmalen zusammen (z.B. Wortlänge und Satzlänge) und ist somit maßgeblich für den *Lesefluss* beim Lesen ohne jegliche inhaltliche Verarbeitungsaktivität. Der Lesefluss ist darüber hinaus von augenscheinlichen, typographischen Gestaltungsmustern beeinflusst, die üblicherweise durch den Anglizismus *Layout* beschrieben werden und von Göpferich (2002, S. 186) in der Dimension „*Perzipierbarkeit*“ repräsentiert sind.

Ausgehend von dieser umfassenden begrifflichen Klärung wird nun zunächst die Seite der Lesenden hinsichtlich relevanter kognitiver Prozesse beim verstehenden Lesen (Abschnitt 3.2.1) und hinsichtlich personenbezogener Voraussetzungen (Abschnitt 3.2.2) näher beleuchtet.

## 3.2 Personenbezogene Abläufe und Voraussetzungen

Lesen kann zu ganz verschiedenen Zwecken (verstehen, kritisieren, reflektieren, unterhalten werden, angeleitet werden, ...) geschehen und je nach Zweck variieren auch die Anforderun-

### 3. Wissensvermittlung anhand von Texten

gen an das Lesematerial sowie an die lesenden Personen (siehe dazu z. B. die Erläuterungen zum Textproduktionsmodell von Susanne Göpferich in Unterkapitel 3.3 auf Seite 55). Der Unterricht bringt nun mitunter den Zweck der deklarativen Wissensaneignung (der Begriff wird im Abschnitt 2.1.1 auf Seite 18 erläutert) mit sich. Dient ein Text der Wissensaneignung, so ist vor allem der spezielle Modus des *verstehenden Lesens* gefragt. Schon rein wörtlich betrachtet verbindet der Ausdruck *verstehendes Lesen* die äußere Tätigkeit des Lesens mit dem geistigen Prozess des Verstehens. Und diese Verbindung besteht nicht nur in sprachlicher Hinsicht: Um verstehend lesen zu können, müssen tatsächlich sensorische mit kognitiven Aktivitäten ineinander greifen. Die konkreten Abläufe, wenn Schriftzeichen zu Schriftzeichenkombinationen, Schriftzeichenkombinationen zu Wörtern, Wörter zu Phrasen, Phrasen zu Sätzen und Sätze zu einer kohärenten, sinnhaltigen, mentalen Repräsentation werden, sind Gegenstand dieses Kapitels. Dabei nehmen individuelle Erfahrungen und Wissensbestände der lesenden Person Einfluss auf das resultierende Verständnis des Texts. In diesem Sinne wird also die lesende Person im Fokus stehen.

Verstehendes Lesen ist ein Sprachverarbeitungsprozess. Es heißt, Bedeutung aus geschriebenem Text zu generieren (Schnotz, 1994, S. 13; vergleiche auch die Begriffsbestimmung auf Seite 41). Es wird im Laufe des folgenden Kapitels klar werden, dass verstehendes Lesen konstruktivistisch aufzufassen ist, da die Bedeutungsgenese für Lesende bedeutet, sich kognitiv aktiv und kreativ mit Text auseinanderzusetzen.

Bei der Beschreibung der kognitiven Sprachverarbeitungsprozesse lassen sich generell zwei Positionen unterscheiden. Sogenannten *reizgesteuerten* Modellen zufolge setzt ein geschriebener Text anhand eines visuellen Reizes eine seriell und autonom ablaufende Sprachverarbeitung bei den Lesenden in Gang (siehe dazu Beyer und Gerlach, 2018, S. 27 ff.): Auf die Worterkennung (phonetische gefolgt von lexikalischer Verarbeitung) baut die Analyse der Syntax auf und daran wiederum schließt sich die semantische Verarbeitung an. Die resultierende Wissensrepräsentation bei den Lesenden entsteht also, indem eine Verarbeitungsstufe auf die nächste folgt. Die Verarbeitungsprozesse werden dabei als unbeeinflusst vom Vorwissen der Lesenden über Inhalte oder Textstrukturen betrachtet. Derartigen Modellen steht eine Auffassung entgegen, die ein *interaktives Zusammenspiel* der einzelnen Verarbeitungsebenen beinhaltet. Das bedeutet, dass die Persönlichkeitsmerkmale, die individuellen Fähigkeiten, die Wissensbestände und die Leseaktivitäten der Lesenden (hierzu zählen: Strategienutzung, Kohärenzbildung, Inferenzziehen u.ä., siehe dazu im Detail Abschnitt 3.2.2) die Sprachverarbeitung mitbestimmen (Lenhard, 2013, S. 14 f.). Solche interaktiven Modelle heißen daher auch *vorwissensgesteuerte* Modelle (Beyer & Gerlach, 2018, S. 28). Diese Arbeit folgt der interaktiven, vorwissensgesteuerten Auffassung. Entsprechende Ausführungen zum Zusammenspiel einzelner Teilprozesse sind Gegenstand des folgenden Abschnitts.

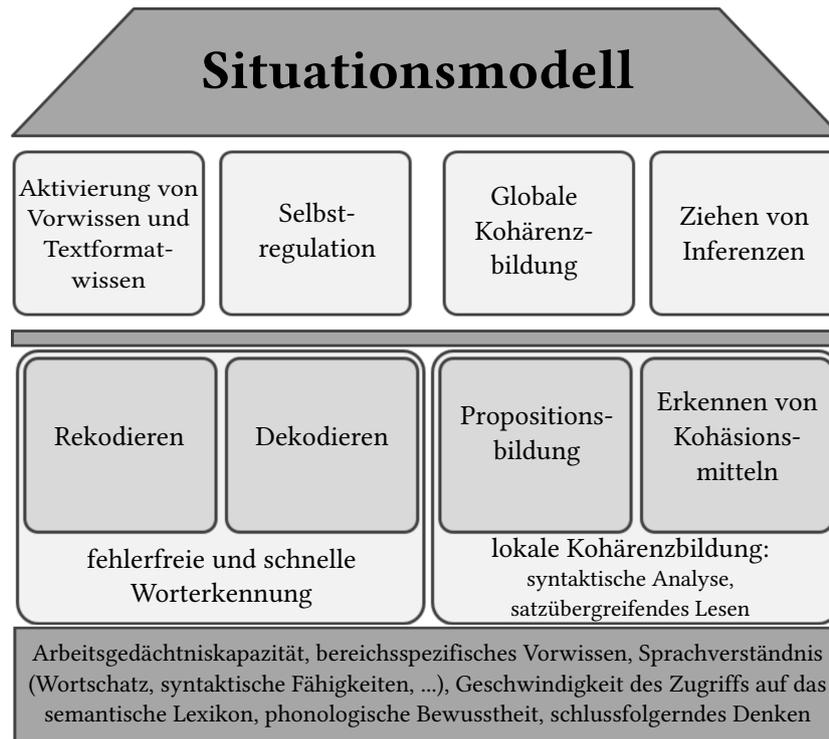


Abbildung 3.2.: Abbildung zum Aufbau eines Situationsmodells nach Lenhard (2013, S. 15). Es ist anzunehmen, dass beim verstehenden Lesen die hierarchieniederen Prozesse (untere „Etage“ in der Abbildung) und die hierarchiehöheren Teilprozesse (obere „Etage“) zu großen Teilen parallel ablaufen. Die Basis des verstehenden Lesens bilden diverse allgemeine kognitive Voraussetzungen der Lesenden.

### 3.2.1 Teilprozesse beim Leseverstehen

Auf grundlegender Ebene erfordert das verstehende Lesen rein mechanische, sensuelle Abläufe. Um jedoch das passiv Rezipierte weiterzuverarbeiten und zu einem (zutreffenden) Textverständnis zu gelangen, müssen Lesende eine aktive Lernhaltung einnehmen und sinnzentriert mit dem Textmaterial arbeiten (Groeben, 1978, S. 12, S. 48). Die Lesedidaktik, die zum Ziel hat das Lesen zu fördern, also „die Lesemenge, die Bereitschaft zum Lesen oder die Lernfähigkeit aus Texten bei Kindern und Jugendlichen zu steigern“ (Rosebrock, 2020), ist daher eng verflochten mit weiteren Wissenschaftsdisziplinen, die unter anderem den Lesevorgang untersuchen. Zu diesen Bezugswissenschaften zählen die Kognitionspsychologie, die kognitive Neurowissenschaft, die Kommunikationswissenschaft (hier ist insbesondere der sozialwissenschaftliche Blickwinkel auf die Lesesozialisation zu nennen) und die Informationswissenschaft (hier werden die computerbasierte Untersuchung – z.B. Eye-Tracking – und auch Imitation des Leseprozesses immer populärer) (vergleiche Schilcher & Wild, 2018).

Im Folgenden werden die Prozesse, die beim verstehenden Lesen ablaufen, detailliert beschrieben. Das heißt in Bezug auf Abbildung 3.1 (Seite 39), die Aufmerksamkeit auf die obere Bildhälfte zu lenken – also auf die Beziehung zwischen Gegenstand, Text und lesender Person in einer Kommunikationssituation. Den Prozess, um von der Textoberfläche zu Begriffen und

### 3. Wissensvermittlung anhand von Texten

Begriffsnetzen zu gelangen bezeichnet man als *Aufbau eines Situationsmodells*. Lenhard (2013) teilt die einzelnen Teilprozesse, die dabei ablaufen, in hierarchieniedere und hierarchiehohe Prozesse ein, je nachdem, ob sie kürzere Textbausteine (Wörter, Satzteile, Sätze, kurze Satzgruppen) betreffen oder zur Erfassung größerer Textbausteine und deren Zusammenhänge dienen. Die Abbildung 3.2 illustriert, wie Lesende ein Situationsmodell aufbauen. Lenhards Ausführungen knüpfen damit an Kintschs *Construction-Integration-Model* an (Kintsch, 1998). Demgemäß wird die Textoberfläche beim Lesen zunächst in eine propositionale Struktur überführt. Ein Satz wird dabei in Bedeutungseinheiten und deren Zusammenhänge zerlegt, indem Wortgruppen gebildet werden. Für die Gruppenbildung sind das Verb und ihm zugeordnete inhaltliche Rollen ausschlaggebend. Solche Rollen sind zum Beispiel Handlungstragende, Objekte, Empfangende der Handlung, Instrumente, Orte und Zwecke<sup>4</sup>. Die Propositionsbildung mündet dann zunächst in ein Begriffsnetz, das nicht in einen größeren thematischen Kontext eingebettet ist (*Construction*). Die noch bestehenden Lücken und Ungereimtheiten werden in der *Integration*-Phase durch Inferenzziehen und Verknüpfungen zum Vorwissen gefüllt beziehungsweise bereinigt. Lenhard (2013) führt dementsprechend detaillierter aus: Ausgehend vom geschriebenen Textmaterial müssen auf der unteren Ebene der Hierarchie Wörter erkannt, Wortgruppen zu Bedeutungseinheiten zusammengefügt und Bezüge zwischen Satzteilen beziehungsweise Sätzen erfasst werden. Zusammenfassend spricht man bei diesen Prozessen von der *lokalen Kohärenzbildung*. Ein Text, der auf dieser Ebene leicht zu verarbeiten ist, weist eine hohe *lokale Kohärenz* auf. Für die lokale Kohärenzbildung spielen Oberflächenstruktur (Wortreihenfolge) und Tiefenstruktur (Syntax) eine Rolle. Parallel dazu, das heißt ohne zeitlichen Versatz, sind auch Prozesse auf hierarchiehoher Ebene aktiv: Unter Nutzung bereits vorhandener Wissensbestände verknüpfen Lesende den Textinhalt mit ihrem Vorwissen und ziehen Inferenzen. Sie wenden darüber hinaus Strategien der Selbstregulation an, um die kürzeren Textbausteine in eine vergrößerte, verdichtete Form zu überführen und um globale Zusammenhänge zu erkennen (Bildung *globaler Kohärenz*) und somit dem Gelesenen eine inhaltliche Struktur einzubeschreiben. Das Endprodukt dieser Prozesse ist das sogenannte Situationsmodell, ein *mentales Modell*, das nicht mehr in einer Schriftzeichenabfolge besteht, sondern in verarbeiteten Textinformationen, die um weitere Wissens Elemente angereichert und in das semantische Netzwerk des Gehirns eingegliedert worden sind.

Die Vorstellungen, die Lenhards Ausführungen zugrunde liegen, sind gemäß der *Fuzzy Trace Theorie* (Brainerd & Reyna, 1988) zutreffend auf das Leseverstehen geübterer Lesender. Etwa ab dem neunten Lebensjahr (Lesealter drei bis vier Jahre) speichern Lesende die Bedeutung eines Texts mehr in abstrakter als in wörtlicher Form ab (Lenhard, 2013, S. 39 f.). Genauer sagt die *Fuzzy Trace Theorie*, dass beide Formen – abstrakt sowie wörtlich – voneinander trennbar und beide stets aktiv sind, jedoch bevorzugen Leseanfänger\*innen die sogenannte *verbatim*, also die wörtliche Form, und fortgeschrittene Leser\*innen die *gist*, also sinngemäße Form.

---

<sup>4</sup>Das folgende Beispiel aus Beyer und Gerlach, 2018, S. 22 veranschaulicht die Propositionsbildung:

„Hans gibt Maria das Buch“ → GEBEN(HANS, MARIA, BUCH)

Dabei ist die propositionale Struktur VERB(HANDLUNGSTRAGENDE, HANDLUNGSEMPFANGENDE, OBJEKT).

Da sie den Inhalt anhand ihrer Wissensbestände elaborierter verarbeiten können, dominiert hier eine sinngemäße Abspeicherung.

Vor dem Hintergrund der hier beschriebenen Teilprozesse, die beim verstehenden Lesen ablaufen, leuchtet ein, dass die lesende Person mit ihren Lesekompetenzen und ihrem Erfahrungshintergrund ganz individuell maßgeblich dafür ist, zu welchem Textverständnis sie gelangt. Das folgende Kapitel beleuchtet daher die Dispositionen der lesenden Person und ihre Rolle beim verstehenden Lesen genauer.

### 3.2.2 Einfluss personenbezogener Dispositionen

Das Modell für Leseverstehen von Lenhard (2013), das im vorangehenden Abschnitt 3.2.1 beschrieben wurde, umfasst neben Prozessbeschreibungen auch Personenmerkmale, die dem verstehenden Lesen vorausgesetzt sind. Den Grundstock für die bereits erläuterten mentalen Prozesse bilden in diesem Modell die folgenden kognitiven Lesendenvoraussetzungen: Arbeitsgedächtniskapazität, bereichsspezifisches Vorwissen, Sprachverständnis, Geschwindigkeit des Zugriffs auf das semantische Lexikon, phonologische Bewusstheit und schlussfolgerndes Denken. Unter der Arbeitsgedächtniskapazität versteht man die Gesamtheit aus dem Fassungsvermögen der Speichereinheit bezüglich sprachlicher Informationen und der Speichereinheit bezüglich visuell-räumlicher Informationen sowie deren Überwachung beziehungsweise Steuerung durch eine zentrale exekutive Gedächtniskomponente (zu diesem weit verbreiteten Arbeitsgedächtnismodell von Baddeley siehe Strobach, 2020, S. 87 ff.). Das bereichsspezifische Vorwissen bezieht sich auf Wissensbestände über die Textinhalte, die Lesende schon vor der Lektüre haben. Das Sprachverständnis umfasst grammatikalische Kompetenzen und Wortschatzwissen (Ennemoser, Marx, Weber & Schneider, 2012). In Bezug auf das Verständnis bildungsbezogener Sprachformen spricht man hier insbesondere von „Cognitive Academic Language Proficiency“ (CALP; der Begriff, der auf Jim Cummins (1979) zurückgeht, wird in Abschnitt 2.1.2 auf Seite 25 erläutert). Der Zugriff auf das semantische Lexikon meint die Erfassung einer Wortbedeutung ausgehend von Schriftzeichen. Dabei wird angenommen, dass dies auf mindestens zwei verschiedenen Wegen ablaufen kann – je nachdem, ob ein Wort direkt als Wortbild erkannt wird, oder ob es indirekt zunächst Buchstabe für Buchstabe in eine Lautfolge übersetzt werden muss (Lenhard, 2013, S. 16 f.). Die entscheidende sprachanalytische Fähigkeit bei diesem Übersetzungsvorgang ist die phonologische Bewusstheit (sie bezieht sich auf die innerliche Übersetzung von Buchstaben in das Lautbild eines Worts; siehe Lenhard, 2013, S. 33). Das schlussfolgernde Denken (auch logisches Schlussfolgern) ist eine von mehreren Facetten der Intelligenz, die in Form von induktivem und deduktivem Denken bei der Auseinandersetzung mit kognitiv anspruchsvollen Problemen zum Tragen kommt (Rost, 2013, S. 52).

Viele dieser Begriffe finden Erwähnung in einem Modell von Ennemoser, Marx, Weber und Schneider (2012), welches die Einflussfaktoren auf Textverständnis von Kindern am Ende der vierten Klasse miteinander in Beziehung setzt. Auf Basis zweier verschiedener

Studien entwickelten Ennemoser, Marx, Weber und Schneider ein statistisch gesehen gut belegtes Modell zur Vorhersage der Lesegeschwindigkeit, des Textverständnisses und der Rechtschreibleistung von Kindern im Grundschulalter. Gemäß diesem Modell lassen sich in stark verkürzter Weise einige Prädiktorvariablen für das Textverständnis am Ende der vierten Klasse nennen: Das Textverständnis ist maßgeblich durch die Lesegeschwindigkeit zum Ende der ersten Klasse und die linguistische Kompetenz bestimmt. Die Lesegeschwindigkeit wiederum hängt von basaleren kognitiven Fähigkeiten wie der Arbeitsgedächtniskapazität, dem IQ, der Zugriffsgeschwindigkeit auf das semantische Lexikon, der linguistischen Kompetenz und der phonologischen Bewusstheit ab<sup>5</sup>.

Wenn auch in der Primarstufe wegweisende Grundlagen für die Lesekompetenzentwicklung gelegt werden und der Entwicklungsverlauf besonders dynamisch ist (Lenhard, 2013, S. 38 f. und S. 43 f.), so bildet sich das verstehende Lesen dennoch im Verlauf der Schulzeit stetig weiter aus. Ab dem frühen Jugendalter betrifft dies vor allem die hierarchiehöheren, selbstregulatorischen Prozesse des Leseverstehens (Artelt et al., 2005, S. 37 f.). Die Aneignung basaler Lesefähigkeiten ist zu diesem Zeitpunkt in der Regel weitestgehend abgeschlossen.

#### **Die besondere Rolle des Textformatwissens und des inhaltlichen Vorwissens**

Zwar wird der Begriff „Vorwissen“ in Verbindung mit dem Textverstehen rege gebraucht, doch fehlen ihm genau genommen scharfe Konturen. Was ist mit Vorwissen im Vergleich zu Wissen gemeint? Gibt es Wissen, das kein Vorwissen ist? Der Begriff betont jedoch, dass gewisse Wissensbestände die Textverarbeitung bereits *im Voraus* mitbestimmen. Dementsprechend sind das (Vor-)Wissen über die vorliegende Textgattung und das bereichsspezifische inhaltliche (Vor-)Wissen zentrale Größen des Leseverstehens.

Derjenige Aspekt des Vorwissens, der die Erwartungen der Lesenden hinsichtlich der Textinhalte und deren Aufbau (sogenannte *Geschichtengrammatiken*, siehe Artelt et al., 2005, S. 16) betrifft, wird bei Lenhard (2013) „Textformatwissen“ genannt. Lesende nehmen dementsprechend eine intuitive Vorstrukturierung vor, welche die sinnvolle Aufmerksamkeitssteuerung beim Lesen unterstützt und somit das Verstehen erleichtert, sofern der Text den üblichen, bekannten Strukturen entspricht (Lenhard, 2013, S. 22; Artelt et al., 2005, S. 27 f.).

Neben dem textstrukturbezogenen Wissen nimmt vor allem auch das inhaltliche Wissen rund um das Thema des Texts Einfluss auf das Textverstehen. Diese Einflussnahme ist erstens theoretisch gut zu begründen und außerdem empirisch oftmals belegt (siehe z. B. Patterson, Roman, Friend, Osborne & Donovan, 2018; Anderson, Reynolds, Schallert & Goetz, 1977). Wie Abschnitt 3.2.1 erläutert, sind die Wissensbestände der Lesenden für die globale Kohärenzbildung sowie das Ziehen von Inferenzen entscheidend. Sie wirken sich außerdem auf die Aufmerksamkeitssteuerung beim Lesen aus, indem Lesende die Relevanz einzelner

---

<sup>5</sup>In die Entwicklung des Modells flossen neben den Messwerten der Observablen auch theoretisch basierte Überlegungen ein. Solche betrafen etwa die Wirkrichtung von Variablen aufeinander oder inwiefern einzelne Variablen direkt oder indirekt aufeinander einwirken. Für die exakten Wirkbeziehungen, die Stärke der Zusammenhänge sowie die detaillierte Variablendefinition sei hier auf die Originalarbeit von Ennemoser, Marx, Weber und Schneider (2012) verwiesen.

Textpassagen mehr oder weniger angemessen bewerten. Es hängt also mitunter vom individuellen inhaltsbezogenen Vorwissen ab, inwiefern metakognitive Strategien erfolgreich zum Leseverstehen beitragen (Artelt et al., 2005, S. 37). Artelt et al. (2005, S. 14 und S. 24) stellen weiter fest, dass bei stark ausgeprägtem inhaltlichen Vorwissen ein anfordernderer Textcharakter durch schwächer ausgeprägte Kohärenz dem Verstehensprozess sogar zuträglich sein kann. Einen anderen Aspekt des Einflusses von Vorwissen bildet die Möglichkeit von Mehrdeutigkeiten: Geschriebener Text bietet oftmals vielseitige Interpretationsmöglichkeiten, die je nach Vorwissen und/oder Kontext zum Tragen kommen. Die kontextbezogene Mehrdeutigkeit wurde in Abschnitt 3.2.1 anhand eines Beispiels zum Wort „Erde“ (Seite 23) erläutert. Es ist anzunehmen, dass die vorwissensbezogene Mehrdeutigkeit zum Beispiel beim Begriff der Kraft zutage treten wird, wenn man erstens eine Physikprofessorin und zweitens ein Kind im Grundschulalter bittet, das Wort „Kraft“ zu beschreiben: Man wird kaum identische Beschreibungen erhalten. Die Genese von Bedeutung und die resultierende mentale Repräsentation der in einem Text vorkommenden Wörter und Ausdrücke variiert also entscheidend mit den vorwissensbezogenen Lesendenvoraussetzungen.

Schließlich soll der Vorwissensbegriff noch etwas weiter gefasst werden: Auch durch Vorerfahrungen geprägte innere Haltungen zählen zum Vorwissen, welches das Textverstehen beeinflusst. Wie in Abschnitt 2.1.1 erwähnt wurde, bedingen sich Wahrnehmung und Sprache gegenseitig. Entsprechend sind auch mentale Repräsentationen von Text durch Vorerfahrungen geprägt. Schweller, Brewer und Dahl (1976) bestätigten in mehreren Experimenten ihre Hypothese, dass die Erinnerung von Testpersonen an zuvor präsentierte Sätze durch Vorerfahrungen beeinflusst wird. Sie erwarteten, dass sich neutral gehaltene Äußerungen in der Erinnerung der Testpersonen hin zu Äußerungen mit mehr Wertung/Aufforderungscharakter verschieben. Und tatsächlich enthielten die Sätze, die die Testpersonen aus ihrer Erinnerung wiedergaben, solche Verben, die mehr Aussage tragen als die ursprünglich verwendeten Verben (in diesem Fall konnte eine Verschiebung von neutralen hin zu illokutionären Verben in erwartungskonformer Weise beobachtet werden).

Zwar ist das Vorwissen – ob bezogen auf den Inhalt, auf die Textstruktur oder auf innere Haltungen – eine Personeneigenschaft, die nicht spontan veränderbar, sondern als gesetzte Ausgangsbedingung für den Textverstehensprozess zu sehen ist. Allerdings kann ein Text durchaus so gestaltet sein, dass das Vorwissen in günstiger Weise aktiviert und eingebunden wird. Es macht einen entscheidenden Teil der Lesendenvoraussetzungen aus und muss beim Textproduktionsprozess berücksichtigt werden (adressat\*innengemäße Kommunikation; siehe auch Abschnitt 2.1.2).

### 3.2.3 Maßnahmen zur Leseförderung

Mit Blick auf die Veranlagungen, Neigungen und inneren Einstellungen von Lesenden lassen sich diverse Förderansätze für das (verstehende) Lesen herausarbeiten. Da erwiesenermaßen insbesondere mangelnde schriftsprachliche Lesekompetenzen zu zukünftigen Schulleistungs-

problemen beitragen (Lenhard, 2013, S. 11 f.), gibt es vielfältige Schulprogramme, die in Form von Lesetrainings der Leseförderung bei Schüler\*innen dienen. Artelt et al. beschreiben ausführlich, welche Ansätze es zur Förderung von Lesekompetenz gibt, wie diese von den Personeneigenschaften in der Zielgruppe abhängen und wie sie im vorschulischen und schulischen Bereich umsetzbar sind beziehungsweise in Form von Länderaktivitäten in Deutschland umgesetzt werden (siehe dazu Artelt et al., 2005, Kap. II.7, Kap. III und Kap. IV). Mit dem Fokus auf Selbstregulation beim Lesen geben Philipp und Schilcher (2012) im Sammelband „Selbstreguliertes Lesen: ein Überblick über wirksame Leseförderansätze“ einen Überblick über den Stand der Forschung sowie verschiedene Lesefördermaßnahmen im gegenseitigen Vergleich. Besonders praxisorientiert ist der Leitfaden für lesefördernde Unterrichtsmethoden, den Holliday und Cain (2012) aus vier großen englischsprachigen bildungswissenschaftlichen Untersuchungen zusammengestellt haben. Vor dem Hintergrund der aktuellen migrationsgesellschaftlichen Lage wird der Sprachförderung für besonders heterogene Zielgruppen eine gesteigerte Bedeutsamkeit zuteil. In diesem Zusammenhang sei beispielshalber auf das große BLK-Projekt FÖRMIG (Förderung von Kindern und Jugendlichen mit Migrationshintergrund) von verwiesen, an dem zahlreiche Projekte in zehn Bundesländern beteiligt sind (genauer dazu bei Gogolin et al., 2011).

Im Allgemeinen sind solche Programme, welche zum Ziel haben, dass lesende Personen ihre Kompetenzen steigern, langfristig anzulegen (Artelt et al., 2005, S. 63). Nicht zuletzt aufgrund der organisatorischen Rahmenbedingungen geht es in dieser Forschungsarbeit nicht um Unterstützungsmöglichkeiten zur persönlichen Verbesserung bei der Arbeit mit Texten, sondern um die Optimierung von Texten. Die Manipulation von Texteigenschaften ist unmittelbar umsetzbar und die Effekte sind kurzfristig messbar. Um einen Orientierungsrahmen für die Variablen der Textoptimierung zu schaffen, werden im Folgenden allgemeine (im Sinne von nicht-physikspezifischen) Kriterien für verständliche Textgestaltung erläutert. Im anschließenden Unterkapitel 3.4 wird anschließend der Stand der Forschung zur Wirkung von Sprache im naturwissenschaftlichen Unterricht zusammengefasst.

### **3.3 Verständliche Textgestaltung im Allgemeinen**

Auf den letzten Seiten standen die personenbezogenen Einflussfaktoren auf das Textverstehen im Fokus. Die nun folgenden Erläuterungen haben die Seite der Texte zum Gegenstand und damit die Frage, wie Texte zu gestalten sind, damit sie verständlich wirken und auch tatsächlich verstanden werden. Solche Texteigenschaften umfassen die inhaltliche Beschaffenheit und Organisation des Texts sowie sprachliche und augenscheinliche Textoberflächenmerkmale.

Insbesondere aus lese- und auch fachdidaktischer Sicht ist ein möglichst günstiger, lernwirksamer Verlauf der Interaktion Lesender mit Texten erstrebenswert. Um hier Einfluss zu nehmen, bietet die Seite der Texte den Vorteil, dass deren Eigenschaften direkter zugänglich sind: Textmaterial kann gezielt gestaltet werden. Dabei ist der Zeitpunkt der Gestaltung los-

gelöst vom Zeitpunkt der Interaktion, sodass schriftliche Sachtexte der verfassenden Person die Möglichkeit wiederholter Kontroll- und Korrekturdurchläufe bieten. Bevor die Texte den Lesenden präsentiert werden, können also insbesondere zentrale Begriffe identifiziert, eine anschlussfähige sprachliche und inhaltliche Umgebung für die Lerninhalte festgelegt und der sich ergebende Text weiteren Personen zur Überprüfung vorgelegt werden. Hierbei gründlich vorzugehen erscheint wichtig, da geschriebener Text beständig ist – er kann in der gleichen Form wieder und wieder rezipiert werden. Es herrscht also Bedarf nach Leitlinien für eine „gute“ Textgestaltung, die verstehendes Lesen bei den Adressat\*innen unterstützt.

#### **Einfach ist nicht gleich gut**

Nicht zuletzt für die Bildungsforschung ist die Suche nach solchen Leitlinien interessant. Speziell die *sprachliche* Gestaltung stellt dabei ein naheliegendes Thema dar. Hierbei sei die Unterscheidung einer *maximalen Vereinfachung* von einer *optimalen Passung* eines Texts für die Adressat\*innengruppe betont. Texte auf sprachlicher Seite so einfach wie möglich zu machen, genügt nicht den fachdidaktischen Ansprüchen, wenn man bedenkt, dass der Umgang mit Fachsprache einerseits anspruchsvoll ist (vergleiche Abschnitt 2.2.2) und andererseits gelehrt und gelernt werden soll (vergleiche Abschnitt 2.1.3). Darüber hinaus ist nicht gesichert, ob eine maximale Vereinfachung zielführend für einen optimalen Lernerfolg ist: Während sich bei Langer, Schulz von Thun und Tausch (2019) eine maximale sprachliche Einfachheit als optimal erwies, stellte Groeben (1978) fest, dass eine mittlere Einfachheit am zuträglichsten für den Lernerfolg ist. Darüber hinaus zeigen empirische Arbeiten zur Textverständlichkeit naturwissenschaftlicher Texte (siehe Unterkapitel 3.4), dass die Annahme, anhand sprachlicher *Vereinfachung* bessere Lernleistungen erzielen zu können, so nicht zutrifft. Mit anderen Worten ist der Gedanke, *möglichst einfach* mit *gut* und dies wiederum mit *verständlich* oder *lernförderlich* gleichzusetzen offenbar zu naiv.

Einzelheiten zu den empirischen, fach- und inhaltsübergreifenden Arbeiten zu den Auswirkungen der sprachlichen Textgestaltung auf die Verständlichkeit sind Gegenstand der Ausführungen auf den nächsten Seiten. Den fachspezifischen Ergebnissen wendet sich anschließend Unterkapitel 3.4 zu.

#### **Modelle für Textverständlichkeit**

Die verschiedenen Modelle und Hinweise zur Textverständlichkeit, die im Folgenden vorgestellt werden, folgen in etwa der folgenden Logik: Beim Verfassen eines Texts richten sich Entscheidungen hinsichtlich der Textgestaltung danach, welche Inhalte mit welchem Vermittlungszweck an welche Adressat\*innen transportiert werden sollen. Die Modelle zeigen dann „Dimensionen“ der Textverständlichkeit auf, die sich vor allem auf die sprachliche Oberfläche beziehen und hinsichtlich deren sich verschiedene Texte vergleichen und vor dem Hintergrund der Vermittlungsabsicht auch bewerten lassen.

Betrachtet man die in den letzten Jahren entwickelten und etablierten Verständlichkeitsmodelle im zeitlichen Verlauf, so lässt sich der folgende Trend erkennen (vergleiche Frickel,

2018): Von einer Verständlichkeit, die an objektiven, quantifizierbaren *Lesbarkeitskriterien* festgemacht wurde, gelangte man über Impulse aus der Kognitionspsychologie zu einer Auffassung des Leseverstehens als interaktiver Prozess (siehe Einleitung zu Kapitel 3) und damit zur Einbeziehung lesendenseitiger Einflussfaktoren auf das Leseverstehen. Neuere Ansätze münden dementsprechend in komplexere Modelle, welche zusätzlich den Einfluss kommunikativer Rahmenbedingungen (kommunikativer Zweck, Mitteilungsabsicht, Adressat\*innenorientierung) abbilden. Dieser zeitliche Verlauf wird auf den folgenden Seiten nachgezeichnet, indem zentrale Arbeiten zur Lesbarkeit, das stark textzentrierte Hamburger Verständlichkeitskonzept sowie die kognitionspsychologisch und sprachwissenschaftlich fundierte Forschung zur Textverständlichkeit von Norbert Groeben vorgestellt werden. Der Abschnitt wird in das umfassende *Karlsruher Rahmenmodell zur Textproduktion* von Susanne Göpferich münden. Es stellt die Grundlage für die Texterstellung in der empirischen Studie dar, die im Rahmen dieser Arbeit entwickelt, durchgeführt und ausgewertet wurde.

#### **Lesbarkeit**

Ein erster Ausgangspunkt für die Identifizierung verständlichkeitsrelevanter Textmerkmale ist die Lesbarkeitsforschung. Die Lesbarkeit eines Texts wird anhand vollkommen textimmanenter, objektiver Merkmale wie der Wortlänge, der Satzlänge, der Vorkommenshäufigkeit von Wörtern, der Silbenanzahl der vorkommenden Wörter etc. festgemacht. Im Lauf des letzten Jahrhunderts wurden diverse Lesbarkeitsindizes entwickelt. Hinter diesen stecken jeweils individuelle, quantifizierbare Texteigenschaften, die mittels Rechenoperationen zu einem Zahlenwert für die Lesbarkeit zusammengeführt werden. Breite Verwendung finden zum Beispiel der Flesch-Index (Flesch, 1948), die verschiedenen Formen der Wiener Sachtextformel (entwickelt von Bamberger und Vanacek, 1984, S. 83 f.), der LIX (Abkürzung für „Lesbarkeitsindex“, ursprünglich für das Schwedische entwickelt und in Björnsson, 1968, explizit für deutsche Texte diskutiert) und der gSMOG (die deutsche – „german“ – Version des SMOG-Index nach Mc Laughlin, 1969, die ebenfalls auf Bamberger und Vanacek, 1984, zurückgeht). Exemplarisch wird hier die Berechnungsformel für den gSmog angegeben (Bamberger & Vanacek, 1984, S. 58 und S. 78):

$$\text{gSmog} = \sqrt{\frac{\text{Prozentsatz der Wörter mit drei oder mehr Silben}}{\text{Zahl der Sätze in 100 Wörtern}} \times 30 - 2}$$

Der gSmog wird in der hier vorgestellten Studie zur Beurteilung der Textschwierigkeiten herangezogen (siehe dazu Unterkapitel 6.2). Lesbarkeitsindizes sind verhältnismäßig einfach und schnell zu berechnen. Da sie auf abzählbaren, rein oberflächlichen Textmerkmalen beruhen, sind sie auch maschinell ohne großen Aufwand ermittelbar<sup>6</sup>. In der Tatsache, dass

---

<sup>6</sup>Damit ist insbesondere gemeint, dass beispielshalber nicht auf moderne, textbasierte *Künstliche Intelligenzen* zurückgegriffen werden muss. Solche sind in der Tat in der Lage, die Ebene der rein quantitativen Schriftzeichen-Eigenschaften zu verlassen und auch die Bedeutungsebene zu evaluieren und in ihre Textanalysen einzubeziehen

die semantische Ebene vollkommen ausgeblendet wird, liegt aber auch die große Schwäche der Lesbarkeitsindizes. Theoretisch lassen sie sich für vollständige Unsinnstexte berechnen; weder müssen die Wörter Bezug zueinander aufweisen noch müssen sich die einzelnen Buchstaben überhaupt zu sinnvollen Wörtern zusammenfügen. Ausgehend von einem prinzipiell sinnvollen Text sind sie aber ein probates, unkompliziertes Maß für dessen sprachliche Schwierigkeit (Björnsson, 1968). Ferner sind die Kriterien für die Lesbarkeitsindizes bei der Textgestaltung unmittelbar zugänglich. Es ist vollständig transparent, welche Kriterien zu dem Zahlenwert beitragen. Dies steht im Gegensatz zu den Kriterien der folgenden Modelle, welche das verstehende Lesen mehr als Interaktionsprozess auffassen und somit von einer rein zahlenbasierten und textzentrierten Perspektive auf das Leseverstehen abweichen. Deren Verständlichkeitsdimensionen sind folglich abstrakterer Natur und entsprechende Textanpassungen bedürfen mehr sprachlichem Gefühl, Kreativität und Geschick der verfassenden Person.

#### **Orientierung an den Lesenden**

Indem die lesende Person mehr und mehr als aktive Größe in die Mechanismen des Textverstehens einbezogen wird, werden die Kriterien für verständliche Texte subjektiver. Sie beruhen nicht mehr rein auf zählbaren Eigenschaften der verwendeten Schriftzeichen, sondern sie beschreiben erwünschte Wirkungen bei den Lesenden. Die Ebene der sprachlichen Mittel, welche den Textautor\*innen zur Verfügung stehen, verschwimmt mit der Zielebene. Zum Beispiel geben Langer, Schulz von Thun und Tausch (2019, S. 32) den Hinweis, dass für „optimal verständliche Texte [...] die Einfachheit am wichtigsten [sei]“ und dass das Optimum bei maximaler Einfachheit liege. Das Merkmal der Einfachheit wiederum lasse sich unter anderem an einer „einfachen Darstellung“ festmachen und daran, ob der Text „anschaulich“ sei. Die Datengrundlage, auf der die Autoren solche Hinweise herausgearbeitet haben, bestand in einer großen Menge subjektiver Textbewertungen. Die Autoren teilen also genau genommen mit, wie ein Text auf Personen wirken soll, damit diese ihn am Ende verständlich finden. Was sie nicht – oder nur stellenweise sagen – ist, was Textautor\*innen *tun* können, damit ein Text so wirkt. Welche *sprachlichen Mittel* machen denn einen Text *anschaulich* oder sorgen für eine *einfache Darstellung*?

Das Modell von Langer, Schulz von Thun und Tausch (auch: Hamburger Modell) beruht auf individuellen Textbewertungen. Das Ziel besteht folglich in der Entwicklung solcher Texte, die Lesende als möglichst verständlich *empfinden*. Eine Überprüfung des Modells hinsichtlich der tatsächlichen Behaltensleistung bestätigte die Wirksamkeit der vier Verständlichkeitsdimensionen auch für kognitive Erfolgskriterien weitestgehend (für Einzelheiten siehe Langer, Schulz von Thun und Tausch, 2019, S. 195 ff.). Die Verständlichkeit wird dabei an vier Dimensionen festgemacht, welche mithilfe faktoranalytischer Methoden<sup>7</sup> aus den Bewertungsdaten herausgearbeitet wurden. Die vier Dimensionen lauten *Einfachheit*, *Gliederung/Ordnung*,

---

<sup>7</sup>für einen Einblick siehe z. B. Moosbrugger und Kelava, 2012, Kap. 13; vergleiche auch die Erläuterungen zur Hauptkomponentenanalyse im Abschnitt 6.4.3

### 3. Wissensvermittlung anhand von Texten

*Kürze/Prägnanz* und *Anregende Zusätze* und werden jeweils durch untergeordnete Merkmale konkretisiert (Langer, Schulz von Thun & Tausch, 2019, S.22 ff.). Die Einfachheit zum Beispiel wird – wie oben erwähnt – unter anderem von der *Anschaulichkeit* und der *Einfachheit der Darstellung* bestimmt; zu Gliederung/Ordnung tragen eine *gute Unterscheidung von Wesentlichem und Unwesentlichem* und die *Übersichtlichkeit* bei. Zur Beurteilung der Textverständlichkeit wird für jede Dimension entschieden, wie stark sie ausgeprägt ist. Die Autoren sehen dafür eine fünfstufige Skala vor und geben das folgende Beurteilungsfenster als Optimum an (Langer, Schulz von Thun & Tausch, 2019, S. 33).

Einfachheit ++	Gliederung/Ordnung ++
Kürze/Prägnanz 0 oder +	Anregende Zusätze 0 oder +

Das Modell erscheint zusammenfassend als bündig und gut handhabbar – auch für Personen ohne fundierten sprachwissenschaftlichen Hintergrund. Es setzt jedoch voraus, dass Textautor\*innen die Eindrücke, die ein Text aufgrund seiner (sprachlichen) Gestaltung weckt, zutreffend einschätzen können. Dazu die Autoren des Modells: „Wir glauben nicht, dass das Zählen irgendwelcher Eigenschaften von Wörtern oder Sätzen eine günstige Methode ist. Wir plädieren für die Schätzmethode, für das Beurteilen von Eindrücken“ (Langer, Schulz von Thun & Tausch, 2019, S. 187). Dieser Ansatz ist prinzipiell vielversprechend, zumal sich die Güte eines Texts immer erst vor dem Hintergrund seiner Rahmendaten wie dem Zweck und eventueller redaktioneller Bestimmungen ergibt (vergleiche dazu insbesondere das auf Seite 55 beschriebene Karlsruher Modell von Göpferich). Auch Frickel kommt im Zuge ihrer Arbeit zur Verständlichkeit literarischer Texte zu dem Schluss, Textschwierigkeit nur relativ zu Lesenden und institutionellem Kontext feststellen zu können. Denn Textmerkmale – so Frickel – wirken ebenso wie die Dispositionen und die Rezeptionssituation der Lesenden auf die Rekonstruktion von Inhalt und Bedeutung eines Texts ein. Autor\*innen für Eindrücke zu sensibilisieren, welche die Verständlichkeit bestimmen, stellt sich als erfolgversprechender Ansatz dar und in den Studien der Hamburger Forschungsgruppe bewährt er sich. Allerdings ist der Anspruch an Textautor\*innen, die richtigen sprachlichen Mittel zu finden und zu wählen, um die erwünschten Eindrücke bei den Lesenden zu erzielen, vor allem bei Fachtexten kein geringer. Die Frage, welche konkreten Mittel die Verstehensleistung bei Schüler\*innen beim Lesen naturwissenschaftlicher Texte tatsächlich beeinflussen, konnte in der Tat noch nicht geklärt werden (vergleiche dazu z. B. Hackemann, Heine & Höttecke, 2022).

#### **Ein theoriegeleitetes Modell**

Darüber hinaus sind die Befunde aus dem Hamburger Modell nicht gänzlich im Einklang mit vergleichbaren Studien. Groeben etwa untersucht ebenfalls die Verständlichkeit von Texten. Er verfolgt dabei eine theoretische Herangehensweise, welche Erkenntnisse der

Kognitions- und Motivationsforschung mit Textcharakteristika in Verbindung bringt. Während das Zielkriterium der Hamburger Forschungsgruppe zunächst eine optimale subjektive Verständlichkeits*beurteilung* war, bemüht sich Groeben in raffinierter, aber auch sehr technischer, aufwendiger Weise, die beiden Gütekriterien *Interesse* und *Behaltensleistung* bei den Lesenden zu operationalisieren und messbar zu machen. Gemäß Groebens theoriebasiertem Ansatz lauten die Verständlichkeitsdimensionen *Stilistik*, *kognitive Strukturierung* und *konzeptueller Konflikt*. Die Stilistik lässt sich dabei unterteilen in die sprachliche Komponente der stilistischen Einfachheit und die bedeutungsbezogene Komponente der semantischen Redundanz. Die kognitive Strukturierung bezieht sich auf die inhaltliche Organisation von Teilinformationen im Text und deren Bezügen zueinander. Zum konzeptuellen Konflikt tragen solche Textelemente bei, die Neugier und Interesse evozieren. In der Empirie erweist sich die inhaltliche Struktur (bestimmt durch kognitive Strukturierung und konzeptuellen Konflikt) als in größtem Maß bestimmend für den Lernerfolg, während semantische und stilistische Einflüsse vernachlässigbar sind. Der optimale Lernerfolg sei mit Texten mittlerer Verständlichkeit zu erreichen, welche sich durch eine Mischform aus kognitiv-strukturierter und konzeptuell-konfliktbehafteter Textgestaltung auszeichnet (Groeben, 1978, S. 113). Vergleicht man nun dieses Modell von Norbert Groeben mit dem Hamburger Modell, lässt sich zunächst eine Ähnlichkeit in den gefundenen Dimensionen feststellen: Die „stilistische Einfachheit“ erinnert an die „Einfachheit“ im Hamburger Modell, die „semantische Redundanz“ an die „Kürze und Prägnanz“, die „kognitive Strukturierung“ an die „Gliederung und Ordnung“ und der „konzeptuelle Konflikt“ an die „Anregenden Zusätze“. Allerdings legt Tergan (1980) ausführlich dar, dass diese Ähnlichkeit ausschließlich oberflächlich besteht und sich bei näherem Hinsehen entscheidende Unterschiede unter anderem hinsichtlich der Gewichtung der Dimensionen für die resultierende Verständlichkeit, hinsichtlich der konkreten Umsetzung der einzelnen Dimensionen in Texten bei der Textproduktion sowie hinsichtlich der Vorhersage des Lernerfolgs bei der Textbewertung auf tun.

#### **Der erweiterte Blick auf die Textproduktion**

Das folgende Rahmenmodell versucht die Verständlichkeit von Texten noch breiter zu erfassen, indem es den gesamten Textproduktionsprozess bedenkt. Das Karlsruher Verständlichkeitsmodell von Göpferich kann hinsichtlich seiner Verständlichkeitsdimensionen als *Erweiterung* sowie *Präzisierung* des Hamburger Modells verstanden werden (Göpferich, 2002, S. 163) und orientiert sich an der praktischen Umsetzung einer schriftlichen Kommunikation aus Sicht eines Autors\* einer Autorin. Es bietet einen umfassenden, gut begründeten und ausführlich erläuterten Rahmen für die Beurteilung und vor allem für die Produktion verständlicher Texte. Die Erstellung der Texte für diese Studie richtete sich daher in weiten Teilen nach Göpferichs Rahmenmodell (siehe dazu vor allem Unterkapitel 6.2). Aufbauend darauf wurden dann sorgsam ausgewählte sprachliche Merkmale variiert und die Auswirkungen untersucht.

Die zentralen Bausteine des Modells werden in den folgenden Zeilen sowie in Abbildung 3.3 zusammengefasst, für Einzelheiten sei auf Göpferich (2002, insbesondere S. 155 ff.) verwiesen.

### 3. Wissensvermittlung anhand von Texten

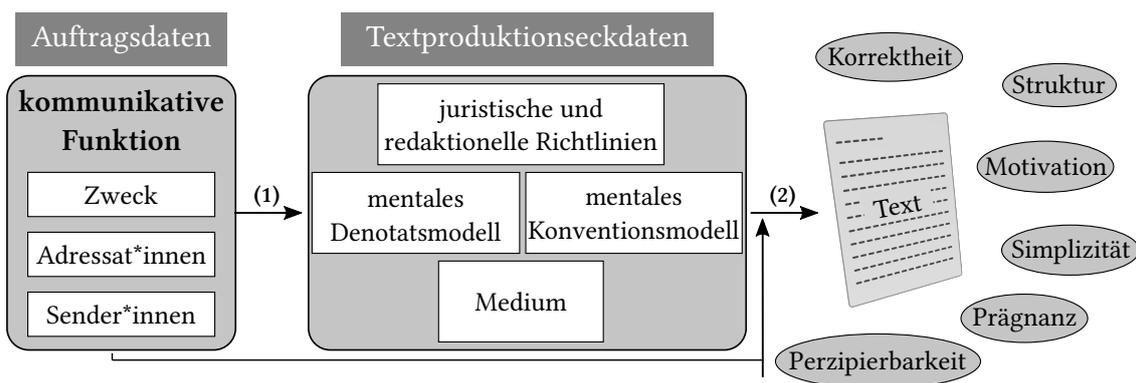


Abbildung 3.3.: Das Karlsruher Modell zu den Rahmenbedingungen und Verständlichkeitsdimensionen bei der Textproduktion und -bewertung. Ausgehend von Auftragsdaten und Textproduktionseckdaten wird ein Text (1) konzipiert und (2) exteriorisiert. Die Verständlichkeit wird an den Dimensionen Korrektheit, Struktur, Motivation, Simplizität, Prägnanz und Perzipierbarkeit festgemacht. (Eigene, leicht verkürzte Darstellung des Originals in Göpferich, 2002, S. 155.)

Indem das Modell Aspekte der kommunikativen Situation und des Textkonzeptions- und -produktionsprozesses berücksichtigt, geht es über reine Textmerkmale hinaus. Es enthält solche Rahmenbedingungen, die zunächst die Konzeption des Texts beeinflussen und als *Auftragsdaten* bezeichnet werden. Die Auftragsdaten ergeben sich gemäß dem Modell aus dem Zweck der Kommunikation, der Mitteilungsabsicht der verfassenden Person und der Orientierung an den Adressat\*innen. Zusammen mit den Auftragsdaten sind redaktionelle sowie konventionsgebundene Richtlinien, das Kommunikationsmedium und die inhaltlichen Bestandteile des Texts maßgebliche Eckdaten beim Verfassen und beim Beurteilen eines Texts. Bezüglich der inhaltlichen Bestandteile benutzt Göpferich den Ausdruck „mentales Denotatsmodell“, das sich als geistiges Abbild der zu vermittelnden Textinhalte (Gegenstände, Sachverhalte und deren Relationen) verstehen lässt und welches bei der Textproduktion in Zeichen umgesetzt wird (Göpferich, 2002, S. 159). Die Textoberfläche an sich wird gemäß dem Modell anhand von sechs Dimensionen charakterisiert. Aus deren jeweiligen Beschreibungen in Göpferich (2002, S. 163 ff.) geht klar hervor, dass sich einzelne Verständlichkeitsdimensionen sowohl aus inhaltlichen wie auch aus sprachlich-oberflächlichen Kriterien zusammensetzen, andere Dimensionen rein auf die Sprachoberfläche („Kodierung“) und wieder andere rein auf den Inhalt bezogen sind. Die erste Dimension ist die *Korrektheit*. Sie umfasst explizit inhaltliche (nötiges Fachwissen der verfassenden Person) sowie auf die Kodierung bezogene Aspekte (Verfälschung des „Gemeinten“ bei der Umsetzung in Textzeichen)<sup>8</sup>. Als zweite Dimension nennt Göpferich die *Motivation*. Sie ist das Vermögen, Interesse bei den Lesenden zu wecken und aufrecht zu erhalten, welches sich einerseits inhaltlich aus dem Denotatsmo-

<sup>8</sup>Diese Unterscheidung ist insofern bemerkenswert, als dass eine Studie von Helzel und Rabe (2023) zeigt, dass für Studierende, die über Sprache reflektieren, die sprachliche und die fachliche Korrektheit nahezu identisch sind. Sie bewerten die Sprache in ähnlicher Weise binär als korrekt oder inkorrekt wie den Inhalt einer Sprachhandlung im Unterricht und verkennen sprachliche Gestaltungsspielräume (vergleiche kontrastierend Abschnitt 2.1.2).

dell (interessante inhaltliche Textbausteine) und andererseits sprachoberflächlich durch die Kodierung (z. B. ansprechender Sprachstil) ergibt. Die Motivation ähnelt den „Anregenden Zusätze“ im Hamburger Modell. Drittens setzt sich auch die *Prägnanz* (angelehnt an die „Kürze und Prägnanz“ im Hamburger Modell) aus inhaltlichen sowie kodierungsbezogenen Aspekten zusammen. Die Detailliertheit des Denotatsmodells sowie der Zeichenaufwand für die vermittelten Inhalte bestimmen diese Dimension. Die vierte Verständlichkeitsdimension ist rein auf die Kodierung bezogen: die *Perzipierbarkeit* bezeichnet das optische Erscheinungsbild des Texts. In der *Simplizität*, mit der die Einfachheit der Wortwahl und des Satzbaus gemeint ist, besteht die fünfte Dimension des Modells. Sie ist angelehnt an die „Einfachheit“ des Hamburger Modells, ergibt sich jedoch bei Göpferich ausschließlich aus sprachoberflächlichen Merkmalen. Schließlich macht die *Struktur* die sechste Dimension aus. Rein auf den Inhalt bezogen sind damit die inhaltlichen Teilaspekte und deren Beziehungen sowie logische Bezüge zwischen Satzteilen und benachbarten Sätzen gemeint. Die inhaltliche Struktur ist angelehnt an die *Gliederung und Ordnung* des Hamburger Modells. Für jede Dimension führt Göpferich konkrete Positiv- und oder Negativbeispiele an und zählt sprachliche Mittel auf, anhand derer die jeweilige Dimension zu beeinflussen sei.

Der Weg führte also von quantitativen, rein textbasierten Indizes hin zu immer umfassenderen Verständlichkeitskonzepten, welche kognitionspsychologische Aspekte sowie die aktive Rolle der lesenden Person mit berücksichtigen. Dies bringt den Nachteil mit sich, dass die Ebene der Textgestaltung mit der Ebene der Textwirkung verschwimmt, da sowohl Textmerkmale als auch Wirkungen auf Lesende als Einflussgrößen auf das Endergebnis *Textverständlichkeit* angesehen werden. An dieser Stelle sei zur Verdeutlichung noch einmal auf das Beispiel der Forderungen nach einer „einfachen Darstellung“ und nach „Anschaulichkeit“ auf Seite 53 verwiesen: Die Forderungen werden zwar an Autor\*innen gestellt, beziehen sich aber auf eine Wirkung auf Lesende. In gleicher Weise beschreibt die Forderung nach „Attraktivität“ eines Lern- beziehungsweise Kommunikationsangebots (siehe Seite 26) eine solche erstrebenswerte *Wirkung* und lässt die Frage nach dafür geeigneten, konkreten Gestaltungsmerkmalen offen. Diese Studie setzt sich zum Ziel, die Ebenen der wünschenswerten *Effekte* in der Text-Lesenden-Interaktion auf der einen Seite und der verfügbaren *sprachlichen Mittel* auf der anderen Seite konsequent voneinander zu trennen. Gemäß dieser Trennung werden sprachliche Gestaltungsmittel einerseits und kognitive und affektive Aspekte der Textverständlichkeit andererseits verortet und somit aus der bestehenden Theorie der Konstruktraum für die Untersuchungsvariablen einer empirischen Studie rekonstruiert (siehe Unterkapitel 5). Besonders die Mittel der Textgestaltung innerhalb dieses Konstruktraums sind dabei neben allgemeinen auch von fachspezifischen Einflüssen geprägt. Diese Einflüsse beziehen sich sowohl auf inhaltliche wie auch auf sprachoberflächliche Charakteristika des spezifischen Fachs. Der Erkenntnisstand zur Textverständlichkeit speziell in den Naturwissenschaften wird daher im folgenden Kapitel aufgezeigt.

### 3.4 Stand der fachspezifischen Leseforschung

Die bis hierhin vorgestellten Kenntnisse zur Wissensvermittlung mit Texten sind allgemein gehalten und glätten über konkrete Inhalte sowie fachspezifische Sprachoberflächenmerkmale hinweg. Der fachdidaktische Blickwinkel bedeutet nun, hier anzuknüpfen und ebensolchen fachlich bedingten, speziellen Ansprüchen Beachtung zu schenken. In den naturwissenschaftlichen Fächern scheinen die Lernbedingungen in Bezug auf den Umgang mit schriftlichem Text besonders verbesserungswürdig zu sein. Als Indiz dafür kann gewertet werden, dass Lernende im Physikunterricht das Schulbuch nur selten zu nutzen scheinen (Härtig, Bernholt, Prechtl & Retelsdorf, 2015; Staraschek, 2003). Eine mögliche Ursache dafür liegt in einer unpassenden Gestaltung der Texte für die Schüler\*innen und zwar derart, dass die Schüler\*innen überfordert sind und/oder von den Texten zu wenig angesprochen werden (vergleiche entsprechende Teilergebnisse einer Lehrkräftebefragung in Merzyn, 1994, S. 107 ff). Betrachtet man genauer, wie naturwissenschaftliche Schulbuchtexte gestaltet sind, fällt auf, dass die Texte Züge naturwissenschaftlicher Fachsprache aufweisen (Härtig, Fraser, Bernholt & Retelsdorf, 2019; Scheller, 2010, S. 65 f.). Zusammenfassend bestehen deren Charakteristika in ungewohntem, fremdsprachlichem Vokabular mit komplexer Morphologie, das mitunter selbst zum Lerngegenstand wird (siehe dazu Seite 27 in Abschnitt 2.1.3); in komplexen und ungeläufigen Satzstrukturen (die ebenfalls in Form zu erlernender Sprechweisen zum eigenständigen Lerngegenstand werden können); und in einem grundlegenden Aufbau, der einer inhaltlich-logischen Gliederung folgt anstelle eine aus alltäglichen Erzählsituationen gewohnte, chronologische Reihenfolge aufzuweisen (siehe Abschnitt 2.2.2 für detaillierte Ausführungen zu fachsprachlichen Auffälligkeiten). Es besteht allgemeiner Konsens darüber, dass diese Charakteristika besonders herausfordernd für Lernende sind (z. B. Strohmaier, Ehmke, Härtig & Leiss, 2023; Fang, 2006; Schmellentin, Dittmar, Gilg & Schneider, 2017; Scheller, 2010). Fang (2006, S. 505 f.) zieht in Bezug auf sprachliche Besonderheiten, die er aus amerikanischen *science textbooks* zusammenträgt, das Fazit:

Together, these features render science texts more turgid, dense, abstract, and distilled than the materials that children have been used to reading during the early years of schooling. Often taken for granted by proficient adult readers, these features can and do present significant comprehension challenges to adolescent students [...]. Knowing what these features are and understanding how they work to construct scientific knowledge, principles, and ideology is an essential part of science literacy.

Um nun Lernende beim Umgang mit schriftlichem, naturwissenschaftlichem Textmaterial zu unterstützen, bieten sich Maßnahmen für den Unterricht an. Holliday und Cain (2012) stellen einen forschungsbasierten Leitfaden für Lehrkräfte zusammen, anhand dessen ein effektiver Unterricht in *science reading comprehension* für Schüler\*innen in naturwissenschaftlichen Fächern gestaltet werden könne. Die Autoren geben Vorschläge für die Wahl und die Vermittlung von Lesestrategien, gehen insbesondere auf die Effektivität des Modelllernens ein

und unterstreichen, dass es besonders wichtig sei, die Schüler\*innen zu motivieren und dabei zu unterstützen, ihr Vorwissen zu aktivieren und nutzen. Evidenz zur Effektivität solcher Maßnahmen liefert eine Studie von Fang et al. (2008). Beim Vergleich einer Treatment-Gruppe mit „reading strategy lessons“ und „home reading programme“ über den Verlauf von sechs Monaten mit einer Kontrollgruppe, die keine Leseförderung erhielt, finden Fang et al. einen signifikanten, positiven Treatment-Effekt auf die Science Literacy (aufgefasst als Komposition aus Lesefähigkeiten und naturwissenschaftlich-inhaltlichem Wissen; Stichprobe: zehn sechste Klassen in den USA).

Besonders interessant sind neben den quantitativen Ergebnissen auch die Ergebnisse auf der Meta-Ebene, die Fang et al. (2008) berichten. Unter anderem sei es für eine erfolgreiche Integration der Leseförderung in den Unterricht wichtig, dass hochwertige, ansprechende Texte zur Verfügung stehen – das Schulbuch würde diesen Ansprüchen in den meisten Fällen nicht genügen. Offensichtlich lohnt auch der Blick auf Möglichkeiten einer lernwirksameren Textgestaltung, um Schüler\*innen beim Verstehen naturwissenschaftlicher Texte zu unterstützen. Vor dem Hintergrund der oben erwähnten, für naturwissenschaftliche Texte charakteristischen Anforderungen erscheint es plausibel, dass evidenzbasierte Erkenntnisse für den speziellen Kontext des naturwissenschaftlichen Fachunterrichts nötig sind. Diesem Erfordernis haben sich in der Vergangenheit zahlreiche Studien zugewandt. Strohmaier, Ehmke, Härtig und Leiss (2023) nutzen die Ergebnisse aus 45 empirischen Studien zur Textverständlichkeit in *STEM*-Fächern (Science, Technology, Engineering and Mathematics) für eine Meta-Analyse und ziehen ein Fazit hinsichtlich der Effekte sprachlicher Modifikationen auf das Textverständnis und das Lernen. Indem die Autoren *Lernen* als Integration von aus dem Text verstandener Information (Textverständnis) in die eigenen Wissensbestände auffassen (Strohmaier, Ehmke, Härtig & Leiss, 2023, S. 2), besteht also die abhängige Variable in einem rein kognitiven Gelingenskriterium verstehenden Lesens. Darüber hinaus wird in der Meta-Analyse untersucht, welche Moderatorvariablen einen signifikanten Einfluss darauf zeigen, wie sich sprachliche Modifikationen auf Textverständnis und Lernen auswirken. Die sprachlichen Modifikationen, die in den in die Analyse eingeschlossenen Einzelstudien vorgenommen wurden, gliedern die Autoren in sechs Kategorien, nämlich „linguistic complexity“, „increasing cohesion“<sup>9</sup> und „increasing clarity and elaboration“ – jeweils differenziert in eine lokale Ebene (Wort- und Satzebene) und einer globale Ebene (Textebene). In aller Kürze lassen sich die Beobachtungen von Strohmaier, Ehmke, Härtig und Leiss (2023) so zusammenfassen, dass *in Einzelstudien* lediglich für eine einzige Kategorie einheitliche Befunde zu finden seien. Für eine Steigerung der Elaboration auf lokaler Ebene – beispielshalber anhand von Worterklärungen oder anhand zusätzlicher, erklärender Einschübe im Satz – wurden in den in die Meta-Analyse einbezogenen Einzelstudien positive, signifikante Effekte auf Verständnis

---

<sup>9</sup>Zur Steigerung der Kohäsion dienen sogenannte Kohäsionsmittel. Meist werden darunter solche sprachlichen Mittel bezeichnet, die auf lokaler Ebene wirksam sind, also die Zusammenhänge innerhalb eines Satzes oder zwischen benachbarten Sätzen zum Ausdruck bringen (z. B. Pronomen, wiederholtes Substantive, logische Konnektiva). In diesem Fall ist auch die globale Kohäsion mit gemeint, die sich auf inhaltliche Zusammenhänge zwischen Textabschnitten bezieht.

und Lernen berichtet (Datengrundlage: drei Studien, die diesen Effekt isoliert betrachtet haben). Hinsichtlich der übrigen fünf Kategorien seien keine einheitlichen beziehungsweise signifikanten Effekte in Einzelstudien zu verzeichnen oder die Effekte seien in Einzelstudien nicht isoliert betrachtet worden. Ein besonderes Augenmerk richteten Strohmaier, Ehmke, Härtig und Leiss auf die Personalisierung und fassten sie als Unterkategorie der Elaboration auf globaler Ebene auf. Einzelstudien würden diesbezüglich ebenfalls tendenziell positive Effekte messen (abhängig von der gemessenen abhängigen Variable und vom Vorwissen). Hinsichtlich der *meta-analytischen Ergebnisse*, welche die unterschiedlichen Einzelstudienresultate zusammenbringen, berichten Strohmaier, Ehmke, Härtig und Leiss (2023) insgesamt einen kleinen, signifikanten Effekt sprachlicher Modifikationen auf das Textverständnis und das Lernen. Dabei würden ausschließlich zwei Kategorien, nämlich die Steigerung der Elaboration auf lokaler Ebene sowie die Steigerung der Elaboration auf globaler Ebene (inklusive Personalisierung) einen signifikanten, positiven Effekt zeigen. Im Allgemeinen würden die Effekte durch das Vorwissen und die Textlänge signifikant moderiert, sodass Lesende mit geringerem Vorwissen stärker von den Modifikationen profitieren und die Effekte mit steigender Textlänge größer werden würden. Sprachliches Vorwissen hingegen habe sich nicht als signifikante Moderatorvariable erwiesen. Ebenso hätten keine unterschiedlichen Effektstärken je nach Level des Textverständnisses (Reproduktion, Anwendung, Transfer) nachgewiesen werden können. Abschließend sei hinsichtlich der Meta-Analyse-Ergebnisse erwähnt, dass die Autoren im Vergleich der Modifikationskategorien für die Unterkategorie *Personalisierungen* die stärksten Effekte auf das Textverständnis und das Lernen finden und dass die Meta-Analyse hinsichtlich *keiner einzigen* der sechs untersuchten Kategorien einen negativen Effekt auf Textverständnis und Lernen verzeichnet.

Einige der Forschungsfragen in der eben beschriebenen Meta-Analyse weisen in eine ähnliche Richtung wie die Leitgedanken, die hinter der Konzeptualisierung der dieser Dissertation zugrundeliegenden Studie stehen. Die Leitgedanken, die sich aus bestehenden Desideraten der fachspezifischen Textverständlichkeitsforschung ergeben, werden ausführlich in Unterkapitel 4.1 auf Seite 66 beschrieben. Ein solches Desiderat liegt in der systematischen Untersuchung der Auswirkungen einzelner, bestimmter sprachlicher Gestaltungsmittel auf das verstehende Lesen beziehungsweise die Textverständlichkeit. Dazu ist es nötig, die *Gestaltungsebene* bei der Produktion eines Texts klar von der *Wirkungsebene* zu trennen und dann ganz bestimmte sprachliche Gestaltungsmittel auszuwählen und deren Effekte isoliert zu beobachten. In vielen Studien wurden ganze Sets an potenziell verständlichkeitsrelevanten Merkmalen zusammengenommen, denen theoriebasiert eine bestimmte Wirkung zuzuschreiben ist (z. B. Mittel zur Modifikation der Kohärenz des Texts bei Härtig, Fraser, Bernholt und Retelsdorf, 2019, oder Mittel zur Modifikation der Komplexität des Texts bei Plath, 2020). Diese Merkmale wurden dann zugleich variiert, sodass ein Rückschluss auf Effekte einzelner Modifikationsmaßnahmen am Ende nicht mehr möglich war. Eine solche Vorgehensweise führte z. B. bei Hackemann, Heine und Höttecke (2022) und Härtig, Fraser, Bernholt und Retelsdorf (2019) zu dem erwartungswidrigen Befund, dass die Textmodifikationen nahezu

keinen Einfluss auf das Textverständnis zeigten. Vermeintlich vereinfachte Texte führten nicht zu einem besserem Verständnis.

Eine andere Vorgehensweise wählte Starauschek (2006), indem er zwei Textversionen verglich, die sich ausschließlich hinsichtlich eines ganz bestimmten lokalen Kohäsionsmittels (der Begriff wird in der Fußnote auf Seite 59 erläutert), nämlich der substantivischen Textkohäsion, unterschieden. Starauschek untersuchte, ob sich je nach Ausprägung dieses Kohäsionsmittels eine unterschiedliche Einschätzung der Verständlichkeit bei Schüler\*innen ergab, fand jedoch keine Unterschiede. Der Messung einer solchen affektiven Komponente der Texttension wurde bisweilen in der fachspezifischen Verständlichkeitsforschung wenig Beachtung geschenkt, sodass hier ein zweites Desiderat liegt. Lediglich in der eben genannten Studie von Hackemann, Heine und Höttecke (2022) wurde neben dem Verständnis, welches ein *kognitives* Gelingenskriterium des verstehenden Lesens bezeichnet, auch die „perceived comprehensibility“, also eine affektive Verständlichkeitseinschätzung mit betrachtet. In dieser Studie wurden inhaltsgleiche Texte anhand der Dimensionen strukturelle Komplexität, Transparenz und Häufigkeit (siehe dazu Heine et al., 2018) in drei verschiedenen Schwierigkeitslevels erstellt. Es zeigte sich einerseits, dass sich für verschiedene Textversionen unterschiedliche Verständlichkeitseinschätzungen durch Lesende ergeben. Andererseits wurden keine signifikanten Unterschiede im tatsächlich gezeigten Verständnis je nach Textversion gemessen. Daraus folgern die Autoren, dass von keinem Zusammenhang zwischen wahrgenommener Verständlichkeit und tatsächlich gezeigtem Verständnis auszugehen sei<sup>10</sup>. Eine weitere Studie, die sich für die Identifikation schwierigkeitsgenerierender Textmerkmale unter anderem auf von den Lesenden *berichtete* (und damit insbesondere bemerkte, wahrgenommene) Verstehensschwierigkeiten stützt, führten Schmellentin, Dittmar, Gilg und Schneider (2017) durch. Auf Basis von Lesendenbeobachtungen und -befragungen erwiesen sich die folgenden Merkmale als problematisch: hohe Fachworddichte, karg erklärte Fachwörter, hohe Informationsdichte, komplexe/unerwartete Satzstrukturen, uneindeutige Kohäsionsmittel, Synonyme, Brüche in der Inhaltsorganisation, untypische Satzstrukturen, uneindeutige Verweise, Probleme mit Abbildungseigenschaften, Bildbezügen und Layout. Bewusst wurde den Lesenden vor allem ein mangelndes Wort- und insbesondere Fachwortverstehen.

Ein letzter Aspekt, der in der bisherigen Textverständlichkeitsforschung nicht systematisch untersucht wurde, ist, inwiefern die Auswirkungen der Textgestaltung auf das Textverständnis oder die wahrgenommene Verständlichkeit durch den Inhalt des Texts moderiert wird. Es sei hier auf einen weiteren Befund der Meta-Analyse von Strohmaier, Ehmke, Härtig und Leiss (2023) hingewiesen, demgemäß sprachliche Modifikationen von Physiktexten im Vergleich aller STEM-Domänen die stärksten Effekte auf das Verständnis und das Lernen zeigten. Die

---

<sup>10</sup>Es geht aus Hackemann, Heine und Höttecke (2022) nicht hervor, ob bei einer anderen Herangehensweise, in der die sprachliche Gestaltung eines Texts außer Acht gelassen wird und stattdessen nur beobachtet wird, ob sich rein auf Grundlage der Verständlichkeitseinschätzung ein bestimmtes Verständnis ergibt, womöglich ein Zusammenhang ergeben hätte. Eine solche Vorgehensweise, welche die affektive Textwahrnehmung für sich genommen als mögliche Ausgangsbedingung für Verständnis betrachtet, wird in der dieser Dissertation zugrundeliegenden Studie verfolgt.

### 3. Wissensvermittlung anhand von Texten

Autoren mutmaßen, dass dies an einem generell höheren inhaltlichen Anspruch physikalischer Themen liegen könnte, sodass das Verbesserungspotenzial der sprachlichen Mittel für die Verständlichkeit steige (ähnlich wie sprachliche Mittel sich bei geringerem Vorwissen als effektiver erwiesen). Ein verwandter Erklärungsansatz ist, dass die Physik besonders abstrakte Themen birgt; dass sie also Begriffe und Konzepte zum Gegenstand hat, die nicht direkt sichtbar und erfahrbar sind, sondern gedanklicher, modellhafter Gestalt sind (siehe dazu auch Seite 20 in Abschnitt 2.1.1). Vor dem Hintergrund, dass das Erlernen solcher abstrakter Begriffe anhand von Sprache geschieht und maßgeblich dadurch beeinflusst ist, wie über die Konzepte gesprochen wird (siehe Abschnitt 2.1.1), erscheint es plausibel, dass die sprachliche Umgebung, in der solche abstrakten Gegenstände vermittelt werden, besonders verständlichkeitsrelevant ist. In Einklang mit dieser Idee besteht das Textthema in der Studie, über die im Folgenden ausführlich berichtet wird, in der Einführung des abstrakten Begriffs der *elektrischen Spannung*.

*Wir sind gleichsam Zwerge, die auf den Schultern von Riesen sitzen, um mehr und Entfernteres als diese sehen zu können.*

---

Bernhard von Chartres

## 4 | FORSCHUNGSINTERESSE

Dieses Kapitel wird nun die Brücke vom Bekannten zum Neuen schlagen. Es wird dargestellt, wie das Forschungsinteresse dieser Arbeit aus dem Forschungsstand zur Wissensvermittlung anhand von Texten und aus dem theoretischen Hintergrund zum Zusammenspiel von Physik und Sprache hervorgeht. Ein treffendes Bild für diese Situation malt der Philosoph Bernhard von Chartres mit seinem oft zitierten Gleichnis zum wissenschaftlichen Fortschritt: Die Schultern von Riesen stehen für das bereits bestehende Wissen, welches umfassend, reichhaltig und viel größer ist, als die eigenen, kleinen Erkenntnisschritte. Indem sich aber die neuen Erkundungen auf das Bekannte stützen, daran anknüpfen und es erweitern, wird es möglich – so, als ob Zwerge auf diesen Schultern säßen – mit Weitblick in wissenschaftliches Neuland zu sehen und dieses (weiter) zu erkunden.

### 4.1 Zentrale Begriffe und Leitgedanken der Studie zur Textverständlichkeit

Bezogen auf diese Arbeit bedeutet das Anknüpfen an Bekanntes, sich die allgemeinen, fach- und inhaltsübergreifenden, oft normativ gehaltenen Vorschläge zur Textoptimierung (Kapitel 3) zunutze zu machen und sie in einem fachspezifischen Kontext um evidenzbasierte Erkenntnisse zu erweitern. Mit Blick auf die besonderen Ausgangsbedingungen und Herausforderungen, welche das Lernen mit Texten in den Naturwissenschaften mit sich bringt (Kapitel 2), erscheinen Untersuchungen zur Wirkung sprachlich modifizierter Texte mit physikalischem Inhalt besonders interessant für den Unterrichtserfolg. Der Umgang mit Fachtexten ist einerseits als zentraler Unterrichtsbaustein zu betrachten, da Lesen und Schreiben ebenso zur Förderung des naturwissenschaftlichen Lernens beitragen, wie umgekehrt (Pearson, Moje & Greenleaf, 2010). Andererseits zählt der Kompetenzaufbau zum rezeptiven wie produktiven Umgang mit naturwissenschaftlichen Texten gemäß den Richtlinien der Kultusministerkonferenz und des Lehrplans zu den erklärten Unterrichtszielen (siehe Kultusministerkonferenz, 2020; sowie für das Bundesland Bayern ISB, 2022b). Nicht nur für den Unterricht ist es wichtig, dass Texte dahingehend günstig *gestaltet* sind (*Handwerkszeug* für Autor\*innen), dass sie

#### 4. Forschungsinteresse

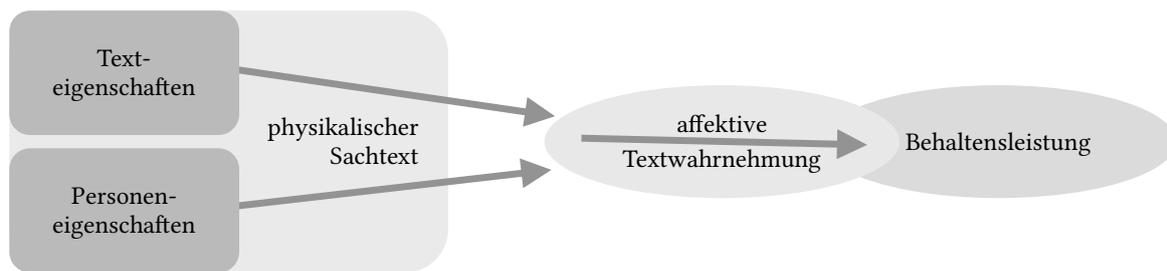


Abbildung 4.1.: In der Abbildung ist die Grundidee des Forschungsinteresses veranschaulicht: Die Ausgangsbedingungen, die ein Text und eine lesende Person mitbringen, führen vor dem Hintergrund eines physikalischen Sachtexts und damit verbundenen sprachlichen Besonderheiten und Erwartungen an den Text zu einer bestimmten affektiven Textwahrnehmung bei lesenden Personen. Diese Wahrnehmung wiederum wirkt – so die Annahme – als mediatorische Einflussgröße auf die Behaltensleistung beim Lesen ein. In dieser Arbeit steht dabei der Text als veränderliche Größe im Fokus. Die Ausgangsbedingungen der Personen werden als unveränderliche Moderatorvariablen betrachtet.

erstens einen Begriff über abstrakte, physikalische Phänomene vermitteln können (bewusster, fachlich korrekter Einsatz von *Sprache im Fach*), dass sie zweitens verständlich auf Schüler\*innen wirken und zu einer aktiven Auseinandersetzung anregen (*affektive* Gesichtspunkte der Textwirkung) und dass sie drittens Verstehensprozesse auch tatsächlich in Gang setzen (*kognitive* Gesichtspunkte der Textwirkung). Die Abbildung 4.1 illustriert graphisch diese Grundidee, auf der das Forschungsinteresse beruht: Im Kontext sprachlicher wie inhaltlicher Charakteristika, die ein physikalischer Text mitbringt, wirken Personen- und Texteigenschaften zusammen auf die affektive Textwahrnehmung der Lesenden ein. Die Wahrnehmung wiederum steht als mediatorische Instanz zwischen den personen- und textbezogenen Ausgangsbedingungen und der Behaltensleistung beim Lesen. Bevor das Forschungsinteresse im Einzelnen näher beschrieben wird, werden in den folgenden Zeilen zunächst die nötigen Begriffe geklärt.

Das Forschungsinteresse dieser Arbeit ist bezogen auf Variationen der *sprachlichen Gestaltung der Texte*. Genauer bilden die drei Gruppen (1.) Wortwahl / morphologische Auffälligkeiten auf Wortebene, (2.) Satzstrukturen / morphologische Auffälligkeiten auf Satzebene und (3.) erzählende/personalisierende Elemente die unabhängige Variable in den Zusammenhangsanalysen zwischen sprachlicher Gestaltung und Textwirkungen. Eine detaillierte Beschreibung und Veranschaulichung dieser und weiterer Gruppen sprachlicher Mittel bietet Unterkapitel 5.1. Wie sie in den Textversionen dieser Studie umgesetzt sind, beschreibt Abschnitt 6.2.2.

Hinsichtlich der abhängigen Variablen liegt unter anderem die *Behaltensleistung beim Lesen* im Forschungsinteresse. Die Behaltensleistung ist aufzufassen als kurzfristig messbarer Lernzuwachs – hier zum Textthema *elektrische Spannung*. Sie wird entsprechend als Zugewinn an Punkten in einem *Wissenstest* vor und nach der Textlektüre operationalisiert. Die Messung

wird ausführlich in Abschnitt 6.3.2 erläutert. Die Studie erhebt explizit nicht den Anspruch, das dahinter liegende psychologische Merkmal *Wissen* über die elektrische Spannung zu behandeln (in ähnlicher Weise wird auch nicht von einer Messung des *Textverstehens* oder *Textverständnis* gesprochen). Dieses wäre als latente Eigenschaft lediglich *schätzbar* über das Antwortverhalten in einem Test und entsprechende statistische Modelle (für Einblicke in die Klassische Testtheorie oder die Item-Response-Theorie sei z. B. auf Moosbrugger und Kelava, 2012, Kapitel 5 beziehungsweise Kapitel 10 verwiesen)<sup>1</sup>.

Die zweite abhängige Variable bildet die *affektive Textwahrnehmung* der Schüler\*innen. Sie setzt sich zusammen aus mehreren einzelnen Facetten der Textwahrnehmung und dem globalen (Verständlichkeits-)Eindruck (siehe dazu weiter unten). Sie wird für die Messung zugänglich gemacht, indem sich die Studienteilnehmenden nach der Textlektüre in einem entsprechenden Fragebogen (*Bewertungsfragebogen*) dazu äußern, wie der Text auf sie gewirkt hat. Die Fragebogendaten sind somit subjektive *Textbewertungen*, welche sich auf den *subjektiven Eindruck* der Lesenden zum Text beziehen und aus welchen Rückschlüsse auf die affektive Textwahrnehmung gezogen werden. Jedes Item besteht in einer Aussage zu einem ganz konkreten Aspekt der affektiven Textwahrnehmung (z. B. „Ich fand den Text auf den ersten Blick übersichtlich“ oder „Beim Lesen brachten mich einzelne Textstellen zum Stocken“).

Die affektive Textwahrnehmung ist als umfassendes, mehrdimensionales Konstrukt zu betrachten. Es ist ein Ziel dieser Studie, die dimensionale Struktur des Konstrukts zu ermitteln und genauer zu beschreiben. Der statistische Ausdruck der *Konstruktdimensionen* bedeutet in der Analyse die extrahierten *Hauptkomponenten* (siehe für Einzelheiten Abschnitt 6.4.3) und im übertragenen Sinne die *Facetten der Textwahrnehmung*. Um den Begriff „Facetten der Textwahrnehmung“ greifbarer zu machen sei hier dem Ergebnisteil 7.2 vorausgegriffen: Es wurden evidenzbasiert sechs Hauptkomponenten extrahiert, sodass sich die affektive Textwahrnehmung als sechsdimensional darstellt und aus den Facetten (1.) *Kognitive Anregung*, (2.) *Lesbarkeit*, (3.) *Logischer Aufbau*, (4.) *Elaboriertheit*, (5.) *Angemessener Bezug zum Vorwissen* und (6.) *Involviertheit* besteht.

Im Zusammenhang mit der affektiven Textwahrnehmung steht auch der *globale (Verständlichkeits-)Eindruck* vom Text. Anhand des sogenannten *globalen Verständlichkeitsitems* im Bewertungsfragebogen wird danach gefragt, wie der Text insgesamt hinsichtlich seiner Verständlichkeit gewirkt hat („Ich fand den Text insgesamt verständlich“). Die Daten zum globalen Verständlichkeitsitem entsprechen also der *globalen Textbewertung*, die als Maß für den globalen Verständlichkeitseindruck fungiert. Im Forschungsinteresse liegt also das Wirkungsgefüge der oben beschriebenen Variablen *sprachliche Gestaltung der Texte* auf der einen Seite und *Behaltensleistung* sowie *affektive Textwahrnehmung* auf der anderen Seite.

---

<sup>1</sup>Wenn im Folgenden an manchen Stellen im Zusammenhang mit den Mess- und Einflussgrößen dieser Studie vom *Vorwissen* die Rede ist, dann ist damit stets das Ergebnis im Pre-Test für das Wissen zur elektrischen Spannung gemeint. Aus Gründen der sprachlichen Ökonomie wird der Begriff stellenweise verwendet werden, obwohl dies bezüglich seiner psychologischen Bedeutung als latentes Konstrukt unpräzise ist.

Die Arbeit zeichnet sich dabei durch die folgenden zugrundeliegenden Leitgedanken aus:

- Leitgedanke der systematischen *Trennung der Gestaltungs- von der Wirkungsebene*.

Diese Arbeit sieht auf Seiten der unabhängigen Variablen ausschließlich die Gestaltungsmittel von Textautor\*innen – die Ebene der sprachlichen Mittel wird daher klar von der Ebene der Textwirkungen getrennt. Es werden also nicht *unterschiedlich schwierige* Texte miteinander verglichen, wobei die (angenommene) Schwierigkeit aufgrund eines Konglomerats aus potenziell verständlichkeitsrelevanten sprachlichen Merkmalen theoretisch plausibel ist (vergleiche z. B. Hackemann, Heine und Höttecke, 2022). Vielmehr werden ganz konkrete, auf der Textoberfläche umsetzbare Gestaltungsmittel aus den allgemeinen wie fachspezifischen Vorarbeiten zur Textverständlichkeit gesammelt und anschließend in übersichtlicher Form gruppiert und geordnet (Unterkapitel 5.1). Strikt getrennt von den sprachlichen Gestaltungsmitteln werden Aspekte der *Wirkungen* von Texten auf Lesende gesammelt, von denen in der Literatur zur Textverständlichkeit die Rede ist. Diese setzen sich – wie oben beschrieben – zur abhängigen Variablen der *affektiven Textwahrnehmung* zusammen.

Die strikte Trennung der Gestaltungs- von der Wirkungsebene stellt einerseits eine Präzisierung der bislang vorhandenen Literatur dar (vergleiche dazu Seite 57) und bildet andererseits den Grundstein dafür, Ursache-Wirkungs-Zusammenhänge der Textgestaltung systematisch untersuchen zu können. Es wird Evidenz hinsichtlich der Frage nach den Auswirkungen ganz bestimmter, individueller Textoberflächenveränderungen auf die Lesenden gesammelt. In dieser Studie wurden für ebendiese isoliert vorgenommenen Veränderungen (1.) wortwahl-bezogene, (2.) satzstruktur-bezogene und (3.) personalisierende sprachliche Mittel ausgewählt (siehe Abschnitt 6.2.2).

- Leitgedanke des *Interesses an affektiver Wahrnehmung*.

In fachübergreifenden Ansätzen spielen affektive Gesichtspunkte der Textwirkung auf Lesende mitunter eine Rolle: Z. B. lassen Langer, Schulz von Thun und Tausch (2019) ihre Gütekriterien für verständliche Texte auf *Bewertungen* von Textbeurteilenden gründen und Groeben (1978) bezieht das „motivationale Erfolgskriterium ‚Interesse‘“ explizit in seine Arbeit mit ein. Aus Unterkapitel 3.4 geht jedoch hervor, dass speziell in naturwissenschaftsbezogenen Arbeiten zur Verständlichkeit von Texten nahezu ausschließlich kognitive Aspekte der Textwirkung betrachtet werden. Der Ansatz, gezielt *Textbewertungen* von Schüler\*innen zu erheben und somit auf deren subjektiven Eindruck zurückzuschließen, ist der physikalischen Fachdidaktik somit neu. Gleichzeitig ist er in Anbetracht von Kapitel 3 (genauer: Seite 43 zum interaktiven Charakter des verstehenden Lesens und Unterkapitel 3.3 zum Forschungsstand verständlicher Textgestaltung) ohne Weiteres plausibel: Entsprechend einer zeitgemäßen Auffassung von Textverstehen ist der Aufbau eines mentalen Modells beim Lesen bestimmt von einer *Interaktion* der Lesenden mit dem Text. Die affektive Wahrnehmung, die Lesende von einem Text haben, erscheint damit als ein plausibler, potenzieller Mediator zwischen der Textgestaltung auf der einen und der kognitiven Wirkung des Texts auf der anderen Seite. Gerade unter didaktischen Gesichts-

punkten ist die affektive Wahrnehmung eine interessante Größe. Denn sie mag durchaus entscheidend dafür sein, inwiefern Lesende sich engagiert fühlen und geneigt zeigen, sich aktiv mit den Textinhalten auseinanderzusetzen und ein echtes Textverständnis aufzubauen.

Härtig, Fraser, Bernholt und Retelsdorf (2019) werfen hierzu die Frage auf, ob nicht womöglich durch sprachliche Manipulation vor allem die propositionale Ebene des Textverstehens betroffen sei, zur Beantwortung von Verständnisfragen aber mitunter hierarchiehöhere Prozesse (vergleiche Unterkapitel 3.2) entscheidend seien und somit die aktive Einbringung von Vorwissen nötig sei. Dass Schüler\*innen dies jedoch nicht leisten (können), stelle eine mögliche Erklärung für die ausbleibenden verstehensbezogenen Effekte bei unterschiedlichen sprachlichen Gestaltungen dar. Unter anderem an solchen Unklarheiten setzen die im Folgenden beschriebenen Forschungsfragen an, indem mit der affektiven Textwahrnehmung eine potenzielle mediatorische Zwischenebene der kognitiven Verarbeitung von Textinformationen adressiert wird.

- Leitgedanke der *Einführung eines abstrakten, physikalischen Begriffs*.

Die Effekte unterschiedlicher sprachlicher Textgestaltungen auf Schüler\*innen sind gerade vor dem Hintergrund eines physikalischen Sachtexts zu einem abstrakten Begriff interessant. In Abschnitt 2.1.1 wird deutlich, inwiefern das inhaltliche und das sprachliche Begriffslernen insbesondere in der Physik stark verknüpft sind. Wie der Abschnitt weiter erläutert, ist aus diesem Grund der Sprachgebrauch beim Erlernen abstrakter physikalischer Begriffe besonders bedeutsam. In *fachinhaltlicher* Hinsicht ist also der spezifische Kontext der Einführung eines abstrakten physikalischen Begriffs (hier: der *elektrischen Spannung*) ein wichtiges Merkmal dieser Studie.

Ebenso sind *fachsprachliche* Aspekte eines abstrakten, physikalischen Kontexts bemerkenswert: dass sprachliche Veränderungen für Schüler\*innen besonders relevant sein könnten, wenn der Sprachstil im Text an sich herausfordernd ist, ist eine naheliegende Annahme. Nun gilt die naturwissenschaftliche Fachsprache als eine Sprachform, die einen außergewöhnlich hohen Anspruch an Lernende aufweist (vergleiche Abschnitte 2.2.1 und 2.2.2). Darüber hinaus wird in Abschnitt 2.1.3 erläutert, dass ein korrekter, fachbezogener Sprachgebrauch und dabei insbesondere rezeptive Sprachfähigkeiten ein Lerngegenstand des Physikunterrichts sind. Dies alles macht den *fachsprachlichen* Kontext, den ein naturwissenschaftlicher Sachtext über einen abstrakten physikalischen Begriff wie die elektrische Spannung mitbringt, zu einem auszeichnenden Merkmal dieser Arbeit.

- Leitgedanke der *Kontrolle der Behaltensleistung*.

Selbst wenn Schüler\*innen ein Lehr-Lern-Mittel verständlich *bewerten*, so ist nicht gesagt, dass es auch tatsächlich zu einem guten Verständnis führt (Kulgemeyer und Wittwer (2023) sprechen hier von der „Verstehensillusion“). Ein wichtiger Leitgedanke ist daher ebenfalls, das kognitive Erfolgskriterium im Sinne einer Behaltensleistung beim Lesen mitzumessen

und im Zusammenhang mit der Textgestaltung sowie der affektiven Textwahrnehmung zu analysieren.

Unter diesen Leitgedanken wurde eine empirische Studie entwickelt, die den Zusammenhang der sprachlichen Gestaltung eines Texts zum Begriff der elektrischen Spannung mit der Behaltensleistung und der affektiven Textwahrnehmung von Schüler\*innen untersucht. Dieses Forschungsinteresse wird im Folgenden anhand von Forschungsfragen präzisiert.

## 4.2 Die Forschungsfragen der Studie zur Textverständlichkeit

Es wurde bereits deutlich, dass sich diese Studie mit kognitiven und affektiven Gelingenskriterien des verstehenden Lesens beschäftigt, indem die Auswirkungen sprachlicher Textmodifikationen auf die Behaltensleistung sowie die affektive Textwahrnehmung von Schüler\*innen untersucht wurden. Dabei ist die Behaltensleistung empirisch anhand eines Wissenstests im Pre-Post-Design gut zugänglich, da sie sich darin manifestiert, wie gut Schüler\*innen nach der Textlektüre im Vergleich zu vorher über die Lerninhalte des Texts Bescheid wissen. Konturloser gestaltet sich jedoch zunächst der Konstruktraum, in dem sich die unabhängige Variable der sprachlichen Gestaltung und die abhängige Variable der affektiven Textwahrnehmung verorten lassen. Daher lautet ein erstes Forschungsdesiderat, dem sich diese Arbeit zuwendet:

Desiderat einer theoretischen Vorarbeit zum Konstruktraum:

Die diffuse Erkenntislage in der Theorie zur Textverständlichkeitsforschung – in fachübergreifender wie in fachspezifischer Hinsicht – bedarf einer systematischen Neuordnung in einen umfassenden Konstruktraum, der einen konsensfähigen, geordneten Rahmen für diese und verwandte Studien zur Wirkung von Texten bietet.

Um also eine empirisch-analytische Herangehensweise an die Untersuchung von Wirkungsgefügen zwischen sprachlicher Textgestaltung und Textwirkung vorzubereiten, wurde der Konstruktraum für diese Studie daher auf der Grundlage einer intensiven Literaturrecherche rekonstruiert: Die oft verschwimmenden Ebenen zwischen Textgestaltung (sprachliche Mittel) und resultierender Textwirkung (resultierende Texteigenschaften, Eindrücke der Lesenden) werden dabei klar aufgetrennt. Einerseits bezieht sich die Rekonstruktionsarbeit auf die unabhängige Variable, die in der sprachlichen Textgestaltung besteht. Die umfassende Sichtung von Literatur zur Textverständlichkeit und zu den sprachlichen Besonderheiten naturwissenschaftlicher Texte mündete somit in eine Sammlung und Gruppierung sprachlicher Mittel, welche den Autor\*innen als Handwerkszeug für die Textproduktion zur Verfügung stehen. Drei Gruppen sprachlicher Mittel wurden schließlich für die Erstellung sprachlich variierten Texte ausgewählt, nämlich die *Wortwahl* beziehungsweise *morphologische Auffälligkeiten auf Wortebene*, die *Satzstrukturen* beziehungsweise *morphologische Auffälligkeiten auf Satzebene*

und die *erzählenden/personalisierenden Elemente* (siehe dazu Unterkapitel 6.2).

Auf der anderen Seite diente die Rekonstruktionsarbeit dazu, die abhängige Variable der affektiven Textwahrnehmung für eine Erhebung zugänglich zu machen. Die einzelnen Facetten der affektiven Textwahrnehmung, die oft implizit in der Beschreibung günstiger oder ungünstiger Texte Erwähnung finden, wurden dafür möglichst ausschöpfend aus einschlägigen Vorarbeiten herausgearbeitet, sortiert und strukturiert.

Das Ergebnis dieser theoretischen Vorarbeit wird ausführlich im Kapitel 5 vorgestellt. Es bildet die Grundlage für den anschließenden empirischen Teil (Kapitel 6 und 7). Es wurden dabei Daten von insgesamt 787 Mittelstufenschüler\*innen aus bayerischen Realschulen und Gymnasien zur Verständlichkeit sprachlich variierten Texte gesammelt und ausgewertet. Die Altersstufe zeichnet sich dadurch aus, dass hier in der Regel basale Lesefähigkeiten abschließend ausgebildet sind und der Leseverstehensprozess nicht mehr vorrangig von technischen Leseprozessen auf hierarchieniederer Ebene bestimmt wird (vergleiche Abschnitt 3.2.2).

Die Studie ist darauf ausgelegt, die folgenden Forschungsfragen zu beantworten:

Forschungsfrage 1:

Inwiefern beeinflussen (fach-)sprachliche Gestaltungsmittel aus den Gruppen Wortwahl / morphologische Auffälligkeiten auf Wortebene, Satzstrukturen / morphologische Auffälligkeiten auf Satzebene und erzählende/personalisierende Elemente die Behaltensleistung von Schüler\*innen beim Lesen eines physikalischen Sachtexts?

Die Studie soll möglichst systematisch untersuchen, inwiefern einzelne sprachliche Modifikationen zu unterschiedlichen Behaltensleistungen bei Schüler\*innen führen. Es werden daher die Effekte solcher sprachlicher Mittel untersucht, die unabhängig voneinander veränderlich sind. Während also solche sprachlichen Merkmale modifiziert werden, die den Überbegriffen *Wortwahl / morphologische Auffälligkeiten auf Wortebene, Satzstrukturen / morphologische Auffälligkeiten auf Satzebene* und *erzählende/personalisierende Elemente* zuzuordnen sind, werden der Textinhalt sowie übrige sprachlichen Merkmale dabei so weit wie möglich konstant gehalten (Einzelheiten zum Design und zur Durchführung der Studie werden in Kapitel 6 erläutert).

Forschungsfrage 2:

Zur affektiven Textwahrnehmung.

Forschungsfrage 2a:

Wie lassen sich die einzelnen Facetten der affektiven Textwahrnehmung auf Grundlage von Textbewertungen durch Schüler\*innen gruppieren und strukturieren? Wie hängen die Facetten der affektiven Textwahrnehmung miteinander und mit einem globalen Verständlichkeitseindruck zusammen?

Forschungsfrage 2b:

Inwiefern beeinflussen (fach-)sprachliche Gestaltungsmittel aus den Gruppen Wortwahl / morphologische Auffälligkeiten auf Wortebene, Satzstrukturen / morphologische Auffälligkeiten auf Satzebene und erzählende/personalisierende Elemente den globalen Verständlichkeitseindruck sowie die einzelnen Facetten der affektiven Textwahrnehmung bei Schüler\*innen?

Wie oben beschrieben ist es Teil der theoretischen Vorarbeit, den Konstruktraum hinsichtlich wahrnehmungsbezogener, affektiver Textwirkungen zu rekonstruieren. Damit soll insbesondere ausschöpfend erfasst werden, welche Facetten der Textwirkung bislang in der Textverständlichkeitsforschung mitgedacht werden. Es bleibt jedoch die Frage offen, wie das Konstrukt *affektive Textwahrnehmung* bei Lesenden tatsächlich realisiert ist. Welche Subfacetten, also Konstruktdimensionen, lassen sich aus den Bewertungsdaten zusammenfassen und voneinander unterscheiden? Die gefundene Struktur soll dann anhand von Zusammenhängen der einzelnen Dimensionen miteinander sowie mit dem übergeordneten, globalen Verständlichkeitseindruck näher beschrieben werden.

Indem der Textwahrnehmung auf diese Art eine evidenzbasierte Struktur einbeschrieben wird, wird der nächste Untersuchungsschritt vorbereitet: Es sollen nun die affektiven Wirkungen der oben genannten sprachlichen Mittel analysiert werden. Konkret bedeutet dies, nach Effekten der Textgestaltung auf erstens den globalen Verständlichkeitseindruck und zweitens auf die im Schritt 2a gefundenen Subfacetten zu fragen.

Forschungsfrage 3:

Inwiefern hängen die Behaltensleistung und der globale Verständlichkeitseindruck beziehungsweise die Facetten der affektiven Textwahrnehmung zusammen?

Die dritte Forschungsfrage bringt nun die beiden abhängigen Variablen Textwahrnehmung und Behaltensleistung zusammen. Die Sachlogik (siehe Abbildung 4.1) begründet einen Ansatz, in dem die Wahrnehmung *zwischen* Textgestaltung und Behaltensleistung liegt. Entsprechend wird in dem Zusammenwirken der beiden Konstrukte die Textwahrnehmung als Regressor für die Behaltensleistung betrachtet. Es soll untersucht werden, inwiefern sich in Abhängigkeit vom subjektiven Eindruck, den eine lesende Person von einem Text hat, eine bessere oder schlechtere Behaltensleistung ergibt.

#### *4.2. Die Forschungsfragen der Studie zur Textverständlichkeit*

In den folgenden Kapiteln werden nun die einzelnen Bestandteile dieser Arbeit und deren Ergebnisse ausführlich dargestellt. Die oben erläuterten Ausführungen zusammenfassend wird es also um die Rekonstruktion bestehender Gütekriterien und Modelle für verständliche Texte (theoretische Vorarbeit, Kapitel 5) gehen sowie um die Durchführung und Auswertung einer empirischen Studie zur Gewinnung evidenzbasierte Erkenntnisse (Kapitel 6 und 7).



## 5 | REKONSTRUKTION DES KONSTRUKTRAUMS FÜR TEXTVERSTÄNDLICHKEIT

Dieses Kapitel ist ein entscheidender Vorbereitungsschritt für die Konzeption und Umsetzung der empirischen Studie gewidmet. Der Schritt bestand darin, einen Konstruktraum zu schaffen, in welchem die Untersuchungsvariablen der Textgestaltung und der Textwirkung<sup>1</sup> präzisiert und in einer umfassenden, konsensfähigen Systematik verortet werden können. Es wurde in der bestehenden Literatur intensiv zur Gestaltung verständlicher Texte beziehungsweise zur Beurteilung der Textverständlichkeit recherchiert und aus der fachübergreifenden sowie naturwissenschaftsspezifischen Theorie ein Konstruktraum für die Untersuchungsgrößen dieser Studie rekonstruiert. Damit ist gemeint, dass möglichst erschöpfend alle Aspekte beziehungsweise Ausprägungs- und Realisierungsformen der sprachlichen Textgestaltung einerseits und der affektiven Textwahrnehmung andererseits zusammengetragen und geordnet wurden. Diese Arbeit hält es somit in gewisser Weise mit dem altgriechischen Dichter Hesiod (siehe das Zitat oben): Den empirischen Untersuchungen ging die Entwicklung eines theoretisch fundierten Gerüsts für die Operationalisierung der Untersuchungsgrößen *sprachliche Textgestaltung* und *affektive Textwahrnehmung* voraus. Dieses Gerüst stellt eine strukturgebende Grundlage für diese Studie und auch für künftige, ähnliche Studien zur Textverständlichkeit bereit.

Die Ausgangslage für die Verortung der textseitig modifizierbaren Einflussfaktoren für die Textverständlichkeit (welche Aspekte der Textgestaltung gibt es?) und der Bestimmungsfaktoren für die Textverständlichkeit, die sich in der Interaktion Lesender mit dem Text einstellen (welche Aspekte der Textwirkung auf Lesende gibt es?), besteht einerseits in Modellen und empirischen Befunden zu diesem Thema. Forschungsergebnisse aus fachübergreifender Sichtweise wurden in Unterkapitel 3.3 detailliert beschrieben. Für den spezifischen Anwendungsfall naturwissenschaftlicher Texte, der eine spezifische Färbung der Gestaltungs- und

---

<sup>1</sup>Unter *Textwirkung* sind gemäß der Auffassung in dieser Arbeit kognitive sowie affektive Aspekte zu fassen (Behaltensleistung und affektive Textwahrnehmung). Auf Seite 68 f. wurde bereits kurz erwähnt, dass eine theoretische Vorarbeit vor allem auf der affektiven Seite als nötig erachtet und vorgenommen wurde. In diesem Kapitel wird es daher auf Seiten der unabhängigen Variablen um die sprachliche Textgestaltung und auf Seiten der abhängigen Variablen um affektive Aspekte gehen.

der Wirkungsaspekte mitbringt, gab Unterkapitel 3.4 einen entsprechenden Überblick zum Erkenntnisstand. Insbesondere für die Gestaltungsmittel zum Zweck gut verständlicher naturwissenschaftlicher Texte sind fachsprachliche Auffälligkeiten relevant, die in Abschnitt 2.2.2 aufgelistet wurden.

Die fachsprachlichen Besonderheiten inklusive deren Anforderungscharakter (vergleiche Seite 58) in die fachübergreifend angelegten Textverständlichkeitskriterien einzuordnen ist nicht trivial. Darauf weisen empirische Ergebnisse der naturwissenschaftsspezifischen Textverständlichkeitsforschung hin, indem die zu erwartenden Wirkungen sprachlicher Textmodifikationen auf den kognitiven Leseerfolg weitestgehend ausbleiben. Die Ergebnisse legen vielmehr nahe, dass Texte mit unterschiedlichen sprachlichen Anforderungsniveaus zu keinen oder nur minimalen Unterschieden im Textverständnis führen (Härtig, Fraser, Bernholt & Retelsdorf, 2019; Hackemann, Heine & Höttecke, 2022; Heine et al., 2018; Plath & Leiss, 2018). Darüber hinaus ist fraglich, inwiefern die fachübergreifend angelegten Verständlichkeitsdimensionen (z. B. nach Göpferich, 2002, siehe Seite 55 in dieser Arbeit oder nach Langer, Schulz von Thun und Tausch, 2019, siehe Seite 53 in dieser Arbeit) auch für Physiktexte gelten. Es erscheint eine plausible Annahme zu sein, dass Schülerinnen bei physikalischen Texten womöglich spezifische Erwartungen haben, die auf bisherigen Erfahrungen mit solchen Texten beruhen (vergleiche dazu die Ausführungen zur Relevanz des Vorwissens für das Textverstehen auf Seite 48). Konkrete, einleuchtende Spekulationen sind, dass Schüler\*innen im Umgang mit Physiktexten an ein spezifisches sprachliches Inventar sowie einen spezifischen logischen Aufbau gewöhnt sein mögen, ohne dass diese Spezifika ausgehend von den Verständlichkeitsmodellen als besonders günstig gelten würden (z. B. für die Physik gängige Fremd- und Fachwörter, Funktionsverbgefüge und einen logisch anstatt temporal orientierten Aufbau mit Vor- und Rückverweisen; siehe für Einzelheiten Abschnitt 2.2.2). Sie mögen auch mit einer speziellen inneren Haltung an die Texte herantreten, welche beispielshalber zur geistigen Aktivierung führt und/oder dazu, dass sie mit einem nüchternen Stil rechnen und mitunter verbende, unterhaltsame Elemente als störend empfinden.

Der im Folgenden vorgestellte Konstruktraum wurde mit dem Ziel erarbeitet, zusammenzutragen, welche Aspekte der Textgestaltung und welche Aspekte der affektiven Textwirkungen bislang in der Literatur (insbesondere für naturwissenschaftliche Sprache/Texte) Erwähnung fanden und wie sich eine systematische Untersuchung der affektiven Auswirkungen der sprachlichen Textgestaltung auf Schüler\*innen veranlassen lässt. Die Systematik der hier vorliegenden Studie ist also nicht darauf ausgerichtet, anhand sprachlicher Modifikationen (vermeintlich) verschiedene Sprachniveaus in den Texten zu generieren und miteinander zu vergleichen. Es sollten stattdessen isolierte Ursache-Wirkungs-Zusammenhänge zwischen sprachlicher Gestaltung und kognitiven sowie affektiven Aspekten der Textwirkung auf Lesende erforscht werden.

## Zum methodischen Vorgehen

Die zentralen Quellen für die Erarbeitung des Konstruktraums waren: Artelt et al. (2005), Fang (2006), Frickel (2018), Göpferich (2002), Heine et al. (2018), Höttecke, Ehmke, Krieger und Kulik (2017), Kernén und Riss (2012), Lenhard (2013), Lutz (2015), Starauschek (2006), Patterson, Roman, Friend, Osborne und Donovan (2018) und Rincke (2010). Aus den aufgelisteten Publikationen wurden sprachliche Gestaltungsmittel einerseits und Aspekte der möglichen Wirkungen auf Lesende andererseits zusammengetragen. Mit der Bearbeitung dieser Literatur stellte sich gewissermaßen eine „theoretische Sättigung“ ein in dem Sinne, dass bei der Durchsicht weiterer einschlägiger Arbeiten (z. B. Langer, Schulz von Thun und Tausch, 2019; Härtig, Fraser, Bernholt und Retelsdorf, 2019; Leisen, 2003) keine neuen Bestandteile für die Sammlung hinzukamen.

Ein besonderes Augenmerk bei dieser Erarbeitung des Konstruktraums lag auf der sorgfältigen Trennung zwischen tatsächlichen, konkreten Mitteln für die sprachliche Textgestaltung (gewissermaßen dem *Handwerkszeug* der Autor\*innen) und den möglichen Wirkungen, die sich beim Lesen eines Texts erzielen lassen<sup>2</sup>. In beiden Ebenen wurden Gruppen gebildet und mit Überbegriffen versehen, sodass sich *acht Gruppen sprachlicher Variablen* ergaben und *zwölf Kategorien zur Beschreibung der affektiven Textwahrnehmung* (vergleiche die Tabellen 5.1 und 5.2). Die Ebenentrennung erforderte mitunter, bestehende Annahmen über Wirkungsgefüge aufzubrechen und getrennt in den Konstruktraum der Gestaltung beziehungsweise in den Konstruktraum der Wirkungen einzugliedern. Dadurch wurde eine Grundlage für Untersuchungen gelegt, welche sprachlichen Gestaltungsmittel tatsächlich für Schüler\*innen beim Lesen eines Physiktexts eine Rolle spielen und deren affektive Textwahrnehmung beeinflussen. Beispielsweise drückt der Hinweis an Autor\*innen, die Kompliziertheit eines Text möglichst zu minimieren, indem Satzstrukturen vereinfacht, überschaubare Einheiten gebildet und eine einheitliche Terminologie verwendet werden (Lutz, 2014, S. 55), eine solche Annahme über ein Wirkungsgefüge aus. Für den Konstruktraum sind die „vereinfachten Satzstrukturen“ als Überbegriff für verschiedene, konkrete sprachliche Mittel (z. B. parataktischer Satzbau, aktivische Formulierungen) zu betrachten. Die „Gliederung in überschaubare Einheiten“ zählt zu den konkreten sprachlichen Mitteln, um die *Textstruktur explizit zu machen* sowie zu den Mitteln der *typographischen Markierungen* und die „einheitliche Terminologie“ wäre ein konkretes Mittel, das dem Überbegriff der *Wortwahl* und auch den *Kohäsionsmitteln* zuzuordnen ist.

In der Ebene der sprachlichen Gestaltungsmittel (in acht Gruppen geordnet) sowie der gegenübergestellten Ebene der affektiven Textwahrnehmung (in zwölf Kategorien eingeteilt) bestand schließlich das Gerüst für die Operationalisierung der Variablen in der empirischen Textverständlichkeitsstudie: Mit der *Wortwahl / morphologischen Auffälligkeiten auf Wortebene*, den *Satzstrukturen / morphologischen Auffälligkeiten auf Phrasen- und Satzebene* und den *erzählenden und personalisierenden Elementen* wurden drei Gruppen sprachlicher Gestal-

---

<sup>2</sup>Die Ebenentrennung ist eine Folge des *Leitgedankens der systematischen Herangehensweise an die Untersuchung von Ursache-Wirkungs-Zusammenhängen*, welcher auf Seite 66 erläutert wird.

tungsmittel ausgewählt, die erstens gut trennbar von den übrigen Kategorien sind – also wenige konkrete Mittel umfassen, die mehreren Gruppen zugeordnet sind – und die zweitens fachsprachlich relevant sind (siehe dazu im Einzelnen Unterkapitel 5.1). Außerdem dienten die Wahrnehmungskategorien anhand ihrer zugeordneten Textwirkungsaspekte (siehe Unterkapitel 5.2) als Grundlage für die Formulierung subjektiv bewertbarer Aussagen zum Text, die schließlich zu einem Fragebogen zur affektiven Textwahrnehmung weiterverarbeitet wurden.

## 5.1 Die Ebene der sprachlichen Mittel

In der oben genannten Literatur kommen zahlreiche ganz konkrete, eng gefasste Gestaltungsmittel zur Sprache, die Autor\*innen bei der Textproduktion zur Hand haben. Die einzelnen Gestaltungsmittel werden dabei oftmals miteinander in Verbindung gebracht, indem sie Überbegriffen zugeordnet werden oder indem Wirkungen beschrieben werden, für welche sie gemeinsam ursächlich sein können. Möglichst alle sprachlichen Mittel zusammenzutragen und gleichzeitig die in der Literatur vorgeschlagenen Verbindungen zu bewahren führte zu dem hier vorgestellten Konstruktraum für sprachliche Mittel, der acht Gruppen umfasst, wobei die Zuordnung der konkreten sprachlichen Mittel zu den Gruppen in den seltensten Fällen in eindeutiger Weise möglich war. So bedeutet beispielshalber das sprachliche Mittel der *Verwendung von Synonymen* (anstelle von *Wortwiederholungen*) eine Veränderung auf *Wortebene* (Artelt et al., 2005, S. 15) und stellt gleichzeitig ein konkretes *Kohäsionsmittel* dar (Heine et al., 2018, S. 83). Genauso zählen Analogien zu den *inhaltlichen Hilfestellungen* (Artelt et al., 2005, S. 28) und gleichzeitig bedienen sie einen anderen *Darstellungsmodus*, indem bildhafte Vorstellungen geschaffen werden sollen (Göpferich, 2002, S. 171). Ein Grund für diese unterschiedlichen Zuweisungsmöglichkeiten liegt darin, dass verschiedene Autor\*innen verschiedene Arten der Sortierung vornehmen (z. B. ordnen sie sprachliche Mittel den Dimensionen eines Verständlichkeitsmodells zu – siehe z. B. Göpferich, 2002 – oder es liegt ein Analyseraster für Texte vor – siehe z. B. Kern und Riss, 2012).

In Tabelle 5.1 sind die acht Gruppen, die aus der Literatur extrahiert wurden, mit exemplarischen, zugeordneten sprachlichen Mitteln aufgelistet. Eine vollständige (mehrere Seiten füllende) Liste mit allen zugeordneten sprachlichen Mitteln für jede Gruppe inklusive Angabe der Textstellen, aus denen die Mittel extrahiert wurden, findet sich im Anhang C. Die Sammlung und Gruppierung der sprachlichen Mittel bietet eine Übersicht über die sprachlichen Mittel, die für Autor\*innen naturwissenschaftlicher Texte zur Verfügung stehen. Für die Praxis stellt sie somit in Anlehnung an den obigen Vergleich der sprachlichen Mittel mit einem *Handwerkszeug für Autor\*innen* eine Art *Werkzeugkoffer* der Textgestaltung dar und zeigt die Bandbreite der sprachlichen Gestaltungsmöglichkeiten auf. Für die Wissenschaft wiederum schafft die Sammlung und Gruppierung sprachlicher Mittel eine wichtige Grundlage für systematische Untersuchungen dazu, mit welchen sprachlichen Modifikationen welche Wirkungen auf Lesende verbunden sind. Dies gilt insbesondere für naturwissenschaftliche

<p>Wortwahl / morphologische Auffälligkeiten auf Wortebene</p> <hr/> <ul style="list-style-type: none"> <li>• Fachlicher Wortschatz/Fachbegriffe</li> <li>• Zusammengesetzte Wörter</li> <li>• Funktionsverbgefüge (Verb-Sinnentleerung)</li> <li>• Jugendslang</li> <li>• Ellipsen (vs. Redundanz)</li> <li>• Metaphorische Ausdrücke</li> </ul> <hr/> 	<p>Satzstrukturen / morphologische Auffälligkeiten auf Satzebene</p> <hr/> <ul style="list-style-type: none"> <li>• Satzlänge (= Wörter pro Satz)</li> <li>• Verbalklammern/Satzklammern</li> <li>• Passivische/aktivische Konstruktionen</li> <li>• Idiome, standardisierte sprachl. <i>Fertigstücke</i> (Phraseologismen)</li> <li>• Ellipsen (vs. Redundanz)</li> <li>• Fragen/Aufgaben zum Text</li> </ul> <hr/> 
<p>Erzählende und personalisierende Textelemente</p> <hr/> <ul style="list-style-type: none"> <li>• unpersönliche Wendungen („man“) vs. persönliche Anrede („du“/„wir“)</li> <li>• Dialoge</li> <li>• Passivische/aktivische Konstruktionen</li> <li>• Jugendslang</li> <li>• Metaphorische Ausdrücke</li> <li>–</li> </ul> <hr/> 	<p>Hilfestellungen zur Textstruktur und Aufmerksamkeitssteuerung</p> <hr/> <ul style="list-style-type: none"> <li>• Verbalisierung des Lesezwecks</li> <li>• (Verbalisierte/nicht-verbalisierte) Gedanken-sprünge/Themenwechsel</li> <li>• Rück- und Vorverweise</li> <li>• (Zwischen-)Überschriften</li> <li>• Analogien</li> <li>• Konjunktionen</li> </ul> <hr/> 
<p>Inhaltliche Hilfestellungen</p> <hr/> <ul style="list-style-type: none"> <li>• inhaltliches Fazit/Zusammenfassung</li> <li>• Redundanz bedeutungsrelevanter Informationen</li> <li>• Abstraktionen/Verallgemeinerungen</li> <li>• Advance Organizers</li> <li>• Fragen/Aufgaben zum Text</li> <li>• Konjunktionen</li> </ul> <hr/> 	<p>Typographische Markierungen</p> <hr/> <ul style="list-style-type: none"> <li>• Schriftfarbe</li> <li>–</li> <li>• (Zwischen-)Überschriften</li> <li>• Fett-/Kursivdruck</li> <li>–</li> </ul> <hr/> 
<p>Kohäsionsmittel</p> <hr/> <ul style="list-style-type: none"> <li>• Wiederaufnahme von Satzteilen</li> <li>• Explizite Bezüge zwischen Textteilen</li> <li>• Wortwiederholungen (nominale vs. synonymiale Verbindungen)</li> <li>• Rück- und Vorverweise</li> <li>• Ellipsen (vs. Redundanz)</li> <li>• Konjunktionen</li> </ul> <hr/> 	<p>Darstellungsmodi und Zeichensysteme</p> <hr/> <ul style="list-style-type: none"> <li>• Tabellen</li> <li>• Diagramme/Graphen</li> <li>• Allegorien</li> <li>• Symbole (i. S. d. Stilmittels)</li> <li>• Analogien</li> <li>• Metaphorische Ausdrücke</li> </ul> <hr/> 

Tabelle 5.1.: Überblick über die acht Gruppen sprachlicher Gestaltungsmittel. Zu jeder Gruppe sind im obersten Listenblock exemplarisch zwei solche Mittel genannt, die nur dieser Gruppe zugeordnet sind. Im zweiten Listenblock sind zwei solche Mittel genannt, die zusätzlich in genau einer weiteren Gruppe vertreten sind, und im dritten Listenblock solche, die mindestens drei Gruppen angehören. Der farbige Balken unterhalb jeder Gruppe gibt der Reihe nach den Anteil eindeutiger Zuordnungen (petrol), zweideutiger Zuordnungen (orange) und vieldeutiger Zuordnungen (hellgrün) wieder. Für eine vollständige Auflistung aller Zuordnungen pro Gruppe siehe Anhang C.

Texte, da fachsprachliche Oberflächenmerkmale bei der Zusammenstellung berücksichtigt wurden. Die Sammlung bietet einen umfassenden Fundus, aus dem einzelne sprachliche Mittel oder Sets an sprachlichen Mitteln ausgewählt werden können und hinsichtlich ihres Einflusses auf die Textwirkung untersucht werden können.

Besonders interessant für diese Studie war, inwiefern sich einzelne Gruppen voneinander trennen lassen. Dies wurde heuristisch anhand der Beobachtung ermittelt, wie viele sprachliche Mittel einer bestimmten Gruppe *eindeutig* genau dieser Gruppe, *zweideutig* dieser und einer weiteren Gruppe oder *drei- und mehrdeutig* zugeordnet wurden. Der farbige Balken in Tabelle 5.1 zeigt daher als gestapeltes Balkendiagramm für jede Gruppe relativ gesehen an, wie viele eindeutige, zweideutige sowie drei- und mehrfache Zuordnung die Gruppe umfasst<sup>3</sup>. Bei einem vergleichenden Blick auf die Balkendiagramme fällt auf, dass sich die Gruppen der *Wortwahl*, der *Satzstrukturen* und der *personalisierenden/erzählenden Elementen* durch anteilmäßig große petrolfarbene Bereiche und kleine hellgrüne Bereiche auszeichnen, also relativ gesehen durch viele eindeutige Zuordnungen charakterisiert sind. Im Gegensatz dazu weisen vor allem die *Hilfestellungen zur Textstruktur / Aufmerksamkeitssteuerung* (hier schwingen Mittel zur Unterstützung der *globalen Kohärenzbildung* mit), die *Typographischen Markierungen* und die *Kohäsionsmittel* (dies ist ein Synonym für Mittel zur Unterstützung der *lokalen Kohärenzbildung*) einen besonders kleinen petrolfarbenen Bereich auf. Mittel, die hier zugeordnet sind, werden also zu großen Teilen in der Literatur auch in Verbindung mit anderen Gruppen genannt. Dies lässt sich vorsichtig als Indiz für ein besonders vielschichtiges Wirkungsspektrum der sprachlichen Mittel dieser Gruppen interpretieren. Eine solche Vielschichtigkeit wiederum könnte einen Erklärungsansatz dafür liefern, warum Veränderungen im Kohärenzgrad naturwissenschaftlicher Texte in empirischen Studien oftmals nicht die erwarteten Wirkungen zeigen (Härtig, Fraser, Bernholt & Retelsdorf, 2019; Starauschek, 2006). Die Konsequenzen von kohärenzbezogenen Textmodifikationen sind offenbar nur unzureichend verstanden.

In der Meta-Analyse von Strohmaier, Ehmke, Härtig und Leiss (2023) (siehe Seite 59) wurde ebenfalls ein Kategoriensystem erstellt, um die in Einzelstudien vorgenommenen sprachlichen Modifikationen einzuordnen. Dadurch gelangten die Autoren zu Aussagen darüber, welche Kategorien sprachlicher Mittel über viele Einzelstudien hinweg welche Auswirkungen auf die gezeigte Verständnisleistung von Schüler\*innen zeigten. Das Kategoriensystem der Meta-Analyse besteht in einem 2×3-Raster, in welchem die drei Dimensionen „linguistic complexity“, „increasing cohesion“ und „increasing clarity and elaboration“ jeweils in die beiden Levels „local“ (Wort- und Satzebene) und „global“ (Textebene) differenziert sind (Strohmaier,

---

<sup>3</sup>Die Balken sind als grobe Orientierung zu betrachten. Von einer quantitativen Auswertung mit präzisen Zahlenangaben wurde bewusst abgesehen um die zusammengetragenen Gestaltungsmittel-Listen statistisch nicht überzubeanspruchen. Es sind erstens selbst die konkreten, den Überbegriffen zugeordneten sprachlichen Mittel nicht ganz überschneidungsfrei miteinander (z. B. sind Nominalisierungsgruppen eine spezielle Art eines Gliedsatzes, die beiden sprachlichen Mittel sind aber beide als eigenständige sprachliche Mittel in der Liste für die *Satzstrukturen* aufgeführt) und zweitens wäre eine Reliabilitätsprüfung für die Gruppeneinteilung sinnvoll, bevor belastbare, quantitative Aussagen getätigt werden.

Ehmke, Härtig & Leiss, 2023, S. 4). Die Meta-Analyse bietet damit ein alternatives Kategoriensystem für sprachliche Mittel. Um beide Kategoriensysteme – das der Meta-Analyse und das der hier vorliegenden Studie (Tabelle 5.1) – zu vergleichen, ist ein Blick auf die den jeweiligen Kategorien zugeordneten konkreten sprachlichen Mittel nötig. Solche Vergleiche werden für die Diskussion der in dieser Studie gefundenen Ergebnisse an geeigneten Stellen vorgenommen (siehe die Diskussionsabschnitte in Unterkapitel 7.1).

Im Gegensatz zur eben erwähnten Meta-Analyse, die sich auf Analysen zu kognitiven Erfolgskriterien des verstehenden Lesens beschränkt, ist es ein Ziel dieser Arbeit, auch Einblicke in das Wirkungsgefüge zwischen Testgestaltung und Textwirkung unter affektiven Gesichtspunkten zu gewinnen. Wie das Konstrukt der affektiven Textwahrnehmung auf der Grundlage bestehender Theorie zur Textverständlichkeit gefasst werden kann, ist daher Gegenstand des folgenden Unterkapitels.

## 5.2 Die Ebene der Textwirkung

In der Literatur zur Textverständlichkeit geht es immer wieder darum, wodurch sich Texte auszeichnen sollten, damit sie verständlich sind. Als verständlichkeitsfördernde Eigenschaften werden teilweise ganz konkrete Aspekte der Textoberfläche genannt, also sprachliche Mittel der Textautor\*innen (siehe das vorangehende Unterkapitel 5.1). Teilweise werden aber auch postulierte oder empirisch belegte *Wirkungen* dieser sprachlichen Mittel beschrieben, die zur Verständlichkeit des Texts beitragen. Dies geschieht zum Beispiel, wenn für Modellbildungen *Dimensionen der Textverständlichkeit* identifiziert werden (z. B. *Gliederung/Ordnung* bei Langer, Schulz von Thun und Tausch, 2019, sowie bei Lutz, 2015, oder *Simplizität* und *Perzipierbarkeit* bei Göpferich, 2002). Und auch wenn der Einfluss der konkreten Mittel auf Teilprozesse beim verstehenden Lesen beleuchtet wird, werden mitunter erstrebenswerte Wirkungen eines Texts auf Lesende beschrieben (z. B. Beschreibung von *Kohärenbildungsprozessen* bei Lenhard, 2013, oder bei Artelt et al., 2005).

In diesem Kapitel wird nun zusammengetragen und in Kategorien eingeteilt, welche Aspekte einer affektiven Textwirkung dabei zur Sprache kommen. Die Rekonstruktionsarbeit aus der Literatur verfolgte also den Anspruch, möglichst ausschöpfend zu erfassen, mit welchen subjektiven Eindrücken Lesender die Verständlichkeit eines Texts in der Theorie verbunden ist. Dies führte zu zwölf Kategorien, die nicht überschneidungsfrei sind und bei denen an vielen Stellen eine Paraphrasierung der in der Literatur genannten Originalbegriffe vorgenommen wurde. Zum Beispiel wurde Göpferichs (2002) *Perzipierbarkeit* hier unter *Lesbarkeit* gefasst und die *Komplexität* von Lutz (2015) genauso wie die *Simplizität* von Göpferich werden zur *Sprachlichen Einfachheit*. Im Folgenden werden die einzelnen Kategorien der Reihe nach ausführlich beschrieben<sup>4</sup>. In Tabelle 5.2 auf Seite 86 sind die Fließtexte der Beschrei-

<sup>4</sup>Für Erläuterungen zum sprachwissenschaftlichen beziehungsweise lesepsychologischen Vokabular, das vereinzelt in den Beschreibungen vorkommt, sei auf Abschnitt 6.2.2 beziehungsweise Unterkapitel 3.2 verwiesen.

bungen zu stichpunktartigen Aussagen kondensiert. Diese Aussagen waren die Grundlage für die Formulierung von Fragebogen-Items, anhand derer die affektive Textwahrnehmung der Physikschrüler\*innen in der empirischen Studie gemessen wurde (siehe Kapitel 6 und insbesondere Abschnitt 6.3.3).

### **Lokale Kohärenz**

Ein Text wirkt lokal umso kohärenter auf Lesende, je einfacher und direkter die Verbindung von einzelnen Wörtern zu sinntragenden Wortgruppen (Propositionen) und von benachbarten Satzteilen und Sätzen zu verbundenen Aussagen gelingt. Dazu werden ausschließlich im Text enthaltene Informationen (Semantik und Syntax) genutzt. Es sind weder inhaltliches Vorwissen noch Inferenzen nötig (Lenhard, 2013, S. 15, S. 18f.). Die lokale Kohärenz ist damit eine Texteigenschaft, die beschreibt, wie zugänglich den Lesenden die phrasen-, satz- und satzgruppenbezogene Tiefenstruktur hinter der Kodierung ist – wie klar also insbesondere logische Bezüge (Gründe, Folgen, Gegensätze,...) und relative Bezüge (Zuordnung von Bezugswörtern) sind (Artelt et al., 2005, S. 24).

### **Globale Kohärenz**

Ein Text wirkt global umso kohärenter auf Lesende, je klarer die Lesenden einzelne Informationen, die hinsichtlich ihrer Position im Text weit voneinander entfernt liegen, aufeinander beziehen können (Starauschek, 2006, S. 130). Es geht also darum, wie zugänglich die Einteilung des Texts in Teilthemen ist und wie ersichtlich die Relationen zwischen den Teilthemen sind (Artelt et al., 2005, S. 24). Solche Relationen können beispielsweise erkannte Ähnlichkeiten, Unterschiede, Inklusionen und Spezifizierungen sein (Groeben, 1978, S.49 f.). Ein Text unterstützt die globale Kohärenzbildung, indem die inhaltliche Strukturierung explizit gemacht wird und indem inhaltliche und optische Strukturierung stimmig zueinander sind. Auf diese Art lassen sich erkannte, separate Inhaltseinheiten mental verdichten und zu einem schlüssigen Gesamtbild zusammensetzen (Lenhard, 2013, S. 15 und S. 20).

### **Informationsdichte**

Die Informationsdichte ist nach Göpferich (2002, S. 163 ff.) das sprachliche Merkmal, welches bestimmt, wie prägnant ein Text wirkt (inhaltlich sei die Prägnanz durch die Reichhaltigkeit der mitgeteilten Informationen im Verhältnis zu den Mitteilungsabsichten bestimmt). Sie lasse sich technisch als *Zeichenaufwand pro Information* fassen. Alternativ kann eine hohe Informationsdichte auch als Abwesenheit von Redundanzen (Kernen & Riss, 2012, S. 3) und Füllwörtern (Langer, Schulz von Thun & Tausch, 2019, S. 56) betrachtet werden. Diese technischen Auffassungen werden jedoch aufgeweicht, wenn man beachtet, dass der gleiche Text je nach inhaltlichem Vorwissen der Adressat\*innen oder auch je nach bestehenden stilistischen Konventionen unterschiedlich dicht wirken kann (Göpferich, 2002, S. 167). Die Informationsdichte manifestiert sich also, wenn Lesende mit dem Text interagieren und ist

damit auch als Kategorie der affektiven Textwahrnehmung zu betrachten. Besonders für naturwissenschaftliche Texte ist eine hohe Informationsdichte charakteristisch. Sie komme mitunter durch die Verwendung von Fachbegriffen, Ellipsen und komplexen Satzglieder zustande (Fang, 2006, S. 493, 505).

### **Sprachliche Einfachheit**

Die (sprachliche) Einfachheit ist eine der Verständlichkeitsdimensionen im Hamburger Modell von Langer, Schulz von Thun und Tausch (2019) (das Modell ist auf Seite 53 beschrieben) und zählt somit zu den relevanten Kategorien hinsichtlich der Beurteilung von Texten durch Lesende. Und auch andere Publikationen rund um die Textverständlichkeit nehmen Bezug auf die sprachliche Einfachheit. So beschreibt beispielshalber Göpferich, 2002, S. 176, dass sich die sprachliche Einfachheit („Simplizität“) anhand verschiedener Aspekte der Kodierung beeinflussen lasse. Sie sei dadurch losgelöst von der Einfachheit/Kompliziertheit des Inhalts zu betrachten. Es gehe also darum, ob die sprachliche Darstellung der inhaltlichen Sachverhalte einfach oder kompliziert wirkt. Als textseitige Einflussfaktoren für die sprachliche Einfachheit werden in der Literatur sowohl Aspekte der Wortwahl (z. B. *Geläufigkeit der Termini*, bei Lutz, 2015, S. 246, oder *Konkretheit vs. Abstraktheit in den Formulierungen* bei Langer, Schulz von Thun und Tausch, 2019, S. 23) als auch der Satzstrukturen (*Hypotaxe vs. Parataxe* oder *Regressivitätsgrad* bei Göpferich, 2002, S. 177) genannt.

### **Lesbarkeit**

Ob ein Text lesbar wirkt, ist textseitig maßgeblich durch äußere Gestaltungsmerkmale auf der Textoberfläche bestimmt. Bei der Lesbarkeit geht es um physikalische und physiologische Vorgänge bei der Texterkennung und nicht um inhaltliche Verarbeitungsprozesse. Es gibt daher Ansätze, die Lesbarkeit vollkommen objektiv anhand von Indizes zu erfassen, die auf zählbaren Wort- und Satzeigenschaften (Silbenanzahl, Zahl der Wörter pro Satz etc., siehe Seite 52) beruhen. Göpferich (2002, S. 186) nennt diese Dimension der Textverständlichkeit „Perzipierbarkeit“ und ordnet ihr noch weitere Eigenschaften zu, die sich auf die formale Textstruktur beziehen (z. B. Auflistungen anstelle von Fließtext). Einen weiteren Aspekt der Lesbarkeit macht die Syntax aus, die ursächlich dafür sein kann, dass ein „vorwärts gerichtetes Lesen“ nicht ohne Weiteres möglich ist, sondern Sprünge nötig sind, um komplexe Satzstrukturen zu erfassen. Die Auswirkungen solcher Textoberflächen- und Syntaxmerkmale auf die Lesbarkeit tragen durchaus subjektiven Charakter, sodass die Lesbarkeit als Facette der affektiven Textwahrnehmung betrachtet werden kann. Auch gemäß dem Hamburger Modell zeigen sich Aspekte der Lesbarkeit als relevant für Textbeurteilungen – allerdings werden sie nicht als eigene Dimension gefasst, sondern unter der Bezeichnung „Äußere Gliederung“ als Teil der *Gliederung und Ordnung* (Langer, Schulz von Thun & Tausch, 2019, S. 24) ins Modell aufgenommen.

### **Strukturelle Komplexität**

Die strukturelle Komplexität ist eine Dimension, die Bestandteil des Modells für Textschwierigkeit von Heine et al. (2018) ist. Darunter verstehen die Autor\*innen, wie kognitiv aufwendig die Textverarbeitung auf der Ebene der Wörter und Sätze für Lesende ist. Der Aufwand sei einem kognitiv-linguistischen Ansatz folgend dadurch bestimmt, wie viele Einzelelemente im Arbeitsgedächtnis gleichzeitig aktiv gehalten, sortiert und ineinander integriert werden müssen, um eine vollständige sinntragende Einheit zu bilden. Das heißt, mit einem hohen Grad an struktureller Komplexität gehen das Gefühl von Anstrengung beim Lesen, Probleme mit der grammatikalischen Aufschlüsselung sowie ein verringertes Lesetempo beziehungsweise stockendes Lesen einher.

### **Vorwissensaktivierung**

Vorwissen ist zunächst ein personenspezifisches Merkmal, welches entscheidend Einfluss auf den Aufbau eines mentalen Modells nimmt (Lenhard, 2013, S. 15, S. 23; die Rolle des Vorwissens für das verstehende Lesen wird ausführlich auch auf Seite 48 erläutert). Die Anforderung an Texte, adressatenorientiert zu sein (siehe z. B. die Ausführungen zur *Textpassung* von Frickel, 2018, oder die *Auftragsdaten* im Karlsruher Modell nach Göpferich, 2002), beinhaltet demnach, dass ein Text an den Vorwissensstand der Lesenden angepasst sein muss. Ein Text kann nicht alle Informationen bereitstellen, die für den Verstehensprozess nötig sind, vielmehr müssen Lesende *aktiv* ihr Vorwissen einbringen (siehe für Einzelheiten Seite 45 in Abschnitt 3.2.1). Die Literatur nennt allerdings mehrere Gestaltungsmerkmale, die dabei behilflich sein können, dass Lesende ihr Vorwissen aktivieren und für das verstehende Lesen nutzen. Dazu zählen beispielshalber Advance Organizer, Analogien oder explizite Elaborationen zur Verknüpfung neuer Informationen mit bestehendem Vorwissen (Artelt et al., 2005, S. 28; Lenhard, 2013, S. 30). Es wurde somit aus der Literatur die Kategorie der Vorwissensaktivierung als Facette der affektiven Textwahrnehmung extrahiert. Sie beschreibt, inwiefern ein Text das Gefühl bei Lesenden hervorruft, angemessene Anforderungen an ihr Vorwissen zu stellen und Bezüge zu diesem Vorwissen aufzuweisen.

### **Kognitive Gliederung**

Mit dieser Kategorie wird beurteilt, wie einleuchtend und geordnet die Textgliederung für Lesende wirkt. Sie ist eng verwandt mit der Facette der *Globalen Kohärenz* (siehe oben). Während sich jedoch ein Text, der global kohärent wirkt, vor allem durch klare *inhaltliche* Bezüge zwischen einzelnen Textstellen auszeichnet, bezieht diese Facette auch die Beurteilung der äußeren Makrostruktur mit ein. Das bedeutet, sie umfasst zusätzlich zur Nachvollziehbarkeit inhaltlicher Bezüge auch, inwiefern die Abfolge der einzelnen Textgliederungspunkte den Konventionen der Textsorte und damit den Erwartungen der Lesenden entspricht (Sander & Rosebrock, 2016, S. 147). Eine solche wäre z. B. die Organisation eines wissenschaftlichen Texts gemäß Einleitung, gefolgt von Theorie, dann Methode, dann Ergebnisse, dann Diskussion

und abschließend Ausblick.

### **Zugänglichkeit der Darstellungsintentionen**

Die Aussagekraft eines Texts geht oftmals über die Tatsachen, die an der Textoberfläche explizit gemacht werden, hinaus. Einen Text zu verstehen heißt, ihn zu deuten. Diese Aufgabe ist zwar vor allem bei literarischen Texten von großer Relevanz, da sich deren Deutungsspielraum im Allgemeinen viel weiter erstreckt als der tatsächlich niedergeschriebene Handlungsrahmen. Entsprechend zählen Sander und Rosebrock (2016, S. 145) die „Anforderungen auf der Ebene der Identifikation der Darstellungsintentionen“ zu den Kategorien, welche die Komplexität literarischer Texte ausmachen. Dennoch ist es auch für das Lernen und Lehren anhand von Sachtexten eine relevante Frage, inwiefern den Schüler\*innen als Adressat\*innen eines Sachtextes ersichtlich ist, was der Text mit ihnen selbst und ihrer Lebenswelt zu tun hat. In diesem Sinne stellt die Kategorie der *Zugänglichkeit der Darstellungsintentionen* für diese Arbeit zur Verständlichkeit naturwissenschaftlicher Sachtexte eine Kategorie der affektiven Textwahrnehmung dar.

### **Einprägsamkeit**

Mit *Einprägsamkeit* bezeichnet Groeben (1978), inwiefern ein Text eine möglichst hohe Behaltensleistung beim verstehenden Lesen unterstützt. Groeben (1978, S. 80 f.) zählt das Behalten zu den Erfolgskriterien verstehenden Lesens und betrachtet es als quantitativ messbar, da sie eine Post-Prozessgröße sei, die in Leistungstests abbildbar ist. Ihre textseitige Bestimmungsgröße, die Einprägsamkeit, ist jedoch auch durch einen affektiven Charakter geprägt. Inwiefern der Text den *Eindruck* auf Lesende erweckt, einprägsam zu sein, wird somit zu den Kategorien der affektiven Textwahrnehmung gezählt.

### **Unterhaltsamkeit**

Kernen und Riss (2012, S. 10) beschreiben Wirkungen, die anhand von erzählenden Elementen und anhand einer Bebilderung eines Texts zu erreichen seien: Es gehe dabei unter anderem um Stimmungswidrigkeit und um die Auflockerung von „Bleiwüsten“. Der Text könne durch solche Elemente also *unterhaltsamer* wirken. Dabei sei jedoch eine zu große Fülle an Erzähl-Elementen beziehungsweise Bildern ungünstig, da Wesentliches von Unwesentlichem unterscheidbar bleiben muss. Zu einem umfassenden Bild der affektiven Textwahrnehmung zählt also auch die Frage, inwiefern ein Text *unterhaltsam* wirkt.

### **Motivation**

Es erscheint sofort einleuchtend, dass eine motivierte Haltung der Lesenden günstig für Verstehensprozesse ist. In Krapps (1993) Strukturmodell für die Lernmotivation wird die *aktualisierte Lernmotivation*, die sich unter anderem aus Bedingungen der Lernsituation und des Lerngegenstands ergibt, als bestimmend für kognitive und emotionale Prozesse der Lernenden

während der Lernhandlung bezeichnet. Speziell für das Lernen anhand von geschriebenem Text beschreibt Groeben (1982, S. 6) das Lesen aus psychologischer Sicht als seriell ablaufende Prozesse: „[Z]unächst wird eine Motivation zum Lesen aufgebaut, die sich zu einem bestimmten Zeitpunkt in vorhandenen Leseinteressen manifestiert; auf dem Hintergrund der Leseinteressen werden dann Texte gelesen, d. h. verstehend verarbeitet [...]; verarbeitete, d. h. verstandene Texte können dann eine bestimmte (nach Textsorten verschiedene) Wirkung haben“. Die *Motivation* ist daher eine oft genannte Dimension der Textverständlichkeit. Entsprechend zählt die Frage, inwiefern ein Text auf Lesende motivierend wirkt zu den hier zu betrachtenden Kategorien der affektiven Textwahrnehmung. Da in der Literatur oftmals Zusammenhänge zwischen einem erzählerischen (nicht-sachlich-nüchternen) Stil und einer motivierenden Wirkung postuliert oder berichtet werden (Fang, 2006, S. 515; Göpferich, 2002, S. 170 f.), ist besonders vor dem Hintergrund des typischerweise unpersönlichen Charakters von naturwissenschaftlicher Fachsprache (siehe dazu Seite 36) interessant, inwiefern Lesende Texte mit physikalischem Thema je nach sprachlicher Gestaltung als motivierend empfinden können. Dazu zählen überraschende, unterhaltende, konfliktevozierene Lesemomente, die zum Nach- und Weiterdenken anregen (Göpferich, 2002; Groeben, 1978).

Wie zu Beginn dieses Unterkapitels 5.2 erwähnt, sollen die beschriebenen Kategorien möglichst vollständig abdecken, welche affektiven Aspekte im Kontext der Wirkung von Texten auf Lesende in der Literatur zur Sprache kommen und somit vermutlich in Verbindung mit dem Gesamteindruck zur Verständlichkeit des Texts stehen. Sie erheben nicht den Anspruch, überschneidungsfrei zu sein. Es bleibt damit eine Aufgabe dieser Arbeit, aus der Zusammenstellung der Kategorien eine auf empirischer Evidenz basierende Struktur für die affektive Textwahrnehmung zu entwickeln. Dadurch lässt sich ein Grundstein für die Unterscheidung unterschiedlicher Facetten der affektiven Textwahrnehmung legen, die im hier vorliegenden Fall auf die Lektüre eines physikalischen Sachtexts durch Schüler\*innen der Mittelstufe bezogen sind.

Die methodische Herangehensweise an diese sowie an die übrigen Forschungsfragen dieser Studie wird im folgenden Kapitel beschrieben.

<p>Lokale Kohärenz</p> <hr/> <ul style="list-style-type: none"> <li>• Einzelne Wörter lassen sich zu sinntragenden Wortgruppen verbinden.</li> <li>• Logische Bezüge zwischen aufeinanderfolgenden Sätzen sind klar.</li> <li>• Relative Bezüge zwischen aufeinanderfolgenden Sätzen sind klar.</li> <li>• Logische Bezüge innerhalb eines Satzes sind klar.</li> <li>• Relative Bezüge innerhalb eines Satzes sind klar.</li> </ul>	<p>Globale Kohärenz</p> <hr/> <ul style="list-style-type: none"> <li>• Der Text folgt einem roten Faden.</li> <li>• Der Text ist inhaltlich nachvollziehbar gegliedert.</li> <li>• Die einzelnen Abschnitte haben jeweils ein klares Thema.</li> <li>• Die Informationen einzelner Abschnitte lassen sich aufeinander beziehen.</li> <li>• Die einzelnen Abschnitte lassen sich zu einem schlüssigen Gesamtbild des Hauptthemas zusammensetzen.</li> </ul>
<p>Informationsdichte</p> <hr/> <ul style="list-style-type: none"> <li>• Einzelne Informationen kommen schrittweise, eine nach der anderen.</li> <li>• Einzelne Informationen kommen nicht zu dicht hintereinander.</li> <li>• Einzelne Informationen werden ausreichend erklärt.</li> <li>• Einzelne Informationen werden mit anderen Worten wiederholt erklärt.</li> <li>• Unklare Begriffe werden ausreichend erklärt.</li> </ul>	<p>Sprachliche Einfachheit</p> <hr/> <ul style="list-style-type: none"> <li>• Im Text werden geläufige Wörter verwendet.</li> <li>• Die Bedeutungen der Wörter im Text sind eindeutig.</li> <li>• Die Sätze sind klar in ihrer Struktur.</li> <li>• Die Sätze sind nicht verschachtelt.</li> <li>• Die Formulierungen sind konkret und nicht abstrakt.</li> <li>• Unklare Begriffe werden ausreichend erklärt.</li> </ul>
<p>Lesbarkeit</p> <hr/> <ul style="list-style-type: none"> <li>• Die Schriftzeichen lassen sich visuell schnell erfassen.</li> <li>• Die Gliederung des Texts wird augenscheinlich deutlich.</li> <li>• Der Text wirkt auf den ersten Blick übersichtlich.</li> <li>• Das Lesetempo ist zügig.</li> <li>• Die Sätze sind bei vorwärts gerichtetem Lesen erfassbar.</li> </ul>	<p>Strukturelle Komplexität</p> <hr/> <ul style="list-style-type: none"> <li>• Den Text zu lesen erfordert volle Konzentration.</li> <li>• Das Lesetempo ist gering.</li> <li>• Die Satzstrukturen sind komplex.</li> <li>• Die Verarbeitung der Sätze erfordert große Anstrengung.</li> <li>• Ein unterbrechungsfreies Lesen von Satzanfang bis -ende ist kaum möglich.</li> <li>• Zusammengehörige Wortteile sind weit im Satz verteilt.</li> <li>• Bezüge zwischen zusammengehörigen Wortteilen sind nicht offensichtlich.</li> <li>• Der Lesefluss gerät oft ins Stocken.</li> <li>• Beim Lesen ist es nötig, mit den Augen vor und zurück zu hüpfen.</li> </ul>
<p>Vorwissensaktivierung</p> <hr/> <ul style="list-style-type: none"> <li>• Der Text nimmt Bezug auf im Voraus bekannte Dinge.</li> <li>• Der Text knüpft mit neuen Informationen an Bekanntes an.</li> <li>• Der Text setzt nur tatsächlich vorhandenes Wissen voraus.</li> <li>• Der Text ruft inhaltsnahes Vorwissen in Erinnerung.</li> <li>• Der Text evoziert Verknüpfungen zu inhaltsfernen Wissensbeständen (themenfremde Analogien).</li> <li>• Die Textgliederung entspricht den Konventionen beziehungsweise Erwartungen.</li> </ul>	<p>Kognitive Gliederung</p> <hr/> <ul style="list-style-type: none"> <li>• Die Gliederung des Texts in seine Abschnitte wirkt sinnvoll.</li> <li>• Die Reihenfolge der Abschnitte ergibt Sinn.</li> <li>• Es fällt leicht, die Beziehungen zwischen einzelnen Abschnitten zu erkennen.</li> <li>• Die Textgliederung entspricht den Konventionen beziehungsweise Erwartungen.</li> <li>• Einzelne Informationen kommen schrittweise, eine nach der anderen.</li> <li>• Der Text ist inhaltlich nachvollziehbar gegliedert.</li> <li>• Die Informationen einzelner Abschnitte lassen sich aufeinander beziehen.</li> <li>• Die einzelnen Abschnitte lassen sich zu einem schlüssigen Gesamtbild des Hauptthemas zusammensetzen.</li> </ul>

Tabelle 5.2.: Kategorien der affektiven Textwahrnehmung (Teil 1; Forts. n. Seite).

## 5. Rekonstruktion des Konstruktraums für Textverständlichkeit

<u>Zugänglichkeit der Darstellungsintentionen</u>	<u>Einprägsamkeit</u>
<ul style="list-style-type: none"><li>• Die Aussagekraft des Texts über den eigentlichen Inhalt hinaus ist klar.</li><li>• Es wird klar, warum es sich lohnt, den Text zu lesen.</li><li>• Es wird klar, inwiefern sich die Textinformationen auf die eigene Lebenswelt beziehen lassen.</li></ul>	<ul style="list-style-type: none"><li>• Die Informationen im Text lassen sich gut im Kopf behalten.</li></ul>
<u>Unterhaltsamkeit</u>	<u>Motivation</u>
<ul style="list-style-type: none"><li>• Der Text wirkt unterhaltsam, nicht langweilig.</li><li>• Das Lesen des Texts macht Freude.</li><li>• Der Text wirkt stellenweise humorvoll.</li><li>• Der Text birgt überraschende Stellen.</li></ul>	<ul style="list-style-type: none"><li>• Der Text spricht Lesende emotional an.</li><li>• Lesende fühlen sich vom Text persönlich angesprochen.</li><li>• Das Lesen des Texts macht Freude.</li><li>• Der Text wirkt unterhaltsam.</li><li>• Der Text regt zum Nachdenken an.</li><li>• Das Lesen des Texts bringt Faszination.</li><li>• Der Text weckt Interesse für sein Thema.</li><li>• Der Text erzeugt gedankliche Konflikte.</li><li>• Der Text birgt überraschende Stellen.</li></ul>

Tabelle 5.2.: Kategorien der affektiven Textwahrnehmung, die aus der Literatur extrahiert wurden. Die Grundlage dafür besteht in Erwähnungen, Berichten und Befunden zu verständlichkeitsrelevanten Wirkungen von Texten auf Lesende in der Theorie der Textverständlichkeit.

*It doesn't matter how beautiful your theory is,  
it doesn't matter how smart you are.  
If it doesn't agree with experiment, it's wrong.*

---

Richard P. Feynman

## 6 | METHODE DER EMPIRISCHEN STUDIE: MATERIAL UND DURCHFÜHRUNG

Bis zu diesem Punkt wurde die Motivation für die Fragestellungen dieser Arbeit ausführlich dargelegt sowie die Rekonstruktion der zugrundeliegenden Theorie erläutert. Es schließt sich nun die Beschreibung des nächsten Schritts an, um die Forschungsfragen nach den Wirkungen variiert sprachlicher Gestaltung auf die Behaltensleistung und die affektive Textwahrnehmung von Schüler\*innen zu beantworten. Der Schritt besteht in der Planung und Durchführung einer Datenerhebung an Schulen. Das einleitende Zitat von Richard P. Feynman bringt die Bedeutsamkeit dieses Forschungsschritts pointiert zum Ausdruck. Anhand einer empirischen Studie zum Lernen mit Texten im Physikunterricht sollen also evidenzbasierte, belastbare Erkenntnisse zur verständlichen Textgestaltung gewonnen werden.

Auf einen groben Überblick über die empirische Studie (Unterkapitel 6.1) folgen zwei Unterkapitel, die sich dem verwendeten Material zuwenden: Die Texte werden in Unterkapitel 6.2 beschrieben, die Testinstrumente in Unterkapitel 6.3. Die statistischen Mittel für die Datenauswertung sind Gegenstand von Unterkapitel 6.4. Unterkapitel 6.5 schließlich wendet sich der Datengewinnung zu, indem die Durchführung der Studie an Schulen sowie die Stichprobenszusammensetzung erläutert werden.

### 6.1 Gesamtkonzeption der empirischen Studie

Das zentrale Interesse der empirischen Studie ist es, anhand von authentischem Datenmaterial Einblicke in die Wirkmechanismen zwischen der sprachlichen Gestaltung eines physikalischen Sachtexts auf der einen Seite und der affektiven Textwahrnehmung sowie der Behaltensleistung Lesender auf der anderen Seite zu gewinnen (die Begriffe werden ausführlich in Kapitel 4 auf Seite 64 geklärt). Statistisch gesprochen bedeutet das, dass

- die unabhängige Variable in sprachlichen Gestaltungsmitteln im Text besteht
- und die abhängigen Variablen erstens kognitionsbezogene Texteffekte (Behaltensleistung) und zweitens die affektive Textwahrnehmung (subjektiver Eindruck) sind.

In Abbildung 6.1 ist dieses Gesamtkonzept schematisch veranschaulicht.

## 6. Methode der empirischen Studie: Material und Durchführung

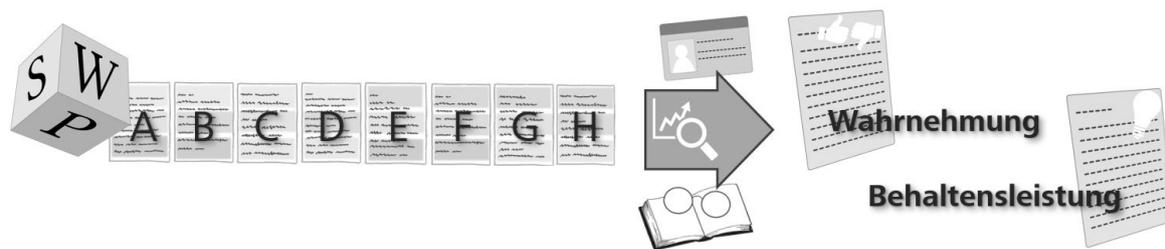


Abbildung 6.1.: Gesamtkonzept der empirischen Studie. Der Würfel bildet die unabhängige Variable ab: Die drei sprachlichen Merkmale „Wortwahl“, „Satzstrukturen“ und „Personalisierungen“ (genauere Erläuterungen im Fließtext) wurden unabhängig voneinander variiert und jeweils in zwei verschiedenen Ausprägungsformen in den Texten realisiert. Es ergibt sich daher ein  $2 \times 2 \times 2$ -Muster, das zu acht verschiedenen Texten führt. Diese acht Texte wurden von Schüler:innen aus achten und neunten Klassen gelesen und bewertet (abhängige Variable 1, dargestellt durch Text mit Daumen-Zeichen: affektive Textwahrnehmung). Die Schüler\*innen absolvierten darüber hinaus einen Lesetest (Moderatorvariable, dargestellt durch Buch mit Lesebrille: Lesefähigkeiten) und einen Wissenstest im Pre-Post-Design (abhängige Variable 2, dargestellt durch Text mit Glühbirne: Lernzuwachs im Sinne einer Behaltensleistung).

Um Gender-, Schular- und Jahrgangsstufeneffekte (die sogenannten *Moderatorvariablen* in dieser Studie) kontrollieren zu können, werden diese Personendaten miterhoben. Ein Lesetest (LGVT 5-12+, siehe Abschnitt 6.3.1 unten) dient der Kontrolle von Einflüssen durch die individuellen Lesefähigkeiten (der Begriff wird in Unterkapitel 3.1 genauer erläutert). Dies fügt der statistischen Struktur der Studie diverse Moderatorvariablen in Form von Personeneigenschaften hinzu<sup>1</sup>.

Um die unabhängige Variable in möglichst analytischer, kontrollierter Weise variieren zu können, wurden drei Gruppen sprachlicher Mittel ausgewählt, die erstens fachsprachliche Relevanz haben und sich zweitens möglichst weitestgehend unabhängig voneinander verändern lassen. Auf Grundlage der theoretischen Vorarbeit (siehe Unterkapitel 5.1) erweisen sich hinsichtlich dieser beider Kriterien die drei Gruppen

- Wortwahl / morphologische Auffälligkeiten auf Wortebene, im Weiteren häufig abgekürzt mit **w** bzw. **W**,
- Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene, Abkürzung **s** bzw. **S**,
- und erzählende und personalisierende Elemente, Abkürzung **p** bzw. **P**

als besonders geeignet. Die fachsprachliche Relevanz zeigt sich in den vielen Überschnei-

<sup>1</sup>Es liegt auf der Hand, dass weitere Einflussgrößen als Moderatorvariablen in Frage kommen, z. B. Tagesform, individuelles Interesse, Professionskompetenzen der Lehrkraft, etc. (auf eine wissenschaftliche Definition und Erläuterung dieser Größen wird hier verzichtet, da sie den Rahmen sprengen und nicht zweckgemäß wären). Davon wurde jedoch aus mehreren Gründen abgesehen: Die Studienökonomie, die sicherlich entscheidend für Teilnahmebereitschaft die Schüler\*innen und Lehrkräfte ist, würde sinken. Der Datenschutz würde bei der Verarbeitung personenbezogener Daten drastisch aufwändiger. Die Auswertung mit zugehörigen Ergebnisberichten und -deutungen würden zunehmend komplexer und zeitintensiver und gerieten an die Grenzen der Möglichkeiten dieser Arbeit.

dungen zwischen den sprachlichen Mitteln, die diesen Gruppen zuzuordnen sind, und den Charakteristika der naturwissenschaftlichen Fachsprache (siehe Abschnitt 2.2.2). In den Abkürzungen entspricht ein Großbuchstabe jeweils einer hohen beziehungsweise anspruchsteigernden Ausprägung der Variable, ein Kleinbuchstabe entsprechend einer schwachen beziehungsweise anspruchsmindernden Ausprägung. Es sei hier noch einmal betont, dass „anspruchsmindernd“ nicht mit „besser“ im Sinne von *lernwirksamer* gleichzusetzen ist (siehe dazu auch Unterkapitel 3.3). Die detaillierten Gruppenzusammensetzungen, die den drei Oberbegriffen *Wortwahl*, *Satzstrukturen* und *Personalisierung* ihre konkreten sprachlichen Mittel zuschreiben und in Unterkapitel 5.1 expliziert sind, dienen jeweils der Realisierung in den Texten. Das Vorgehen zur Erstellung der Texte – inhaltlich und sprachlich – wird im Folgenden genau beschrieben.

## 6.2 Texterstellung

Um sich empirisch der Frage zu nähern, inwiefern die sprachliche Gestaltung eines physikalischen Sachtexts Einfluss auf die Behaltensleistung und auf die affektive Textwahrnehmung durch Schüler\*innen nimmt, wurde geeignetes Textmaterial eigens für diese Studie erstellt. Das Vorgehen war dabei, einen physikalischen Sachtext zum Thema *Die elektrische Spannung* zu entwerfen (siehe Abschnitt 6.2.1), der gemessen an den Empfehlungen von Göpferich (2002), Langer, Schulz von Thun und Tausch (2019) und Groeben (1978) als verständlich gilt und diesen Sachtext im Anschluss zu variieren (siehe Abschnitt 6.2.2). Inwiefern diese Modifikationen sich auf den Lernerfolg im Sinne der affektiven Textwahrnehmung und der Behaltensleistung auswirken, war der Untersuchungsgegenstand dieser Studie. Es sei hier daran erinnert, dass eine Unterscheidung von *maximaler Einfachheit* und *optimaler Passung* eines Texts für Zweck und Lesende wichtig ist (siehe Seite 51). Die simple Vorstellung, dass Texte für das Lernen besser werden, indem man den Anspruchsgrad gewisser Textmerkmale mindert, bleibt zu überprüfen (siehe die fachspezifischen Forschungsergebnisse in Unterkapitel 3.4).

Unter physikalischen Sachtexten versteht Baurmann (2006) Texte, die Fakten vermitteln und der Wissenserweiterung der Lesenden dienen. Sie seien an die kognitiven und sprachlichen Fähigkeiten der Zielgruppe angepasst. Der Inhalt könne deskriptiver, argumentativer und narrativer Art sein und wird in einem Umfang von wenigen Sätzen bis hin zu ganzen Sachbüchern wiedergegeben. Die Sprache sei unter anderem durch Präzision, wenig Redundanz, fachsprachliche Elemente, Neigung zum Nominalstil und zu elliptischem Satzbau geprägt und das graphische Beiwerk umfasse Listen, Tabellen, Grafiken u. Ä. Der Ursprungstext, aus dem die Sachtexte für die an der Erhebung teilnehmenden Schüler\*innen hervorgingen, lässt sich in Bezug auf die Sachtext-Eigenschaften nach Baurmann (2006) und in Bezug auf die Auftrags- und Produktionsdaten nach Göpferich (2002) (siehe dazu im Detail Seite 55 in Unterkapitel 3.3) wie folgt beschreiben:

- Die kommunikative Funktion des Texts ist es, Wissen zu einer bisher unbekanntem

physikalischen Größe, nämlich der elektrischen Spannung, zu vermitteln. Der Text richtet sich an Schüler\*innen der achten oder neunten Klasse an Gymnasien oder Realschulen, die zuvor im Unterricht noch nichts über die elektrische Spannung gehört haben.

- Die Anpassung an die sprachlichen Fähigkeiten der Lesenden wurde anhand des sprachlichen Inventars im Text (und auch in den Varianten) vorgenommen. Eine Überprüfung erfolgte mithilfe des Textanalysetools RATTE<sup>2</sup>. Der Sprachschwierigkeitsindex *gSmog* (siehe Unterkapitel 3.3 und Mc Laughlin, 1969, für das *SMOG*-Grading beziehungsweise Bamberger und Vanacek, 1984, S. 58 f., für die deutschsprachige Anpassung der Formel) ist so normiert, dass sein Zahlenwert der Schulklassenstufe entspricht, für die der Text passend ist. Im hier vorliegenden Fall rangieren die *gSmog*-Indizes der acht Textvarianten zwischen 5.60 und 9.53, sodass die sprachlichen Fähigkeiten von Schüler\*innen der achten und neunten Klasse je nach Variante unterschritten oder gut getroffen werden. Die Ergebnisse der RATTE-Analyse sind in Tabelle 6.4 auf Seite 100 dargestellt. Im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> sind darüber hinaus die vollständigen Outputs des Analyseprogramms als Screenshots abgebildet.
- Indem die Inhalte und Begriffe im Text an bekannten Alltagsphänomen beziehungsweise am Vorunterricht ausgerichtet sind, werden die kognitiven Fähigkeiten im Sinne des Vorwissens berücksichtigt. Entsprechend dem Curriculum sind der Begriff des elektrischen Leiters, der elektrischen Stromstärke und der elektrischen Ladung bereits bekannt und die Schüler\*innen sind in der Lage, einfache elektrische Schaltkreise aufzubauen und zu erklären (ISB, 2022b; ISB, 2022c).
- Der Text ist vorwiegend deskriptiver Natur, indem er die elektrische Spannung als physikalische Größe einführt und beschreibt. Vereinzelt Textstellen mögen narrativ wirken, um Bezüge zur Lebenswelt oder bildhafte Vergleiche herzustellen und die abstrakte Größe der *elektrischen Spannung* konkret zu machen (z. B. „Auf Batterien findest du Angaben wie 1,5 V oder 6 V, wobei das V für die Einheit Volt steht“ oder „Stell dir die elektrische Spannung als ‚Tritt in den Hintern‘ für den elektrischen Strom vor“<sup>3</sup>).
- Der Umfang des Ursprungstexts und damit auch aller sprachlichen Varianten richtet sich nach einer typischen Schulbuchseite.
- Die obenstehende Beschreibung der sprachlichen Charakteristika von Sachtexten nach Baurmann (2006) ist sicherlich hinsichtlich der Präzision und der fachsprachlichen Elemente zutreffend auf den Ursprungstext und damit auf alle Studientexte. In Hinblick auf Baurmanns übrige sprachliche Merkmale entsprechen einige Textversionen der Beschreibung und andere nicht (siehe dazu ausführlich Abschnitt 6.2.2). Denn auf der Verwendung fachsprachlicher Ausdrucksweisen und ihrer Variation liegt bei den Studientexten ein besonderes Augenmerk: Einerseits sind fachsprachliche Wendungen Gegenstand der sprachlichen Variation, da die Wortwahl in ihrer vermittelnden, kommunikativen Funkti-

---

<sup>2</sup>Wild, J. & Pissarek, M. (o. J.). Ratte. Regensburger Analysetool für Texte. Version 2.0.

<https://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/downloads/ratte/index.html> (Zugriff: 23.03.2023)

<sup>3</sup>siehe Textversion D bzw. F im digitalen Anhang unter <http://doi.org/10.5283/epub.58778>

on einer der Untersuchungsgegenstände ist. Andererseits dürfen diejenigen Ausdrücke und Sprechweisen, die direkt mit der elektrischen Spannung in Verbindung stehen und somit als zu erlernen gelten, nicht variiert werden (siehe auch die Funktion der Sprache als Konstruktionsmittel des Denkens und als eigenständiger Lerngegenstand; Unterkapitel 2.1). Es wurde also im Vorfeld genau festgelegt, in welcher sprachlichen Umgebung ein Begriff von der elektrischen Spannung aufgebaut werden soll und an welchen Stellen fachsprachliche Ausdrücke für die Variation verwendet werden können. Dies wird im Detail in Abschnitt 6.2.2 beschrieben.

- Textbegleitende Grafiken kommen in dieser Studie bewusst nicht zum Einsatz, da die Gestaltung illustrierender Repräsentationen ein eigenständiges Forschungsgebiet mit vielen offenen Fragen ist (siehe dazu z. B. Berger & Staraschek, 2009). Insbesondere war wichtig zu verhindern, dass die Auswirkungen der sprachlichen Variation von Effekten der Bildrezeption und des Bildverstehens sowie von den Herausforderungen durch Darstellungswechsel (Rau, 2017) überlagert werden.

Die konkrete Gestaltung der acht Textvarianten ist Gegenstand der folgenden beiden Unterkapitel. Zunächst wird im Kapitel zur inhaltlichen Konzeption erläutert, wie der Begriff der elektrischen Spannung im Sachtext didaktisch rekonstruiert ist; das heißt, welcher inhaltlichen Struktur die Studientexte folgen, anhand welcher Analogien und Beispiele die elektrische Spannung zugänglich gemacht wird etc. Anschließend geht das Kapitel zur Erstellung der Textvarianten darauf ein, wie ausgehend vom sprachlichen Code des Ursprungstexts acht sprachliche Varianten realisiert wurden.

### 6.2.1 Inhaltliche Konzeption der Texte

Die *inhaltliche* Konzeption von Texten im Allgemeinen und damit auch im speziellen Fall der Texterstellung für diese Studie steht nicht im Zentrum der hier beschriebenen Forschungsarbeit. Aus diesem Grund war eine pragmatische Vorgehensweise angemessen: Inhaltliche Entscheidungen und iterative Verbesserungen im Zuge der Texterstellung fußten auf den Ratschlägen erfahrener, sachkundiger Personen (Lehrkräfte, Physikdidaktiker\*innen, Physikstudierende). Das Karlsruher Modell (Göpferich, 2002, S. 155 ff.) wurde genutzt, um eine sinnvolle Abfolge von Entscheidungsschritten einzuhalten, welche insbesondere die Festlegung auf rahmengebende Eckdaten beinhaltet (siehe dazu auch Seite 89). Die praktischen Tipps von Reinhard Tausch in Langer, Schulz von Thun und Tausch (2019, S. 178 f.) unterstützten den Aufbau einer logischen Struktur und die Festlegung auf konkrete Inhalte. Im hier vorliegenden Fall wurde dies in Form einer reichhaltigen Mindmap erarbeitet, welche in die Textkonzeption mündete. Ein besonders zentraler Schritt der Textkonzeption war, sich auf konkrete Lernziele für den Text festzulegen. Diese lauten:

- Die Lernenden benennen analoge Beziehungen zwischen der elektrischen Spannung und der Höhendifferenz bei einem Gebirgsfluss.
- Die Lernenden verstehen die Batterie als Bauteil, das elektrische Spannung bereitstellt.

## 6. Methode der empirischen Studie: Material und Durchführung

- Die Lernenden unterscheiden die elektrische Spannung vom elektrischen Strom, indem sie diese als dessen Antrieb verstehen.
- Die Lernenden verbinden den Begriff der elektrischen Lage von einzelnen Punkten im Stromkreis mit der elektrischen Spannung.
- Die Lernenden verbinden die Begriffe „Pluspol“ und „Minuspol“ einer Batterie mit der elektrischen Spannung.
- Die Lernenden unterscheiden korrekte von falschen Sprechweisen in Verbindung mit der elektrischen Spannung.
- Die Lernenden ordnen der elektrischen Spannung das Formelzeichen  $U$  zu.
- Die Lernenden ordnen der elektrischen Spannung die Einheit *Volt* mit Abkürzung  $V$  zu.

Die Formulierung der Lernziele deutet bereits an, dass die elektrische Spannung in den Lesetexten anhand der Analogie zu einem Gebirgsfluss, in welchem das Wasser aufgrund der Höhendifferenz zwischen Berg und Tal zum Strömen gebracht wird, eingeführt wird. Die Gegenüberstellungen

- Elektrische Spannung und Höhendifferenz,
- Elektrisches Potenzial (im Text „Elektrische Lage“) und Höhe und
- Elektrischer Strom und Wasserstrom

ziehen sich konsequent durch den Text. Um das neue Konzept der elektrischen Spannung in Verbindung zu Bekanntem zu bringen, wird das Beispiel einer batteriebetriebenen Taschenlampe an den Textanfang gestellt. Damit soll begreiflich werden, inwiefern die elektrische Spannung Relevanz in der Lebenswelt der Lernenden hat.

Die elektrische Spannung soll also als Potentialdifferenz identifiziert werden und als Antrieb des elektrischen Stroms verstanden werden. Damit geht einher, dass die Spannung als Größe verstanden wird, die zwischen zwei Punkten im Stromkreis besteht, insbesondere zwischen zwei Polen einer Batterie. Diese Erkenntnis bildet auch die korrekte Sprechweise ab („Spannung liegt zwischen zwei Punkten an“ im Gegensatz etwa zu der fehlerhaften Sprechweise „Spannung haben“). Weitere Möglichkeiten, die elektrische Spannung zu definieren – etwa als Leistung pro Stromstärke oder als Energie pro Ladung – werden bewusst im Text nicht erwähnt.

An die Formulierung der Lernziele und die Wahl einer Analogie, anhand derer der elektrische Spannungsbegriff in Bekanntes eingebettet werden kann, schloss sich ein weiterer didaktischer Arbeitsschritt an, nämlich die Berücksichtigung typischer Fehlvorstellungen. Häufige Fehlvorstellungen aus dem Bereich der Elektrizitätslehre finden sich kompakt zusammengefasst bei Hopf, Schecker, Höttecke und Wiesner (2022, S. 34 ff.). Indem fachlich anschlussfähige sprachliche Umgebungen gewählt werden, soll ein fachlich korrekter Konzeptaufbau ermöglicht und in Gang gesetzt werden. Gängigen Fehlvorstellungen wird stellenweise auch konfrontierend begegnet, indem fehlerhafte Sprech- und Denkweisen explizit als falsch bezeichnet beziehungsweise direkt berichtigt werden („Spannung herrscht nicht an *einem* Punkt.“, „Auch ohne elektrischen Strom bleibt die elektrische Spannung bestehen.“).

Es wurde also ein *Ursprungstext* mit den genannten Inhalten entworfen, der einen Sachtext

nach den oben beschriebenen Kriterien darstellt. Ausgehend davon wurden acht Textvarianten mit jeweils unterschiedlicher sprachlicher Gestaltung erstellt. Dies wird im folgenden Abschnitt ausführlich dargestellt.

## 6.2.2 Erstellung der sprachlich variierten Textversionen

Der Erstellung der acht Textvarianten liegt das Grundprinzip der Effektmaximierung zugrunde: Ausgehend von einem Ursprungstext wurden erstens drei *möglichst unabhängige* sprachliche Merkmalsgruppen variiert und zweitens die Variationen in *möglichst stark ausgeprägter Weise* umgesetzt. Die theoretische Vorarbeit hatte zu dem Ergebnis geführt, dass sich die drei Gruppen **Wortwahl / morphologische Auffälligkeiten auf Wortebene**, **Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene** und **erzählende und personalisierende Elemente** einerseits besonders gut von anderen Gruppen sprachlicher Mittel abgrenzen lassen und andererseits naturwissenschaftsspezifische Aspekte aufweisen. Die Variation wird realisiert, indem für jede der drei Gruppen die zugeordneten sprachlichen Mittel (siehe Unterkapitel 5.1) in zwei verschiedenen Ausprägungsgraden in den Texten verwendet werden. Die Tabellen 6.1, 6.2 und 6.3 stellen für jede der drei Merkmalsgruppen konkrete Unterschiede übersichtlich dar. Auf diese Art und Weise wird also jede der drei Gruppen W, S und P jeweils einmal in einer hohen beziehungsweise anspruchsteigernden und einmal in einer niedrigen beziehungsweise anspruchsmindernden Ausprägung umgesetzt. Es entstehen damit  $2 \times 2 \times 2 = 8$  Textvarianten. Hinsichtlich des Inhalts und hinsichtlich nicht variiertes sprachlicher Merkmale sind die Varianten – im Rahmen unvermeidlicher, geringfügiger Abweichungen durch die Textoberflächenvariation – nahezu identisch mit dem Ursprungstext und damit auch miteinander. Für alle Textvarianten gilt, dass sie eine klar erkennbare Absatztrennung mit Zwischenüberschriften aufweisen, dass sie konfliktevozierende und aufmerksamkeitssteuernde Fragesätze genauso wie eine Vorstrukturierung (*Advance Organizer*) und eine Zusammenfassung beinhalten, dass genau ein Thema pro Absatz behandelt wird, dass logische Bezüge explizit gemacht werden und dass typographische Hervorhebungen verwendet werden. Die Textvarianten berücksichtigen damit die üblichen, konsensuellen Gütekriterien für verständliche Texte (siehe Unterkapitel 3.3). Im Folgenden wird beschrieben, inwiefern sie sich bezogen auf die einzelnen Gruppen *Wortwahl / morphologische Auffälligkeiten auf Wortebene*, *Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene* und *erzählende und personalisierende Elemente* voneinander unterscheiden.

### Variation W vs. w

Die Variation der Variable *Wortwahl / morphologische Auffälligkeiten auf Wortebene* betrifft die folgenden sprachlichen Mittel. Entsprechende Textbeispiele sind jeweils aus Tabelle 6.1 zu entnehmen<sup>4</sup>.

<sup>4</sup>Die vollständigen Textvarianten finden sich im digitalen Anhang unter <http://doi.org/10.5283/epub.58778>; Anhang A.2.2 enthält eine Tabelle mit allen W-Variationen (Tabelle A.1).

## 6. Methode der empirischen Studie: Material und Durchführung

- Verwendung beziehungsweise Vermeidung von Fachbegriffen, technischem/akademischem Wortschatz, mehrdeutigen Wörtern und Fremdwörtern.

Aus einer Gruppe von Synonymen wird derjenige Vertreter gewählt, der im alltäglichen Gebrauch üblich und kein Fremdwort ist sowie eine einzige Bedeutung hat (w) oder im Gegensatz dazu zum physikalischen Fachjargon zählt beziehungsweise dem technisch-akademischen Sprachregister entstammt und/oder ein Fremdwort und/oder mehrdeutig ist (Variante W). Hierbei sei einschränkend bemerkt, dass solche (fach)sprachlichen Wendungen, die explizit zum Lerngegenstand zählen, nicht variiert werden. Der Begriff der elektrischen Spannung wird in einer bewusst gewählten sprachlichen Umgebung in allen Texten gleichermaßen eingeführt (die elektrische Spannung als „Antrieb“ des elektrischen Stroms, die elektrische Spannung als „Unterschied hinsichtlich der elektrischen Lage“, Spannung „liegt zwischen zwei Polen an“ und wird „zwischen zwei Punkten gemessen“).

- Verwendung beziehungsweise Vermeidung von nicht bekannten/wenig geläufigen Wörtern.

Aus einer Gruppe von Synonymen wird ein bestimmter Vertreter gewählt, wobei das Auswahlkriterium die Geläufigkeit des Worts für die Adressat\*innen ist. Für die Variante w werden geläufigere Wörter gewählt als für Variante W.

- Verwendung beziehungsweise Vermeidung von zusammengesetzten und langen Wörtern. Die Ausprägung der Wortwahl-Variable lässt sich auch an der Wortlänge festmachen, indem einzelne Wörter zu einem zusammengesetzten Wort verbunden werden (Variante W) und aus Synonymgruppen dasjenige Wort mit der höheren (W) oder niedrigeren (w) Silbenzahl ausgewählt wird.

- Wortwiederholungen.

Gegenüber der Verwendung von Wortwiederholungen (Variante w) werden in der anspruchsteigernden Variante W mehrere Synonyme für den gleichen Begriff verwendet.

### Variation S vs. s

Um die Gruppe *Satzstrukturen / morphologische Auffälligkeiten auf Satzebene* zu variieren, werden die folgenden sprachlichen Mittel herangezogen. Beispiele sind wiederum in Tabelle 6.2 enthalten<sup>5</sup>.

- Satzlänge und Satzgefüge.

Ein parataktischer Satzbau (Aneinanderreihung) aus kurzen, eigenständigen Sätzen ohne bedeutungsleere Füllwörter wird in Texten der Variante s umgesetzt. Dem gegenüber steht der hypotaktischen Aufbau (hierarchische Verbindung) langer Hauptsatz-Nebensatz-Gefüge, die vermehrt bedeutungsleere Füllwörter enthalten, in Variante S.

- Verwendung beziehungsweise Vermeidung von Passivkonstruktionen.

Passivkonstruktionen verändern die Satzstrukturen in anspruchsteigernder Form (S), indem Verbalklammern entstehen und dabei der bedeutungstragende Bestandteil, näm-

<sup>5</sup>Die vollständigen Textvarianten finden sich im digitalen Anhang unter <http://doi.org/10.5283/epub.58778>; Anhang A.2.2 enthält eine Tabelle mit allen S-Variationen (Tabelle A.2).

	w	W
fachsprachliche/ mehrdeutige Ausdrücke, Fremdwörter	in Zahlen leuchten Lage Zeichen in der Form... schreiben	quantitativ Licht emittieren Niveau Symbol die Notation... verwenden
Geläufigkeit der Wörter	brauchen verbinden genauso ähnlich nötig	bedürfen koppeln gleichermaßen vergleichbar vonnöten
Wortlänge	Begriff nennen bestehen Drähte in der Lampe	Bezeichnung betiteln fortbestehen Lampendrähte
Wortwiederholungen	fließen Lage	fließen / strömen / sich bewegen / laufen Lage / Ort / Niveau / Punkt

Tabelle 6.1.: Sprachliche Variationen hinsichtlich der W-Variable (Wortwahl / morphologische Auffälligkeiten auf Wortebene). Es werden für jedes sprachliche Mittel, das zur Textvariation hinsichtlich der Wortwahl benutzt wurde, exemplarische Textstellen gegenübergestellt. Die Zuordnung von Textstellen zu sprachlichen Mitteln ist dabei nicht ein-eindeutig (eindeutig in beiden Richtungen). Eine Liste aller Modifikationen zwischen w- und W-Texten findet sich im Anhang A.2.2.

lich die Verbform im Partizip, am Satzende steht. Um den Anspruch hinsichtlich der Satzkonstruktionen zu vermindern (s) werden daher vorwiegend Aktivformen verwendet. Passivkonstruktionen können auch den Grad an Personalisierung beeinflussen, wenn kein Handlungsträger genannt wird (die sogenannte Deagentivierung wird bei den Personalisierungs-Variationen näher erläutert, siehe dazu auch Seite 36 in Abschnitt 2.2.2).

- Nominalisierungen.

Nominalisierungen tragen zu einer anspruchsteigernden Satzstruktur bei, indem bedeutungstragende Verben als Prädikat verwendet (Variante s) oder in Nominalform gebracht werden (Variante S). Die Nominalisierung ist wie die Passivkonstruktionen Teil der Deagentivierung (siehe auch Seite 36 in Abschnitt 2.2.2).

- Verbalklammern.

Wenn die Bestandteile eines Verbkomplexes<sup>6</sup> an getrennten Positionen im Satz stehen und somit weitere Satzteile umklammern, erhöht sich der Anspruch der Satzstruktur (S). Im Gegensatz zeichnet sich eine anspruchsmindernde Satzstruktur durch eingliedrige Verben aus.

- Variation des Regressivitätsgrads.

In regressiven Phrasen ist die entscheidende Information an den Schluss gestellt<sup>7</sup>, wodurch die Belastung des Arbeitsgedächtnisses steigt. In den hier vorzunehmenden Textvariationen ist vor allem das finite Verb für den Regressionsgrad entscheidend. Eine Stellung am Satzende entspricht der anspruchsteigernden Version S, eine frühe Nennung des Hauptverbs der Version s.

- Partizipialkonstruktionen.

Auch anhand von Partizipialkonstruktionen (S) beziehungsweise deren Vermeidung (s) wird der Anspruchsgrad der Satzstrukturen verändert. Partizipialkonstruktionen sind solche Satzbausteine, die sich aus zwei oder mehreren Bestandteilen zusammensetzen und in deren Bedeutungszentrum ein Partizip<sup>8</sup> steht.

- Mehrwortkomplexe.

Partizipialkonstruktionen und Nominalisierungen lassen sich häufig zu umfassenden, satzgliedwertigen Mehrwortkomplexen ausbauen. Diese Wortverbindungen, die sich aus zwei oder mehr Bestandteilen zusammensetzen und als Ganzes ein Satzglied darstellen (z. B. *die auf dem Gehweg dahinschlendernde junge Frau*), zählen zu den Merkmalen für anspruchsteigernde Satzstrukturen (S). Für die anspruchsgeminderte Variante s werden entsprechende Informationen als eigenständige Hauptsätze wiedergeben.

---

<sup>6</sup>Zu einem mehrgliedrigen Verbkomplex gehören diejenigen Bestandteile eines Satzes, die dem Verb zugeordnet sind. Das sind ein finites Verb und zugehörige infinite Teile wie weitere Verben, Subjunktionen,... (siehe den Eintrag zu *Satzklammern* in Glück und Rödel, 2016). Beispiele sind: „kann... kochen“, „wird... bearbeiten“ und „schreibt... ab“.

<sup>7</sup>Vergleiche z. B. die regressive deutsche Sprache mit der progressiven französischen:  
*Das auffallend lange, besonders schön glänzende, im Sonnenlicht schimmernde Haar* gegenüber  
*Les cheveux étonnamement longs, particulièrement brillants et chatoyants au soleil.*

<sup>8</sup>Bildung aus einem Verb mit Endung {...-(en)d} oder mit {(ge-)...-en} oder {(ge-)...-t} (siehe den Eintrag zu *Partizip* in Glück und Rödel, 2016). Beispiele sind „schleichend“ oder „geschenkt“.

	s	S
Satzgefüge	Dieser Text erklärt .... Dazu betrachtet man ....	In diesem Text wird erläutert, was man sich unter ... vorstellen kann, wobei ... betrachtet wird.
Passiv	Das Zeichen $U$ bezeichnet die elektrische Spannung.	Die elektrische Spannung wird mit dem Zeichen $U$ bezeichnet.
Nominalisierungen	Die Batterie stellt ... bereit.	Die Bereitstellung der Batterie...
Verbalklammern	Dazu erinnert man sich an...	Dazu kann man sich an ... erinnern.
Regressionen	Das Zeichen $U$ bezeichnet in Formeln und Rechnungen die elektrische Spannung.	In Formeln und Rechnungen wird die elektrische Spannung mit dem Zeichen $U$ bezeichnet.
Partizipialkonstruktionen	Wasser strömt von der höheren Lage in eine tiefere.	von der höheren Lage in eine tiefere Lage strömendes Wasser
Mehrwortkomplexe	Damit ..., muss elektrische Ladung durch die Drähte in der Lampe fließen.	Damit ..., ist durch die Drähte in der Lampe strömende elektrische Ladung nötig.

Tabelle 6.2.: Sprachliche Variationen hinsichtlich der S-Variable (Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene). Es werden für jedes sprachliche Mittel, das zur Textvariation hinsichtlich der Satzstrukturen benutzt wurde, exemplarische Textstellen gegenübergestellt. Die Zuordnung von Textstellen zu sprachlichen Mitteln ist dabei nicht ein-eindeutig. Eine Liste aller Modifikationen zwischen s- und S-Texten findet sich im Anhang A.2.2.

### Variation P vs. p

Die Gruppe *erzählende und personalisierende Elemente* wurde anhand der folgenden sprachlichen Mittel variiert. Die Umsetzung in den Texten zur elektrischen Spannung ist anhand von Textauszügen in Tabelle 6.3 dargestellt<sup>9</sup>.

- Persönliche Anrede.  
Sätze in der du/wir-Form erhöhen die Personalisierung (P) gegenüber Sätzen, in denen Passiv- oder unpersönliche man-Konstruktionen verwendet werden (p).
- Deagentivierung.  
Anhand von passiven oder nominalisierten Verbformen im Gegensatz zu einer aktiven Formulierung lässt sich der Handlungsträger einer Aktion verschleiern („Die Spannung wird gemessen“ – *Wer misst?*). Diese sog. Deagentivierung wird in unpersönlichen Texten (p) verwendet. Den Handlungsträger klar zu benennen trägt zum erzählerischen und personalisierten Charakter bei (P) (vergleiche Seite 36 in Abschnitt 2.2.2).
- Dichte bedeutungstragender Einheiten.  
Fachsprachliche Texte zeichnen sich gegenüber alltagssprachlichen Texten unter anderem durch eine hohe Dichte an Informationen aus (siehe Abschnitt 2.2.2). Im Umkehrschluss tragen gewisse Füllwörter und -phrasen einen narrativen Charakter. Daher werden in Textvarianten mit einem hohen Grad an erzählenden Elementen und Personalisierungen (P) die Sätze mit solchen erzählenden Elementen angereichert. Für Textvarianten mit gering ausgeprägter Personalisierung (p) wird auf diese verzichtet.
- Werbende Attribute.  
Werbende Attribute, also positiv konnotierte Zusätze (meist: Adjektive), mit denen die Lesenden bewusst auf die Neuheit oder Interessantheit des Texts aufmerksam gemacht werden, zählen zu den erzählenden und personalisierenden Elementen. P-Texte werden damit angereichert, p-Texte nicht.
- Humorvolle Elemente.  
Die erzählenden und personalisierenden Elemente umfassen unter anderem humorvolle Elemente. Hier werden daher für Texte mit hoher Ausprägung der Personalisierungsvariable (P) Wortspiele und unerwartete Querverweise integriert, während in Texten mit geringer Ausprägung (p) darauf verzichtet wurde.
- Wiedergabe von Stimmungen.  
Indem Stimmungen wiedergegeben werden, kann der erzählende/personalisierende Charakter eines Texts verändert werden. Texte mit vielen erzählenden und personalisierenden Elementen (P) enthalten daher entsprechende Einschübe, weniger erzählerische/personalisierte Texte (p) nicht.

---

<sup>9</sup>Die vollständigen Textvarianten finden sich im digitalen Anhang unter <http://doi.org/10.5283/epub.58778>; Anhang A.2.2 enthält eine Tabelle, die alle p-variierten Textstellen gegenüberstellt (Tabelle A.3).

	<b>p</b>	<b>P</b>
Persönliche Anrede	Eine Folge ... ist ... Dazu betrachtet man ...	Du kennst nun eine Folge ... Dazu betrachten wir...
Deagentivierung	Wie wird über ... gesprochen? Zum Ingangsetzen des Ladungsstroms in der Taschenlampe wird eine Batterie benötigt.	Wie sprechen wir über ...? ... muss elektrische Ladung fließen. Dazu benötigen wir eine Batterie.
Dichte bedeutungstragender Einheiten	Warum fließt ein Fluss in eine bestimmte Richtung?	Warum fließt ein Fluss eigentlich in eine bestimmte Richtung?
Werbende Attribute	Die Gemeinsamkeiten sind:  Noch deutlicher wird das am Beispiel einer Taschenlampe.	Folgende spannende Gemeinsamkeiten kennen wir:  Noch schöner zeigt uns das eine Taschenlampe.
Humorvolle Elemente	Die elektrische Spannung treibt den elektrischen Strom an.  /	Stell dir die elektrische Spannung als „Tritt in den Hintern“ für den elektrischen Strom vor.  Und auch wenn es hier um Pole geht: Am Plus- oder Minuspol findest du sicherlich keine Eisbären oder Pinguine.
Wiedergabe von Stimmungen	... ist die elektrische Spannung der Antrieb des elektrischen Stroms.  Dafür gibt es noch weitere Modelle zur elektrischen Spannung.	... ist die elektrische Spannung der Antrieb des elektrischen Stroms. Und diesen Job macht sie richtig gut!  Dafür gibt es noch weitere „spannende“ Spannungs-Modelle. Diese wirst du in Zukunft noch kennenlernen.

Tabelle 6.3.: Sprachliche Variationen hinsichtlich der P-Variable (erzählende und personalisierende Elemente). Es werden für jedes sprachliche Mittel, das zur Textvariation hinsichtlich der Personalisierung benutzt wurde, exemplarische Textstellen gegenübergestellt. Die Zuordnung von Textstellen zu sprachlichen Mitteln ist dabei nicht ein-eindeutig. Eine Liste aller Modifikationen zwischen p- und P-Texten findet sich im Anhang A.2.2.

Text	sprachl. Gestaltung	gSmog	Wortzahl	Satzzahl	Ø Wortlänge	Ø Satzlänge	Füllwörter
A	wsp	5.65	720	82	5.87	8.78	37
B	wsP	5.60	807	93	5.83	8.68	52
C	wSp	8.74	837	51	5.92	16.41	59
D	wSP	8.49	957	57	5.77	16.79	79
E	Wsp	6.49	725	82	6.14	8.84	33
F	WsP	6.41	800	89	6.09	8.70	47
G	WSp	9.53	833	51	6.15	16.33	55
H	WSP	9.36	955	57	5.97	16.75	74

Tabelle 6.4.: Auszug aus den RATTE-Outputs. Mit W/w, S/s und P/p sind die sprachlichen Variablen der Textgestaltung – *Wortwahl*, *Satzstrukturen* und *Personalisierung* – abgekürzt. Ein Großbuchstabe steht dabei für eine hohe/anspruchsteigernde, ein Kleinbuchstabe für eine schwache/anspruchsmindernde Ausprägung. Die durchschnittliche Lesedauer für alle Texte liegt dem Analyseprogramm zufolge jeweils bei fünf bis sechs Minuten für durchschnittliche Lernende der neunten Jahrgangsstufe. Der gSmog (Simple Measure of Gobbledygook – german) ist ein Maß für das Lesealter in Schulklassenstufen, für welches der Text optimal geeignet ist. Screenshots, die die vollständigen Outputs zeigen, sind im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> zu finden.

### Kombination der drei sprachlichen Merkmalsgruppen

In den drei vorangehenden Abschnitten ist beschrieben, wie jede der drei sprachlichen Merkmalsgruppen *Wortwahl / morphologische Auffälligkeiten auf Wortebene*, *Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene* und *erzählende und personalisierende Elemente* variiert wurde. Für jeden der acht Texte ergeben sich daraus Richtlinien für die sprachliche Gestaltung, je nach umzusetzender Ausprägung W oder w, S oder s und P oder p. Die Ergebnisse dieser sprachlichen Gestaltung sind die Lesetexte, wie sie die Schüler\*innen schließlich in der empirischen Untersuchung erhalten haben. Sie sind im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> vollständig einzusehen.

Die Tabelle 6.4 stellt außerdem gegenüber, wie sich die beschriebenen Variationen in quantitativen Textkennwerten niederschlagen. Sie basiert auf einer Textanalyse mit dem Programm RATTE<sup>10</sup>. Es lässt sich unter anderem beobachten, wie sich die Satzlänge mit der Variable S verändert und die Anzahl an Wörtern, insbesondere Füllwörtern, sowohl mit P als auch mit S. Für die W-Variable ist eine leichte Tendenz zu einer größeren durchschnittlichen Wortlänge für W-Texte im Vergleich zu w-Texten zu erkennen. Die vollständigen Textanalyseergebnisse sind im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> zu finden.

Die Texte – beziehungsweise deren Ausprägung hinsichtlich Wortwahl, Satzstrukturen und Personalisierung – stellen die unabhängige Variable dieser Studie dar. Die Testinstrumente für die Messung der Moderator- und der abhängigen Variablen werden im Folgenden dargestellt.

<sup>10</sup>Wild, J. & Pissarek, M. (o. J.). Ratte. Regensburger Analysetool für Texte. Version 2.0.

<https://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/downloads/ratte/index.html> (Zugriff: 23.03.2023)

## 6.3 Testinstrumente

Die Studie zur Textverständlichkeit wurde mit Schüler\*innen der achten oder neunten Klasse an bayerischen Realschulen und Gymnasien durchgeführt. Um statistisch aussagekräftige Zusammenhangsanalysen entsprechend dem Forschungsinteresse durchführen zu können, war ein großer Stichprobenumfang vorausgesetzt. Die Studie war daher auf die Kooperation möglichst vieler Schulleitungen und Physiklehrkräfte angewiesen. Somit rückt für die Testerstellung das Gütekriterium der Durchführungsökonomie besonders in den Fokus, um den Mehraufwand, den die Studienteilnahme für die Lehrkräfte mit sich bringt, möglichst gering zu halten.

Die Studie sieht auf Seiten der abhängigen Variablen einerseits das kognitive Konstrukt *Behaltensleistung* und andererseits die subjektiven Eindrücke der Lesenden, also die *affektive Textwahrnehmung* vor. Um analysieren zu können, wie sich die sprachliche Gestaltung der Texte auf diese beiden abhängigen Variablen auswirkt, müssen die Effekte weiterer (potenzieller) Einflussfaktoren (Moderatorvariablen) kontrolliert werden. Dementsprechend werden die Personeneigenschaften Gender, Schulart und Jahrgangsstufe auf dem Teilnahmebogen schriftlich abgefragt beziehungsweise bei der Datendigitalisierung mit abgespeichert.

Als weitere Moderatorvariable werden die sprachlichen Fähigkeiten der Teilnehmenden anhand eines standardisierten Lesetests bestimmt: Der LGVT 5-12+ liefert Kennwerte für die Lesegeschwindigkeit, das Leseverständnis und die Lesegenauigkeit. Der folgende Abschnitt 6.3.1 dient einer kurzen Beschreibung des Tests.

Für die beiden abhängigen Variablen *Lernzuwachs* im Sinne einer *Behaltensleistung* und *affektive Textwahrnehmung* wurde jeweils eigens für diese Studie ein Testinstrument entworfen (siehe die nachfolgenden Abschnitte 6.3.2 und 6.3.3). Die Entwicklung der Tests verlief dem Testentwicklungsleitfaden von Brandt und Moosbrugger (2020) zufolge in drei Schritten. Aufbauend auf einigen Vorüberlegungen zur Merkmalsart (quantitativ/qualitativ, unidimensional/multidimensional, zeitlich stabil/veränderlich), zur dementsprechend geeigneten Testart (Leistungstest, Persönlichkeitstest, ...) und zur Zielgruppe wurden jeweils die Teststruktur, die Testlänge und -zeit und die Durchführungsmodalitäten festgelegt. Im zweiten Schritt wurde die Planung umgesetzt und ein vorläufiger Test konstruiert. Dabei stehen im Allgemeinen unterschiedliche Konstruktionsstrategien zur Verfügung (intuitiv, kriteriumsorientiert, ...). Die Konstruktion beinhaltet die Festlegung auf bestimmte Aufgabentypen und Antwortformate, die Formulierung konkreter Items und deren Anordnung sowie die Aufstellung konkreter Durchführungs- und Testanweisungen für teilnehmende Schüler\*innen und durchführende Lehrkräfte. Im dritten und letzten Schritt kamen verschiedene Maßnahmen zur Erprobung und Überarbeitung der vorläufigen Testversion zum Einsatz (Pilotierungsstudie mit dem Ziel einer statistischen Itemanalyse, Teilnehmendenbefragung,...). Der Aufbau der Abschnitte 6.3.2 und 6.3.3, in denen es um die Messung der Behaltensleistung beziehungsweise der affektiven Textwahrnehmung geht, richtet sich nach den eben genannten drei Testentwicklungsschritten. Vorangestellt ist jeweils eine Beschreibung des

endgültigen Tests und seiner Durchführung. Die verwendeten statistischen Methoden zur Erprobung werden in den entsprechenden Abschnitten jeweils kurz und in qualitativer Weise erklärt; eine ausführlichere Beschreibung findet sich im anschließenden Abschnitt 6.4.3.

### 6.3.1 Messung der sprachlichen Fähigkeiten: der LGVT

Der LGVT 5-12+ (Lesegeschwindigkeits- und -verständnistest für die Jahrgangsstufen 5 bis 12<sup>11</sup>, Schneider, Schlagmüller und Ennemoser, 2017) ist ein bewährter, standardisierter Test zur Erfassung des Leseverständnisses, der Lesegeschwindigkeit und der Lesegenauigkeit. Seine Anwendung als ökonomischer Gruppentest und seine kurze Durchführungszeit von ca. zwölf Minuten (davon sechs Minuten reine Bearbeitungszeit) sind für diese Studie angemessen. Der Test ist geeignet für die fünfte bis dreizehnte Jahrgangsstufe, wobei gesicherte Normen in den Jahrgangsstufen fünf bis zehn vorliegen.

Der Test in der hier ausgewählten Version „Laufburschentext“ besteht aus einem Lückentext, der 2155 Wörter und 47 zu füllende Stellen umfasst. An diesen Stellen wählen die teilnehmenden Schüler\*innen durch Unterstreichung aus drei angebotenen Füllwörtern diejenige Alternative aus, die in den Textzusammenhang passt. Außerdem umkreisen sie nach Ablauf der sechsminütigen Bearbeitungszeit das letzte Wort, das sie gelesen haben.

Indem die Anzahl der gelesenen Wörter und die Anzahl der insgesamt bearbeiteten, der korrekt bearbeiteten und der falsch bearbeiteten Textlücken nach Auswertungsvorgabe verrechnet werden, werden Rohwerte für das Leseverständnis, die Lesegeschwindigkeit und die Lesegenauigkeit ermittelt. Anhand der jeweiligen zum „Laufburschentext“ zugehörigen Prozentrang-Normtabellen lassen sich bezogen auf die Normstichprobe des Tests (7 142 Schüler\*innen aus 13 deutschen Bundesländern) Gruppen stärkerer und schwächerer Lesender bilden.

Die Retest-Reliabilitäten des LGVT 5-12+ liegen für das Leseverständnis im Bereich 0.82 bis 0.89, für die Lesegeschwindigkeit im Bereich 0.80 bis 0.82 und für die Lesegenauigkeit im Bereich 0.72 bis 0.82. Die Validität des Tests ist am Vergleich mit verwandten Testergebnissen/Kriterien festgemacht (z. B. PISA 2000-Leseverständnistest, KFT, Deutschnote). Die Korrelationskoeffizienten (*Pearson-Produkt-Moment-Korrelation*) liegen dabei überwiegend im Bereich 0.20 bis 0.40. Für detailliertere Informationen sei auf das Testmanual von Schneider, Schlagmüller und Ennemoser (2017) verwiesen.

### 6.3.2 Messung der Behaltensleistung: Entwicklung eines Wissenstests

Bei dem Wissenstest handelt es sich um einen Papier-und-Bleistift-Test, der acht Aufgaben umfasst. Jede Aufgabe ist genau einem der acht Lernziele zugeordnet, die in den Texten

---

<sup>11</sup> Der Test ist über die Testzentrale des Hogrefe-Verlags verfügbar:  
[www.testzentrale.de/shop/lesegeschwindigkeits-und-verstaendnistest-fuer-die-klassen-5-12.html](http://www.testzentrale.de/shop/lesegeschwindigkeits-und-verstaendnistest-fuer-die-klassen-5-12.html)

vermittelt werden (siehe Tabelle 6.5 und vergleiche Abschnitt 6.2.1). Es werden pro Aufgabe fünf Aussagen zum entsprechenden Aspekt präsentiert. Die Teilnehmenden müssen diejenigen Aussagen ankreuzen, die sie für korrekt halten. In den Aufgaben 1 bis 6 schließen sich die Aussagen nicht gegenseitig aus, die Teilnehmenden müssen also eine ihnen unbekannte Anzahl an Aussagen als korrekt identifizieren (*multiple choice*). In den Aufgaben 7 und 8 ist aus logischen Gründen klar und darüber hinaus explizit angegeben, dass nur eine Aussage korrekt sein kann, die Teilnehmenden müssen also diese Aussage aus den fünf Alternativen auswählen (*single choice*). Der Test ist unter Kapitelnummer B.1 angehängt.

In der Hauptstudie bearbeiten die Teilnehmenden den Test zweimal: Im ersten Teil der Erhebung als Vorwissenstest und nach der Textlektüre als Test für die Behaltensleistung. Die Behaltensleistung (kurzfristig messbarer Lernzuwachs) wird als Differenz von Pre-Test-Ergebnis und Post-Test-Ergebnis aufgefasst. Die Verwendung dieser Differenz, auch *Gain-Score* genannt, als Maß für die Behaltensleistung wird in Abschnitt 7.1.2 diskutiert. Wie Tabelle 6.8 in Abschnitt 6.5.1 zu entnehmen ist, liegen damit mindestens sieben Tage zwischen den beiden Testzeitpunkten. Diese Zeitspanne soll der Vermeidung von Sequenzeffekten wie z. B. Testwiederholungseffekten dienen (auch dies wird im Einzelnen in Abschnitt 7.1.2 aufgegriffen). Sie soll auch dazu beitragen, dass sich die Teilnehmenden zum Zeitpunkt der Textlektüre nicht mehr präzise an den Vorwissenstest erinnern und somit in ihrer Aufmerksamkeitssteuerung beim Lesen nicht durch den Vorwissenstest beeinflusst sind.

### Testkennwerte und Testinhalt

Für die Berechnung der Testkennwerte dienten Pilotierungsdaten einer Stichprobe aus  $n_{\text{PilotierungWT}} = 71$  Teilnehmenden, die den Test als Behaltensleistungstest im Anschluss an die Textlektüre bearbeiteten. Die Tabelle D.1 mit allen in Betracht gezogenen Kennwerten ist im Anhang D.1 zu finden. In der Testauswertung wurden aufgabenweise Credits vergeben (im Folgenden *Task-Scores*), die zu einem Gesamtestwert (im Folgenden *Test-Score* oder *Gesamtscore*) addiert wurden. Ein Ergebnis der Pilotierung war es, ein Schema für die Bepunktung festzulegen, welches einerseits starke und schwache Testergebnisse auflösen kann und insbesondere möglichst keine Boden- und Deckeneffekte erzeugt und andererseits zwischen wahllosem Raten und bewusstem Auswählen korrekter Aussagen differenzieren kann. Darüber hinaus wurde die Analyse der Pilotierungsdaten genutzt, um die Distraktoren zu evaluieren. Eine Itemselektion im Sinne des Identifizierens und Entfernens ganzer Aufgaben wurde nicht vorgenommen, da die Aufgaben nicht redundant konstruiert wurden und jede einzelne Aufgabe inhaltlich an genau einen Lernaspekt der Texte gebunden ist.

Die Testkennwerte lassen insgesamt den Schluss zu, dass der Test geeignet ist, die Behaltensleistung zu messen. Er wurde in der Hauptstudie auch als Pre-Test vor der Textlektüre eingesetzt, um das themenspezifische Vorwissen zu erheben. Als Textkennwerte wurden die Item-Schwierigkeit, die Varianz und die Item-Trennschärfe betrachtet. Der Kennwert für die Itemschwierigkeit dient als Indiz für die Selektivität des Tests, das heißt, inwiefern gute von schwachen Behaltensleistungen im Test unterscheidbar sind. Mathematisch betrachtet sagt

Aufgabe	Inhalt
Aufgabe 1	Analoge Bezüge zwischen Gebirgsfluss und Stromkreis
Aufgabe 2	Batterie als Bereitsteller von elektrischer Spannung
Aufgabe 3	Zusammenhänge der beiden Größen elektrische Spannung und elektrische Stromstärke
Aufgabe 4	Elektrische Spannung als Unterschied der elektrischen Lage
Aufgabe 5	Zusammenhang zwischen den Polen einer Batterie und elektrischer Spannung
Aufgabe 6	Sprechweisen („Elektrische Spannung liegt zwischen zwei Punkten an“)
Aufgabe 7	Die Einheit der elektrischen Spannung
Aufgabe 8	Das Formelzeichen für die elektrische Spannung

Tabelle 6.5.: Inhaltlicher Bezug zwischen den Aufgaben im Wissenstest und den Lernzielen im Lesetext.

der Kennwert aus, welcher Anteil am maximal erzielbaren Task-Score im Mittel durch die Stichprobe erzielt wurde. Je näher er also am Wert 1 ist, desto besser wird die Aufgabe gelöst. Entsprechend der Empfehlung von Moosbrugger und Kelava (2020, S. 155) liegt der Kennwert bestenfalls im Bereich 0.2 bis 0.8, wobei auch besonders einfache Items im Bereich 0.8 bis 0.95 und besonders schwierige im Bereich 0.05 bis 0.2 vertreten sein können. Da der empfohlene Bereich durch die Aufgaben im Wissenstest optimal abgedeckt ist, ist eine zufriedenstellende Selektivität des Test anzunehmen. Die Varianz – ein Maß für die Breite der Verteilung – ist unauffällig und liegt für die Aufgaben 1 bis 6 im Bereich 1.2 bis 2.1. Erwartungsgemäß ist die Varianz für die beiden Single-Choice-Aufgaben (Aufgabe 7 und Aufgabe 8, möglicher Task-Score jeweils entweder 0 oder 2) deutlich geringer als für die übrigen Aufgaben, bei denen Task-Scores zwischen 0 und 5 erzielt werden können. Die Item-Trennschärfe ist die Korrelation des Task-Scores zum Summenscore aus allen übrigen Aufgaben und gibt daher einen Hinweis auf die inhaltliche Validität der jeweiligen Aufgabe. Gemessen an den Richtwerten in Bortz und Döring (2006, S. 220) (mindestens 0.3 für eine mittelmäßige und mindestens 0.5 für eine hohe Trennschärfe) erweisen sich die Item-Trennschärfen dieses Tests als nicht optimal, aber akzeptabel – zumal die Aufgaben eigenständige Lerninhalte repräsentieren und daher eine nur mäßig starke Korrelation gut begründet ist. Die Item-Trennschärfen der Aufgaben 2 bis 8 liegen im Bereich 0.3 bis 0.5. In Aufgabe 1 (besonders geringe Trennschärfe von 0.2) wurden die Distraktoren für die Hauptstudie daher überarbeitet (siehe dazu *Erprobung und Überarbeitung* unten).

Tabelle 6.5 gibt die Zuordnung der Aufgaben zu den Lerninhalten wieder. Da alle Aufgaben eigenständige Lerninhalte repräsentieren, ist die Berechnung von Maßen innerer Konsistenz (etwa Cronbachs  $\alpha$  für den gesamten Test) nicht sinnvoll. Es wurden stattdessen Korrelationsanalysen *innerhalb* der einzelnen Aufgaben genutzt, um nicht-funktionierende Distraktoren zu identifizieren: Wenn zwei Aussage-Alternativen übermäßig häufig gemeinsam oder über-

mäßig selten gemeinsam angekreuzt werden, ist dies ein Indiz dafür, dass sie redundant sind beziehungsweise sich gegenseitig ausschließen. Ein weiteres Mittel der Distraktorenanalyse war die Betrachtung des Schwierigkeitsindex' jedes einzelnen Distraktors, um festzustellen, ob gewisse Distraktoren auffällig selten angekreuzt werden und damit von den Testpersonen zu leicht als falsche Aussage erkannt werden. Die Ergebnisse werden im Abschnitt zur Erprobung und Überarbeitung weiter unten dargestellt.

### **Testplanung**

Um zu messen, wie viel Schüler\*innen inhaltlich aus den Texten zur elektrischen Spannung lernen, wurde ein Prä-Post-Vergleich in Form eines Vorwissens- und eines Behaltenstests angestellt. Beide Tests waren dabei identisch. Zwischen dem Zeitpunkt des Vorwissenstests und dem Zeitpunkt der Textlektüre und des Behaltenstests lagen mindestens sieben Tage. Zum Zweck einer ökonomischen Durchführung gemeinsam mit den anderen Erhebungsteilen war ein Gruppentest im Klassenverbund im Papier-und-Bleistift-Format naheliegend. Das zu messende Konstrukt lässt sich als *Lernzuwachs im Sinne einer Behaltensleistung auf Basis der Textlektüre bezüglich der elektrischen Spannung* beschreiben. Es kann in dem Sinne als acht-dimensional aufgefasst werden, als dass sich acht einzelne Lerninhalte im Text klar voneinander trennen lassen (siehe Abschnitt 6.2.1). Die inhaltliche Validität des Tests wurde dementsprechend von Beginn an sichergestellt, indem zu jedem inhaltlichen Lernaspekt in den Texten genau eine Aufgabe im Wissenstest konstruiert wurde. Die Passung der Aufgabeninhalte und des Aufgabendesigns (eine Mischung aus Multiple-Choice- und Single-Choice-Aufgaben) zum Text wurde durch zwei erfahrene Physiklehrkräfte bestätigt.

### **Testkonstruktion**

Der zentrale Leitgedanke bei der Testkonstruktion war die inhaltliche Abbildung der acht Lerninhalte in den Testaufgaben. Die elektrische Spannung ist ein abstrakter physikalischer Begriff, über den zahlreiche Fehlvorstellungen bei Schüler\*innen bestehen. Beispiele sind eine fehlende Unterscheidung von elektrischem Strom und elektrischer Spannung oder die Annahme, die Batterie sei ein Speicher für elektrischen Strom (für diese und weitere siehe z. B. Hopf, Schecker, Höttecke und Wiesner, 2022, S. 34 ff.). Ein Auswahlaufgaben-Format mit richtigen und falschen Aussagen zu den Lerninhalten, in welchem sich die Distraktoren an den gängigen Fehlvorstellungen orientieren, erweist sich damit als gut geeignet für die hier vorliegende Studie.

Für die Formulierung der korrekten Aussage-Alternativen wurde darauf geachtet, eine identische Wortwahl und damit keine unnötigen Synonyme im Vergleich zu den Lesetexten zu verwenden. Die Distraktoren orientieren sich inhaltlich an typischen Fehlvorstellungen, wann immer dies möglich ist. Die Aufgaben 1 bis 6 sind dabei als „echte“ Multiple-Choice-Aufgaben konstruiert; das heißt, die Distraktoren sollen sich gegenseitig nicht ausschließen. Die Aufgaben 7 und 8 sind Single-Choice-Aufgaben, da die Frage nach der Einheit beziehungsweise den Formelzeichen für die elektrische Spannung offensichtlich nur *eine* richtige

Antwort hat. In der Pilotierung zeigte sich, dass dennoch einige Schüler\*innen mehr als eine Antwortalternative in Aufgabe 7 und 8 ankreuzten, sodass für die Hauptstudie ein entsprechender Hinweis („In den letzten beiden Aufgaben ist **genau eine** Aussage richtig. Kreuze bitte diejenige Aussage an, die du für richtig hältst.“) eingefügt wurde.

### Erprobung und Überarbeitung

Die Pilotierungsdaten dienten einerseits dazu, ein angemessenes Bewertungsschema für den Task-Score festzulegen und andererseits dienten sie der Distraktorenanalyse. Im hier vorliegenden Fall müssen bei den sechs Multiple-Choice-Aufgaben fünf korrekte Entscheidungen getroffen werden, um die Aufgabe vollständig richtig zu bearbeiten, da jeder Distraktor sowohl richtig als auch falsch sein kann. In den beiden Single-Choice-Aufgaben muss nur aus den fünf Alternativen die eine korrekte Aussage ausgewählt werden. Aus diesem Grund ist ein höherer maximaler Task-Score bei den ersten sechs Aufgaben zu vergeben. Wenn man die Bewertung, die eine Person erhalten sollte, die sich bei *einer* Aussage in einer Multiple-Choice-Aufgabe sicher ist und vier rät, gleichsetzt mit dem Score der richtigen Entscheidung in einer *Single-Choice*-Aufgabe, so sind für den maximalen Single-Choice-Taskscore 2 Credits anzusetzen<sup>12</sup>. Auf Grundlage der Pilotierungsdaten stellte sich im Vergleich verschiedener denkbarer Bewertungsschemata außerdem heraus, dass das ratekorrigierte *LOTSE-2-Schema* („Lernzielorientierte Testaufgaben zur standardisierten Erfolgsbeurteilung“, Version 2) (Schulz-Gerlach, 2022) zu guten Test- und Itemkennwerten führt (siehe Abschnitt *Testkennwerte und Testinhalt* oben sowie Tabelle D.1 im Anhang D.1). Das Schema sieht einerseits eine differenzierte Bewertung von Teilleistungen vor, das heißt *einige* korrekte Multiple-Choice-Entscheidungen mit einer angemessenen Anzahl Credits zu bewerten. Andererseits berücksichtigt die Ratekorrektur, dass durch Raten nicht zu viele Credits erzielt werden. Genauer im hier vorliegenden Fall: Die Teilnehmenden der Pilotierung hatten den Text gelesen und im Mittel einen Gesamtscore von 17.7 (Standardabweichung 5.2). Der Erwartungswert bei rein zufälligem Raten liegt im *LOTSE-2-Schema* für diesen Test bei 10.2 Credits und unterscheidet sich daher – wie gewünscht – deutlich vom Mittelwert der Stichprobe. Das Schema setzt für fünf Antwortalternativen die folgende Bepunktung an:

- 5 Credits für fünf korrekte Entscheidungen,
- 3 Credits für vier korrekte Entscheidungen und eine falsche,
- 2 Credits für drei korrekte Entscheidungen und zwei falsche,
- 1 Credits für zwei korrekte Entscheidungen und drei falsche,
- 0 Credits für eine oder null korrekte Entscheidungen und vier beziehungsweise fünf falsche.

Eine korrekte Entscheidung bedeutet dabei entweder, eine korrekte Aussage anzukreuzen,

---

<sup>12</sup>Das passende mathematische Modell ist ein Bernoulli-Experiment mit  $p = 0.5$ . Um den Erwartungswert bei einer sicher richtigen und vier geratenen Entscheidungen zu berechnen, werden die Taskcores für eine, zwei, drei, vier und fünf korrekte Antworten mit den jeweiligen Wahrscheinlichkeiten für genau null, einen, zwei, drei beziehungsweise vier Treffer beim Raten multipliziert und die Produkte addiert. Das Ergebnis sind 2.06 Credits.

oder eine falsche Aussage nicht anzukreuzen. Bei den Multiple-Choice-Fragen wurde die Nicht-Bearbeitung von der Entscheidung nichts anzukreuzen unterschieden, indem von einer Nicht-Bearbeitung ausgegangen wird, wenn in zwei oder mehr Aufgaben in Folge keine einzige Aussage angekreuzt wurde. Bei den Single-Choice Aufgaben gab es genau dann 2 Credits, wenn die eine richtige Antwortalternative angekreuzt wurde. Andernfalls wurden 0 Credits vergeben.

Für die Distraktorenanalyse in den Multiple-Choice-Aufgaben ist maßgeblich, dass die Distraktoren ihren Zweck als scheinbar plausible, jedoch tatsächlich unzutreffende Antwortalternative erfüllen und dass sie sich nicht gegenseitig ausschließen oder redundant sind. Indem der Schwierigkeitsindex jedes einzelnen Distraktors betrachtet wurde, konnte festgestellt werden dass lediglich die Aussage *„Die elektrische Spannung ist vergleichbar mit der Breite des Flussbetts.“* in Aufgabe 1 (Schwierigkeitsindex 0.99) nahezu immer korrekterweise als falsch erkannt wird. Alle weiteren Schwierigkeitsindizes liegen im Bereich 0.30 bis 0.90 (vergleiche Tabelle D.2 im Anhang D.1) und legen keine weiteren Überarbeitungen nahe. Ob sich einzelne Distraktoren gegenseitig ausschließen wurde überprüft, indem Aufgabe für Aufgabe für alle denkbaren Kombinationen zweier Distraktoren jeweils betrachtet wurde, wie häufig beide Distraktoren gleichzeitig angekreuzt wurden, wie häufig genau einer und wie häufig keiner von beiden angekreuzt wurde. Es fiel dadurch auf, dass sich in Aufgabe 2 die Distraktoren teilweise gegenseitig ausschließen. Dies motivierte die folgenden Änderungen:

- Aufgabe 1: Ein nicht-funktionierender Distraktor wird durch eine als zutreffend zu identifizierende Antwortalternative ersetzt, die nicht im Widerspruch zu den übrigen Aussagen steht.

„Die elektrische Spannung ist vergleichbar mit der Breite des Flussbetts.“

→ „Punkte im Stromkreis, die elektrisch hoch liegen, sind vergleichbar mit Orten, die hoch am Berg liegen.“

- Aufgabe 2: Ein Distraktor, der im Widerspruch zu einem anderen steht, wird durch einen verbesserten Distraktor ersetzt, der nicht im offensichtlichen Widerspruch zu den übrigen Aussagen steht.

„Die Batterie wandelt elektrischen Strom in elektrische Spannung um.“

→ „Die Batterie nutzt elektrischen Strom, um Spannung zu erzeugen.“

Auf Grundlage dieser Überarbeitungen darf davon ausgegangen werden, dass der Test ein valides und reliables Messinstrument für die Hauptstudie darstellt.

Es sei an dieser Stelle bereits erwähnt, dass die in der Hauptstudie gefundenen signifikanten Unterschiede zwischen dem mittleren Pre-Test-Score und dem mittleren Post-Test-Score sowie zwischen allen mittleren Pre-Test-Task-Scores und mittleren Post-Test-Task-Scores ein weiteres Indiz für die Sensitivität des Tests auf die Lektüre des Texts sind. Dieser Befund sowie alle weiteren Hauptstudienresultate werden ausführlich in Kapitel 7 dargestellt.

### 6.3.3 Messung der Textwahrnehmung: Entwicklung eines Bewertungsfragebogens

In Unterkapitel 6.1 wurde bereits erwähnt, dass ein zentrales Interesse der Studie die Erhebung der affektiven Textwahrnehmung von Schüler\*innen ist. Als Messinstrument für die affektive Textwahrnehmung wurde ein Rating-Fragebogen entworfen. Es handelt sich dabei um einen Papier-und-Bleistift-Test mit einem Item zur globalen Verständlichkeitsbeurteilung und 44 Items zur Bewertung differenzierterer Aussagen zur Textwirkung. Die bipolare, symmetrische Ratingskala umfasst fünf Stufen von „Stimme gar nicht zu“ bis „Stimme voll zu“. Der Test ist im Anhang B.2 zu finden.

#### Testkennwerte und Testinhalt

Die Testkennwerte erweisen sich als zufriedenstellend. Sie wurden auf Basis einer Pilotierungsstichprobe mit  $n_{\text{PilotierungFB}} = 99$  Teilnehmenden berechnet und sind in Tabelle D.3 im Anhang D.1 zusammengefasst. Es wurden der Item-Mittelwert, die Item-Schwierigkeit, die Pearson-Produkt-Moment-Korrelation jedes Items mit dem Globalurteil und die Varianz der einzelnen Items berechnet. Die Item-Mittelwerte zeigen keine Ausreißer-Items, die mit auffällig großer Zustimmung beziehungsweise Ablehnung bewertet wurden. Die Item-Schwierigkeit (siehe z. B. Kelava & Moosbrugger, 2020, S. 146 ff.) ist im Fall eines Bewertungsfragebogens die mittlere Bewertung eines Items in Relation zur Maximalbewertung. Sie ist somit als eine Art „standardisierter Mittelwert“ zu verstehen. Kelava und Moosbrugger (2020, S. 155) empfehlen Items, deren Schwierigkeiten bei 0.05 bis 0.95 liegen, wobei der Bereich von 0.2 bis 0.8 besonders häufig vertreten sein sollte. Die hier berechneten Itemschwierigkeiten entsprechen der Empfehlung, wenn auch mittlere Schwierigkeiten um 0.5 öfter vertreten sein dürften. Da der Item-Trennschärfe-Kennwert für das vielschichtige Konstrukt *affektive Textwahrnehmung* nicht aussagekräftig ist, wurde als Ersatzkennwert die Korrelation der einzelnen Items mit der globalen Verständlichkeitsbeurteilung berechnet. Sie funktioniert als Indiz dafür, ob die einzelnen Items in erwarteter Richtung mit dem globalen Verständlichkeitseindruck zusammenhängen und somit eventuell vom Test ausgeschlossen werden müssen. Je nach theoretischem Hintergrund zu dem jeweiligen Textwirkungsaspekt sind für die einzelnen Items positive oder negative Korrelationen erwartbar. Zum Beispiel wird für das Item „2. Ich fand die Sätze im Text kompliziert.“ eine negative und für das Item „17. Ich fand die Einteilung des Texts in seine Abschnitte passend.“ eine positive Korrelation erwartet. Die Korrelationen fallen in 41 von 44 Fällen erwartungskonform aus. Lediglich bei den drei Items „25. Der Text rief gedankliche Konflikte bei mir hervor.“ ( $r = -0.07$ ), „26. Es gab Stellen im Text, die überraschend für mich waren.“ ( $r = -0.12$ ) und „33. Ich habe Wissen, das nichts mit Elektrizitätslehre zu tun hat, gebraucht um den Text zu verstehen.“ ( $r = -0.26$ ) wurde im Widerspruch zu einer erwarteten positiven Korrelation eine schwache, negative Korrelation gefunden. Trotz dieses Befunds wurden die Items 25, 26 und 33 nicht aus dem Test gestrichen. Die erwartungswidrigen Korrelationen wurden als Indiz für eine potenzielle Missverständlichkeit

Wahrnehmungsfacette	$\alpha$ (Pilotierung)	$\alpha$ (Hauptstudie)
Lokale Kohärenz	.79	.80
Globale Kohärenz	.85	.84
Informationsdichte	.71	.82
Sprachliche Einfachheit	.75	.81
Lesbarkeit	.73	.77
Strukturelle Komplexität	.62	.84
Vorwissen	.42	.49
Kognitive Gliederung	.81	.74
Darstellungsintentionen	.22	.54
Einprägsamkeit	–	–
Unterhaltsamkeit	.70	.66
Motivation	.83	.79

Tabelle 6.6.: Cronbachs  $\alpha$  als Kennwert für die innere Konsistenz derjenigen Itemgruppen, die gemäß der theoretisch abgeleiteten Wahrnehmungsfacetten gebildet werden können.

gewertet und die betroffenen Items daher in der Hauptstudie mit besonderem Augenmerk analysiert (siehe Abschnitt 7.2.1). Die Varianz der Items ist unauffällig: alle Varianzen liegen im Bereich 0.80 bis 1.80 und es lassen sich keine Ausreißer mit besonders enger oder breiter Bewertungsstreuung erkennen.

Als Maß für die innere Konsistenz des Fragebogens dient Cronbachs  $\alpha$  (Gäde, Schermelleh-Engel & Werner, 2020, S. 314 ff.). Die theoretisch abgeleiteten Wahrnehmungsfacetten (siehe Unterkapitel 5.2) dienten als inhaltliche Grundlage für die Itementwicklung und können daher einen vorsichtigen Anhaltspunkt geben, inwiefern gewisse Itemgruppen innerlich konsistent sein sollten. Tabelle 6.6 zeigt, dass die Werte für Cronbachs  $\alpha$  sowohl auf Grundlage der Pilotierungsdaten als auch in der Hauptstudie für viele Kategorien akzeptabel sind (üblicher unterer Grenzwert für akzeptable innere Konsistenz bei psychometrischen Verfahren:  $\alpha \geq 0.70$ , siehe z. B. Taber, 2018;  $\alpha$ -Werte unter 0.50 gelten als nicht akzeptabel und Werte ab 0.80 gelten gemeinhin als gut, siehe Field, Miles und Field, 2012, S. 799<sup>13</sup>). Bei der Berechnung von Cronbachs  $\alpha$  wurde berücksichtigt, dass die Items 2, 9, 12, 15, 18, 19, 21, 23, 27, 31, 34, 38 und 41 invertiert formuliert waren. Welche Items jeweils den Facetten zugeordnet sind, ist in Tabelle 6.7 festgehalten. Die Facetten *Darstellungsintentionen* und *Vorwissen* fallen durch eine besonders geringe innere Konsistenz auf. Es ist jedoch einschränkend anzumerken, dass die Eindimensionalität der Facetten fraglich ist und die Anzahl der Items pro Facette unterschiedlich, sodass die einzelnen  $\alpha$ -Werte nicht direkt vergleichbar sind. Aus diesem Grund werden die Cronbachs- $\alpha$ -Werte zu Zwecken des Vergleichs von Datenlage und Theorie abgebildet, jedoch nicht für die Itemselektion genutzt.

<sup>13</sup>Die Richtwerte sind mit Vorsicht zu betrachten:  $\alpha$  steigt mit der Anzahl der eingeschlossenen Items (Field, Miles & Field, 2012, S. 799). Dieser Trend ist auch in den Daten dieser Studie zu erkennen. Für genauere Analysen sind daher Item-Item-Korrelationen mit in Betracht zu ziehen. Dies wurde im Rahmen der Hauptkomponentenanalyse in der Hauptstudie beachtet.

Facette	Zugeordnete Items <sup>14</sup>
Lokale Kohärenz	<p>9 Ich fand den Aufbau der einzelnen Sätze verwirrend.</p> <p>20 Für mich war beim Lesen eines Satzes stets klar, wie er sich auf den vorherigen Satz bezieht.</p> <p>24 Mir erschienen aufeinanderfolgende Sätze inhaltlich logisch miteinander verknüpft.</p> <p>44 Mir war klar, wie die Wörter im Satz miteinander in Beziehung stehen.</p>
Globale Kohärenz	<p>12 Ich fand die Informationen im Text wirr.</p> <p>13 Ich konnte gut nachvollziehen, wie die Textinhalte aufeinander aufbauen.</p> <p>29 Für mich war beim Lesen eines Textabschnitts stets klar, wie er sich auf die übrigen Absätze bezieht.</p> <p>35 Ich konnte im Text durchgehend einen roten Faden erkennen.</p> <p>36 Die einzelnen Abschnitte setzen sich für mich logisch zu einem Gesamtbild der elektrischen Spannung zusammen.</p> <p>37 Mir war in jedem Textabschnitt klar, was das Thema dieses Abschnitts ist.</p>
Informationsdichte	<p>12 Ich fand die Informationen im Text wirr.</p> <p>18 Der Text erschien mir sehr voll an Informationen.</p> <p>19 Ich hätte bei schwierigen Aspekten noch eine Erklärung aus einer anderen Sicht gebraucht.</p> <p>27 Um den Text zu verstehen, hätte ich mehr Erklärung zu den einzelnen Aspekten der elektrischen Spannung gebraucht.</p> <p>34 Schwierige Begriffe im Text haben mir Verständnisprobleme bereitet.</p> <p>38 Ich hätte mehr Erklärung zu den Fachbegriffen im Text gebraucht.</p>
Sprachliche Einfachheit	<p>2 Ich fand die Sätze im Text kompliziert.</p> <p>9 Ich fand den Aufbau der einzelnen Sätze verwirrend.</p> <p>15 Ich wusste am Satzende nicht mehr, wie der Satz begonnen hatte.</p> <p>34 Schwierige Begriffe im Text haben mir Verständnisprobleme bereitet.</p> <p>38 Ich hätte mehr Erklärung zu den Fachbegriffen im Text gebraucht.</p>
Lesbarkeit	<p>9 Ich fand den Aufbau der einzelnen Sätze verwirrend.</p> <p>10 Ich konnte dem Satzbau beim Lesen problemlos vom Satzanfang bis zum Satzende folgen.</p> <p>15 Ich wusste am Satzende nicht mehr, wie der Satz begonnen hatte.</p> <p>21 Ich musste mit den Augen zwischen einzelnen Textstellen vor und zurück hüpfen anstatt den Text flüssig zu lesen.</p> <p>22 Ich fand den Text auf den ersten Blick übersichtlich.</p> <p>41 Ich musste den Text in einem langsamen Tempo lesen.</p>

<sup>14</sup>Die Nummer entspricht der Itemnummer in derjenigen Fragebogenversion, wie sie im Anhang B.2 zu finden ist. In der Studie wurden vier Versionen mit verschiedener Itemabfolge verwendet (siehe dazu Seite 113).

Strukturelle Komplexität	<p>2 Ich fand die Sätze im Text kompliziert.</p> <p>9 Ich fand den Aufbau der einzelnen Sätze verwirrend.</p> <p>10 Ich konnte dem Satzbau beim Lesen problemlos vom Satzanfang bis zum Satzende folgen.</p> <p>15 Ich wusste am Satzende nicht mehr, wie der Satz begonnen hatte.</p> <p>21 Ich musste mit den Augen zwischen einzelnen Textstellen vor und zurück hüpfen anstatt den Text flüssig zu lesen.</p> <p>23 Ich musste meine volle Aufmerksamkeit auf den Text richten, um ihm Information zu entnehmen.</p> <p>31 Beim Lesen brachten mich einzelne Textstellen zum Stocken.</p> <p>41 Ich musste den Text in einem langsamen Tempo lesen.</p>
Vorwissen	<p>11 Ich fand die Gliederung des Texts typisch für Physiktexte.</p> <p>14 Der Text hat mich an mein Vorwissen zum Thema Elektrizitätslehre erinnert.</p> <p>30 Ich konnte mein Wissen über Elektrizitätslehre um neue Informationen aus dem Text erweitern.</p> <p>33 Ich habe mein Wissen, das nichts mit Elektrizitätslehre zu tun hat, eingesetzt um den Text zu verstehen.</p> <p>42 Die Informationen im Text knüpften an mein bisheriges Wissen an.</p> <p>43 Mein Vorwissen über Elektrizitätslehre reichte aus, um den Text zu verstehen.</p>
Kognitive Gliederung	<p>4 Ich fand die Reihenfolge der einzelnen Textabschnitte sinnvoll.</p> <p>11 Ich fand die Gliederung des Texts typisch für Physiktexte.</p> <p>12 Ich fand die Informationen im Text wirr.</p> <p>13 Ich konnte gut nachvollziehen, wie die Textinhalte aufeinander aufbauen.</p> <p>17 Ich fand die Einteilung des Texts in seine Abschnitte passend.</p> <p>29 Für mich war beim Lesen eines Textabschnitts stets klar, wie er sich auf die übrigen Absätze bezieht.</p> <p>36 Die einzelnen Abschnitte setzen sich für mich logisch zu einem Gesamtbild der elektrischen Spannung zusammen.</p>
Darstellungsintentionen	<p>28 Mir war klar, warum es sich lohnt, den Text zu lesen.</p> <p>39 Mir war klar, inwiefern der Text mir Erfahrungen aus meiner Lebenswelt erklärt.</p>
Einprägsamkeit	<p>32 Ich konnte mir die Informationen aus dem Text merken.</p>
Unterhaltsamkeit	<p>1 Ich fand den Text humorvoll gestaltet.</p> <p>5 Es hat mir Freude gemacht, den Text zu lesen.</p> <p>8 Ich fand den Text unterhaltsam.</p> <p>26 Es gab überraschende Stellen im Text.</p>

## 6. Methode der empirischen Studie: Material und Durchführung

Motivation	3 Der Text hat mich emotional angesprochen.
	5 Es hat mir Freude gemacht, den Text zu lesen.
	6 Der Text hat mich zum Nachdenken angeregt.
	7 Ich war fasziniert von den Textinhalten.
	8 Ich fand den Text unterhaltsam.
	16 Der Text hat Interesse für das Thema <i>elektrische Spannung</i> bei mir geweckt.
	25 Der Text rief gedankliche Konflikte bei mir hervor.
	26 Es gab überraschende Stellen im Text.
	40 Ich habe mich beim Lesen persönlich angesprochen gefühlt.

---

Tabelle 6.7.: Verteilung der Fragebogenitems auf die aus der Literatur abgeleiteten Wahrnehmungsfacetten. Die Theorie liefert keine Grundlage für eine eindeutige Zuteilung, sodass einzelne Items repräsentativ für mehrere Facetten sein können.

### Testplanung

Der Ansatz, im Rahmen der Verständlichkeitsforschung für (physikalische) Lehrtexte die Lernenden explizit und in differenzierter Weise nach deren Eindruck vom Text zu fragen, ist innovativ und bisher unerforscht. Da dementsprechend nicht auf Vorarbeiten oder einschlägige Erhebungsinstrumente zurückgegriffen werden konnte, wurde für diese Studie ein Fragebogen zur affektiven Textwahrnehmung entwickelt. In der vorliegenden Situation ist somit eine intuitive, definitionsgeleitete Annäherung an den Forschungsgegenstand eine adäquate Strategie für die Testentwicklung (Moosbrugger & Brandt, 2020, S. 71).

Den Grundstein für diese Herangehensweise legt die in Unterkapitel 5.2 beschriebene, eigene theoretische Vorarbeit zum Konstrukt „affektive Textwahrnehmung“. Die aus der Literatur gewonnenen Facetten der affektiven Textwahrnehmung mit deren konkreten Beschreibungen liefern eine umfangreiche, theoriegeleitete Basis für differenzierte, bewertbare Aussagen, die sich auf den Eindruck beziehen, den Lesende vom Text haben. Der Messgegenstand *affektive Textwahrnehmung* erscheint zusammengefasst als multidimensionales Konstrukt, das sich anhand von differenzierten Textbewertungen auf einer äquidistant geschätzten Bewertungsskala intervallskaliert beschreiben lässt. Die tatsächliche Anzahl der Dimensionen war zum Zeitpunkt der Testplanung unklar (Bestimmung erst im Rahmen der Hauptkomponentenanalyse, siehe Abschnitt 7.2.2), jedoch ließ die eigene Auslegung der Literatur zwölf Dimensionen als Richtwert für die Maximalanzahl vermuten (siehe die Tabellen 6.6 und 6.7). Darüber hinaus wird davon ausgegangen, dass die affektive Textwahrnehmung zeitlich konstant ist, dass sie also so, wie sie zum Messzeitpunkt erhoben wird, auch an einem anderen Zeitpunkt, erhoben werden würde (sofern der andere Zeitpunkt nicht erhebliche andere individuelle Lernstands- und Lesevoraussetzungen bedeutet). Ein Rating-Fragebogen, der als Gruppentest im Klassenverbund durchgeführt werden kann, wird daher als geeignetes Messinstrument eingeschätzt. Die Bearbeitungszeit sollte weniger als zehn Minuten betragen, damit das Antwortverhalten der Schüler\*innen möglichst geringfügig durch sinkende Aufmerksamkeit verzerrt wird und

die Durchführung gemeinsam mit den weiteren Studienmaßnahmen problemlos in einer Schulstunde möglich ist. Konkret lassen sich damit etwa 50 Items umsetzen<sup>15</sup>. Indem der Test in Papier-und-Bleistift-Form realisiert wurde, konnten die Testmaterialien zwischen Universität und Schule per Post versandt werden und für die Lehrkräfte ergaben sich weder großer organisatorischer noch technischer Zusatzaufwand.

### Testkonstruktion

Die endgültige Testversion enthält 44 differenzierte Aussagen zu Textwirkungen als bewertbare Items. Diese wurden im Rahmen der Erprobung (siehe den nächsten Abschnitt weiter unten) aus ursprünglich 53 Items herausgearbeitet. Es ist darüber hinaus ein Item zur *globalen Textbewertung* („Ich fand den Text insgesamt verständlich“) vorgesehen.

Tabelle 6.7 zeigt, wie sich die Items auf die aus der Literatur abgeleiteten Wahrnehmungsfacetten verteilen. Für die Testkonstruktion stand im Fokus, dass die Items alle Facetten ausschöpfend abbilden. Eine eindeutige Zuordnung in dem Sinne, dass sich die Facetten inhaltlich nicht überlappen, konnte auf Grundlage der Theorie nicht vorgenommen werden. Die Items evidenzbasiert neu zu gruppieren, also den differenzierten Aussagen zur Textwahrnehmung eine empirisch belegte Struktur einzubeschreiben, ist eines der Forschungsziele dieser Arbeit (siehe dazu Kapitel 4 sowie Abschnitt 7.2.2).

Die Itemformulierungen richten sich eng nach den Beschreibungen der einzelnen Kategorien der Textwirkung (siehe Unterkapitel 5.2). Bei der Umwandlung der Beschreibungen in bewertbare Aussagen zu einem Text wurden etablierte Formulierungsrichtlinien für Fragebogen- und Testitems berücksichtigt (für die sogenannten *10 Gebote der Frageformulierung* siehe Porst, 2014, S. 99 ff.). Die Aussagen sollten auf einer fünfstufigen Skala mit den Niveaus „Stimme gar nicht zu“, „Stimme eher nicht zu“, „Habe keine Tendenz“, „Stimme eher zu“ und „Stimme voll zu“ bewertet werden. Eine in der vorläufigen Testversion vorgesehene „Weiß ich nicht“-Kategorie erwies sich in der Erprobung als unbrauchbar, da die Kategorie nicht quantitativ mit den übrigen Bewertungen verrechnet werden kann und zu erheblichen Hürden in der Datenanalyse (*missing-value*-Probleme) führt. Die ohnehin zur Verfügung stehende Mittelkategorie bündelt somit alle solchen Bewertungen, die keine Zustimmung oder Ablehnung zur Aussage ausdrücken (dies kann der Fall sein, wenn eine Person der Aussage entweder weder zustimmen noch widersprechen *möchte* oder dies z. B. aufgrund von Verständnisproblemen nicht *kann*). Um systematischen Verzerrungen im Antwortverhalten aufgrund von Ermüdungseffekten u. ä. vorzubeugen, wurde die Itemreihenfolge im Bewertungsfragebogen randomisiert. Es wurden daher vier verschiedene Fragebogenversionen erstellt. Die Itemnummerierung in Tabelle 6.7 bezieht sich auf diejenige der vier Versionen, die im Anhang B.2 zu finden ist.

<sup>15</sup>Die Einschätzung folgt der folgenden, großzügigen Hochrechnung: Etwa zehn Sekunden Arbeitszeit pro Item addieren sich bei 50 Items zu etwa 8 Minuten. Dazu werden zwei Minuten für einleitende Worte durch die Lehrkraft eingeplant.

## Erprobung und Überarbeitung

Eine erste Augenscheinvalidierung durch drei Physikdidaktiker\*innen führte zu Überarbeitungen, die in erster Linie sicherstellten, dass alle Items subjektiv aus Sicht der Lesenden formuliert waren (z. B. „Im Text war durchgehend ein roter Faden erkennbar.“ → „Ich konnte im Text durchgehend einen roten Faden erkennen.“), dass redundante Items identifiziert und entfernt wurden und dass Inversionen verwendet wurden, wo diese der Präzision wegen sinnvoll waren (z. B. „Schwierige Aspekte wurden in anderen Worten wiederholt.“ → „Ich hätte bei schwierigen Aspekten noch eine Erklärung aus einer anderen Sicht gebraucht.“).

Eine weitere Erprobungsstrategie bestand in einer Paraphrasierungsstudie mit drei Jugendlichen im Zielgruppenalter (dreizehn bis fünfzehn Jahre). Mit jeder der drei teilnehmenden Personen wurde ein Gespräch geführt, in dem der\*die Teilnehmende die Fragebogenitems jeweils in eigenen Worten wiedergeben sollte. Anhand der sinngemäßen Übereinstimmung von intendierter Bedeutung und (durch die Paraphrasierung zutage tretender) rezipierter Bedeutung kann überprüft werden, ob die Formulierungen für die Zielgruppe verständlich und eindeutig sind. Es erwiesen sich drei Items als problematisch, da zwei von drei Paraphrasierungen als nicht zutreffend zu bewerten waren:

- Ich konnte mir unter den Formulierungen im Text nur vage etwas vorstellen.
- Der Text war abwechslungsreich.
- Ich konnte die Wörter im Text flüssig lesen (ohne auf den Inhalt zu achten).

Im Zuge einer Pilotierungsstudie, in der vier achte Klassen den Fragebogen zur affektiven Textwahrnehmung bearbeiteten, wurden Daten für eine klassische Itemanalyse gesammelt. Die 99 Schüler\*innen lasen dazu eine Vorgängerversion derjenigen Lesetexte zur elektrischen Spannung, die schließlich in der Hauptstudie verwendet wurden, und bearbeiteten direkt im Anschluss den Fragebogen. Die Testkennwerte, wie sie oben zusammengefasst und im Anhang D.1 (Tabelle D.3) aufgeführt sind, gaben keinen Anlass für die Aussortierung einzelner Items. Erwartungswidrige Korrelationen einzelner Items mit dem Globalurteil traten sehr selten auf (lediglich bei zwei von 44 Items, siehe *Testkennwerte und Testinhalte* oben) und waren sehr schwach ausgeprägt. Mangels einer fundierten Theorie über eindimensionale Teilkonstrukte, welche eine innere Konsistenz bestimmter Itemgruppen induzieren würde (vergleiche Tabelle 6.6, die einen vorläufigen, theoriegeleiteten Gruppierungsversuch zeigt), bietet auch die Berechnung von Cronbachs  $\alpha$  keine Grundlage für die Itemselektion.

Das Vorgehen bei der Itemanalyse stützte sich stattdessen auf iterativ wiederholte Hauptkomponentenanalysen (Abkürzung *PCA* für die englische Übersetzung *Principal Component Analysis*) zur Identifikation und Eliminierung schwacher Items (siehe Abbildung 6.2). Das im folgenden beschriebene Vorgehen richtet sich in weiten Teilen nach Brandt (2020). Näheres zur PCA zum Zweck einer Dimensionsanalyse wird in Abschnitt 6.4.3 beschrieben. Da die Datenmenge ( $n_{\text{PilotierungFB}} = 99$ ) gemessen an der Anzahl der zu untersuchenden Items ( $n_{\text{ItemsPilotierung}} = 53$ ) sehr gering war, ergaben sich die folgenden Konsequenzen:

- Eine PCA basiert auf Item-Item-Kovarianzen. Die Kovarianz zwischen zwei Items kann nur auf vollständigen Datensätzen berechnet werden, das heißt der Umgang mit ein-

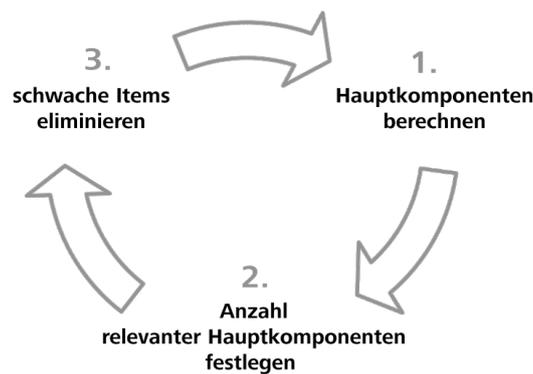


Abbildung 6.2.: Einzelschritte einer Iteration im Itemanalyseprozess mithilfe einer PCA (Principal Component Analysis). Auf dem Datensatz von  $n_1$  Items werden auf Grundlage von Kovarianzen  $n = n_1$  Hauptkomponenten berechnet. Aus diesen werden  $k < n$  relevante Hauptkomponenten ausgewählt. Gängige Kriterien hierfür sind im Fließtext beschrieben. Die Verteilung der Items auf die Hauptkomponenten bildet die Grundlage für eine Itemselektion. Auf dem neuen Datensatz von  $n_2 < n_1$  Items startet die nächste Iteration.

zelen fehlenden Datenpunkten ist entscheidend. In diesem Fall wurde der Ansatz der sogenannten *pairwise deletion* gewählt. Der Ansatz sei an einem Beispiel erläutert: Person A hat zu Item 1 keine Bewertung abgegeben, zu allen anderen Items schon. Es soll nun die Korrelation zwischen Item 1 und Item 2 berechnet werden. Dazu werden alle Bewertungen zu Item 1 mit den entsprechenden Bewertungen zu Item 2 verglichen. Da Person A zu Item 1 keine Bewertung abgegeben hat, muss Person A aus der Korrelationsberechnung von Item 1 und Item 2 entfernt werden, kann aber für alle Korrelationsberechnungen, die sich nicht auf Item 1 beziehen, verwendet werden. Das heißt: Für jedes einzelne Itempaar müssen genau die Personen-Datensätze entfernt werden, die auf einem der beiden Items einen fehlenden Wert aufweisen. Die oben erwähnte „Weiß ich nicht“-Kategorie verursachte eine große Menge solcher fehlender Werte (335 von 5346 Werten, also etwas mehr als 6 %, nur 18 von 99 Personen hatten einen vollständigen Datensatz geliefert) und wurde daher für die Hauptstudie entfernt. Da obendrein der Stichprobenumfang in der Pilotierung sehr gering war, waren einfacher zu implementierende *listwise deletion*- oder alternative *imputation*-Methoden für den Umgang mit *missing values* nicht sinnvoll.

- Das zentrale Ziel der PCA ist die Dimensionsreduktion, das heißt, man versucht die Varianz, die durch die Daten erzeugt wird, anhand weniger Hauptkomponenten zu erklären. Eine Hauptkomponente ist dabei eine Linearkombination aus einzelnen Items und fasst somit Items zu Gruppen zusammen. Zunächst werden bei der Analyse genauso viele Hauptkomponenten wie Items gebildet, die Varianz ist damit zu 100 % erklärt. Es wird dann auf explorativem Weg entschieden, wie viele Hauptkomponenten *relevant* sind. Das bedeutet festzulegen, auf wie viele Dimensionen die Darstellung reduziert werden kann, um die Daten noch zutreffend genug widerzuspiegeln. Dabei gibt es keine präzisen Regeln,

sondern die analysierende Person legt die Anzahl auf der Grundlage unterstützender, unscharfer Entscheidungskriterien fest: Das Eigenwertkriterium (auch: Kaiser-Kriterium), das Ellbogenkriterium (auch: Scree-Critereon) und das Kriterium der minimal erklärten Varianz (jeweils genauer erläutert in Brandt und Moosbrugger, 2020, S. 590 ff.) geben Richtlinien für eine plausible Anzahl an Hauptkomponenten. In dieser Darstellung zeigt sich nun, welche Items gut in die Datenstruktur „passen“ und welche nicht. Diejenigen Items, die entweder kaum durch die Hauptkomponenten repräsentiert sind (schwache *Faktorladungen* des Items auf allen Komponenten; entsprechender Kennwert: *Kommunalität*) oder über alle Hauptkomponenten hinweg „verschmiert“ sind (*Faktorladungen* vergleichbarer Stärke auf mehreren Komponenten; entsprechender Kennwert: *Komplexität*), stören das Ziel, die Daten durch wenige Hauptkomponenten hinreichend präzise wiederzugeben. Solche Items werden als schwach identifiziert und aus dem Datensatz eliminiert.

Um die geringe Datenmenge in der Pilotierung bestmöglich als Entscheidungsgrundlage zu nutzen, wurden jeweils mehrere Möglichkeiten für eine sinnvolle Dimensionsanzahl angenommen und die Iterationsschritte der Itemanalyse somit auf verzweigten Pfaden weitergeführt (etwa: Im ersten Iterationsschritt erscheinen 9 oder 15 Komponenten sinnvoll; beide Anzahlen werden für die Elimination schwacher Items weitergeführt. Auf den zwei verschiedenen, verbleibenden Itemsets werden erneut *jeweils* eine PCA durchgeführt, um weitere Items zu eliminieren). Nach drei PCA-Eliminierungs-Iterationen auf jedem Pfad der Analyse wurde schließlich verglichen, welche Items in *jedem Pfad* schwach erscheinen.

Zusammengefasst wurden zusätzlich zu den oben genannten Items, die sich in der Paraphrasierung als problematisch darstellten, weitere sechs Items entfernt:

- Ich fand die Sätze im Text lang.
- Die Wörter im Text waren mir fremd.
- Ich fand die Schrift leserlich.
- Ich fand die Wörter im Text mehrdeutig.
- Mir war klar, inwiefern der Text relevant für mich ist.
- Ich musste mir zusammengehörige Wortteile im Satz zusammensuchen.

Weitere Testanpassungen etwa auf Grundlage der Testkennwerte (siehe *Testkennwerte und Testinhalte* oben) wurden nicht vorgenommen. Der Fragebogen, wie er in der Hauptstudie eingesetzt wurde, ist im Anhang B.2 zu finden. Ohne das Ziel einer erneuten Test-Überarbeitung, sondern zu Vergleichs- und Kontrollzwecken, wurden die gleichen Testkennwerte wie in der Pilotierung auch in der Hauptstudie berechnet. Die Kontrolle fiel zufriedenstellend aus. Die Kennwerte auf Basis der Hauptstudien Daten ( $n_{\text{HauptstudieFB}} = 781$ ) geben keinen Anlass, die Validität oder Reliabilität des Fragebogens erneut in Frage zu stellen. Die Item-Mittelwerte, die Item-Schwierigkeiten, die Pearson-Produkt-Moment-Korrelationen der Items mit dem Globalurteil und die Varianzen der einzelnen Items sind in Tabelle D.5 im Anhang D.2 dargestellt. Erneut zeigen sich die erwartungswidrigen Korrelationen bei den Items „25. Der

Text rief gedankliche Konflikte bei mir hervor.“ ( $r = -0.26$ ) und „26. Es gab Stellen im Text, die überraschend für mich waren.“ ( $r = -0.02$ ). Die Korrelation des Items „33. Ich habe Wissen, das nichts mit Elektrizitätslehre zu tun hat, gebraucht um den Text zu verstehen.“ fällt nun nicht mehr entgegen der Erwartung aus ( $r = 0.01$ ), jedoch zeigt sich nun bei Item „11. Ich fand die Gliederung des Texts typisch für Physiktexte.“ unerwartet eine (schwache) negative Korrelation ( $r = -0.12$ ).

## 6.4 Statistische Datenauswertung

Dieser Studie liegt die Digitalisierung, statistische Aufbereitung und softwaregestützte Analyse von je 787 Lesetests, Bewertungsfragebögen, Wissens-Pre-Tests und Wissens-Post-Tests zugrunde. Alle Daten lagen zunächst in Papierform vor und sind quantitativer Natur. Dieses Kapitel stellt sukzessive dar, wie ausgehend von Papierdaten die digitale Verarbeitung vorbereitet wurde (Abschnitt 6.4.1), welche Software zum Einsatz kam (Abschnitt 6.4.2) und anhand welcher statistischer Methoden die Erkenntnisse gewonnen wurden (Abschnitt 6.4.3).

### 6.4.1 Organisatorische Vorbereitungen für die Datenverarbeitung

Da die Datenanalyse unter anderem die Auswertung eines Fragebogens mit mehr als 40 Items einschließen würde, war schon im Vorfeld der Hauptuntersuchung klar, dass ein großer Stichprobenumfang von mindestens 600 Teilnehmenden nötig ist. Für die Digitalisierung einer solch großen Menge an analogen Daten aus Papier- und Bleistift-Tests mussten wohlüberlegte Vorbereitungsmaßnahmen getroffen werden. Erstens musste die manuelle Eingabe der Daten möglichst schnell umsetzbar sein und zweitens musste ein algorithmisierbarer (und damit automatisierbarer) Einlesevorgang der digitalen Datentabellen in die Statistikprogramme ermöglicht werden. Um die Digitalisierung auf mehrere Personen aufteilen zu können und dabei die Einheitlichkeit der Dateneingabe sicherzustellen, wurden exakte schrittweise Arbeitsanweisungen inklusive eines ausführlichen Codebooks erstellt<sup>16</sup> sowie jeweils leicht handhabbare Eingabemasken für den Bewertungsfragebogen, den Wissenstest und den Lesetest erstellt. Gemeinsam mit präzisen Durchführungshinweisen für die teilnehmenden Lehrkräfte (genauer dazu im folgenden Unterkapitel 6.5<sup>17</sup>) wurde so die Grundlage für zeitlich optimierte Eingabevorgänge und einen automatisierten Datenimport von den Eingabemasken in das Statistikprogramm geschaffen.

<sup>16</sup>Das Codebook ist im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> einsehbar.

<sup>17</sup>Die vollständigen Durchführungshinweise sind im digitalen Anhang (<http://doi.org/10.5283/epub.58778>) hinterlegt

## 6.4.2 Verwendete Software

Die Datenorganisation und Datenanalyse wurde überwiegend mithilfe der Programmiersprache PYTHON (Version 3.10., Python Software Foundation, www.python.org) durchgeführt, wobei die Bibliotheken PANDAS (Version 1.3.5, The pandas development team, 2020; Wes McKinney, 2010), SCIPY (Version 1.8.1, Virtanen et al., 2020), STATSMODELS (Version 0.13.2, Seabold und Perktold, 2010), PINGOUIN (Version 0.5.3, Vallat, 2018), RESEARCHPY (Version 0.3.5, Bryant, 2018) und SCIKIT-LEARN (Version 1.3.0, Pedregosa et al., 2011) verwendet wurden. Graphische Auswertungen wurden mit MATPLOTLIB (Version 3.5.2, Hunter, 2007) und PGF/TIKZ (Version 3.1.9a, HTML-Version der Dokumentation: tikz.dev) erzeugt.

Für die ANOVA (Analysis of Variance) wurde das *psych*-Paket der Statistiksoftware *R* mit den vorimplementierten Rotationsmethoden des *GPArotation*-Pakets verwendet. Die Berechnung des ordinalen Kennwerts  $\alpha_{\text{ord}}$  für die innere Konsistenz einer Skala erfolgte ebenfalls mithilfe des *psych*-Pakets in *R*<sup>18</sup>.

## 6.4.3 Statistische Methoden im Einzelnen

Bei der Erprobung der Testinstrumente in der Pilotierungsstudie sowie zum Zweck der Erkenntnisgewinnung in der Hauptstudie kamen verschiedene statistische Methoden zum Einsatz. Diese seien hier im Einzelnen erläutert.

Unter einem Mittelwert  $M$  (auch:  $\bar{x}$ ) wird jeweils das arithmetische Mittel einer Datenreihe verstanden. Aus den Abweichungen einzelner Werte vom Mittelwert lässt sich die Varianz, ein Maß für die Streuung der Messwerte, berechnen:  $\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ . Der Laufindex  $i$  steht dabei für eine Testperson, die Stichprobengröße ist  $n$ . Die Wurzel aus der Varianz ist die (für die Population geschätzte) Standardabweichung  $\hat{\sigma}_x$ <sup>19</sup>.

Die Schwierigkeit eines Items  $k$  in einem Test wird anhand des Schwierigkeitsindex'  $p_k = \frac{\sum_{i=1}^n x_{ki}}{x_{\text{max},k} \cdot n}$  berichtet (vergleiche Bortz & Döring, 2006, S.218 f.). Dabei ist  $x_{ki}$  jeweils der Datenwert von Person  $i$  bei Item  $k$  und  $x_{\text{max},k}$  der maximal erreichbare Itemwert. Der Index  $p_k$  liegt zwischen 0 und 1 und nähert sich 1, wenn die einzelnen Bewertungen in vielen Fällen im Datensatz nahe der Maximalbewertung liegen – im Falle eines Leistungstests also dann, wenn das Item insgesamt gut bearbeitet wird. Der Schwierigkeitsindex ist damit für einfachere Items größer.

Der Zusammenhang zweier quantitativer Variablen auf Intervallskalenniveau  $x$  und  $y$ , die für jeden Merkmalsträger  $i$  bestimmte Ausprägungswerte  $x_i$  und  $y_i$  annehmen, wird mit der Produkt-Moment-Korrelation nach Pearson angegeben (vergleiche Rasch, Friese,

---

<sup>18</sup>Der Grund für eine stellenweise Verwendung von *R* liegt darin, dass Python nicht spezifisch für humanwissenschaftliche Zwecke optimiert und ausgelegt ist. Aufwendigere Funktionen für solche statistischen Auswertungsbausteine wie rotierte Faktorladungsmatrizen oder polychore Korrelationsmatrizen, die im Allgemeinen seltener benutzt werden, sind daher nicht vorimplementiert.

<sup>19</sup>Das Zeichen  $\hat{\phantom{x}}$  zeigt an, dass es sich um eine aus den vorliegenden Stichprobendaten geschätzte Größe für die gesamte Population handelt. Von dieser Schätzung rührt auch das „-1“ im Nenner her (vergleiche Rasch, Friese, Hofmann & Naumann, 2014a, S. 14).

Hofmann & Naumann, 2014a, S. 85 f.). Der zugehörige Korrelationskoeffizient  $r = \frac{cov(x,y)}{\hat{\sigma}_x \cdot \hat{\sigma}_y}$  ist die standardisierte Kovarianz der beiden Variablen. Dabei gilt für die Kovarianz  $cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$ , dass sie eine Aussage über die gemeinsame Variation von  $x$  und  $y$  macht und mit steigender Stärke des Zusammenhangs größer wird.

Insbesondere wird im Zusammenhang mit Testitems die *Pearson-Produkt-Moment-Korrelation*  $r_{kt}$  zwischen einem Wert  $x_k$  für das Item  $k$  und dem Gesamt-Testwert  $t$  (falls dieser sinnvoll berechnet werden kann, etwa als Summe aller Teilleistungen) als Item-Trennschärfe bezeichnet (vergleiche Bortz & Döring, 2006, S. 219):  $r_{kt} = \frac{cov(x_k, t)}{\hat{\sigma}_{x_k} \cdot \hat{\sigma}_t}$ . Der Item-Trennschärfe-Koeffizient ist ein Zahlenmaß dafür, wie gut ein bestimmter Item-Wert das Gesamt-Testergebnis vorhersagen kann. Im hier vorliegenden Fall sind die Gesamt-Testwerte jeweils part-whole-korrigiert (vergleiche Kelava & Moosbrugger, 2020, S. 154). Das bedeutet, dass der Gesamt-Test-Score  $t$  aus allen Test-Items *außer* dem Item  $k$  ermittelt wird. Auf diese Weise vermeidet man eine positive Verzerrung des Trennschärfe-Koeffizienten, die dadurch zustande käme, dass eine gute Bearbeitung des Items  $k$  allein aus mathematischen Gründen ein besseres Testergebnis  $t$  bewirkt.

Von der *Pearson-Produkt-Moment-Korrelation* macht auch der Testkennwert  $\alpha$  Gebrauch, welcher üblicherweise als Maß für die innere Konsistenz dient. Der Kennwert geht auf Lee J. Cronbach (1951) zurück und wird daher auch Cronbachs  $\alpha$  genannt. Da  $\alpha$  auf Inter-Item-Korrelationen beruht und dabei den Anteil an Varianz aller einbezogenen Items angibt, welcher durch *eine* Merkmalsdimension erzeugt wird, steigt der Wert von Cronbachs  $\alpha$  mit steigender Homogenität der Items (vergleiche Bortz & Döring, 2006, S. 198 f.). Es gilt  $\alpha = \frac{m}{m-1} \left( 1 - \frac{\sum_{k=1}^m \sigma_{x_k}^2}{\sigma_t^2} \right)$  mit der Anzahl aller einbezogenen Items  $m$  und den Bezeichnungen wie bei der Item-Trennschärfe  $r_{kt}$  oben. Cronbachs  $\alpha$  wächst darüber hinaus auch mit größer werdender Anzahl der einbezogenen Items – dies ist bei der Interpretation zu beachten. Für ordinale Daten, die keine kontinuierliche Skala aufweisen, sind oftmals korrigierte, sogenannte ordinale Korrelationskoeffizienten zutreffender. Denn für diskrete Skalen unterschätzen die regulären Korrelationskoeffizienten im Allgemeinen den tatsächlichen Zusammenhang (Gadermann, Guhn & Zumbo, 2019). Ein ordinale Cronbachs  $\alpha$  kann mithilfe der *polychoric*-Funktion des *psych*-Pakets der Statistik-Software R berechnet werden.

Für die Erkenntnisgewinnung in der Hauptstudie stand im Zentrum, den Einfluss der sprachlichen Gestaltung (unabhängige Variable) auf verschiedene abhängige Variablen zu untersuchen. In diesem Zusammenhang sind in vielen Analyseschritten die Frage nach der Zufälligkeit (Signifikanz) und nach dem Ausmaß der Bedeutsamkeit (Effektstärke) von Mittelwertsunterschieden zu klären. Für den Fall, dass Unterschiede zwischen genau zwei Gruppen betrachtet werden, liefert der sogenannte t-Test gefolgt von einer Effektstärkeberechnung Antworten auf diese Fragen (z. B. Rasch, Friese, Hofmann & Naumann, 2014a, Kap. 3). Bei mehreren Gruppen, die zu vergleichen sind, gibt eine ANOVA (Analysis Of Variance) Auskunft darüber, ob signifikante Mittelwertsunterschiede vorliegen (z. B. Rasch, Friese, Hofmann & Naumann, 2014b, Kap. 5 und 6). Eine grundlegende Voraussetzung für ANOVAs ist, dass die zu vergleichenden Gruppen hinsichtlich jedes Faktors unabhängig sind. Dies ist durch

das Design der Studie gegeben: ein Text ist *entweder* der anspruchsgeminderten, *oder* anspruchsgesteigerten Ausprägung einer sprachlichen Variable zugeordnet. Die ANOVA sieht das RESEARCHPY-Paket mit dem Befehl *anova()* für pythonbasierte Statistik vor; für die Berechnung der *sum-of-squares* wurde für diese Arbeit der Typ III gewählt. Dieser eignet sich nach Field, Miles und Field (2012, S. 475 f.) am besten für die Berechnung von Haupteffekten, wenn diese im Zentrum des Interesses liegen, aber Interaktionseffekte nicht auszuschließen sind. Die Frage nach der Signifikanz wird in dieser Arbeit stets auf dem Level einer  $\alpha$ -Fehler-Wahrscheinlichkeit von  $\alpha = 5\%$  entschieden. Entsprechend werden Ergebnisse mit  $0.01 < p \leq 0.05$  als signifikant (\*), solche mit  $0.001 < p \leq 0.01$  als hoch signifikant (\*\*) und solche mit  $p \leq 0.001$  als höchst signifikant (\*\*\*) bezeichnet. Falls mehrere Mittelwertsvergleiche auf dem gleichen Datensatz durchgeführt werden, wird der  $\alpha$ -Fehler-Kumulierung gemäß der Bonferroni-Holm-Korrektur entgegengewirkt. Dabei werden die Kontraste vom größten zum kleinsten sortiert und sequentiell jeweils gegen größer werdende Signifikanzlevel  $\lambda$  ( $\lambda = 0, 1, 2, \dots$ ) mit einer Fehlerwahrscheinlichkeit von  $\alpha_\lambda = \frac{\alpha}{m-\lambda}$  ( $m =$  Anzahl der Kontraste) getestet (vergleiche dazu z. B. Bortz und Schuster, 2010, S. 232). Als Effektstärkemaß wird in dieser Arbeit der Anteil  $\omega^2$  der partiellen Effektvarianz an der Gesamtvarianz berichtet; nach Rasch, Friese, Hofmann und Naumann (2014a, S. 65, für t-Tests) und Rasch, Friese, Hofmann und Naumann (2014b, S. 23 f., für die Varianzanalyse) ist diese Größe ein geeigneter Schätzer, um die Effektstärke in der Population zu beziffern. Um Effektstärken hinsichtlich ihrer Größe einzuordnen, werden die Richtwerte 0.01 für kleine Effekte, 0.06 für mittlere Effekte und 0.14 für große Effekte – siehe Rasch, Friese, Hofmann und Naumann, 2014a, S.65 – herangezogen. Der Vergleich mit Richtwerten kann im Allgemeinen nur als sehr grobe Orientierung für die größenmäßige Einordnung der Effektstärke dienen. Aus diesem Grund werden im Ergebnisteil stets die quantitativen Rohwerte berichtet. Die Gültigkeit der Aussagen, die sich aus einem t-Test oder einer ANOVA ableiten lassen, unterliegt den Bedingungen, dass das untersuchte Merkmal Intervallskalenniveau aufweist, dass es in der Population normalverteilt ist und dass die Varianzen in den zu vergleichenden Gruppen homogen, also in etwa gleich groß sind. Intervallskalenniveau liegt für die abhängigen Variablen dieser Studie (Credits im Wissenstest beziehungsweise Bewertungen im Fragebogen) generell vor. Darüber hinaus sind t-Test und ANOVA unter den in dieser Studie vorliegenden Versuchsbedingungen robust gegenüber Voraussetzungsverletzung: Wenn die zu vergleichenden Stichprobenumfänge ungefähr gleich sind und größer als dreißig ( $n_i \approx n_j > 30 \forall$  Personen  $i, j$ ), gelten die Informationen aus den Tests dennoch als zuverlässig (Rasch, Friese, Hofmann & Naumann, 2014a, S. 43; Rasch, Friese, Hofmann & Naumann, 2014b, S. 30 f.). Aus diesem Grund wird jeweils angenommen, dass t-Test und ANOVA aussagekräftig sind, solange die Robustheitsanalysen keine erhebliche Voraussetzungsverletzung anzeigen. Um sicherzugehen, dass nicht-signifikante Ergebnisse mit großer Wahrscheinlichkeit darauf zurückzuführen sind, dass tatsächlich keine Unterschiede zwischen den Gruppen vorliegen, wird mit einem Post-Hoc-Sensitivitätstest mittels des Programms *G\*Power* (Faul, Erdfelder, Buchner & Lang, 2009) überprüft, ob der Test in der Lage gewesen wäre, einen kleinen Effekt zu finden (vgl. Rasch, Friese, Hofmann

& Naumann, 2014b, S. 57). Hierbei wurde die entsprechende Fehlerwahrscheinlichkeit dem Richtwert von Rasch, Frieze, Hofmann und Naumann folgend auf  $\beta = 10\%$  festgelegt.

Darüber hinaus lag es im Forschungsinteresse dieser Studie, die affektive Textwahrnehmung von Schüler\*innen möglichst differenziert zu beschreiben und zu analysieren. Es wurde zu diesem Zweck eine Hauptkomponentenanalyse (Abkürzung PCA für Principal Component Analysis) vorgenommen, die der Textwahrnehmung auf Grundlage der Fragebogendaten eine empirisch gewonnene Struktur einbeschreibt<sup>20</sup>. Auf diese Art konnte mit einer handlicheren (das heißt: in ihren Dimensionen reduzierten), inhaltlich schlüssigen und auf Evidenz beruhenden Datenstruktur weitergearbeitet werden. Der mathematische Algorithmus (siehe dazu Abschnitt 6.3.3) ist iterativ angelegt und sieht vor, dass die erste Hauptkomponente möglichst viel Gesamtvarianz erklärt. Entsprechend werden viele Fragebogenitems in ihr abgebildet. Die restliche Varianz wird dann möglichst ausschöpfend durch die zweite Hauptkomponente erklärt. Bezogen auf die Items ist die Komponente somit am zweitwichtigsten. Dies lässt sich in einer Datenmenge aus  $n$  Datensätzen genau  $n$  mal fortsetzen. Alle  $n$  Komponenten erklären dann 100 % der Varianz und tragen von der ersten bis zur  $n$ -ten Komponente immer weniger Bedeutsamkeit. Wird jedoch das Komponentensystem rotiert, so lässt sich erreichen, dass sich die Fragebogenitems gleichmäßiger auf die Hauptkomponenten verteilen und dass nicht eine oder zwei Komponenten entstehen, die nahezu die gesamte Information in sich tragen. Die Ergebnisse in Form der extrahierten Hauptkomponenten werden besser inhaltlich interpretierbar (Field, Miles & Field, 2012, S. 764 f.). Für diese Arbeit wurde daher eine PCA mit Oblimin-Rotation durchgeführt. Dazu diente der Befehl *principal()* mit *oblimin*-Option des *psych*-Pakets in R, da in Python keine vorimplementierten Rotationsmechanismen zur Verfügung stehen. Die Oblimin-Rotation ist eine schiefe Rotationsmethode. Das bedeutet, dass die berechneten Hauptkomponenten nicht zwingend orthogonal zueinander (im übertragenen Sinne: potenziell korreliert) sind. Im hier vorliegenden Fall ist anzunehmen, dass die gefundenen Hauptkomponenten (im übertragenen Sinne: Facetten der Textwahrnehmung) inhaltlich nicht gänzlich unabhängig voneinander sind, sodass schiefe Rotationsmethoden die Realität besser abbilden als orthogonale (vergleiche Costello & Osborne, 2005).

## 6.5 Durchführung und Stichprobe

An der Studie haben insgesamt 787 Schüler\*innen der Mittelstufe an bayerischen Gymnasien und Realschulen teilgenommen. Das Erkenntnisinteresse der Forschungsarbeit bezieht sich auf Zusammenhänge zwischen der sprachlichen Gestaltung eines physikalischen Lehrtexts und seiner Wirkung auf Lesende (affektive Textwahrnehmung sowie Behaltensleistung). Es ist davon auszugehen, dass in diesem Wirkungsgefüge den Lesefähigkeiten der Teilnehmenden eine mediatorische Rolle zuteil wird. Es werden daher die Lesefähigkeiten gemessen,

<sup>20</sup>Die PCA wurde in ihrem Zweck für eine Itemanalyse bereits in Abschnitt 6.3.3 erläutert. Die Ausführungen an dieser Stelle beleuchten die PCA vor dem Hintergrund einer Dimensionsreduktion in großen Datenmengen.

ein Wissenstest im Pre-Post-Design durchgeführt und die Textwahrnehmung in Form eines Bewertungsfragebogens erhoben. Weitere Begleitvariablen, deren statistische Kontrolle dem Erkenntnisinteresse dient und die daher miterhoben werden, sind die Genderzugehörigkeit, die Schulart und die Jahrgangsstufe der einzelnen Teilnehmenden (siehe dazu auch Unterkapitel 6.1 oben).

### **6.5.1 Zeitlicher und organisatorischer Ablauf der Studie in Schulen**

Die Studie wurde auf Grundlage einer Erhebungsgenehmigung für öffentliche Schulen in Bayern durchgeführt. Diese Genehmigung wurde durch das Bayerische Staatsministerium für Unterricht und Kultus erteilt. Im Vorfeld der Datenerhebung an den einzelnen Schulen wurde die Erlaubnis der jeweiligen Schulleitung eingeholt und Lehrkräfte von infrage kommenden Klassen wurden per Email detailliert über den inhaltlichen, organisatorischen und datenschutzrechtlichen Kontext der Studie informiert und ihre Teilnahmebereitschaft erfragt. Über die teilnahmebereiten Lehrkräfte wurden weitere Informationsschreiben an die Schüler\*innen und deren Erziehungsberechtigte ausgehändigt. Der Brief an die Schulleitungen ist im digitalen Anhang (<http://doi.org/10.5283/epub.58778>) einsehbar; die Informationen für die Lehrkräfte und Erziehungsberechtigten folgten im Wortlaut stark dem Schreiben an die Schulleitungen. Für die Schüler\*innen wurden die Formulierungen in eine Form gebracht, die möglichst kindgerecht ist und dennoch den rechtlichen Ansprüchen genügt. An die Lehrkräfte wurden darüber hinaus präzise Durchführungsanweisungen verschickt (ebenfalls unter <http://doi.org/10.5283/epub.58778> einsehbar) und in persönlichen Telefongesprächen beziehungsweise E-mailkorrespondenzen besprochen. Auf dieser Grundlage trugen sie dafür Sorge, dass die Studie in den Klassen plangemäß durchgeführt wurde und nur Daten von Schüler\*innen, die gemeinsam mit ihren Erziehungsberechtigten explizit der Studienteilnahme zugestimmt haben, zur Verwendung für das Forschungsvorhaben an die Universität zurückgeschickt wurden.

Die Erhebung erfolgte anhand eines zweiteiligen Testhefts vollständig in Papier- und Bleistift-Form im regulären Physikunterricht unter Anleitung der jeweiligen Physiklehrkraft. Versuchsleitereffekten oder systematischen Verzerrungen in den Daten aufgrund der Klassenzugehörigkeit wurde vorgebeugt, indem die Teilnahmeinstruktionen standardisiert und schriftlich in den Testheften festgehalten wurden und die Studienteilnahme der einzelnen Klassen an die Bedingung geknüpft war, dass die Schüler\*innen noch keinen Vorunterricht zur elektrischen Spannung erfahren hatten oder ein solcher Vorunterricht mehr als ein Jahr zurücklag. Die Texte stellten also gewissermaßen einen „ersten Berührungspunkt“ mit der physikalischen Größe der elektrischen Spannung dar; klassenverbundsspezifisches inhaltliches Vorwissen war nicht zu erwarten. Indem die für die Schüler\*innen gewohnten Lehrkräfte die Durchführung anleiteten, blieb eine realistische Unterrichtssituation erhalten und die Rückmeldung der Schüler\*innen wurde nicht durch Effekte der sozialen Erwünschtheit beeinflusst. Der zeitliche Umfang betrug dabei zwei Unterrichtsstunden. Der genaue Ablauf ist in Tabel-

<b>Erste Schulstunde</b>	
Die Lehrkraft leitet die Konstruktion der Pseudonyme für die Teilnehmenden an und händigt den ersten Teil des Testgehefts aus.	Dauer ca. 5 Min.
Unter dem jeweiligen Pseudonym bearbeiten die Teilnehmenden den ersten Teil des Erhebungsgehefts. Dieser Teil umfasst: <ul style="list-style-type: none"> <li>• Ausfüllen des Geheft-Deckblatts,</li> <li>• Pre-Test Wissen,</li> <li>• Lesefähigkeitstest.</li> </ul>	Dauer insgesamt ca. 20 Min.: <ul style="list-style-type: none"> <li>• 1 Min.</li> <li>• 6 Min.</li> <li>• 12 Min.</li> </ul>
<b>Zweite Schulstunde</b> (sieben bis achtundzwanzig Tage später)	
Die Lehrkraft verteilt den zweiten Teil des Testgehefts entsprechend der Pseudonyme.	Dauer < 1 Min.
Unter dem jeweiligen Pseudonym bearbeiten die Schülerinnen und Schüler den zweiten Teil des Testgehefts. Dieser Teil umfasst: <ul style="list-style-type: none"> <li>• Textlektüre,</li> <li>• Bewertungsfragebogen,</li> <li>• Posttest Wissen.</li> </ul>	Dauer insgesamt ca. 35 Min.: <ul style="list-style-type: none"> <li>• 10 Min.</li> <li>• 12 Min.</li> <li>• 6 Min.</li> </ul>

Tabelle 6.8.: Zeitlicher Ablauf der empirischen Studie an den Schulen. Beide Teile waren im zeitlichen Rahmen jeweils einer Schulstunde durchführbar.

le 6.8 dargestellt. Der erste Teil enthielt ein Deckblatt, auf dem die Personendaten einzutragen waren, den Wissens-Pre-Test (siehe Abschnitt 6.3.2) sowie den Lesetest (siehe Abschnitt 6.3.1). Nach sieben bis achtundzwanzig Tagen fand der zweite Erhebungsteil statt, der die Textlektüre (zur Gestaltung der Texte siehe Unterkapitel 6.2), den Bewertungsfragebogen zur Textwahrnehmung (siehe Abschnitt 6.3.3) und den Post-Wissenstest (siehe Abschnitt 6.3.2) umfasste. Die Phase der Textlektüre war nicht zeitlich begrenzt: die Lehrkräfte sollten den Schüler\*innen so viel Zeit einräumen, wie jede\*r Einzelne zum aufmerksamen Lesen benötigte. Damit zusammengehörige Testgeheft-Teile einander zugeordnet werden konnten, vermerkten die Teilnehmenden ein (unter Anleitung der Lehrkraft) selbst konstruiertes Pseudonym auf beiden Teilen. Bei der Digitalisierung wurden die Pseudonyme nicht gespeichert, sodass eine Reidentifizierung ausgeschlossen war und die weitere Datenverarbeitung anonym erfolgte.

Den Abschluss der Kooperation mit den Lehrkräften bildete der Versand individueller, schlaglichtartiger Ergebnisberichte für einzelne Klassen. Diese enthielten neben dem Text-Ranking hinsichtlich der Behaltensleistung und demjenigen hinsichtlich der Globalbewertung in der Gesamtstichprobe auch schulklassenspezifische Ergebnisse bezüglich der Leseleistung und des erreichten Lernstands über die elektrische Spannung. Technisch wurde dies umgesetzt, indem eine Briefvorlage erstellt wurde, welche automatisiert mit den klassenspezifischen Ergebnissen gefüllt wurde. Die Ergebnisse wurden dabei aus einer passend strukturierten Datendatei eingelesen. Ein Musterexemplar dieser Ergebnisberichte ist im digitalen Anhang (<http://doi.org/10.5283/epub.58778>) eingefügt.

## 6. Methode der empirischen Studie: Material und Durchführung

Gesamt	Gender	Schulart	Jahrgangsstufe	Textversion
787	w: 475 (60 %) m: 282 (36 %) d: 30 (4 %)	Gym: 492 (63 %) RS: 295 (37 %)	8: 147 (19 %) 9: 640 (81 %)	A: 95 (12 %) B: 91 (12 %) C: 97 (12 %) D: 87 (11 %) E: 99 (13 %) F: 102 (13 %) G: 105 (13 %) H: 111 (14 %)

Tabelle 6.9.: Überblick über die Zusammensetzung der Gesamtstichprobe inklusive der Verteilung der gelesenen Textversion (die Versionen sind in Tabelle 6.4 erläutert). Zu den übrigen Abkürzungen: *m*, *w* und *d* stehen für männlich, weiblich und divers; *Gym* und *RS* stehen für Gymnasium und Realschule.-

### 6.5.2 Beschreibung der Stichprobe

Im Zuge der Stichprobenrekrutierung konnten das Einverständnis von 21 Schulleitungen und die Teilnahmebereitschaft von 45 Schulklassen gewonnen werden. Das bedeutete den Versand von 1207 Testheften. Insgesamt nahmen schließlich 787 Schüler\*innen an der Datenerhebung teil, wobei „Teilnahme“ bedeutet, dass sowohl der erste als auch der zweite Teil des Testhefts bearbeitet wurden und das Einverständnis zur Auswertung erteilt worden war.

Die Textversionen wurden über die gesamte Stichprobe hinweg vollständig randomisiert verteilt. Es ist dadurch sichergestellt, dass sich die verschiedenen Textversionen nicht systematisch in einzelnen Schulen oder Klassen häufen.

*If we have data, let's look at data.  
If all we have are opinions, let's go with mine.*

---

James L. Barksdale  
Ehemaliger CEO der *Netscape* Communications Corporation.

## 7 | ERGEBNISSE UND DISKUSSION

Mit einem kleinen Augenzwinkern soll das obige Zitat von James L. Barksdale dieses Kapitel einleiten. Es werden die Ergebnisse der empirischen Textverständlichkeitsstudie dargestellt sowie Interpretationsansätze angeboten und diskutiert. Der Aufbau orientiert sich entlang der Forschungsfragen. Zuerst werden demnach die Ergebnisse aus dem Wissenstest vorgestellt (Unterkapitel 7.1), anschließend wenden sich die Ausführungen in Unterkapitel 7.2 dem Bewertungsfragebogen zu und im dritten und letzten Teil werden wissens- und wahrnehmungsbezogene Ergebnisse zusammengebracht und hinsichtlich ihrer Zusammenhänge analysiert (Unterkapitel 7.3). Die beiden umfangreicheren Unterkapitel 7.1 und 7.2 sind wiederum in Abschnitte gegliedert, um die Analysen nochmals kleinschrittiger zu strukturieren. Immer am Ende eines solchen Analyseschritts (mit Ausnahme rein deskriptivstatistischer Ergebnisberichte) wird direkt eine *topikale Diskussion* angeschlossen. Diese bezieht sich ausschließlich auf die direkt zuvor berichteten Ergebnisse und soll jeweils dazu dienen, die Einordnung in den aktuellen Forschungsstand und die Bedeutsamkeit für die Unterrichtsrealität durchgehend im Auge zu behalten. Zum Ende jedes Ergebniskapitels (7.1, 7.2 und 7.3) findet sich schließlich eine bündige Ergebniszusammenfassung mit zugehöriger kritischer, *summarischer Diskussion*.

### 7.1 Ergebnisse aus dem Wissenstest

Um die Forschungsfrage 1 (Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung der Schüler\*innen; siehe Unterkapitel 4.2) zu beantworten, wurde der Wissenstest ausgewertet. Das Testinstrument ist in Abschnitt 6.3.2 beschrieben. Die Ergebnisse werden in die folgenden vier Abschnitte unterteilt: Abschnitt 7.1.1 fasst die deskriptive Statistik zusammen. Im nächsten Abschnitt 7.1.2 werden Ergebnisvergleiche zwischen Pre- und Post-Test dargestellt. Insbesondere werden solche Fälle näher betrachtet, in denen das Post-Test-Ergebnis schlechter als das Pre-Test-Ergebnis ausfällt. Abschnitt 7.1.3 wendet sich der Frage nach dem Einfluss sprachlicher Gestaltungsmittel auf die Behaltensleistung zu. Im vierten und letzten Abschnitt 7.1.4 der Wissenstest-Auswertung werden schließlich Analysen der Behaltensleistung unter der Berücksichtigung von Moderatorvariablen dargestellt, indem

Subgruppen je nach den einzelnen Personeneigenschaften (Gender, Schulstatus, Lesefähigkeiten und Vorwissen) vergleichend betrachtet werden. Mit einer Zusammenfassung inklusive summarischer Diskussion (Abschnitt 7.1.5) wird das Kapitel abgeschlossen.

### 7.1.1 Deskriptivstatistische Ergebnisse aus dem Wissenstest

Für die Analyse wurden  $n_{\text{HauptstudieWT}} = 782$  Datensätze der insgesamt  $n_{\text{Hauptstudie}} = 787$  Datensätze verwendet. Bei den fünf aussortierten Wissenstest-Datensätzen waren entweder im Pre- oder im Post-Test mindestens vier der acht Testaufgaben nicht bearbeitet worden. Ein sinnvoller Prä-Post-Vergleich war in solchen Fällen nicht möglich. Da der Wissenstest aus 8 Aufgaben bestand (siehe Abschnitt 6.3.2), wurden also mit 782 Datensätzen  $782 \times 8 = 6256$  Aufgaben ausgewertet. Dabei trat 35 Mal der Fall auf, dass kein einziges Kreuz in den Ankreuz-Kästchen gesetzt wurde. In einem solchen Fall war es an der Stelle der Auswertung nicht möglich zu entscheiden, ob die Testperson die Aufgabe nicht hatte bearbeiten wollen oder können (und somit 0 Credits erhalten sollte) oder ob die Testperson bewusst kein Kreuz gesetzt hatte, da sie die Aussagen allesamt als nicht zutreffend eingeschätzt hatte (und somit für jede korrekte Nicht-Zutreffend-Entscheidung entsprechende Credits erhalten sollte). Es wurde entschieden, jede Aufgabe als bearbeitet anzusehen und nicht pauschal 0 Credits für Aufgaben mit keinem gesetztem Kreuz zu vergeben. Dieses Auswertungsproblem mag die Testvalidität schmälern, jedoch ist davon auszugehen, dass das Ausmaß der Beeinträchtigung recht gering ist: In nur 35 von 6256 ausgewerteten Aufgaben *könnte* auf Grundlage dieser Entscheidung eine unzutreffende Bepunktung angelegt worden sein.

Eine ausführliche Aufstellung der deskriptiven Teststatistik bietet Tabelle D.4 im Anhang D.2. Es wurden analog zur Pilotierungsstudie (siehe dazu Abschnitt 6.3.2 im Methodenteil) die Schwierigkeit, die Varianz sowie die Trennschärfe der einzelnen Aufgaben berechnet. Zusammengefasst ergab sich ein ähnliches Bild wie in der Pilotierung: Der Bereich von 0.42 bis 2.08, in dem die Varianzen liegen, ist vergleichbar mit den Ergebnissen aus der Pilotierung und ist unauffällig. Wie erwartet tendierten die beiden Einfach-Wahl-Aufgaben 7 und 8 zu einer geringeren Varianz. Die Item-Schwierigkeiten decken den empfohlenen Bereich (siehe Seite 104) optimal ab. Der Kennwert ist darüber hinaus für jede Aufgabe im Post-Test höher als im Pre-Test, was dafür spricht, dass jede einzelne Aufgabe nach der Textlektüre besser bearbeitet wurde als vorher. Die Item-Trennschärfen des Pre-Tests sind tendenziell schwach (0.14 bis 0.23). Im Post-Test rangieren die Trennschärfen jedoch im Bereich von 0.3, was als akzeptabler Wert gilt (Bortz & Döring, 2006, S. 220). Diese Entwicklung hin zu größeren Trennschärfe-Kennwerten ist insofern plausibel, als dass die Versuchspersonen zum Zeitpunkt des Pre-Tests im Unterricht noch nichts über die elektrische Spannung gelernt hatten und somit vermutlich stärker zum Raten tendierten. Nach der Textlektüre jedoch beantworteten die Testpersonen die Fragen auf Grundlage von Wissen, das sie aus dem Text behalten konnten. Da nun der Trennschärfe-Kennwert im Korrelationskoeffizienten zwischen Task-Score und *part-whole-korrigiertem* Test-Score besteht (siehe Seite 119 im Methodenteil),

ist es schlüssig, dass der Trennschärfe-Kennwert bei dem verstärkt zufälligen Antwortverhalten vor der Textlektüre kleiner als beim systematischeren, wissensbasierten Antwortverhalten nach der Textlektüre ist.

Der Vergleich der Kennwerte für die Item-Schwierigkeit und die Item-Trennschärfe war ein erstes Indiz für das folgende Ergebnis:

Die Teilnehmenden lernten anhand der Textlektüre im Mittel hinsichtlich jedes Lerninhalts dazu.

Inwiefern allgemeine Sequenzeffekte anstelle eines echten Lernzuwachses die beobachteten Veränderungen verursacht haben könnten, wird im folgenden Abschnitt 7.1.2 erörtert. Es sei hier bereits erwähnt, dass sich die Annahme eines echten Lernzuwachses als plausibel erweisen wird.

### 7.1.2 Prä-Post-Vergleiche im Wissenstest

Um zu überprüfen, ob die Testpersonen anhand des Texts dazugelernt haben und ob sich dies im Testinstrument widerspiegelt, wurden Mittelwertsvergleiche anhand von t-Tests zwischen Pre-Test-Score und Post-Test-Score durchgeführt sowie aufgabenweise der mittlere Pre-Task-Score und der mittlere Post-Task-Score verglichen. Alle Analysen wurden für abhängige Stichproben gerechnet, da je ein Post-Ergebnis mit einem zugehörigen Pre-Ergebnis zu vergleichen ist. Die Robustheitsanalysen der Daten für t-Test-Vergleiche fielen positiv aus. Sie sind im Anhang D.3 einsehbar. Die Ergebnisse der t-Tests sind als belastbar anzusehen. Sie sind übersichtlich in Tabelle 7.1 aufgelistet. Im Studiendesign war keine Kontrollgruppe vorgesehen. Damit konnte kein Vergleich dazu angestellt werden, inwiefern sich die Testleistung von Schüler\*innen auch ohne Textlektüre von Pre-Test zu Post-Test verändert hätte. Es werden daher im Diskussionsteil alternative Argumente dafür angeführt werden, dass es tatsächlich die Textlektüre war, welche die Ursache für den Lernzuwachs war.

#### Pre-Scores vs. Post-Scores

Die Ergebnisse des t-Tests zum Prä-Post-Vergleich der (Gesamt-)Test-Scores und der einzelnen Task-Scores sind neben zugehörigen deskriptivstatistischen Daten der Tabelle 7.1 zu entnehmen. Für den Pre-Test-Score im Vergleich zum Post-Text-Score ergab sich:

Das mittlere Test-Ergebnis nach der Textlektüre liegt mit höchst signifikantem Unterschied höher als vor der Textlektüre (Teststatistik  $T = 32.98$  und p-Wert  $p \ll 0.001$ )<sup>1</sup>. Die Effektstärke ist sehr groß ( $\omega^2 = 0.58$ , siehe Seite 120 im Methoden-Abschnitt 6.4.3).

<sup>1</sup>Wie ausführlich im Anhang D.3 erläutert wird, ist der t-Test auf den vorliegenden Daten robust gegenüber Voraussetzungsverletzung. Dennoch wurde ergänzungshalber der verteilungsfreie Wilcoxon-Signed-Rank-Test berechnet, der nicht auf Normalverteilung oder Varianzhomogenität angewiesen ist (Bortz, Lienert & Boehnke, 2008, S. 200 ff.). Sein Ergebnis (Teststatistik  $W = 14507$  und  $p < 0.001$ ) deckt sich erwartungsgemäß mit dem des t-Tests.

		Pre	Post	t-Test	
Test-Score	mean	11.5 ( $\pm$ 4.2)	17.7 ( $\pm$ 5.2)	p	< 0.001
	min	3	3	T	32.98***
	max	30	34	$\omega^2$	0.58***
A1	mean	1.5 ( $\pm$ 1.3)	3.1 ( $\pm$ 1.4)	p	< 0.001
	min	0	0	T	25.21***
	max	5	5	$\omega^2$	0.45***
A2	mean	1.9 ( $\pm$ 1.2)	2.7 ( $\pm$ 1.4)	p	< 0.001
	min	0	0	T	12.40***
	max	5	5	$\omega^2$	0.16***
A3	mean	1.0 ( $\pm$ 1.1)	1.7 ( $\pm$ 1.4)	p	< 0.001
	min	0	0	T	12.26***
	max	5	5	$\omega^2$	0.16***
A4	mean	1.2 ( $\pm$ 1.2)	2.0 ( $\pm$ 1.4)	p	< 0.001
	min	0	0	T	12.48***
	max	5	5	$\omega^2$	0.17***
A5	mean	1.7 ( $\pm$ 1.0)	2.6 ( $\pm$ 1.3)	p	< 0.001
	min	0	0	T	16.20***
	max	5	5	$\omega^2$	0.25***
A6	mean	1.7 ( $\pm$ 1.4)	2.3 ( $\pm$ 1.4)	p	< 0.001
	min	0	0	T	8.82***
	max	5	5	$\omega^2$	0.09***
A7	mean	1.4 ( $\pm$ 0.9)	1.8 ( $\pm$ 0.6)	p	< 0.001
	min	0	0	T	9.73***
	max	2	2	$\omega^2$	0.11***
A8	mean	1.0 ( $\pm$ 1.0)	1.6 ( $\pm$ 0.8)	p	< 0.001
	min	0	0	T	15.16***
	max	2	2	$\omega^2$	0.23***

Tabelle 7.1.: Datentabelle für die Prä-Post-Vergleiche im Gesamttest sowie in den einzelnen Aufgaben A1 bis A8 (siehe dazu den Wissenstest im Anhang B.1) der  $n_{\text{HauptstudieWT}} = 782$  Testpersonen. Es sind jeweils der Mittelwert der gemessenen Punkteverteilung mit der Standardabweichung in runden Klammern angegeben, das Minimum der Verteilung sowie das Maximum. In der letzten Spalte sind die resultierenden t-Test-Ergebnisse für den Vergleich zwischen Pre- und Post-Wissenstest enthalten: Neben der Teststatistik  $T$  mit ihrem zugehörigen p-Wert  $p$  ist vor allem die Effektstärke  $\omega^2$  aussagekräftig für die Bedeutsamkeit der Mittelwertsunterschiede. Die p-Werte sind gemäß der Bonferroni-Holm-Korrektur (siehe Abschnitt 6.4.3) für multiple t-Tests korrigiert. Signifikanz wird anhand der Sternchen angezeigt (\*\*\*) für  $p < .001$ , \*\* für  $p < .01$  und \* für  $p < .05$ ).

Der Unterschied im mittleren Test-Score vor der Textlektüre und nach der Textlektüre ist also höchst signifikant. Dies war bei einer derart zahlenstarken Stichprobe zu erwarten, sodass vor allem die Effektstärke aussagekräftig ist. Wie im Methodenteil 6.4.3 auf Seite 120 erläutert wurde, wird hier  $\omega^2$  berichtet. Mit  $\omega^2 = 0.58$  werden durch die Zugehörigkeit eines Testergebnisses zur Gruppe der Pre-Test-Scores oder zur Gruppe der Post-Test-Scores 58 % der Gesamtvarianz erklärt.

Hinsichtlich einzelner Aufgaben ergab sich anhand entsprechender Prä-Post-Vergleiche in aller Kürze (siehe Tabelle 7.1 für einzelne Werte):

In allen Aufgaben liegen die mittleren Post-Task-Scores mit höchst signifikantem Unterschied höher als die jeweiligen Pre-Task-Scores. Die Effektstärken sind groß ( $\omega^2 = 0.09$  bis  $\omega^2 = 0.45$ , siehe Seite 120 im Methoden-Abschnitt 6.4.3).

Da hier für jede Person mehrere t-Tests durchgeführt wurden, war eine  $\alpha$ -Fehler-Korrektur angemessen; dies ist in der Ergebnistabelle 7.1 berücksichtigt. Alle p-Werte waren auch nach der Bonferroni-Holm-Korrektur (siehe Abschnitt 6.4.3) deutlich kleiner als 0.001 und zeigen somit höchst signifikante Unterschiede an. Es werden je nach Aufgabe zwischen 9 % und 45 % der Gesamtvarianz erklärt<sup>2</sup>.

### Ein Maß für den Lernzuwachs: Der Gain-Score

Für die folgenden Analysen wurden Werte für den sogenannten *Gain-Score* in den Blick genommen. Unter dem Gain-Score wird die Differenz vom Pre-Test-Score zum Post-Test-Score gefasst. In der hier vorliegenden Anwendung beziffert der Gain-Score somit den Lernzuwachs hinsichtlich der elektrischen Spannung im Sinne einer Behaltensleistung beim Lesen<sup>3</sup>.

Die Verteilung des Gain-Scores in der Gesamtstichprobe mit den zugehörigen deskriptivstatistischen Informationen ist in Abbildung 7.1 gezeigt. Es ergab sich eine recht breite Verteilung (Standardabweichung 5.2 Credits) um den Mittelwert (6.2 Credits). Die besten Gain-Scores lagen bei 21 Credits (drei Mal), die schwächsten bei -9 Credits (zwei Mal). Ein negativer Gain-Score bedeutet, dass eine Person nach der Textlektüre schlechter im Wissenstest abschnitt als zuvor. Dies war bei 62 Testpersonen der Fall (Bereich links der grau-gepunkteten Linie in der Abbildung).

Ein genauerer Blick auf diese Fälle ist interessant. Der folgende Abschnitt wendet sich daher den negativen Gain-Scores zu.

<sup>2</sup>Es fällt auf, dass der Anteil an erklärter Varianz in den Task-Scores deutlich geringer ist als für den Test-Score. Dies ist insofern erwartbar, da hinsichtlich einzelner Aufgaben die unsystematische Varianz in den Daten größer ist als im Gesamttest-Ergebnis (der summarische Test-Score glättet über Stärken und Schwächen hinsichtlich der Einzelaspekte in den Aufgaben hinweg). Dennoch sind die Effektstärken überwiegend im Bereich starker und teilweise mittlerer Effekte einzuordnen.

<sup>3</sup>Für eine eingehendere Diskussion, inwiefern der Gain-Score in dieser Studie ein passendes Maß für die Behaltensleistung darstellt, siehe die Vorüberlegung in Abschnitt 7.1.3.

## 7. Ergebnisse und Diskussion

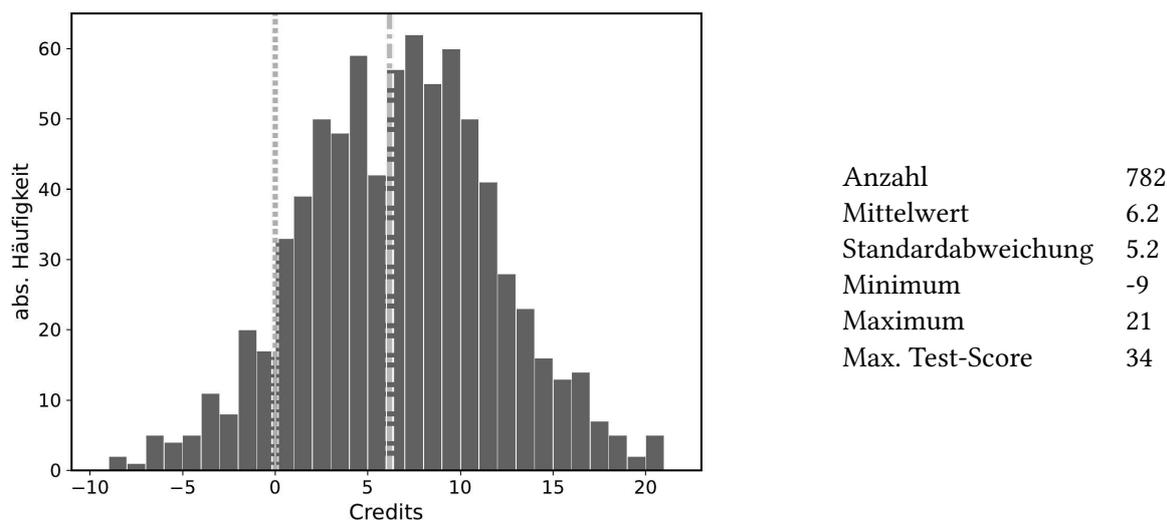


Abbildung 7.1.: Häufigkeitsverteilung für den Gain-Score in der Gesamtstichprobe ( $n_{\text{HauptstudieWT}} = 782$ ). Die gestrichelte Linie markiert den Mittelwert und die gepunktete Linie den Wert 0. Links der Null-Linie liegen diejenigen Datensätze mit einem schwächeren Post-Test-Ergebnis als Pre-Test-Ergebnis. Die nebenstehende Tabelle enthält zentrale deskriptivstatistische Verteilungskennwerte.

	A1	A2	A3	A4	A5	A6	A7	A8
mittl. Task-Gain-Score	1.64	0.72	0.69	0.73	0.89	0.54	0.36	0.61
# neg. Task-Gain-Scores	78 (.1)	167 (.21)	162 (.21)	158 (.2)	123 (.16)	196 (.25)	46 (.06)	42 (.05)
Itemschwierigk. Pre-Test	0.29	0.39	0.21	0.25	0.33	0.34	0.7	0.5
Itemschwierigk. Post-Test	0.62	0.53	0.35	0.39	0.51	0.45	0.88	0.81

Tabelle 7.2.: Aufgabenspezifische Betrachtung des Gain-Scores. Die Tabelle listet auf, welcher mittlere Gain-Score in der Gesamtstichprobe ( $n_{\text{HauptstudieWT}} = 782$ ) in den einzelnen Aufgaben A1 bis A8 im Wissenstest erzielt wurde und wieviele Personen in den einzelnen Aufgaben einen negativen Task-Gain-Score hatten. In runden Klammern sind die zugehörigen relativen Häufigkeiten angegeben. Als interessanter Bezugswert sind zudem die Itemschwierigkeiten aus dem Pre-Test und aus dem Post-Test mit dargestellt.

### Zu den Fällen mit negativem Gain-Score

Da in 62 von 782 Fällen (das entspricht knapp 10 % der Fälle) ein negativer Gain-Score erzielt wurde, stellte sich die Frage, ob dies...

- ...an bestimmten Aufgaben lag: Wurden einzelne Aufgaben besonders häufig im Post-Test schlechter bearbeitet als im Pre-Test?
- ...an bestimmten Texten lag: Trat bei der Lektüre bestimmter Textversionen besonders häufig der Fall auf, dass der Post-Test-Score schwächer als der Pre-Test-Score ist?
- ...an bestimmten Personeneigenschaften lag: Erzielten Personen mit einer bestimmten Vorwissensausprägung, mit bestimmten Lesefähigkeiten oder mit bestimmtem Gender besonders häufig im Post-Test schlechtere Ergebnisse als im Pre-Test?

Eine allzu detailreiche Analyse solcher Faktoren, die zwar die Behaltensleistung beim Lesen

beeinflussen mögen, jedoch jenseits der Textgestaltung liegen, geht über die Forschungsfragen dieser Studie hinaus. Daher werden im Folgenden nur deskriptive und augenscheinliche Untersuchungsansätze vorgestellt. Die Grundlage bilden dabei Tabelle 7.2 und die Verteilungsvergleiche in Abbildung 7.2.

Im Hinblick auf die *Aufgaben* kristallisieren sich nicht einzelne Aufgaben heraus, die auffallend häufig nachher schlechter bearbeitet wurden als vorher. Es lässt sich allenfalls bemerken, dass die Aufgabe 1 sowie die beiden Einzelwahl-Aufgaben 7 und 8<sup>4</sup> besonders selten einen negativen Punktezuwachs verzeichnen lassen. Die negativen Gain-Scores wurden also nicht maßgeblich durch einzelne, wenige Aufgaben hervorgerufen. Es scheinen sich keine Aufgaben dadurch auszuzeichnen, anhand der Textinformationen schlecht zu lösen zu sein.

Hinsichtlich der Frage nach dem Zusammenhang mit gelesenen *Textversionen* und/oder *Personeneigenschaften* liefert der augenscheinliche Vergleich der jeweiligen Merkmalsverteilungen in Abbildung 7.2 interessante Einblicke: Es wird jeweils gegenübergestellt, wie sich das zu untersuchende Merkmal in der Gesamtstichprobe verteilt und wie es sich in der Teilstichprobe mit negativem Gain-Score verteilt. Aus dem Vergleich lässt sich festhalten:

- Hinsichtlich der gelesenen Textversionen fallen keine deutlichen Verteilungsunterschiede auf. In der Gruppe der Personen mit negativem Gain-Score wurden Text B (wsP) und E (Wsp) tendenziell häufiger gelesen und Text H (WSP) seltener. Ein starker Zusammenhang zwischen gelesener Textversion und negativen Gain-Score-Ergebnis wird hier dennoch nicht augenscheinlich.
- Hinsichtlich des Vorwissens fällt deutlich auf, dass in der Gruppe der Personen mit negativem Gain-Score ein größerer Anteil an Personen ein großes Vorwissen hatte. 85 % aller Personen mit negativem Gain-Score erzielten im Pre-Test einen Test-Score, der oberhalb des Gesamtstichproben-Mittelwerts lag. Ein Zusammenhang ist somit naheliegend.
- Hinsichtlich der Lesefähigkeiten zeigen sich deutliche Verteilungsunterschiede beim Leseverständnis und der Lesegenauigkeit. Bei den Personen mit negativem Gain-Score sind deutlich mehr Schüler\*innen zu finden, die über ein schwaches oder sehr schwaches Leseverständnis sowie über eine schwache oder sehr schwache Lesegenauigkeit verfügen. Es deutet sich somit an, dass ein Mangel an diesen beiden Fähigkeiten das Lernen aus den Texten beeinträchtigte. Die Lesegeschwindigkeit erscheint weniger relevant.
- Hinsichtlich der Gender-Verteilung fällt auf, dass in der Gruppe der Personen mit negativem Gain-Score anteilig viel weniger weibliche, aber dafür deutlich mehr männliche und vor allem Schüler\*innen diversen Genders vertreten sind. Gender und negative Gain-Score-Ergebnisse scheinen einen Zusammenhang aufzuweisen.

Auf Grundlage der Daten aus Tabelle 7.2 und Abbildung 7.2 lässt sich zusammenfassend festhalten:

---

<sup>4</sup>Für Erläuterungen zum Wissenstest und den einzelnen Aufgaben siehe Abschnitt 6.3.2

## 7. Ergebnisse und Diskussion

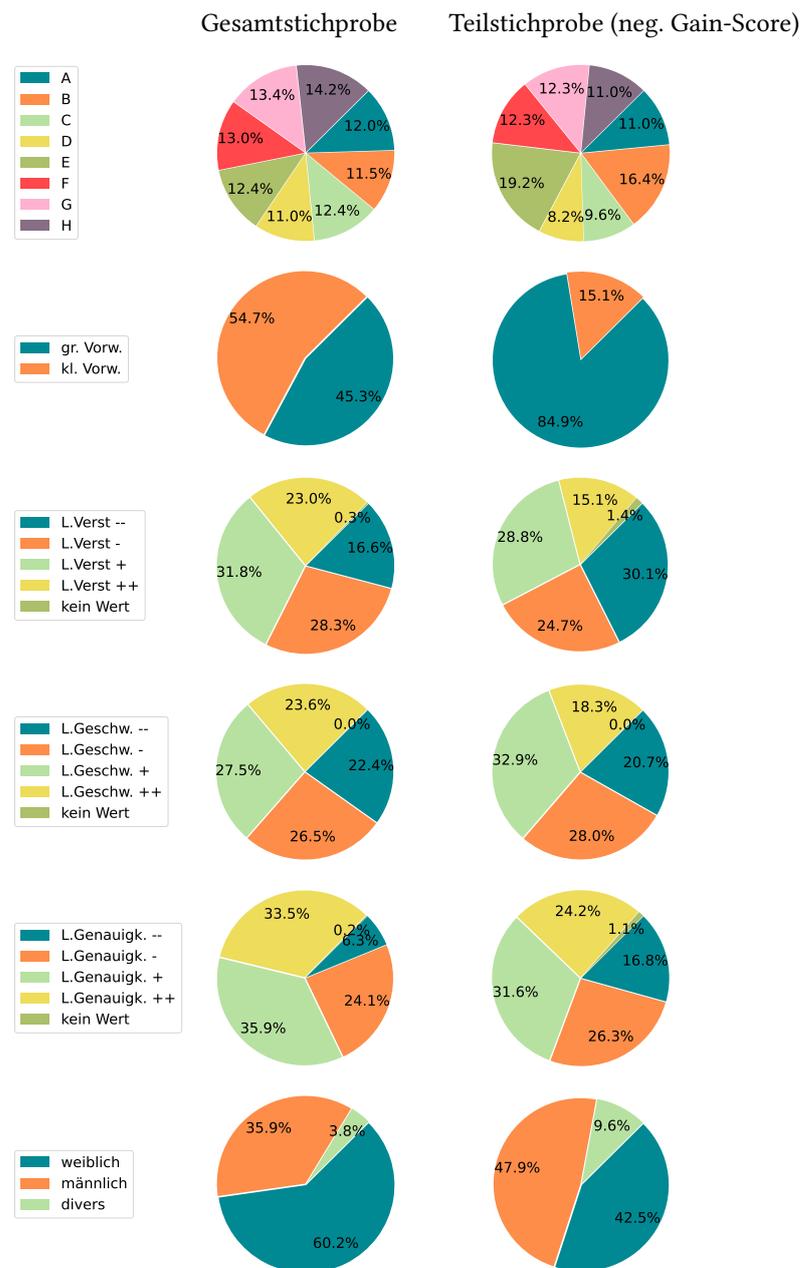


Abbildung 7.2.: Tortendiagramme zur Gegenüberstellung einzelner Merkmalsverteilungen in der Gesamtstichprobe ( $n_{\text{HauptstudieWT}} = 782$ ) und in der Teilstichprobe derjenigen Personen mit negativem Gain-Score ( $n_{\text{HauptstudieWT\_neg}} = 62$ ). In den obersten Diagrammen ist zu sehen, wie sich die gelesenen Textversionen A bis H verteilen. Die zweite Zeile bildet ab, welcher Anteil an Personen im Pre-Test einen Test-Score erzielt hat, der über beziehungsweise unter dem Gesamtstichproben-Mittelwert liegt (in der Legende mit „gr. Vorw.“ beziehungsweise „kl. Vorw.“ bezeichnet). Zeile drei bezieht sich auf das Leseverständnis, Zeile vier auf die Lesegeschwindigkeit und Zeile fünf auf die Lesegenauigkeit, wie sie jeweils im LGVT erhoben wurden. Für jede Lesefähigkeit wird die Merkmalsverteilung in vier Gruppen (sehr schwach -- bis sehr stark ++) geplottet. In der untersten Zeile ist die Gender-Verteilung zu sehen.

Etwa 10 % der Schüler\*innen schnitten im Post-Test schlechter ab als im Pre-Test. Es deutet sich an, dass für solche Negativ-Ergebnisse ein sehr gutes Vorwissen, ein schwach ausgeprägtes Leseverständnis und eine mangelnde Lesegenauigkeit eine Rolle spielten. Auch das Gender scheint damit zusammenzuhängen.

### **Topikale Diskussion: Prä-Post-Vergleiche im Wissenstest**

Die Prä-Post-Vergleiche, die in diesem Abschnitt 7.1.2 berichtet wurden, führten zu den Ergebnissen, dass die teilnehmenden Schüler\*innen im Test-Score sowie in den einzelnen Task-Scores nach der Textlektüre im Mittel deutlich bessere Ergebnisse erzielten als zuvor. Die Schüler\*innen konnten also Wissen zur elektrischen Spannung aus dem Text behalten und dies bildete sich in den Wissenstestergebnissen ab. Dennoch traten in 10 % der Fälle negative Gain-Scores auf, die Personen schnitten also im Post-Test schlechter ab als im Pre-Test. Dies war besonders häufig für Personen der Fall, die über ein gutes Vorwissen verfügten (hoher Pre-Test-Score), die einen schwachen Leseverständnis- und/oder einen schwachen Lesegenauigkeitswert hatten (LGVT, siehe Abschnitt 6.3.1) und/oder die ein männliches oder diverses Gender angegeben hatten.

Um diese Ergebnisse hinsichtlich ihrer Bedeutsamkeit und Aussagekraft zu interpretieren, ist zunächst zu bemerken, dass Testergebnisse im Allgemeinen durch Boden- oder Deckeneffekte verzerrt sein können (siehe z. B. Bortz & Döring, 2006, S. 182 f.). Der Einfluss solcher Effekte sollte in diesem Fall geringfügig sein, da die Item-Schwierigkeiten im Optimalbereich liegen (siehe Abschnitt 7.1.1) und der Bereich möglicher Test-Scores (0 bis 34 Credits) von den Testpersonen gut abgedeckt wird: In keinem Fall wurde ein Test-Score von 0 Credits und in nur drei Fällen – jeweils im Post-Test – wurde die Maximalpunktzahl von 34 Credits erreicht. Fehlinterpretationen könnten außerdem aufgrund von Sequenzeffekten entstehen. Denn es gibt neben der Textlektüre auch ganz generelle Gründe, aus denen Schüler\*innen im Allgemeinen in einem wiederholten Test bessere Ergebnisse erzielen als bei der ersten Testdurchführung. Die Schüler\*innen kennen beispielshalber den Test und seine Modalitäten und sie werden allgemein im Fach Physik besser. Betrachtet man nun aber die Effektstärke  $\omega^2 = 0.58$  des Vergleichs der mittleren Post-Test-Scores gegenüber den Pre-Test-Scores, so lässt sich im Vergleich mit den gängigen Richtwerten (0.01 für kleine Effekte, 0.06 für mittlere Effekte und 0.14 für große Effekte, siehe Seite 120 im Methoden-Abschnitt 6.4.3) ein sehr großer Effekt konstatieren. Insbesondere lohnt der Vergleich mit empirischen Vergleichsmaßstäben aus ähnlichen Forschungskontexten. Hier stellen Brunner, Stallasch und Lüdtke (2023) anschließend an die US-amerikanische Arbeit von Hill, Bloom, Black und Lipsey (2008) eine weitreichende Analyse von Effektstärken im Zusammenhang mit bildungsbezogenen Interventionsstudien in Deutschland zur Verfügung. Unter anderem ermittelten sie eine mittlere Effektgröße für den Lernzuwachs von Schuljahr zu Schuljahr in den Naturwissenschaften. In den Jahrgangsstufen 8 beziehungsweise 9 werden die Schüler\*innen im Lauf eines ganzen Jahres demnach um  $0.22 (\pm 0.01)$  beziehungsweise  $0.15 (\pm 0.02)$  Standardabweichungen besser.

Diesen „natürlichen“ Lernzuwachs übersteigt der hier gefundene Effekt einer Verbesserung um ca. 1.32 Standardabweichungen über eine Zeit von einer bis vier Wochen eklatant. Ein weiteres Indiz dafür, dass die Verbesserungen auf die Textlektüre zurückzuführen sind und nicht auf allgemeine Verbesserungseffekte, liefert die folgende Beobachtung zum Gain-Score: Wenn der Gain-Score maßgeblich vom Lernzuwachs beim *Lesen* herrührte, dann sollten *Lesefähigkeiten* entscheidend mit dem Gain-Score korrelieren. Wenn jedoch Sequenzeffekte die entscheidende Ursache für die Verbesserung waren, sollten Lesefähigkeiten einen kaum nachweisbaren Zusammenhang mit dem Gain-Score haben. Die Pearson-Korrelationen (siehe Abschnitt 6.4.3) zwischen Lesefähigkeiten und Gain-Score berechnen sich zu: Leseverstehen  $\rho = 0.18^{***}$ , Lesegenauigkeit  $\rho = 0.27^{***}$  und Lesegeschwindigkeit  $\rho = -0.02$ . Zwischen dem Gain-Score auf der einen Seite und dem Leseverstehen sowie der Lesegenauigkeit auf der anderen Seite zeigen sich also deutliche Korrelationen und damit Zusammenhänge mit der Behaltensleistung. Dass wiederum die Lesegeschwindigkeit unerheblich ist, ist insofern erwartbar, als bei der Textlektüre in der Hauptstudie kein Zeitlimit für die Schüler\*innen gesetzt war. Das Ergebnis, dass die Teilnehmenden aufgrund der Textlektüre einen Lernzuwachs verzeichneten, wird also durch die Prä-Post-Vergleiche der Test-Scores gestützt.

Die Schüler\*innen gewannen also anhand der Texte Wissen über die elektrische Spannung dazu und zeigten dies im Wissenstest. Ein ganz grundlegender Schluss, der sich für die Unterrichtsrealität ziehen lässt, ist, dass umsichtig konstruierte Texte einen Lernzuwachs hervorrufen. Im Fall dieser Studie wurde dies anhand solcher Texte, deren Erstellung vor allem am Karlsruher Modell für Textproduktion und -bewertung sowie an den Hamburger Verständlichkeitsdimensionen ausgerichtet war (siehe Unterkapitel 6.2) und anhand einer Messung des Lernzuwachses in einem Wissenstest zur Behaltensleistung nachgewiesen.

Die Ergebnisse zu den Fällen mit negativem Gain-Score zeigen jedoch, dass diese Aussage zwar im Mittel, jedoch nicht für alle Einzelfälle gilt. Es gab durchaus Personen, die durch die Textlektüre keinen Lernzuwachs über die elektrische Spannung verzeichnen konnten oder deren Lernzuwachs nicht zutreffend im Gain-Score abgebildet wurde. Erklärungsansätze dafür, dass vor allem einzelne Personen mit starkem *Vorwissen* im Post-Test schlechter abschnitten als im Pre-Test, liegen im Effekt der *Regression zur Mitte* und im *Expertise-Reversal-Effekt*. Die Regression zur Mitte ist das folgende Phänomen (vergleiche Barnett, van der Pols und Dobson, 2005): Wenn Personen ein sehr starkes Pre-Test-Ergebnis erzielen, so kommt dieses im Allgemeinen neben einem hohen Vorwissen auch durch zufälliges, gutes Raten zustande. Die Wahrscheinlichkeit, dass sie dann im Post-Test weniger Credits durch Raten erzielen, ist somit groß. Wenn die Personen also nach der Textlektüre ähnlich viel oder marginal mehr wissen, jedoch schlechter raten, werden sie schlechter als im Pre-Test abschneiden. Eine weitere Erklärung liefert der Expertise-Reversal-Effekt (eine gute Übersicht bietet das Review von Kalyuga, Ayres, Chandler und Sweller, 2003, speziell für Texte siehe z. B. McNamara, Kintsch, Songer und Kintsch, 1996). Dieser Effekt führt dazu, dass Personen mit gutem Vorwissen nicht gefordert sind und dadurch nicht engagiert, aktiv und letztlich nicht *verstehend* lesen; sie lernen somit anhand des Texts wenig oder nichts *dazu*. Die Relevanz der *Lesefähig-*

*keiten* bei negativen Gain-Scores wird mitunter durch das Studiendesign plausibel: Bei der Textlektüre in der Hauptstudie war kein Zeitlimit vorgegeben; die Texte gut zu verstehen sollte daher maßgeblicher von den Fähigkeiten Leseverstehen und Lesegenauigkeit abhängen. Dies spiegelt sich in den Verteilungsvergleichen wider. Die augenscheinlichen Unterschiede in Bezug auf die *Genderverteilung* in der Gesamtstichprobe und in der Teilstichprobe mit negativem Gain-Score zu interpretieren fällt schwer. Hier tiefgehend einzusteigen führt am Forschungsinteresse dieser Studie vorbei. Mögliche Denk- und Analyseansätze wären etwa, dass das Gender mit den Lesefähigkeiten korrelieren könnte und sich daher an dieser Stelle *moderierte* Zusammenhänge andeuten. Darüber hinaus kommen auch weitere, in den Geschlechtern nicht gleichverteilte Personenmerkmale wie Interesse, Lesemotivation, Sorgfalt, etc. in Frage, die Einfluss auf das Erzielen eines negativen Gain-Scores nehmen könnten und somit einen Moderatoreffekt des Genders hervorrufen.

In Hinblick auf die Unterrichtsrealität deuten diese Ergebnisse an, dass die Ausprägung bestimmter Personenmerkmale (hier: thematisches Vorwissen, Lesefähigkeiten und Gender) einen Einfluss auf den Lernzuwachs beim verstehenden Lesen hat. Für Anschlussstudien wäre also ein interessanter Forschungsansatz, solche potenziell einflussreichen Personenmerkmale fokussiert zu untersuchen. Die hier vorliegende Studie geht darauf nur am Rande ein.

So wird auch im folgenden Abschnitt der Zusammenhang zwischen sprachlicher Gestaltung der Texte und der Behaltensleistung beim Lesen zunächst in der Gesamtstichprobe und anschließend in Teilstichproben je nach Ausprägung verschiedener Personenmerkmale beleuchtet.

### **7.1.3 Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung (ohne Moderatorvariablen)**

Für die Prä-Post-Vergleiche des vorangehenden Abschnitts wurden Analysen für *abhängige* Stichproben vorgenommen, da je ein Post-Score mit einem zugehörigen Pre-Score zu vergleichen war. Die folgenden Analysen beziehen sich nun auf *unabhängige* Stichproben, da die Behaltensleistungen verschiedener Personen miteinander zu vergleichen sind. Gemäß Forschungsfrage 1 soll untersucht werden, inwiefern sich die Behaltensleistung in Abhängigkeit von der sprachlichen Gestaltung der Texte hinsichtlich Wortwahl (W), Satzstruktur (S) und Personalisierung (P) beziehungsweise in Abhängigkeit von der Textversion veränderte. Eine bestimmte Textversion kann hierbei gewissermaßen als „Konglomerat“ mit einer ganz bestimmten Zusammensetzung aus den Variablen W, S und P betrachtet werden (siehe Abschnitt 6.2.2 für Einzelheiten).

Die zufällige Verteilung der Texte an die teilnehmenden Schüler\*innen (siehe Abschnitt 6.5.1) hatte zu einer zufriedenstellenden Gleichverteilung der Testpersonen auf die zu vergleichenden Gruppen geführt (siehe dazu Tabelle 7.3).

Anzahl Personen (sprachliche Gestaltung)			Anzahl Personen (Textversion)			
w: 367	s: 383	p: 393	A: 94	B: 90	C: 97	D: 86
W: 415	S: 399	P: 389	E: 97	F: 102	G: 105	H: 111

Tabelle 7.3.: Verteilung der Testpersonen ( $n_{\text{HauptstudieWT}} = 782$ ) je nach gelesenem Text in absoluten Zahlen: links ist die Verteilung je nach sprachlicher Gestaltung hinsichtlich *Wortwahl*, *Satzstrukturen* und *Personalisierung* aufgeführt, rechts die Verteilung je nach Textversion (Texte A bis H; siehe Abschnitt 6.2.2 für Einzelheiten).

### Vorbereitende Überlegung: Der Gain-Score als Maß für die Behaltensleistung

Um die Forschungsfrage 1 (und auch Forschungsfrage 3; siehe Unterkapitel 7.3) zu den Einflussfaktoren auf die Behaltensleistung zu beantworten, war ein Maß für die abhängige Variable *Behaltensleistung* nötig. Als solches wurde die Differenz aus Post-Test-Score und Pre-Test-Score gewählt, der sogenannte *Gain-Score* (siehe dazu auch den entsprechenden Unterpunkt in Abschnitt 7.1.2 auf Seite 129). Eine alternative Möglichkeit, die Einflüsse auf die Behaltensleistung zu analysieren, wäre es, den Post-Test-Score (oder äquivalent dazu die mittlere Post-Test-Itemschwierigkeit) als abhängige Variable zu betrachten und den Post-Test-Score hinsichtlich des Pre-Test-Scores zu kontrollieren. Statistisch wäre das in ANCOVA-Modellen mit entsprechender Kovariate „Pre-Test-Score“ oder in Multilevel-Modellen umsetzbar. Für eine Verwendung des Gain-Scores als Maß für die Behaltensleistung spricht jedoch, dass der Gain-Score die Realität in der Stichprobe gut abbildet. Denn eine ANCOVA und eine Gain-Score-Analyse haben verschiedene Fragen zum Gegenstand: Eine ANCOVA mit Kovariate „Vorwissen“ beantwortet die Frage, ob die unabhängige Variable (hier: die sprachliche Textgestaltung) zu signifikant unterschiedlichen Post-Test-Ergebnissen führt, *wenn gleiche Vorwissens-Bedingungen vorliegen*. Das bedeutet, die ANCOVA untersucht eine fiktive Situation, in der alle Individuen die gleichen „Startbedingungen“ haben (Smolkowski, 2019). In der hier vorliegenden Situation ist daher der Gain-Score dasjenige Maß, das besser zur Forschungsfrage passt. Er glättet nicht über unterschiedliche Startbedingungen hinweg, sondern berücksichtigt explizit die Heterogenität der Schüler\*innen und deren Lernvoraussetzungen. Der Gain-Score ist damit im Allgemeinen stark vom Vorwissen abhängig. Dies wird mit folgender Überlegung schnell klar: bei sehr gutem Pre-Test-Ergebnis ist wenig Raum nach oben für den Post-Test und es kann nur noch ein geringer Gain-Score erreicht werden. Textlernerfekte sind also nicht alleine maßgeblich für den Gain-Score. Doch dies ist Teil der Unterrichtsrealität und damit auch des Textverstehens in der Schule. Die Population, für die in dieser Studie Inferenzen gezogen werden, ist eine vielfältige Schülerschaft. Der Gain-Score ist somit in dieser Studie ein geeignetes Zahlenmaß für die Behaltensleistung der Testpersonen.<sup>5</sup>

<sup>5</sup>Dennoch kann auch der leicht anders gelagerte Standpunkt einer ANCOVA (Analysis of Covariance) mit der Kontrollvariable *Vorwissen* – also der Suche nach Effekten der Textgestaltung bei gleichen Startbedingungen in der Stichprobe – interessant sein. Eine entsprechende Analyse wurde nach dem Vorbild von Hilbert, Stadler, Lindl, Naumann und Bühner (2019) durchgeführt und ist im Anhang D.5 einzusehen.

unabhängige Variable	F (7,781)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
W	5.68	.052 .017	.01	5.75 (W) 6.65 (w) +0.90 ( $\Delta$ )
S	1.00	.556 0.35	.00	6.07 (S) 6.29 (s) +0.12 ( $\Delta$ )
P	0.00	1.00 .972	.00	6.17 (P) 6.18 (p) +0.01 ( $\Delta$ )
W : S	0.31	.578	.00	
W : P	0.00	.968	.00	
S : P	0.00	.979	.00	
W : S : P	0.10	.755	.00	

Tabelle 7.4.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen *Wortwahl*, *Satzstrukturen* und *Personalisierung* und deren Interaktionen („:“). Es werden die Teststatistik F für die Modellparameter (7,781) und der p-Wert in unkorrigierter Form sowie in  $\alpha$ -Fehler-korrigierter Form berichtet. Nach einer Korrektur gemäß Bonferroni-Holm (siehe Abschnitt 6.4.3) ist kein Effekt signifikant. Entsprechend ist keiner der Effektstärke-Werte  $\omega^2$  mit Signifikanz-Sternchen markiert. Ergänzend sind die Gain-Score-Mittelwerte für W-Texte gegenüber w-Texten, für S-Texte gegenüber s-Texten und für P-Texte gegenüber p-Texten dargestellt. Mit  $\Delta$  ist der Unterschied „anspruchsgeminderte / schwach ausgeprägte Form“ minus „anspruchsgesteigerte / ausgeprägte Form“ bezeichnet. Stichprobengröße  $n_{\text{HauptstudieWT}} = 782$ .

### Zusammenhangsanalysen in der Gesamtstichprobe

Für die statistische Analyse des Zusammenhangs zwischen der sprachlichen Gestaltung der Texte und der Behaltensleistung durch die Lesenden wurden ANOVA-Modelle berechnet. Die Daten erwiesen sich als robust für diese Analysen (die Normalverteilungsannahme und die Varianzhomogenitätsannahme erscheinen zutreffend in den Teilstichproben je nach sprachlicher Gestaltung der gelesenen Texte; siehe dazu ausführlich Anhang D.3). In einem ersten Modell (Tabelle 7.4) bestanden die unabhängigen Variablen in den sprachlichen Gestaltungsvariablen, also in den drei Faktoren W (*Wortwahl / morphologische Auffälligkeiten auf Wortebene*), S (*Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene*) und P (*erzählende und personalisierende Elemente*). Es ist dabei für die Gültigkeit der ANOVA wichtig, dass die drei Faktoren unabhängig voneinander sind. Dies ist aufgrund des Studiendesigns gegeben. Genauer: da die sprachlichen Textmodifikationen hinsichtlich W, S und P jeweils isoliert realisiert wurden (siehe für Einzelheiten Abschnitt 6.2.2). Im zweiten ANOVA-Modell (Tabelle 7.5) war die unabhängige Variable die Textversion (bezeichnet mit den Buchstaben A bis H<sup>6</sup>). Für beide Modelle bestand die abhängige Variable in der Behaltensleistung in Form

<sup>6</sup>Die sprachliche Gestaltung der Textversionen A bis H ist ausführlich in Abschnitt 6.2.2 beschrieben. Für einen schnellen Überblick sei hier kurz festgehalten:

unabhängige Variable	F (7,781)	p	$\omega^2$	Mittelwerte
Textversion	0.92	.49	.00	A: 6.80, B: 6.93, C: 6.47, D: 6.40 E: 5.82, F: 5.70, G: 5.69, H: 5.81

Tabelle 7.5.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der Textversion (Stichprobengröße  $n_{\text{HauptstudieWT}} = 782$ ). Es werden die Teststatistik F für die Modellparameter (7,781), der p-Wert und der Effektstärke-Wert  $\omega^2$  berichtet. Ergänzend sind in der letzten Spalte die Gain-Score-Mittelwerte für alle Textversionen A bis H dargestellt.

des Gain-Scores.

Zusammenfassend erscheinen hinsichtlich der sprachlichen Gestaltung nach einer  $\alpha$ -Fehler-Korrektur gemäß Bonferroni-Holm (für eine Erläuterung siehe Abschnitt 6.4.3; hier: drei paarweise Vergleiche W vs. w, S vs. s und P vs. p) keine Haupt- oder Interaktionseffekte signifikant. Ferner sind alle Effektstärken im Bereich sehr kleiner oder sogar kaum nachweisbarer Effekte.

Ebenso erweist sich der Effekt der Textversion nicht signifikant bei einer nicht nachweisbaren Effektstärke (hier finden keine multiplen, paarweisen Vergleiche statt, daher keine  $\alpha$ -Fehler-Korrektur.).

Die quantitativen Ergebnisse aus den ANOVAs werden anhand von Abbildung 7.3 um einen visuellen Eindruck ergänzt. Die Grafiken zeigen den mittleren Gain-Score in Abhängigkeit von der sprachlichen Gestaltung (W, S und P; links in der Abbildung) beziehungsweise der Textversion (rechts in der Abbildung). Die Zahlenwerte für die geplotteten Mittelwerte sind zusammen mit ihren Standardabweichungen in die Abbildung mit aufgenommen. Die Grafiken ermöglichen insbesondere einen überblicksartigen Vergleich der (Haupt-)Effekte durch die sprachliche Gestaltung beziehungsweise der Effekte durch die Textversion. Dabei fällt auf, dass jeder der Texte A bis D zu einer besseren Behaltensleistung führte als alle Texte E bis H. Die Texte A bis D zeichnen sich gegenüber den Texten E bis H in der Ausprägung der Wortwahl-Variablen aus. Entsprechend ist auch in der linken Grafik der Abbildung 7.3 der Unterschied W vs. w der größte im Vergleich mit S vs. s und P vs. p. Es ist dennoch zu beachten, dass alle Effekte *deutlich* innerhalb der Bereiche der Standardabweichungen liegen und somit im Verhältnis zu interpersonellen Unterschieden beim Lesen *einer Textsorte* gering ausfallen. Die entsprechenden Fehlerbalken, die jeweils einen Bereich von plusminus ca. 5 Credits einschließen würden, sind aus Gründen der optischen Darstellbarkeit nicht in den Grafiken abgebildet.

Obwohl die ANOVA quantitativ zeigt, dass Interaktionseffekte zwischen den einzelnen Variablen der sprachlichen Gestaltung (W, S und P) nicht signifikant werden, soll auch in

A: wsp	B: wsP	C: wSp	D: wSP
E: Wsp	F: WsP	G: WSp	H: WSP

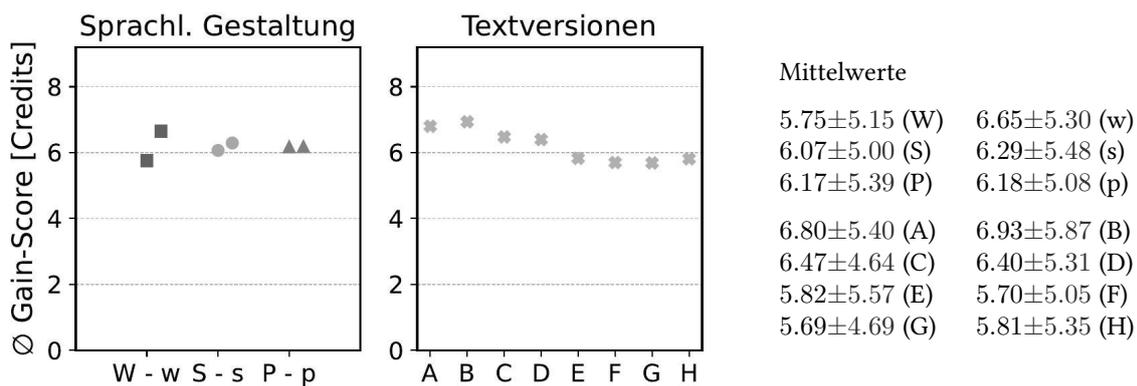


Abbildung 7.3.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung (links) beziehungsweise von der Textversion (rechts) (Stichprobengröße  $n_{\text{HauptstudieWT}} = 782$ ). Es sind aus Gründen der optischen Darstellbarkeit keine Fehlerbalken abgebildet. Die Standardabweichungen sind in der nebenstehenden Tabelle jeweils in grau angegeben.

diesem Zusammenhang ein visueller Eindruck gegeben werden; siehe dazu Abbildung 7.4. Parallel verlaufende Verbindungslinien in den Profilplots sprechen gegen das Vorliegen von Interaktionseffekten. Je weniger parallel die Linien verlaufen, desto stärker ist die Interaktion der betrachteten Variablen<sup>7</sup>. Im vorliegenden Fall ist der Linienvverlauf in vier der sechs Subplots nahezu perfekt parallel. Lediglich für die Interaktion W : S beziehungsweise S : W (links in der Abbildung) ist grafisch eine Tendenz zur Interaktion erkennbar.

Augenscheinlich ist ausschließlich für das Variablenpaar W und S ein leichter Interaktionseffekt zu erkennen und zwar derart, dass bei anspruchsgeminderter Wortwahl die Satzstruktur einen größeren Einfluss auf den Gain-Score hatte als bei anspruchsgesteigerter Wortwahl. In jedem Fall war eine anspruchsgeminderte Satzstruktur die günstigere.

Die bisher berichteten Ergebnisse beziehen sich auf die Gesamtstichprobe. Diese ist als Abbild für die Population zu betrachten, das heißt für eine vielfältige Schülerschaft mit unterschiedlichsten Personenmerkmalen und Lernvoraussetzungen. Bei der Suche nach Effekten der sprachlichen Gestaltung erzeugen all diese heterogenen Voraussetzungen ein „Rauschen“ in den Daten. Der folgende Abschnitt verfolgt das Ziel, die Effekte der sprachlichen Gestaltung zu isolieren, also den Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung „klinischer“, das heißt fokussierter, aber dafür weniger realitätsnah zu analysieren.

<sup>7</sup>Eine Linie verbindet den Gain-Score-Mittelwert, der sich ergibt, wenn eine Variable  $a$  in Ausprägungsform  $A1$  vorliegt und eine Variable  $b$  in Ausprägungsform  $B1$ , mit dem Gain-Score-Mittelwert für Variable  $a$  in Ausprägungsform  $A2$  und unveränderter Variable  $b$ , also  $B1$ . Verläuft nun die Vergleichs-Linie für die andere Ausprägungsform  $B2$  von Variable  $b$  (sie verbindet entsprechend den Mittelwert für  $A1/B2$  mit  $A2/B2$ ) schief liegend zur ersten Verbindungslinie, so ändern sich die Mittelwerte des Gain-Scores mit  $a$  offensichtlich andersartig, je nachdem ob  $b$  in Ausprägungsform  $B1$  oder  $B2$  vorliegt.

## 7. Ergebnisse und Diskussion

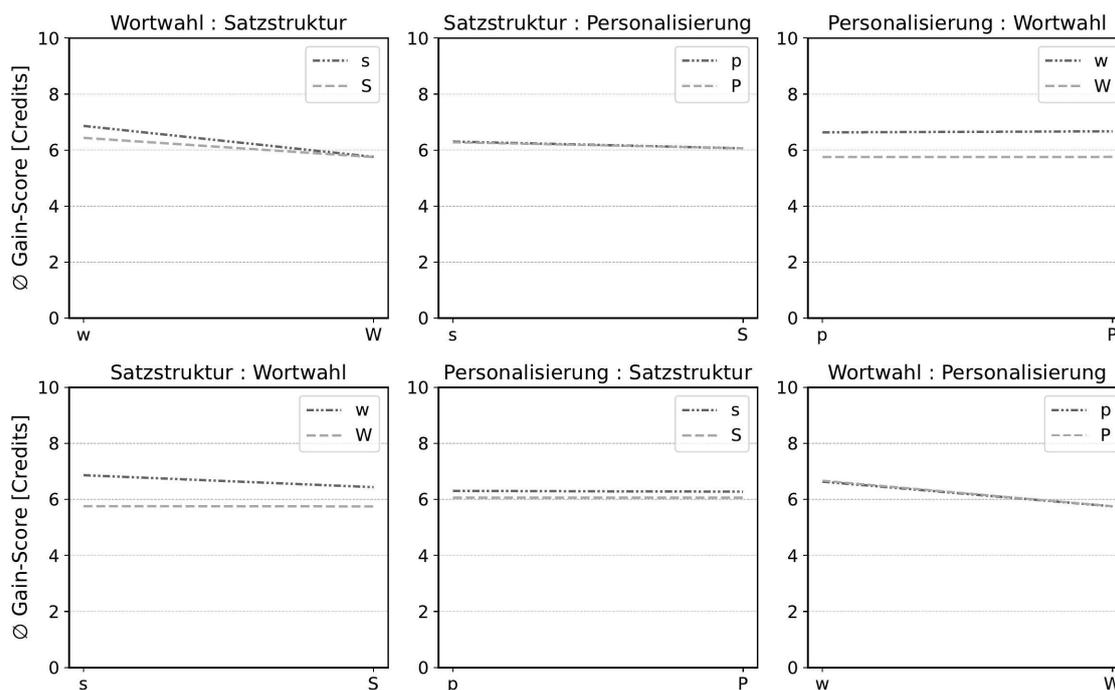


Abbildung 7.4.: Profilplots zur grafischen Darstellung der paarweisen Interaktionen zwischen den Variablen der sprachlichen Gestaltung **Wortwahl**, **Satzstrukturen** und **Personalisierung** (Stichprobengröße  $n_{\text{HauptstudieWT}} = 782$ ). In einem Profilplot für die Variablen  $a$  und  $b$  stellt eine Linie dabei (etwas unpräzise, aber anschaulich gesprochen) die Änderung des Mittelwerts des Gain-Scores dar, wenn Variable  $a$  ihre Ausprägungsform ändert, während die Ausprägung von  $b$  konstant bleibt. (Nahezu) parallele Linien bedeuten, dass sich  $a$  unabhängig von der Ausprägungsform von  $b$  auf den Gain-Score auswirkt. Die Legenden in den Plots zeigen jeweils, für welche Variablenausprägung die gepunktet-gestrichelte Linie ( $\cdots-$ ) und die gestrichelte Linie ( $-$ ) stehen.

### Zusammenhangsanalysen beschränkt auf Personen mit positivem Gain-Score

Den in den nächsten Zeilen berichteten Analysen liegt die folgende Annahme zugrunde: Für einen negativen Gain-Score waren andere Einflussgrößen entscheidender als die sprachliche Gestaltung der Texte. In Abschnitt 7.1.2 zu den Prä-Post-Vergleichen (genauer: auf Seite 130 ff.) wurden bereits Anzeichen dafür aufgeführt, welche anderen Einflussgrößen in Form von Moderatorvariablen hier eine Rolle gespielt haben könnten. Es deutet sich dabei unter anderem an, dass weniger die Textsorte ausschlaggebend war, sondern mitunter das Vorwissen und damit erfolgreiches Raten (siehe Seite 134). Es ist darüber hinaus denkbar und sogar naheliegend, dass das Set der in dieser Studie erhobenen und betrachteten Moderatorvariablen nicht ausschöpfend ist, sondern dass noch weitere Faktoren auf das Erzielen negativer Gain-Scores einwirkten. Es liegt auf der Hand, dass im Allgemeinen auch Faktoren wie Aufregung, Tagesform, Gewissenhaftigkeit etc. bei der Bearbeitung des Wissenstests relevant sind. Mit solchen Faktoren vor Augen erscheint es plausibel, dass zur Erklärung *negativer* Gain-Scores nicht die sprachliche Gestaltung der Texte in erster Linie entscheidend ist. Wäre die Textlektüre die entscheidende Ursache für einen negativen Gain-Score, so müss-

unabhängige Variable	F(7,708)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
W	7.65	.017 .006	.01*	6.71 (W) 7.63 (w) +0.92 ( $\Delta$ )
S	3.97	.093 .047	.00	6.83 (S) 7.48 (s) +0.65 ( $\Delta$ )
P	0.03	.890 .890	.00	7.14 (P) 7.14 (p) 0.00 ( $\Delta$ )
W : S	0.59	.442	.00	
W : P	1.13	.288	.00	
S : P	0.05	.824	.00	
W : S : P	1.65	.199	.00	

Tabelle 7.6.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen *Wortwahl*, *Satzstrukturen* und *Personalisierung* und deren Interaktionen („:“) in der Teilstichprobe von Personen mit positivem Gain-Score (Stichprobengröße  $n_{\text{HauptstudieWT\_pos}} = 709$ ). Es werden die Teststatistik F für die Modellparameter (7,708) und der p-Wert in unkorrigierter Form sowie in  $\alpha$ -Fehler-korrigierter Form berichtet. Nach einer Korrektur gemäß Bonferroni-Holm (siehe Abschnitt 6.4.3) ist ausschließlich der Haupt-Effekt der Wortwahl signifikant (Markierung mit \*\*\* für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$  in der Spalte für die Effektstärke  $\omega^2$ ). Ergänzend sind in der letzten Spalte die Gain-Score-Mittelwerte für W-Texte gegenüber w-Texten, für S-Texte gegenüber s-Texten und für P-Texte gegenüber p-Texten dargestellt. Mit  $\Delta$  ist der Unterschied „anspruchsgeminderte / schwach ausgeprägte Form“ minus „anspruchsgesteigerte / ausgeprägte Form“ bezeichnet.

te eine Person aufmerksam ihren Text lesen, dabei ihr Wissen über die elektrische Spannung (z. B. via kategorischer Missverständnisse) *verringern* und somit im Post-Test schlechter abschneiden als im Vortest. Dies mag in Einzelfällen vorgekommen sein, im Allgemeinen folgen die Analysen in den nächsten Zeilen aber der Annahme, dass das Ausmaß eines negativen Gain-Scores *nicht* maßgeblich durch die sprachliche Gestaltung eines Texts bestimmt war. Indem also nur die Teilstichprobe der Personen mit positivem Gain-Score betrachtet wird, kann – so die Annahme – ein Teil des zufälligen „Rauschens“ in den Daten entfernt werden und sprachliche Einflüsse treten deutlicher zutage.

Analog zum Vorgehen in der Gesamtstichprobe (Seite 137 ff.) wurden je ein ANOVA Modell für die sprachliche Gestaltung in Form der unabhängigen Variablen *W*, *S* und *P* und eines für die unabhängige Variable *Textversion* berechnet. Der *Gain-Score* stellte die abhängige Variable dar. Die Ergebnisse sind in Tabelle 7.6 beziehungsweise 7.7 festgehalten.

unabhängige Variable	F (7,708)	p	$\omega^2$	Mittelwerte
Textversion	2.01	.042	.01*	A: 7.73, B: 8.49, C: 7.17, D: 7.21 E: 7.34, F: 6.53, G: 6.43, H: 6.62

Tabelle 7.7.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der Textversion in der Teilstichprobe von Personen mit positivem Gain-Score (Stichprobengröße  $n_{\text{HauptstudieWT\_pos}} = 709$ ). Es werden die Teststatistik F für die Modellparameter (7,708), der p-Wert und der Effektstärke-Wert  $\omega^2$  berichtet. Ergänzend sind in der letzten Spalte die Gain-Score-Mittelwerte für alle Textversionen A bis H dargestellt.

In der Teilstichprobe der Personen mit positivem Gain-Score erscheint bei  $\alpha$ -Fehler-korrigierten p-Werten der Haupteffekt der Wortwahl signifikant bei kleiner Effektstärke  $\omega^2 = 0.01$  (vergleiche Seite 120 im Methoden-Abschnitt 6.4.3; die  $\alpha$ -Fehler-Korrektur nach Bonferroni-Holm wird in Abschnitt 6.4.3 erläutert; hier finden drei paarweise Vergleiche statt: W vs. w, S vs. s und P vs. p). Weitere signifikante Effekte der sprachlichen Gestaltung sind nicht zu beobachten.

Im zweiten Modell ist der Effekt der Textversion signifikant. Auch hier ist die Effektstärke als klein zu bezeichnen ( $\omega^2 = 0.01$ ).

Da die Textversion einen signifikanten Effekt zeigt, wurden in einer Post-Hoc-Analyse paarweise t-Tests durchgeführt. Vor einer  $\alpha$ -Fehler-Korrektur gemäß der Bonferroni-Holm-Methode (siehe Abschnitt 6.4.3) erwiesen sich fünf Mittelwertsunterschiede als hoch signifikant beziehungsweise signifikant. Diese sind untenstehend aufgelistet. Mit  $\Delta$  ist jeweils die Differenz der zweitgenannten zur erstgenannten Textversion bezeichnet. Nach der Korrektur liegt kein p-Wert mehr unterhalb der 5%-Marke.

- B (wsP) vs. G (WSp):  $p=.002 \xrightarrow{\alpha\text{-korr.}} p=0.06$ ,  $\Delta = 2.06$
- B (wsP) vs. F (WsP):  $p=.005 \xrightarrow{\alpha\text{-korr.}} p=0.14$ ,  $\Delta = 1.96$
- B (wsP) vs. H (WSP):  $p=.007 \xrightarrow{\alpha\text{-korr.}} p=0.19$ ,  $\Delta = 1.87$
- A (wsp) vs. G (WSp):  $p=.046 \xrightarrow{\alpha\text{-korr.}} p=1.00$ ,  $\Delta = 1.30$
- B (wsP) vs. C (wSp):  $p=.049 \xrightarrow{\alpha\text{-korr.}} p=1.00$ ,  $\Delta = 1.32$

Bei acht Textversionen, die miteinander zu vergleichen sind, ergeben sich 28 paarweise Vergleiche. Bei einer solch großen Anzahl an t-Tests sind die p-Werte in einer  $\alpha$ -Fehler-Korrektur deutlich nach oben zu berichtigen, sodass hier die Post-Hoc-Analyse keine signifikanten Unterschiede nachweisen konnte. In Anbetracht der p-Werte liegt dennoch nahe, dass vor allem die Unterschiede zwischen den Texten B (wsP) und G (WSp), B und F (WsP) sowie B und H (WSP) zu dem signifikanten ANOVA-Ergebnis (Tabelle 7.7) führen.

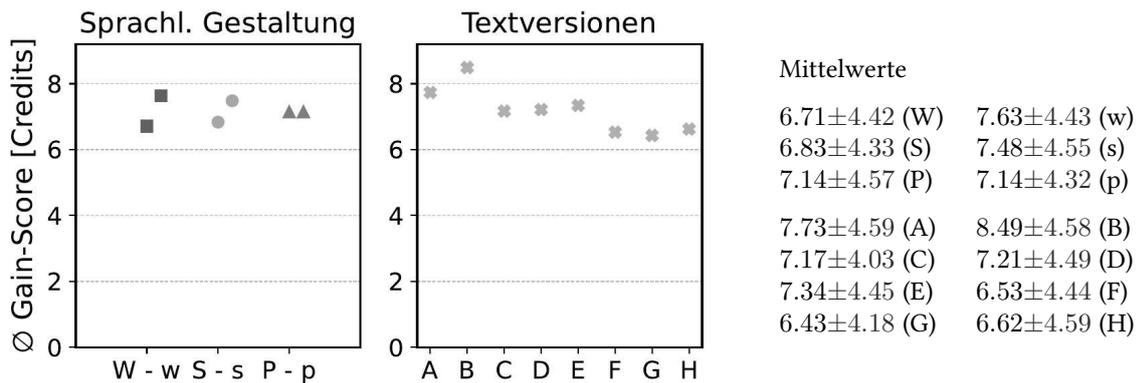


Abbildung 7.5.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung (links) beziehungsweise von der Textversion (rechts) in der Teilstichprobe der Personen mit positivem Gain-Score ( $n_{\text{HauptstudieWT\_pos}} = 709$ ). Es sind aus Gründen der optischen Darstellbarkeit keine Fehlerbalken abgebildet. Die Standardabweichungen sind in der nebenstehenden Tabelle jeweils in grau angegeben.

Die Textversion B (wsP) hebt sich in paarweisen Vergleichen am deutlichsten von den übrigen Textversionen ab. Es deutet sich an, dass dieses Ergebnis überzufällig ist und somit die Kombination *anspruchsgeminderte Wortwahl* mit *anspruchsgeminderten Satzstrukturen* mit *stark ausgeprägter Personalisierung* bei kleiner Effektstärke ( $\omega^2 = 0.01$ , vergleiche Seite 120 im Methoden-Abschnitt 6.4.3) günstiger für die Behaltensleistung war als andere Kombinationen.

Abbildung 7.5 ergänzt die Ergebnisse dieses Abschnitts um eine Visualisierung der Mittelwerte je nach sprachlicher Gestaltung (links in der Abbildung) beziehungsweise je nach Textversion (rechts in der Abbildung) in der Teilstichprobe der Personen mit positivem Gain-Score. Dabei bestätigt sich der Eindruck aus den vorherigen Analysen in der Gesamtstichprobe: Die Wortwahl scheint den deutlichsten Einfluss gehabt zu haben; die Texte A bis D (anspruchsgeminderte Wortwahl) führten zu den besseren Gain-Score-Ergebnissen als die Texte F bis G (anspruchsgesteigerte Wortwahl). Entgegen diesem Trend liegt der mittlere Gain-Score für Text E (Wsp) ebenfalls tendenziell hoch. Über den Unterschied zwischen W- und w-Texten hinaus fällt die Lage des Mittelwerts von Text B auf, der deutlich oberhalb der übrigen Textversionsmittelwerte liegt. Der Text B zeichnet sich also durch einen besonders hohen mittleren Gain-Score aus. Die Unterschiede zwischen den Mittelwerten liegen wie bei den Gesamtstichproben-Betrachtungen *deutlich* innerhalb der Bereiche der Standardabweichungen (jeweils zwischen 4 und 5 Credits; siehe die rechtsstehenden Angaben in Abbildung 7.5) und sind somit verglichen mit interpersonellen Unterschieden von minderm Ausmaß.

Insgesamt lassen sich die Befunde aus der Analyse der Gesamtstichprobe in der Teilstichprobe der Personen mit positivem Gain-Score wiederfinden. Dabei verdeutlichen sich die Effekte. Die Analysen führen also zu dem Ergebnis, dass ausschließlich der Haupteffekt der Wortwahl-Variable die Behaltensleistung bei kleiner Effektstärke signifikant beeinflusste.

### **Topikale Diskussion: Sprachliche Gestaltung – Behaltensleistung (ohne Moderatorvariablen)**

Bevor die Ergebnisse dieses Abschnitts 7.1.3 zusammengefasst und diskutiert werden, seien hinsichtlich der Deutung der Ergebnisse auf dem verkürzten Datensatz der Personen mit positivem Gain-Score drei Aspekte bemerkt: Erstens folgt die abhängige Variable *Gain-Score* auf dieser Teilstichprobe nicht mehr einer Normalverteilung<sup>8</sup>. Zwar sprechen die Stichprobengröße mit ca. 350 Personen pro zu vergleichender Gruppe (W vs. w, S vs. s und P vs. p) beziehungsweise ca. 100 für den Textversions-Vergleich (A vs. B vs. C usw.; siehe Tabelle 7.3 für die genauen Zahlen) und das Vorliegen von Varianzgleichheit<sup>9</sup> für die Robustheit der ANOVA-Analysen, dennoch ist die Normalverteilungsannahme erheblich verletzt. Zweitens sorgt das Vorgehen, den Datensatz künstlich auf eine Teilstichprobe zu verkürzen, dafür, dass der Datensatz nicht mehr die empirisch vorgefundene Realität widerspiegelt. Der betrachtete Ausschnitt der Stichprobe ist kein repräsentatives Abbild der Gesamtpopulation. Drittens jedoch ist bemerkenswert, dass die Signifikanz der Effekte steigt, *obwohl* die Stichprobe durch die Einschränkung auf Personen mit positivem Gain-Score kleiner wird. Der Statistik des Zufalls folgend wäre genau die gegenteilige Beobachtung zu erwarten; Ergebnisse werden in kleinen Stichproben *seltener* signifikant. Dies ist ein Indiz dafür, dass das Vorgehen dieser künstlichen Einschränkung der Stichprobe geeignet dafür ist, den Einfluss solcher Effekte, die nicht von der sprachlichen Gestaltung herrühren (das sogenannte *Rauschen* in den Daten), zu reduzieren und die Effekte sprachlicher Gestaltung zu isolieren und hervorzuheben.

Insgesamt lassen sich die Ergebnisse dieses Abschnitts 7.1.3 auf der Datengrundlage der gesamten sowie der eingeschränkten Stichprobe in dem Befund zusammenfassen, dass die sprachliche Textgestaltung einen nachweisbaren, jedoch sehr geringen Unterschied in der Behaltensleistung der Studienteilnehmenden bewirkte. Dies wird daraus geschlossen, dass in der Gesamtstichprobe (spätestens nach einer  $\alpha$ -Fehler-Korrektur) keine signifikanten Unterschiede im mittleren Gain-Score je nach sprachlicher Textgestaltung gemessen wurden. Am ehesten tendierte die Wortwahl dazu, signifikante Effekte auf die Behaltensleistung zu zeigen, jedoch ist die Effektstärke mit  $\omega^2 = 0.01$  ohnehin als sehr klein zu betrachten (siehe Seite 120 im Methoden-Abschnitt 6.4.3). Eine Einschränkung der Stichprobe auf solche Personen, die einen positiven Gain-Score verzeichnet hatten, ließ den Effekt der Wortwahl deutlicher hervortreten: sowohl wurde der Effekt der Wortwahl-Variable auf den Gain-Score signifikant

<sup>8</sup>Dies ist insofern erwartbar, als dass die Verteilung schlicht in einem *Ausschnitt* der Gesamtstichprobenverteilung besteht. Grafisch ist sie Abbildung 7.1 zu entnehmen, indem nur der Teil rechts der 0-Credits-Linie betrachtet wird.

<sup>9</sup>Levene-Test-Ergebnis:  $L = 0.43$ ,  $p = 0.83$

als auch zeigte die Textversion einen signifikanten Effekt. In paarweisen Vergleichen zeigte sich, dass dafür die Ausprägung der Wortwahl ursächlich war. Es manifestierte sich also der Befund, dass die Wortwahl hinsichtlich der Behaltensleistung einflussreicher als die beiden übrigen Variablen der sprachlichen Gestaltung – Satzstrukturen und Personalisierung – war. Erwähnenswert ist darüber hinaus, dass sich die Textversion B mit ihrer sprachlichen Gestaltung wsP am deutlichsten von den übrigen Textversionen abhebt, indem sie in mehreren paarweisen Textversionsvergleichen zu einem signifikant besseren mittleren Gain-Score in der Teilstichprobe der Personen mit positivem Gain-Score führte.

Da sich auch Interaktionseffekte als nicht signifikant erwiesen, ist ihre Bedeutsamkeit für die Realität vermutlich zu vernachlässigen. Die grafische Analyse lässt dennoch eine Tendenz erkennen, dass der Effekt der Satzstruktur auf die Behaltensleistung unwichtiger wurde, wenn die Wortwahl anspruchsgesteigert ausgeprägt war. Dies könnte bedeuten, dass eine anspruchsgesteigerte Wortwahl „dominant“ auf den Gain-Score einwirkte, sodass der Effekt der Satzstruktur nur bei anspruchsgeminderter Wortwahl durchschlug (während die Personalisierung in keinem der Fälle einen merkbaren Einfluss zeigte). Es sei jedoch erneut darauf hingewiesen, dass dieser in der Grafik sichtbare Interaktionseffekt möglicherweise zufällig auftrat ( $p = 0.58$ , siehe Tabelle 7.4).

Für die Praxis im Unterricht scheint es also angeraten, bei der Beurteilung von Texten in erster Linie auf eine anspruchsgeminderte Wortwahl und in zweiter Instanz auf anspruchsgeminderte Satzstrukturen sowie eine ausgeprägte Personalisierung zu achten. Dennoch werden in der Unterrichtsrealität mit all den übrigen Einflussfaktoren, die jenseits des Beobachtungsspektrums dieser Studie liegen (z. B. Tagesform, individuelles Interesse, Sorgfalt und Aufregung der Lernenden sowie Professionskompetenz der Lehrkraft<sup>10</sup>) weder die Wortwahl im Text noch die Satzstrukturen oder die Ausprägung der Personalisierung die Schlüsselgrößen zum kognitiven Leseerfolg sein.

Ein Vergleich der hier ermittelten Ergebnisse mit den Resultaten der Meta-Analyse von Strohmaier, Ehmke, Härtig und Leiss (2023) (siehe Unterkapitel 3.4) ist interessant: In der Meta-Analyse wurden erstens für die Personalisierung, zweitens für die Steigerung der Klarheit/Elaboration auf Wort- und Satzebene und drittens für die ebendiese Steigerung auf Textebene positive Effekte auf die Behaltensleistung gefunden. In der hier vorliegenden empirischen Studie hingegen blieb ein messbarer Effekt der Personalisierung aus. Dies mag daran liegen, dass sich die Effekte einer gesteigerten Personalisierung stärker in Transfer-Aufgaben niederzuschlagen scheinen (vergleiche Strohmaier, Ehmke, Härtig & Leiss, 2023, S. 5), der Wissenstest hier aber die Behaltensleistung anhand von textnah formulierten Aufgaben maß. Der Befund dieser Studie über einen schwachen, positiven Effekt der anspruchsgeminderten Wortwahl fügt sich in die Meta-Analyse-Ergebnisse. Eine anspruchsgeminderte Wortwahl ist zwar sicherlich vor allem der Meta-Analyse-Kategorie der „reduzierten Komplexität auf Wort-Ebene“ zuzuordnen, für welche *keine* Verständniseffekte nachgewiesen wurden. Sie ist

<sup>10</sup>Ein intuitiver Begriff dieser Beispiele genügt an dieser Stelle für die Konkretisierung der denkbaren, nicht erhobenen „Einflussfaktoren“; von einer wissenschaftliche Definition und Erläuterung wird daher abgesehen.

jedoch durchaus auch mit einer gesteigerten lokalen Elaboriertheit, wie Strohmaier, Ehmke, Härtig und Leiss (2023, S. 5) sie fassen, verbunden (z. B. Wortwiederholungen, Erklärungen zu schwierigen Wörtern). Eine hohe lokale Elaboriertheit erwies sich in der Meta-Analyse als günstig für das Verständnis.

#### 7.1.4 Zusammenhang zwischen sprachlicher Gestaltung und Behaltensleistung (mit Moderatorvariablen)

Aufbauend auf die bisher berichteten Ergebnisse bleibt noch eine interessante Frage offen. Nämlich die, inwiefern die Textgestaltung in unterschiedlicher Weise auf die Behaltensleistung von Schüler\*innen mit bestimmten Personenmerkmalen wirkte. In dieser Studie wurden dafür das Gender, die Jahrgangsstufe, die Schulart sowie die Lesefähigkeiten erhoben (siehe Unterkapitel 6.1). Um darüber hinaus in der Stichprobe hinsichtlich des Vorwissens zu differenzieren, wurde das Pre-Test-Ergebnis herangezogen. Es ist anzunehmen, dass die Wirkung der sprachlichen Textgestaltung je nach Ausprägung des Vorwissens, der Lesefähigkeiten, womöglich auch je nach Gender und je nach Schulart/Jahrgangsstufe, variiert. Für die Untersuchungen zu dieser Fragestellung wurde die Gesamtstichprobe im Folgenden jeweils in Gruppen hinsichtlich der Moderatorvariablen eingeteilt. Für die a priori kontinuierliche (also: nicht natürlich-kategoriale) Variable *Vorwissen* erfolgte dies anhand eines Mediansplits. Hinsichtlich der Lesefertigkeiten konnte auf die Prozentrangtabellen der Testnormierungsdaten zurückgegriffen werden (siehe Schneider, Schlagmüller und Ennemoser, 2017) und die Testpersonen damit der stärkeren beziehungsweise der schwächeren Populationshälfte zugeteilt werden. In den Subgruppen wurden dann jeweils ANOVAs für den Einfluss der sprachlichen Gestaltung hinsichtlich der Variablen Wortwahl (W), Satzstruktur (S) und Personalisierung (P) berechnet und verglichen<sup>11</sup>. Die grundlegende Überlegung für die Schlussfolgerungen aus den Vergleichen ist dabei die folgende: Falls eine sprachliche Variable  $X$  ( $X \in \{W, S, P\}$ ) in einer ersten Subgruppe hinsichtlich der Moderatorvariablen  $\Lambda$  ( $\Lambda \in \{\text{Vorwissen, Leseverständnis, ...}\}$ ) einen signifikanten Einfluss auf die Behaltensleistung hat und in einer zweiten  $\Lambda$ -Subgruppe nicht, dann erweist sich ausschließlich für Personen der ersten  $\Lambda$ -Subgruppe die Variable  $X$  als bedeutsam. Daraus folgt wiederum, dass die Moderatorvariable  $\Lambda$  relevant für das Wirkungsgefüge zwischen  $X$  (sprachliche Gestaltung) und Behaltensleistung ist: die Moderatorvariable  $\Lambda$  interagiert mit der sprachlichen Variablen  $X$ .

Tabelle 7.8 fasst die Ergebnisse der Analysen in den Subgruppen je nach Moderatorvariable bündig zusammen. Es wurde für jede Subgruppe hinsichtlich der einzelnen Moderatorvariablen eine ANOVA mit den unabhängigen Variablen  $W$ ,  $S$  und  $P$  und der abhängigen Variablen *Gain-Score* berechnet. Die vollständigen ANOVA-Tabellen sind im Anhang D.6 aufgeführt; dort findet sich auch eine Übersicht über die Teilstichprobengrößen (Tabelle D.7). Aus diesen

---

<sup>11</sup>Auf erneute Robustheitsanalysen wurde verzichtet. Auf Grundlage entsprechender Tests für die Analysen ohne Moderatorvariablen (Anhang D.3) ist auch in den Subgruppen von Robustheit auszugehen.

	$\omega^2$					
	W		S		P	
Gender	weiblich		männlich		divers	
	.01	.00	.00	.00	.00	.00
					-.02	-.01   -.03
Lesefähigkeiten	niedrig			hoch		
Verständnis	.02*	.00	.00	.00	.00	.00
Geschwindigkeit	.02**	.00	.00	.00	.00	.00
Genauigkeit	.01	.00	.00	.00	.00	.00
Schulart	Realschule			Gymnasium		
	.03**	.00	.00	.00	.00	.00
Jahrgangsstufe	8. Klasse			9. Klasse		
	.00	-.01	.00	.01	.00	.00
Vorwissen	niedrig			hoch		
	.02**	.00	.00	.00	.00	.00

Tabelle 7.8.: Effektstärken ( $\omega^2$ ) aus ANOVA-Modellen für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen *Wortwahl*, *Satzstrukturen* und *Personalisierung*. Es wurde für Teilstichproben bezüglich des Genders, bezüglich der Lesefähigkeiten (siehe dazu Abschnitt 6.3.1), bezüglich des Schulstatus und bezüglich des Vorwissens jeweils eine ANOVA berechnet. Für die einzelnen Teilstichprobengrößen siehe Tabelle D.7 im Anhang. Signifikante Effekte sind mit Sternchen markiert (\*\*\*) für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$ ); dabei wurde eine  $\alpha$ -Fehler-Korrektur nach Bonferroni-Holm (siehe ebenfalls Abschnitt 6.4.3) angewandt.

einzelnen ANOVAs werden der Übersichtlichkeit halber jeweils lediglich die Effektstärken der Haupteffekte gezeigt und miteinander verglichen. Dabei sind signifikante Effekte mit Sternchen markiert (die p-Werte wurden zuvor  $\alpha$ -Fehler-korrigiert nach Bonferroni-Holm; siehe Abschnitt 6.4.3). Interaktionseffekte sind in der Tabelle nicht aufgeführt; sie waren in nahezu keinem Fall signifikant<sup>12</sup>. Zusammengefasst lässt sich aus Tabelle 7.8 ablesen:

Signifikante Effekte zeigt ausschließlich die Wortwahl-Variable. Satzstrukturen und Personalisierung wirkten sich in keiner Subgruppe signifikant auf die Behaltensleistung aus. Die Effektstärken der Wortwahl-Variablen sind als klein bis mittelgroß zu bezeichnen ( $0.01 \leq \omega^2 \leq 0.06$ , vergleiche Seite 120 im Methoden-Abschnitt 6.4.3). Die Ergebnisse zeigen, dass ausschließlich für Personen mit geringerem Leseverstehen, für Personen mit geringerer Lesegeschwindigkeit, für Realschüler\*innen und für Personen mit geringerem Vorwissen die Wortwahl einen signifikanten Effekt auf die Behaltensleistung hatte.

Die folgenden Unterabschnitte ergänzen die Ergebnistabelle 7.8 um die zugehörigen quantitativen sowie graphischen Mittelwertsvergleiche. Die Grafiken und Zahlenwerte werden jeweils kurz hinsichtlich ihrer Auffälligkeiten beschrieben. Es sei bereits hier das Ergebnis vorweggegriffen, welches sich in den einzelnen Unterabschnitten zu den Mittelwertsvergleichen zeigen wird:

In jedem Fall führte die Wortwahl in ihrer anspruchsgeminderten Form zu einem höheren oder mindestens gleich großem Gain-Score. Eine anspruchsgeminderte Wortwahl erwies sich also niemals als ungünstig, sondern entweder als irrelevant oder sogar als günstig für die Behaltensleistung.

### **Der Gain-Score je nach Gender**

Teilt man die Stichprobe je nach Gender in Teilstichproben auf, so wirkte sich die sprachliche Gestaltung in keiner der Subgruppen signifikant auf die Behaltensleistung aus (siehe Tabelle 7.8).

Abbildung 7.6 zeigt, dass über die sprachliche Gestaltung hinweg der Gain-Score der weiblichen Teilnehmerinnen tendenziell am höchsten liegt. Hinsichtlich der Sprachgestaltungsvergleiche führte die Wortwahl in ihrer anspruchsgeminderten Form jeweils zu den höheren Gain-Scores. Die Satzstrukturen zeigen keine klare Tendenz über alle Subgruppen hinweg. Im Vergleich weiblich vs. männlich zeigte sich ein stärkerer Einfluss der Satzstrukturen bei den weiblichen Teilnehmerinnen (anspruchsgeminderte Form günstiger). Die Personalisierung

<sup>12</sup>Genau genommen liegt der p-Wert für den Interaktionseffekt W:S in der Teilgruppe niedrigen Vorwissens mit .033 unterhalb der 5 %-Schwelle (siehe dazu im Einzelnen Tabelle D.14 im Anhang D.6). Der Wert ist jedoch nicht  $\alpha$ -Fehler-korrigiert, die zugehörige Effektstärke ist mit  $\omega^2 = 0.01$  klein und die Interaktion zeigt in keiner anderen Subgruppe einen signifikanten Effekt. Aus diesem Grund wird auf diese Beobachtung nicht weiter eingegangen.

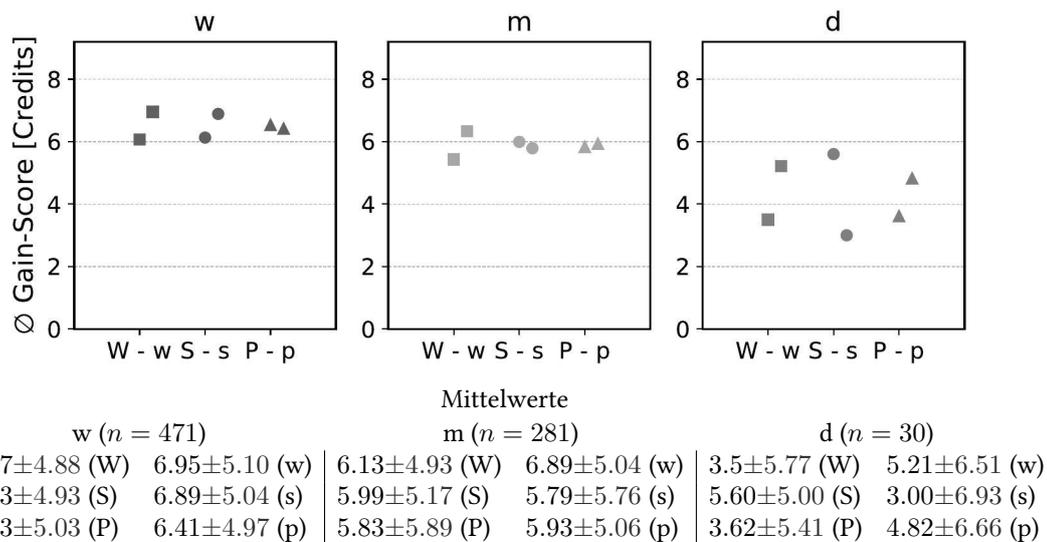


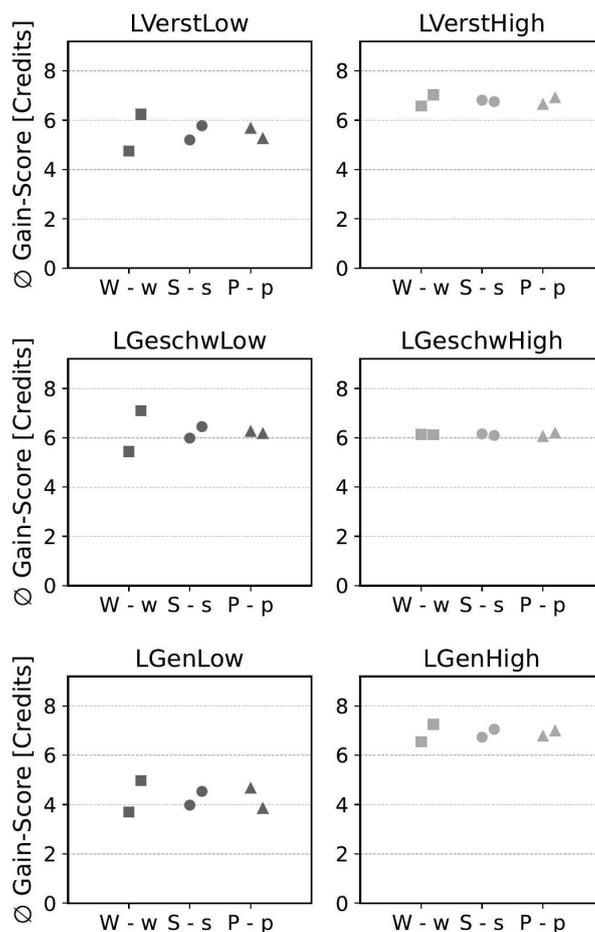
Abbildung 7.6.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung je nach Gender: weiblich (w) links, männlich (m) in der Mitte und divers (d) rechts. Die Teilstichprobengrößen sowie die Zahlenwerte der abgebildeten Mittelwerte inklusive der Standardabweichungen (in grau) sind der Tabelle unterhalb der Grafiken zu entnehmen.

weist die geringsten Unterschiede auf. Die Mittelwertsverteilung in der divers-Subgruppe sticht am meisten ins Auge, da sie sich deutlich von den Grafiken für die weiblichen und die männlichen Studienteilnehmenden unterscheidet: Die Subgruppe der Personen mit Genderangabe *divers* erzielte erstens – unabhängig von der sprachlichen Gestaltung – schwächere Gain-Scores als Personen mit Angabe *weiblich* oder *männlich*. Zweitens sind die Mittelwerte je nach Textgestaltung zwar deutlich unterschiedlich ausgeprägt, allerdings war die Subgruppengröße mit 30 Personen so klein und die Breite der Gain-Score-Verteilungen mit Standardabweichungen bei etwa 6 Credits (siehe die untenstehenden Angaben in Abbildung 7.6) so groß, dass die Unterschiede sich mit großer Wahrscheinlichkeit zufällig so darstellen dürften (p-Werte für Wortwahl, Satzstruktur und Personalisierung: 0.47, 0.28 und 0.76; siehe auch Tabelle D.8 im Anhang D.6).

### Der Gain-Score je nach Lesefähigkeiten

In den Subgruppen je nach Lesefähigkeiten (Leseverständnis, Lesegeschwindigkeit und Lesegenauigkeit, siehe Abschnitt 6.3.1) waren signifikante beziehungsweise hoch signifikante Effekte der Wortwahl-Variable in den Subgruppen niedrigen Leseverständnisniveaus beziehungsweise niedrigen Lesegeschwindigkeitsniveaus zu messen (die p-Werte wurden gemäß der Bonferroni-Holm-Methode  $\alpha$ -Fehler-korrigiert; siehe dafür Abschnitt 6.4.3). Die Effektstärken sind als klein bis mittelgroß zu betrachten (für die Richtwerte siehe Seite 120 im Methoden-Abschnitt 6.4.3; für Leseverständnis und -geschwindigkeit ist jeweils  $\omega^2 = 0.02$ ). In den übrigen Subgruppen zeigt die sprachliche Gestaltung keine signifikanten Effekte.

## 7. Ergebnisse und Diskussion



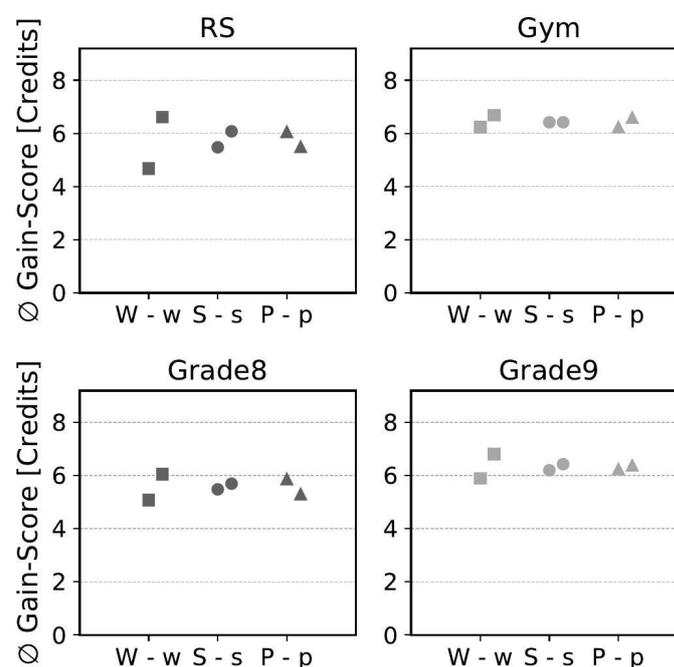
### Mittelwerte

LVerst niedrig ( $n = 351$ )		LVerst hoch ( $n = 429$ )	
$4.75 \pm 5.30$ (W)	$6.25 \pm 5.18$ (w)	$6.57 \pm 4.88$ (W)	$7.03 \pm 5.37$ (w)
$5.20 \pm 4.84$ (S)	$5.78 \pm 5.74$ (s)	$6.81 \pm 5.02$ (S)	$6.75 \pm 5.21$ (s)
$5.67 \pm 5.45$ (P)	$5.26 \pm 5.12$ (p)	$6.64 \pm 5.30$ (P)	$6.91 \pm 4.94$ (p)
LGeschw niedrig ( $n = 424$ )		LGeschw. hoch ( $n = 358$ )	
$5.44 \pm 4.95$ (W)	$7.10 \pm 5.02$ (w)	$6.13 \pm 5.38$ (W)	$6.12 \pm 5.57$ (w)
$5.99 \pm 4.73$ (S)	$6.45 \pm 5.34$ (s)	$6.16 \pm 5.30$ (S)	$6.09 \pm 5.65$ (s)
$6.26 \pm 5.04$ (P)	$6.17 \pm 5.06$ (p)	$6.05 \pm 5.84$ (P)	$6.19 \pm 5.12$ (p)
LGen niedrig ( $n = 207$ )		LGen hoch ( $n = 573$ )	
$3.70 \pm 5.18$ (W)	$4.97 \pm 4.59$ (w)	$6.55 \pm 4.92$ (W)	$7.62 \pm 5.40$ (w)
$3.98 \pm 4.55$ (S)	$4.54 \pm 5.28$ (s)	$6.74 \pm 4.96$ (S)	$7.06 \pm 5.38$ (s)
$4.67 \pm 4.84$ (P)	$3.85 \pm 5.06$ (p)	$6.78 \pm 5.48$ (P)	$6.99 \pm 4.84$ (p)

Abbildung 7.7.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung je nach Lesefähigkeiten. Die Stichprobe wurde per Mediansplit für das Leseverständnis (LVerst), die Lesegeschwindigkeit (LGeschw) und die Lesegenauigkeit (LGen) jeweils in zwei Teilstichproben (Niveau niedrig: Low und Niveau hoch: High) geteilt. Die Teilstichprobengrößen sowie die Zahlenwerte der abgebildeten Mittelwerte inklusive der Standardabweichungen (in grau) sind der Tabelle unterhalb der Grafiken zu entnehmen.

*In der Lesefähigkeitsmessung gab es fehlende Werte. In diesen Fällen wurden einzelne Personen keiner Teilstichprobe zugeteilt und die Summe der beiden Teilstichprobengrößen (Level niedrig plus Level hoch) entspricht nicht der Gesamtstichprobengröße von 782.*

Abbildung 7.7 zeigt, dass die Mittelwerte für unterschiedliche sprachliche Gestaltungen generell in den Subgruppen niedrigerer Lesefähigkeitsniveaus größere Abstände haben. In diesen Subgruppen war die günstigere Form hinsichtlich der Wortwahl die anspruchsgeminderte Realisierung (w günstiger als W), hinsichtlich der Satzstrukturen ebenfalls die anspruchsgeminderte Realisierung (s günstiger als S) und hinsichtlich der Personalisierung der ausgeprägtere Grad (P günstiger als p). Es fällt darüber hinaus auf, dass die Gain-Score-Mittelwerte bei hohem Leseverständnis und hoher Lesegenauigkeit ganz allgemein höher liegen als bei niedrig ausgeprägten Fähigkeiten. Die Lesegeschwindigkeit bewirkte diesbezüglich keine Unterschiede<sup>13</sup>.



Mittelwerte

Realschule ( $n = 293$ )		Gymnasium ( $n = 489$ )	
4.67±5.10 (W)	6.61±5.26 (w)	6.23±5.11 (W)	6.68±5.34 (w)
5.48±4.90 (S)	6.08±5.63 (s)	6.42±5.04 (S)	6.42±5.39 (s)
6.06±5.49 (P)	5.49±5.06 (p)	6.04±5.35 (P)	6.60±5.06 (p)
Jahrgangsstufe 8 ( $n = 147$ )		Jahrgangsstufe 9 ( $n = 635$ )	
5.07±5.19 (W)	6.05±5.08 (w)	5.89±5.14 (W)	6.81±5.35 (w)
5.48±4.60 (S)	5.69±5.68 (s)	6.20±5.09 (S)	6.43±5.43 (s)
5.86±5.21 (P)	5.30±5.09 (p)	6.24±5.44 (P)	6.38±5.07 (p)

Abbildung 7.8.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung je nach Schulstatus: Realschule (RS) gegenüber Gymnasium (Gym) sowie achte Jahrgangsstufe (Grade 8) gegenüber neunter Jahrgangsstufe (Grade 9). Die Teilstichprobengrößen sowie die Zahlenwerte der abgebildeten Mittelwerte inklusive der Standardabweichungen (in grau) sind der Tabelle unterhalb der Grafiken zu entnehmen.

<sup>13</sup>Es sei hier erneut betont, dass die Phase der Textlektüre für die Schüler\*innen nicht zeitlich begrenzt war (vergleiche Abschnitt 6.5.1).

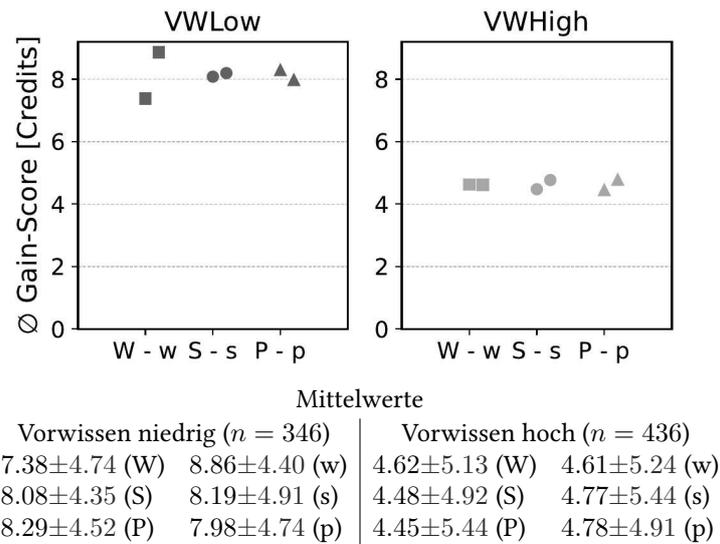


Abbildung 7.9.: Mittelwerte für den Gain-Score in Abhängigkeit von der sprachlichen Gestaltung je nach Vorwissen. Die Gesamtstichprobe wurde per Mediansplit in zwei Teilstichproben geteilt (VWLow und VWHigh). Die Teilstichprobengrößen sowie die Zahlenwerte der abgebildeten Mittelwerte inklusive der Standardabweichungen (in grau) sind der Tabelle unterhalb der Grafiken zu entnehmen.

### Der Gain-Score je nach Schulstatus

In den Subgruppen je nach Schulstatus (Schulart und Jahrgangsstufe) wurde ein hoch signifikanter Effekt der Wortwahl-Variable in der Subgruppe der Realschüler\*innen verzeichnet (die p-Werte wurden gemäß der Bonferroni-Holm-Methode  $\alpha$ -Fehler-korrigiert; siehe dafür Abschnitt 6.4.3). Die Effektstärke ist als klein bis mittelgroß zu betrachten ( $\omega^2 = 0.03$ , vergleiche Seite 120 im Methoden-Abschnitt 6.4.3 für die Richtwerte). In den übrigen Subgruppen zeigte die sprachliche Gestaltung keine signifikanten Effekte.

Abbildung 7.8 zeigt, dass die Mittelwerte für unterschiedliche sprachliche Gestaltungen generell in der Realschule größere Abstände haben als im Gymnasium. Für die unterschiedlichen Jahrgangsstufen ist dies nicht zu beobachten. Tendenziell erwiesen sich allgemein eine anspruchsgeminderte Wortwahl, anspruchsgeminderte Satzstrukturen und eine ausgeprägte Personalisierung als günstiger für den Gain-Score. Es fällt darüber hinaus auf, dass in Jahrgangsstufe 9 ganz allgemein *leicht* höhere Gain-Scores erzielt wurden als in Jahrgangsstufe 8 und ebenso im Gymnasium im Vergleich zur Realschule.

### Der Gain-Score je nach Vorwissen

In den Subgruppen je nach Vorwissensniveau zeigt die Wortwahl-Variable einen hoch signifikanten Effekt in der Subgruppe der Personen mit geringerem Vorwissen (die p-Werte wurden gemäß der Bonferroni-Holm-Methode  $\alpha$ -Fehler-korrigiert; siehe dafür Abschnitt 6.4.3). Die Effektstärke ist als klein bis mittelgroß zu betrachten ( $\omega^2 = 0.02$ , vergleiche Seite 120 im Methoden-Abschnitt 6.4.3 für die Richtwerte). In der Subgruppe mit hohem Vorwissen zeigte

die sprachliche Gestaltung keine signifikanten Effekte.

In Abbildung 7.9 ist zu sehen, dass die Mittelwerte für unterschiedliche Wortwahl-Ausprägung generell in der Gruppe mit niedrigerem Vorwissensniveau größere Abstände haben als in der Gruppe mit höherem Vorwissensniveau. Bei geringerem Vorwissen erscheint eine anspruchsgeminderte Wortwahl deutlich günstiger als eine anspruchsgesteigerte. Es fällt darüber hinaus auf, dass bei geringerem Vorwissen ganz allgemein deutlich höhere Gain-Scores erzielt wurden als bei größerem Vorwissen<sup>14</sup>.

### **Topikale Diskussion: Sprachliche Gestaltung – Behaltensleistung (mit Moderatorvariablen)**

Die folgende Diskussion wird sich auf das zusammenfassende Fazit aus den spezifischen Subgruppenvergleichen (Seite 148) beziehen. Es fällt beim Blick auf Tabelle 7.8 auf, dass in keinem Fall signifikante Effekte in den leistungsstärkeren Subgruppen (jeweils rechte Hälfte der Tabelle mit Ausnahme der Gender-Zeile) vorliegen, sondern jeweils nur in den Subgruppen mit niedrigerem Leistungsniveau. Personen mit geringerem Leseverstehen, geringerer Lesegeschwindigkeit, Personen aus der Realschule und Personen mit geringerem Vorwissen scheinen also sensibler auf die sprachliche Gestaltung eines Text hinsichtlich der Wortwahl reagiert zu haben als Personen mit jeweils entsprechendem höheren Leistungsniveau. Dabei sei bemerkt, dass sich Signifikanz offensichtlich nicht nur als Folge der Gruppengröße einstellte (aus mathematischen Gründen werden Effekte mit steigender Gruppengröße schneller signifikant), denn im jeweiligen Vergleich liegen keineswegs die signifikanten Effekte stets in den zahlenmäßig größeren Subgruppen (für die exakten Teilstichprobengrößen siehe Tabelle D.7 im Anhang D.6).

Dabei ist es stets die Wortwahl, die – sofern ein Effekt der sprachlichen Gestaltung signifikant wurde – diesen verursacht hat. Für die oben genannten leistungsschwächeren Subgruppen erwies sich also eine anspruchsgeminderte Wortwahl mit signifikantem, kleinem bis mittelgroßem Unterschied als günstig für die Behaltensleistung beim Lesen. Für die übrigen Subgruppen sowie für die sprachlichen Gestaltungsvariablen Satzstrukturen und Personalisierung wurden keine signifikanten Effekte gemessen.

In der Meta-Analyse von Strohmaier, Ehmke, Härtig und Leiss (2023) wurden das inhaltliche sowie das sprachliche Vorwissen als potenzielle Moderatorvariablen betrachtet. Dabei fanden die Autoren, dass Personen mit geringerem inhaltlichem Vorwissen mehr von sprachlichen Modifikationen profitieren. Dies steht im Einklang zu den im vorangegangenen Abschnitt vorgestellten Ergebnissen dieser empirischen Studie. In der Meta-Analyse zeigte sich kein signifikanter Effekt des sprachlichen Vorwissens, während in der hier vorliegenden Studie das Leseverstehen und die Lesegeschwindigkeit durchaus relevant gewesen zu sein scheinen. Jedoch argumentieren die Autoren der Meta-Analyse, dass ein solcher sprachlicher Vorwissenseffekt durchaus existieren könnte und lediglich in der Meta-Analyse verborgen blieb,

<sup>14</sup>In diesem Zusammenhang sei auf die vorangehende Diskussion zum Zusammenhang von Vorwissen und Gain-Score im Kontext der negativen Gain-Scores verwiesen (siehe Seite 134 in Abschnitt 7.1.2).

da er womöglich anhand des Designs und der Methode nicht nachweisbar war (Strohmaier, Ehmke, Härtig & Leiss, 2023, S. 13).

In Hinsicht auf die Praxis im Unterricht lässt sich aus diesem Abschnitt 7.1.4 schließen, dass Texte mit einer anspruchsgeminderten Wortwahl zum Zweck des inhaltlichen, kognitiven Lernerfolgs (im Sinne einer Behaltensleistung) günstiger zu sein scheinen als Texte mit anspruchsgesteigerter Wortwahl. Denn einzelne Subgruppen profitieren hinsichtlich der Behaltensleistung von der anspruchsgeminderten Wortwahl, für andere Subgruppen ist sie irrelevant.

In den vorangehenden Abschnitten 7.1.2, 7.1.3 und 7.1.4 wurden die Ergebnisse in Bezug auf Forschungsfrage 1 (Zusammenhang von sprachlicher Gestaltung und Behaltensleistung) kleinschrittig dargestellt. Der folgende Abschnitt wird nun das Gesehene zusammenfassen und summarisch im Kontext von Unterrichts- und Textproduktionspraxis diskutieren.

### **7.1.5 Zusammenfassung und summarische Diskussion der Wissenstest-Ergebnisse**

Um sich der Forschungsfrage 1 zuzuwenden und die Behaltensleistung der Schüler\*innen beim Lesen des Texts zur elektrischen Spannung in Zahlen abzubilden, wurde der Gain-Score als Differenz zwischen Pre-Test und Post-Test für die Behaltensleistung betrachtet. Wie in Abschnitt 7.1.2 gezeigt wurde, ist davon auszugehen, dass die Schüler\*innen anhand der Textlektüre tatsächlich einen Lernzuwachs (im Sinne einer spontan messbaren Behaltensleistung) verzeichneten, sodass der Gain-Score ein geeignetes Maß für die Behaltensleistung beim Lesen ist. Es traten auch Fälle auf (ca. 10 % der Studienteilnehmenden), die einen negativen Gain-Score verzeichneten. Eine Erklärung hierfür können Effekte liefern, die auf der Statistik des Zufalls beruhen (Regression zur Mitte), die psychologischer Natur sind (Expertise-Reversal-Effekt) und die mit einer mangelnden Leseleistung zu tun haben (bei schwachem Leseverständnis und schwacher Lesegenauigkeit werden auffallend oft negative Gain-Scores erzielt).

Die Zusammenhänge zwischen der sprachlichen Textgestaltung hinsichtlich *Wortwahl / morphologische Auffälligkeiten auf Wortebene, Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene* und *erzählende und personalisierende Elemente* auf der einen Seite und der Behaltensleistung auf der anderen Seite traten nicht sofort und ohne Weiteres in Erscheinung (siehe Abschnitt 7.1.3). Es deutete sich bei der Betrachtung der Gesamtstichprobe an, dass allenfalls die Wortwahl einen überzufälligen Effekt kleiner Effektgröße auf die Behaltensleistung zeigen könnte. Anhand weiterführender Analysen (Betrachtung der Teilstichprobe mit positiven Gain-Scores sowie die Betrachtung von Subgruppen mit spezifischen Personenmerkmalen; für letzteres siehe Abschnitt 7.1.4) verfestigte sich dieser erste Eindruck: in mehreren Subgruppen erwies sich die Wortwahl-Variable in ihrer anspruchsgeminderten Form signifikant als günstiger für die Behaltensleistung (in anderen Subgruppen bleibt die sprachliche Gestaltung ohne messbaren Effekt auf die Behaltensleistung). Diese Subgruppen,

für die eine anspruchsgeminderte Wortwahl relevant für einen höheren Gain-Score war, sind neben der Teilstichprobe der Personen mit positivem Gain-Score die Subgruppe der Personen mit geringerem Leseverstehen, der Personen mit geringerer Lesegeschwindigkeit, der Realschüler\*innen und der Personen mit geringerem Vorwissen. Ein nachteiliger Effekt der Wortwahl auf die Behaltensleistung wurde in keinem Fall zu verzeichnet. Generell konnten außerdem keine signifikanten Interaktionseffekte der Variablen Wortwahl, Satzstrukturen und Personalisierung nachgewiesen werden.

### **Summarische Diskussion**

Die bisher gefundenen Ergebnisse sind konsistent mit ähnlich angelegten Studien, die sich mit dem Zusammenhang der Textgestaltung und des Lernzuwachs von Schüler\*innen beim Lesen naturwissenschaftlicher Texte beschäftigen: So maßen Hackemann, Heine und Höttecke (2022), dass die Lösungswahrscheinlichkeit für Wissenstest-Items unabhängig davon ist, welches Anspruchsniveau der zugehörige Informationstext aufwies. Auch Härtig, Fraser, Bernholt und Retelsdorf (2019) fanden in drei miteinander verwandten Teilstudien wiederholt jeweils keinen signifikanten Effekt der sprachlichen Textgestaltung auf das Abschneiden der Schüler\*innen im Verständnistest. Ein entscheidender Vorteil der dieser Arbeit zugrundeliegenden Studie ist, dass das Konglomerat an sprachlichen Veränderungen, welches die Textversionen ausmacht, in den statistischen Analysen aufgebrochen werden konnte. Denn die acht unterschiedlichen Textversionen weisen jeweils eine spezifische Ausprägung hinsichtlich jeder der drei Variablen W, S und P auf (vergleiche Abschnitt 6.2.2). Auf diese Art und Weise konnten die Untersuchungen zu Textversionseffekten auf die Analyse hinsichtlich einzelner Variablen der sprachlichen Gestaltung ausgeweitet werden. Dadurch konnte die besondere Rolle der Wortwahl vor allem in einzelnen Subgruppen (siehe die Zusammenfassung oben) ans Licht gebracht werden. Im Gegensatz zu den Meta-Studienergebnissen von Strohmaier, Ehmke, Härtig und Leiss (2023) (siehe Unterkapitel 3.4) trat in dieser Studie der in der Meta-Analyse aufgedeckte kognitionsbezogene Effekt der Personalisierung nicht auf, sondern es erwies sich die Wortwahl, also in erster Linie eine Variable der sprachlichen Komplexität, als einflussreicher. Über die Gründe hierfür kann an dieser Stelle nur gemutmaßt werden. Da sich Einzelstudien zur Textverständlichkeit unter anderem in den ganz konkreten sprachlichen Modifikationsmaßnahmen, im Textthema sowie in den Messmethoden für die abhängige Variable unterscheiden (vergleiche dazu Strohmaier, Ehmke, Härtig und Leiss, 2023, S. 6) sind vollkommen einheitliche Studienergebnisse nicht zu erwarten. Die gefundenen (kleinen) Effektstärken in dieser Studie sowie der Befund, dass Personen mit geringerem Vorwissen stärker von Textmodifikationen profitieren, stehen im Einklang mit den Meta-Analyse-Resultaten.

Eine Implikation der Ergebnisse für Autor\*innen ist, dass sie ihren kreativen Neigungen hinsichtlich der Wortwahl, der Satzstrukturen und der Personalisierung beim Textproduktionsprozess frei folgen können, ohne dadurch eine starke, negative Beeinträchtigung der

Behaltensleistung bei Schüler\*innen befürchten zu müssen<sup>15</sup>. Es ist auf Grundlage dieser Stichprobendaten anzunehmen, dass in der Realität einer vielfältigen Schülerschaft mit ihren heterogenen Lernvoraussetzungen die sprachliche Gestaltung eines physikalischen Sachtexts keinen ausschlaggebenden Einfluss auf die Behaltensleistung der Lernenden beim Lesen hat. Dennoch scheint es in keinem Fall zu schaden, sondern unter Umständen zu nützen, den Anspruch vor allem hinsichtlich der Wortwahl gering zu halten sowie darüber hinaus anspruchsgeminderte Satzstrukturen zu verwenden.

Betrachtet man nun entsprechend den bisher berichteten Ergebnissen die Wortwahl als die einflussreichste Größe im Vergleich zu Satzstrukturen und Personalisierung, so ist Folgendes in Betracht zu ziehen: Womöglich waren es nicht die Lesenden, die auf eine variierte Wortwahl stärker reagieren als auf variierte Satzstrukturen oder Personalisierung, sondern es waren die Texte, die sich in der Wortwahl untereinander am stärksten unterscheiden. Diese Arbeit hält keine endgültige Antwort auf diese Überlegung bereit. In Abschnitt 6.2.2 wurde jedoch ausführlich dargelegt, inwiefern hinsichtlich *jeder* Variable das Prinzip der Effektmaximierung das Ziel war und somit jeweils ein möglichst niedriges gegenüber einem möglichst hohen Anspruchs- beziehungsweise Ausprägungsniveau realisiert wurde. Die verwendeten Textversionen als Ganzes sowie eine übersichtliche Auflistung aller variierten Textstellen sind im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> bzw. im Kapitel A einsehbar, sodass die Methode vollkommen transparent ist, eine individuelle Einschätzung hinsichtlich der oben aufgeworfenen Überlegung ermöglicht wird und Anschlussstudien angeregt werden können.

Es ist außerdem bei der Suche nach Zusammenhängen zwischen Textgestaltung und Behaltens- beziehungsweise Verständnisleistungen zu bedenken, dass die sprachlichen Veränderungen womöglich durchaus eine Wirkung bei den Lesenden erzeugten, diese jedoch anhand des verwendeten Messinstruments für die Behaltens- beziehungsweise Verständnisleistungen nicht zutage trat. Härtig, Fraser, Bernholt und Retelsdorf (2019) bringen diesbezüglich den Erklärungsansatz vor, „dass die [in den Texten] durchgeführten Veränderungen im Kern den Übergang in das propositionale Modell des Texts unterstützen sollen. Es lässt sich jedoch nicht ausschließen, dass für die Beantwortung der Fragen das erfolgreiche Herstellen eines Situationsmodells viel entscheidender ist“ (die Begriffe der Propositionen und des Situationsmodells werden in Abschnitt 3.2.1 erläutert). Tatsächlich zeigen Bernholt, Härtig und Retelsdorf (2023), dass sich Schüler\*innen durch die Lektüre eines physikalischen Sachtexts deutlicher in solchen Aufgaben steigerten, die anhand von Code-basierter Erinnerung zu beantworten sind als in solchen, welche Inferenzen erfordern und die Situationsmodell-Ebene adressieren. Diesem Befund steht jedoch erstens entgegen, dass sich in der umfangreichen Meta-Analyse von Strohmaier, Ehmke, Härtig und Leiss (2023) (siehe Unterkapitel 3.4) das

---

<sup>15</sup>Offensichtlich ist ein gewisses Mindestmaß an sprachlicher Passung für die Adressat\*innen dennoch nötig. Alle Texte dieser Studie genügten in weiten Teilen gängigen Verständlichkeitskriterien (siehe dazu Unterkapitel 6.2), sodass nicht davon auszugehen ist, dass solche Kriterien auf Grundlage der Ergebnisse dieser Studie als unnötig zu betrachten seien.

Level, auf welchem „Verständnis“ gemessen wird (Reproduktion, Anwendung, Transfer), nicht als signifikante Einflussgröße für die Effekte sprachlicher Modifikationen zeigte. Zudem gilt für diese Studie, dass die konkreten Formulierungen im Wissenstest sehr eng an den Formulierungen im Text orientiert sind, sodass anzunehmen ist, dass für eine gute Testbearbeitung tatsächlich auch erfolgreiche Leseprozesse rein auf propositionaler Ebene bedeutsam waren. Dass also nicht erst eine Vernetzung mit verwandten Wissensbereichen oder über den Text hinausreichende Schlussfolgerungen zu guten Testergebnissen führten. Dennoch ist nicht auszuschließen, dass sich die sprachlichen Veränderungen in den Texten bei den Lesenden auf einer Ebene niederschlagen, die der Wissenstest nicht abzubilden vermochte.

Gewissermaßen an diesem Punkt setzt die zweite Forschungsfrage dieser Studie an, welche einen innovativen, bisher in der Forschung kaum verfolgten Gedanken darstellt (vergleiche dazu auch Kapitel 4, insbesondere Seite 66): Es wird untersucht, inwiefern die affektive Textwahrnehmung eine solche Ebene darstellt, die beeinflussbar durch sprachliche Gestaltung ist und als mediatorische Instanz zwischen Textoberfläche und Verständnis der Lesenden liegt. Diesem Interesse wendet sich nun in aller Ausführlichkeit das folgende Kapitel zu.

## **7.2 Ergebnisse aus dem Bewertungsfragebogen**

Zur Beantwortung der Forschungsfragen 2a (Strukturierung der affektiven Textwahrnehmung in Subfacetten) und 2b (Zusammenhang zwischen sprachlicher Gestaltung und affektiver Textwahrnehmung; siehe für die Forschungsfragen Unterkapitel 4.2) dienen die Angaben der Schüler\*innen im Bewertungsfragebogen. Das Testinstrument ist in Abschnitt 6.3.3 beschrieben. Die vier folgenden Abschnitte sind der Reihenfolge nach den deskriptivstatistischen Angaben zum Fragebogen (Abschnitt 7.2.1), der inneren Struktur der affektiven Textwahrnehmung (Abschnitt 7.2.2), dem Zusammenhang zwischen sprachlicher Gestaltung und dem globalen Verständlichkeitsurteil der Schüler\*innen (mit und ohne Betrachtung der Moderatorvariablen; Abschnitt 7.2.3) und dem Zusammenhang zwischen sprachlicher Gestaltung und den einzelnen Subfacetten der affektiven Textwahrnehmung (ausschließlich ohne Moderatorvariablen; Abschnitt 7.2.4) gewidmet. Abschnitt 7.2.5 schließt die Ergebnisse zur Forschungsfrage 2 mit einer bündigen Zusammenfassung und zugehöriger summarischer Diskussion ab.

### **7.2.1 Deskriptivstatistische Ergebnisse aus dem Bewertungsfragebogen**

Der Bewertungsfragebogen in der Hauptstudie bestand aus 44 Items mit unterschiedlichen Aussagen zur Textwirkung auf die Lesenden und einem zusätzlichen Global-Item, welches den Gesamteindruck zur Verständlichkeit abfragte. Davon waren 13 Items invertiert formuliert (siehe Tabelle D.5 im Anhang D.2). Der gesamte Fragebogen ist im Anhang B.2 einsehbar.

Für die im folgenden beschriebenen Analysen wurden 41 der 44 Items verwendet (3 Items wurden aufgrund potenzieller Missverständnisse aus der Analyse ausgeschlossen; siehe dazu die folgenden Abschnitte).

Aus den Datensätzen der insgesamt  $n_{\text{Hauptstudie}} = 787$  Teilnehmenden in der Hauptstudie konnten  $n_{\text{HauptstudieBewFB}} = 781$  Fragebogen-Datensätze zur Auswertung verwendet werden. Bei den 6 aussortierten Fragebogen-Datensätzen wurden weniger als zwei Drittel aller Items beantwortet, das heißt es lagen mehr als 15 unbeantwortete Items bei insgesamt 45 Items vor und die Fragebögen wurden daher als zu wenig aussagekräftig eingestuft.

Deskriptivstatistische Kennwerte in Form der Anzahl der nicht-fehlenden Werte für jedes Item, der einzelnen Item-Mittelwerte, der Item-Korrelation mit dem globalen Verständlichkeitssitem (im direkten Vergleich zur erwarteten Korrelationsrichtung), der Itemschwierigkeit und der Itemvarianz sind in Tabelle D.5 im Anhang D.2 aufgeführt. Es lässt sich zusammenfassend feststellen, dass die Itemvarianzen, die in einem Bereich zwischen 0.88 und 1.70 liegen, nicht weiter statistisch auffällig sind: Sie lassen auf eine Breite der Verteilung schließen, bei der 68 % aller Daten in einem Bereich von 0.9 bis 1.3 Credits um den Mittelwert liegen. Die Schwierigkeiten der Items (die in diesem Fall als mittleres Bewertungsurteil im Verhältnis zur Maximalbewertung zu interpretieren sind) liegen gemessen an den Richtlinien von Kelava und Moosbrugger (2020, S. 155) (siehe auch Abschnitt 6.3.3) in einem optimalen Bereich ohne extrem schwierige oder extrem leichte Items. Diese deskriptiven Kennwerte sprechen insgesamt für die Reliabilität des Fragebogens.

Wie auch schon in Abschnitt 6.3.3 im Zusammenhang mit der Testplanung und -pilotierung erläutert wurde, ergab sich für alle Items theoriebasiert jeweils eine Erwartung hinsichtlich der Korrelationsrichtung mit dem Globalurteil. In 41 von 44 Fragebogenitems in der Hauptstudie fiel die Itemkorrelation mit dem Globalurteil in erwartungskonformer Richtung aus. Für die drei Items 11 („Ich fand die Gliederung des Texts typisch für Physiktex-te“;  $r = -0.12$ ), 25 („Der Text rief gedankliche Konflikte bei mir hervor“;  $r = -0.26$ ) und 26 („Es gab Stellen im Text, die überraschend für mich waren“;  $r = -0.02$ ) ließ die Theorie positive Korrelationen erwarten, es fanden sich aber negative Werte. Das Item zur typischen Gliederung zielte auf die Anwendbarkeit von Textformatwissen ab (siehe Seite 48 in Abschnitt 3.2.2), die Items zu den gedanklichen Konflikten und zu den überraschenden Stellen entstanden mit Bezug auf die vermeintlich verstehensförderliche Wirkung von kognitivem Konflikt und Überraschungsmomenten beim Lesen (vergleiche die Textverständlichkeitsdimensionen Groeben, 1978, siehe Seite 54). Dass diese Items jedoch bei den Lesenden eher mit *negativen Erfahrungen mit Physiktex-ten* beziehungsweise *Verwirrung* anstelle von *Verständlichkeit* assoziiert wurden, somit von den Schüler\*innen anders verstanden wurden als bei der Konstruktion des Fragebogens intendiert und dass sich daher eine negative Korrelation zwischen den Items und dem globalen Verständlichkeitseindruck ergibt, erscheint plausibel. Dies wurde zum Anlass genommen, die Items bereits *vor* dem faktoranalytischen Verfahren zur Strukturierung der affektiven Wahrnehmung (Hauptkomponentenanalyse, siehe Abschnitt 7.2.2 unten) aus dem Datensatz zu entfernen.

Eine weitere Prüfung der Reliabilität des Messinstruments für die affektive Textwahrnehmung bestand in der Bestimmung der inneren Konsistenz des Fragebogens. Die Fragebogenitems waren auf Grundlage der Beschreibungen der zwölf aus der Theorie gewonnenen Wahrnehmungsfacetten konstruiert worden (siehe Kapitel 5 für die theoretische Vorarbeit und Abschnitt 6.3.3 für die Testentwicklung). Es ist daher zu erwarten, dass diejenigen Items, die sich theoretisch einer gemeinsamen Kategorie zuordnen lassen (siehe Tabelle 6.7 auf Seite 112), einen Zusammenhang in Form der inneren Konsistenz zeigen. Ein Maß dafür ist Cronbachs  $\alpha$  (siehe Abschnitt 6.4.3). Die Ergebnisse wurden bereits im Methodenteil (Tabelle 6.6, Seite 109) dargestellt. Es sei erneut erwähnt, dass die Werte für Cronbachs  $\alpha$  nicht überinterpretiert werden dürfen, da  $\alpha$  unter anderem von der Anzahl der eingeschlossenen Items abhängt. Zusammenfassend lässt sich jedoch festhalten, dass die  $\alpha$ -Werte keinen Anlass geben, an der generellen Aussagekraft und Reliabilität der einzelnen Items zu zweifeln. Denn alle Skalen erwiesen sich in akzeptablem Ausmaß innerlich konsistent. Am schwächsten schneiden die *Darstellungsintentionen* ( $\alpha = 0.54$ ) und das *Vorwissen* ( $\alpha = 0.49$ ) ab. Die Kategorie *Darstellungsintentionen* umfasst aber lediglich zwei zugeordnete Items, daher war ein geringer Wert für  $\alpha$  erwartbar. Die Kategorie *Vorwissen* wiederum schließt das Item 11 ein, welches aufgrund seiner potenziellen Missverständlichkeit später aus der Analyse ausgenommen wurde; ohne Item 11 ergab sich  $\alpha = 0.59$  für die Kategorie *Vorwissen*.

Zusammengefasst sind die deskriptiven Kennwerte, die sich aus den Hauptstudientaten für den Bewertungsfragebogen ergaben, zufriedenstellend. Lediglich die erwartungswidrigen Korrelationen der Items 11, 25 und 26 legten eine Selektion nahe, sodass diese Items nicht in die im folgenden beschriebenen Analyseschritte einbezogen wurden.

## 7.2.2 Innere Struktur der affektiven Textwahrnehmung

Die Forschungsfrage 2a bezieht sich auf die innere Struktur der affektiven Textwahrnehmung, welche die Studienteilnehmenden in Form der Textbewertungen im Fragebogen zum Ausdruck gebracht haben. Zur Bestimmung einer solchen inneren Struktur wurde ein faktoranalytisches Verfahren angewandt. Mithilfe einer Hauptkomponentenanalyse (kurz: PCA für den englischen Ausdruck *Principal Component Analysis*; Einzelheiten zu dieser Methode werden im Methodenteil auf Seite 121 erläutert) wurden die über vierzig Fragebogenitems auf wenige Dimensionen reduziert mit dem Ziel, die vorliegenden Daten bei hinreichender Präzision einerseits und angemessener Übersichtlichkeit für anschließende Untersuchungsschritte andererseits abzubilden. Das Ergebnis dieser Dimensionsreduktion hat erstens für sich genommen inhaltliche Aussagekraft (Welche Subfacetten der affektiven Textwahrnehmung lassen sich identifizieren und welche Zusammenhänge zeigen sich innerhalb der Struktur?) und zweitens ist es nutzbar als eine auf Evidenz beruhende, handhabbare Datenstruktur für die Weiterarbeit (Welche Zusammenhänge der affektiven Textwahrnehmung zeigen sich mit der sprachlichen Gestaltung sowie mit der Behaltensleistung?).

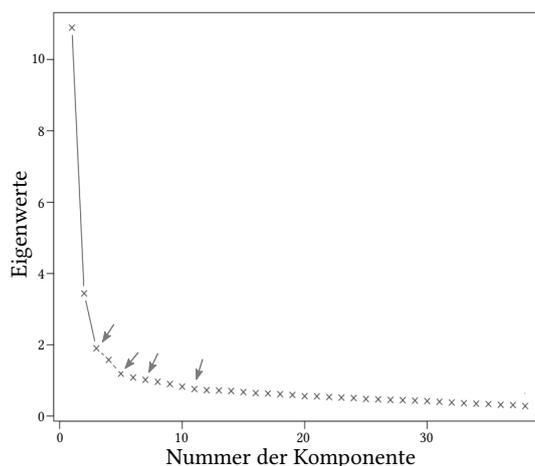


Abbildung 7.10.: Scree-Plot zur Bestimmung der Anzahl der zu extrahierenden Faktoren auf der Datengrundlage von  $n_{\text{HauptstudieBewFB}} = 781$  Fragebögen mit 41 Items. Es werden die Eigenwerte der einzelnen Komponenten absteigend nach Größe aufgetragen. Stellen, an denen ein leichter Knick im Verlauf der Plot-Linie erkennbar ist, sind mit Pfeilen markiert.

### Identifikation einzelner Facetten der affektiven Textwahrnehmung

Damit eine PCA als aussagekräftig gelten kann, müssen bestimmte Voraussetzungen an die Stichprobengröße und an die Item-Item-Korrelationen erfüllt sein. Die hier vorliegende Datengrundlage ( $n_{\text{HauptstudieBewFB}} = 781$  mit fehlenden Werten durchsetzte Schüler\*innenbewertungen zu 41 Items und dem Global-Item) erfüllt diese Voraussetzungen (Stichprobenumfang ausreichend groß und ausreichend starke Korrelationen zwischen den Items; für Einzelheiten zu den quantitativen und augenscheinlichen Analyseschritten zur Voraussetzungsüberprüfung siehe Anhang D.4).

Bereits vor der Durchführung der PCA wurden die Items 11, 25 und 26 entfernt, da sie eine erwartungswidrige Korrelation mit dem Globalurteil aufweisen (siehe oben). Hinsichtlich der *Anzahl der Hauptkomponenten*, die aus den 41 Fragebogenitems extrahiert werden könnten, führten die üblichen Kriterien (siehe auch Seite 115 im Methodenteil) auf kein eindeutiges Ergebnis: Das Kaiser-Kriterium (Eigenwerte der Komponenten größer als 1) ließe bis zu neun Komponenten zu. Eindeutige Knickstellen im Scree-Plot (Abbildung 7.10) waren nicht zu erkennen. Leichte Knickstellen zeigen sich beim dritten, fünften, siebten und elften Datenpunkt, sodass es sinnvoll erschien, entweder zwei, vier, sechs oder zehn Komponenten zu extrahieren. Es wurden zunächst alle vier Möglichkeiten verfolgt, also PCAs mit  $k$  oblimin-rotierten<sup>16</sup> Komponenten durchgeführt ( $k \in \{2, 4, 6, 10\}$ ). Es wurden dann jeweils die resultierende kumulativ erklärte Varianz der Komponenten und die Verteilung der Items auf die Komponenten betrachtet und anhand von Kommunalität und Komplexität<sup>17</sup> bewertet. Darüber hinaus wurde ermittelt, inwiefern sich die erklärte Varianz und die Itemkennwerte verbessern

<sup>16</sup>siehe Seite 121 im Methodenteil für eine Erläuterung

<sup>17</sup>siehe Seite 116 im Methodenteil für eine Erläuterung

	PC1	PC2	PC3	PC4	PC5	PC6
erklärte Varianz	0.11	0.11	0.11	0.11	0.07	0.05
kumulativ erklärte Varianz	0.11	0.22	0.33	0.44	0.51	0.56

Tabelle 7.9.: Anteil der erklärten Varianz und der kumulativ erklärten Varianz an der Gesamtvarianz im Bewertungsfragebogen für die einzelnen Hauptkomponenten. Datengrundlage:  $n_{\text{HauptstudieBewFB}} = 781$ .

lassen, indem einzelne Items aus dem Datensatz eliminiert werden. Schließlich wurde die resultierende Verteilung der Items auf die  $k$  Komponenten auf inhaltliche Interpretierbarkeit überprüft.

Als geeignetstes Ergebnis der PCA für die Bewertungsfragebogendaten wurde schließlich eine 6-dimensionale Struktur bei Elimination der Items 20, 22, 33, 37 und 39 ausgewählt (siehe Tabelle 7.10). Die Hauptkomponenten lassen sich mit den Überbegriffen *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit* (im Sinne ausführlicher Erläuterungen und sorgfältiger Erklärungen), *Angemessener Bezug zum Vorwissen* und *Involviertheit* als Facetten der affektiven Textwahrnehmung betiteln.

Der Tabelle 7.9 sind die Varianzanteile, die durch die sechs Komponenten erklärt werden, zu entnehmen. Insgesamt wird ein zufriedenstellender Anteil von 56 % der Gesamtvarianz durch die dimensionreduzierte Datenstruktur erklärt.

Facette	Zugeordnete Items
Kognitive Anregung $\alpha = .83$ $\alpha_{\text{ord}} = .85$ $\bar{r} = .41$ $\text{var} = .65$	5 Es hat mir Freude gemacht, den Text zu lesen. 6 Der Text hat mich zum Nachdenken angeregt. 7 Ich war fasziniert von den Textinhalten. 16 Der Text hat Interesse für das Thema Elektrische Spannung bei mir geweckt. 28 Mir war klar, warum es sich lohnt, den Text zu lesen. 29 Für mich war beim Lesen eines Textabschnitts stets klar, wie er sich auf die übrigen Absätze bezieht. 30 Ich konnte mein Wissen über Elektrizitätslehre um neue Informationen aus dem Text erweitern. 32 Ich konnte mir die Informationen aus dem Text merken.
Lesbarkeit $\alpha = .83$ $\alpha_{\text{ord}} = .86$ $\bar{r} = .40$ $\text{var} = .68$	2 ( <i>inv</i> ) Ich fand die Sätze im Text kompliziert. 9 ( <i>inv</i> ) Ich fand den Aufbau der einzelnen Sätze verwirrend. 10 Ich konnte dem Satzbau beim Lesen problemlos vom Satzanfang bis zum Satzende folgen. 15 ( <i>inv</i> ) Ich wusste am Satzende nicht mehr, wie der Satz begonnen hatte. 21 ( <i>inv</i> ) Ich musste mit den Augen zwischen einzelnen Textstellen vor und zurück hüpfen anstatt den Text flüssig zu lesen. 23 ( <i>inv</i> ) Ich musste meine volle Aufmerksamkeit auf den Text richten, um ihm Information zu entnehmen. 31 ( <i>inv</i> ) Beim Lesen brachten mich einzelne Textstellen zum Stocken. 41 ( <i>inv</i> ) Ich musste den Text in einem langsamen Tempo lesen.
Logischer Aufbau $\alpha = .86$ $\alpha_{\text{ord}} = .89$ $\bar{r} = .43$ $\text{var} = .52$	4 Ich fand die Reihenfolge der einzelnen Textabschnitte sinnvoll. 13 Ich konnte gut nachvollziehen, wie die Textinhalte aufeinander aufbauen. 17 Ich fand die Einteilung des Texts in seine Abschnitte passend. 24 Mir erschienen aufeinanderfolgende Sätze inhaltlich logisch miteinander verknüpft. 29 Für mich war beim Lesen eines Textabschnitts stets klar, wie er sich auf die übrigen Absätze bezieht. 35 Ich konnte im Text durchgehend einen roten Faden erkennen. 36 Die einzelnen Abschnitte setzen sich für mich logisch zu einem Gesamtbild der elektrischen Spannung zusammen. 44 Mir war klar, wie die Wörter im Satz miteinander in Beziehung stehen.

Elaboriertheit	12 ( <i>inv</i> ) Ich fand die Informationen im Text wirr.
$\alpha = .82$	18 ( <i>inv</i> ) Der Text erschien mir sehr voll an Informationen.
$\alpha_{\text{ord}} = .84$	19 ( <i>inv</i> ) Ich hätte bei schwierigen Aspekten noch eine Erklärung aus einer anderen Sicht gebraucht.
$\bar{r} = .42$	27 ( <i>inv</i> ) Um den Text zu verstehen, hätte ich mehr Erklärung zu den einzelnen Aspekten der elektrischen Spannung gebraucht.
$\text{var} = .72$	34 ( <i>inv</i> ) Schwierige Begriffe im Text haben mir Verständnisprobleme bereitet.
	38 ( <i>inv</i> ) Ich hätte mehr Erklärung zu den Fachbegriffen im Text gebraucht.
Angemessener Bezug zum Vorwissen	14 Der Text hat mich an mein Vorwissen zum Thema Elektrizitätslehre erinnert.
$\alpha = .72$	42 Die Informationen im Text knüpften an mein bisheriges Wissen an.
$\alpha_{\text{ord}} = .77$	43 Mein Vorwissen über Elektrizitätslehre reichte aus, um den Text zu verstehen.
$\bar{r} = .47$	
$\text{var} = .77$	
Involviertheit	1 Ich fand den Text humorvoll gestaltet.
$\alpha = .71$	3 Der Text hat mich emotional angesprochen.
$\alpha_{\text{ord}} = .77$	8 Ich fand den Text unterhaltsam.
$\bar{r} = .38$	40 Ich habe mich beim Lesen persönlich angesprochen gefühlt.
$\text{var} = .70$	
Eliminierte Items	11 Ich fand die Gliederung des Texts typisch für Physiktex-te.
	20 Für mich war beim Lesen eines Satzes stets klar, wie er sich auf den vorherigen Satz bezieht.
	22 Ich fand den Text auf den ersten Blick übersichtlich.
	25 Der Text rief gedankliche Konflikte bei mir hervor.
	26 Es gab Stellen im Text, die überraschend für mich waren.
	33 Ich habe mein Wissen, das nichts mit Elektrizitätslehre zu tun hat, eingesetzt um den Text zu verstehen.
	37 Mir war in jedem Textabschnitt klar, was das Thema dieses Abschnitts ist.
	39 Mir war klar, inwiefern der Text mir Erfahrungen aus meiner Lebenswelt erklärt.

Tabelle 7.10.: Verteilung der Fragebogenitems auf die extrahierten Hauptkomponenten, die sich als Subfacetten der affektiven Textwahrnehmung interpretieren lassen. Es sind auch die Cronbachs  $\alpha$  Werte ( $\alpha$  regulär und  $\alpha_{\text{ord}}$  ordinal) sowie die mittleren Item-Item-Korrelationskoeffizienten ( $\bar{r}$ ) als Maß für die innere Konsistenz der Facettenzusammensetzung dargestellt. Auch die Varianzen ( $\text{var}$ ) in den einzelnen Facetten sind angegeben. Datengrundlage:  $n_{\text{HauptstudieBewFB}} = 781$ .

Tabelle 7.10 enthält neben der Zuordnung der Items zu den Komponenten auch die Benennung der Komponenten als unterschiedliche Subfacetten der affektiven Textwahrnehmung: PC1 wurde als *Kognitive Anregung* bezeichnet, PC2 als *Lesbarkeit*, PC3 als *Logischer Aufbau*, PC4 als *Elaboriertheit* (aufgefasst als Ausführlichkeit und Sorgfalt in den Erläuterungen und

Erklärungen im Text), PC5 als *Angemessener Bezug zum Vorwissen* und PC6 als *Involviertheit*. Mit den Benennungen ist jeweils gemeint, inwiefern Lesende einen Text hinsichtlich des entsprechenden Aspekts subjektiv beurteilen; ob sie ihn also kognitiv anregend, als einfach lesbar etc. *empfinden*. Die Itemzuordnung erfolgte gemäß ihrer höchsten Faktorladung<sup>18</sup> (siehe Anhang D.7 für die Faktorladungsmatrix), die Benennung der Facetten erfolgte auf Grundlage der Inhalte der zugeordneten Items. Es sind darüber hinaus auch Cronbachs  $\alpha$ , das ordinale Cronbachs  $\alpha$ , die mittlere Pearson-Item-Item-Korrelation und die Varianz für die einzelnen Komponenten angegeben (alle Kennwerte sind im Methoden-Abschnitt 6.4.3 näher erläutert). Die Kennwerte sprechen insgesamt für eine sehr gute innere Konsistenz der extrahierten Faktoren ( $\alpha$ -Werte ab 0.8 gelten gemeinhin als gut, siehe Field, Miles und Field, 2012, S. 799; und für den Pearson-Korrelationskoeffizienten  $r$  spricht man ab 0.3 von mittleren und ab 0.5 von starken Korrelationen, siehe Cohen, 1988<sup>19</sup>).

Einzeln betrachtet bilden die extrahierten Faktoren also eine statistisch sowie inhaltlich schlüssige Strukturierung der affektiven Textwahrnehmung in sechs Subfacetten. Im Folgenden werden die Facetten in ihrer Gesamtheit betrachtet, das heißt es werden ihre Zusammenhänge miteinander analysiert.

### Zusammenhänge zwischen den Subfacetten

Die Zusammenhänge zwischen den Subskalen wurden anhand von Pearson-Korrelationen (siehe Abschnitt 6.4.3) bemessen, das Ergebnis ist in Tabelle 7.11 gezeigt. Es wurde dazu für jeden der 781 Datensätze aus den Bewertungsfragebögen zunächst berechnet, wie die jeweils gelesene Textversion hinsichtlich der sechs Facetten bewertet wurde. Dazu wurde das arithmetische Mittel der den einzelnen Facetten zugeordneten Items gebildet<sup>20</sup>. Aus diesen Facetten-Bewertungen wurde die 6×6-Korrelationsmatrix berechnet.

---

<sup>18</sup>Lediglich Item 8 lädt mit .44 auf die *Involviertheit* und nur marginal stärker mit .45 auf *kognitive Anregung* und wurde aus inhaltlichen Gründen dennoch der Facette *Involviertheit* zugeordnet. Die inneren Konsistenzen und mittleren Korrelationen in den beiden Komponenten *Involviertheit* und *kognitive Anregung* wurden probenhalber verglichen, je nachdem, ob Item 8 der einen oder der anderen Komponente zugeordnet wurde. Es zeigt sich dabei, dass die Cronbachs- $\alpha$ -Werte sowie die mittleren Item-Item-Korrelationen innerhalb der Komponenten beinahe identisch sind, wobei eine Hinzunahme von Item 8 zur *Involviertheit* einen positiven Effekt auf die mittlere Item-Item-Korrelation dieser Komponente hat. Item 8 zur *Involviertheit* zu zählen war somit inhaltlich sowie statistisch plausibel.

<sup>19</sup>Wie bereits im Methodenteil in Zusammenhang mit der Effektgröße  $\omega^2$  berichtet (Seite Seite 120 im Methoden-Abschnitt 6.4.3), sind solche Richtwerte vorsichtig zu interpretieren. Dennoch werden sie als Ergänzung zu den rohen Zahlenwerten als grober Orientierungsrahmen herangezogen.

<sup>20</sup>Die Itembewertungen fanden auf einer Skala von 1 bis 5 statt; siehe Abschnitt 6.3.3.

Betrachte also z. B. die *Involviertheit*: Angenommen, Person xy bewertet Item 01 mit 1, Item 02 mit 3, Item 08 mit 3 und Item 40 mit 4, dann wurde die Bewertung von Person xy für die *Involviertheit* zu  $\frac{1+3+3+4}{4} = 2.25$  berechnet.

Die Bewertungen invertiert formulierter Items gingen auch invertiert in die Mittelwertbildung ein: z. B. bedeutet eine Bewertung des Items „Ich fand die Informationen im Text wirr“ mit 4 („Stimme tendenziell zu“), dass für die Mittelwertbildung in der Facette *Elaboriertheit* der Wert 2 verwendet wird.

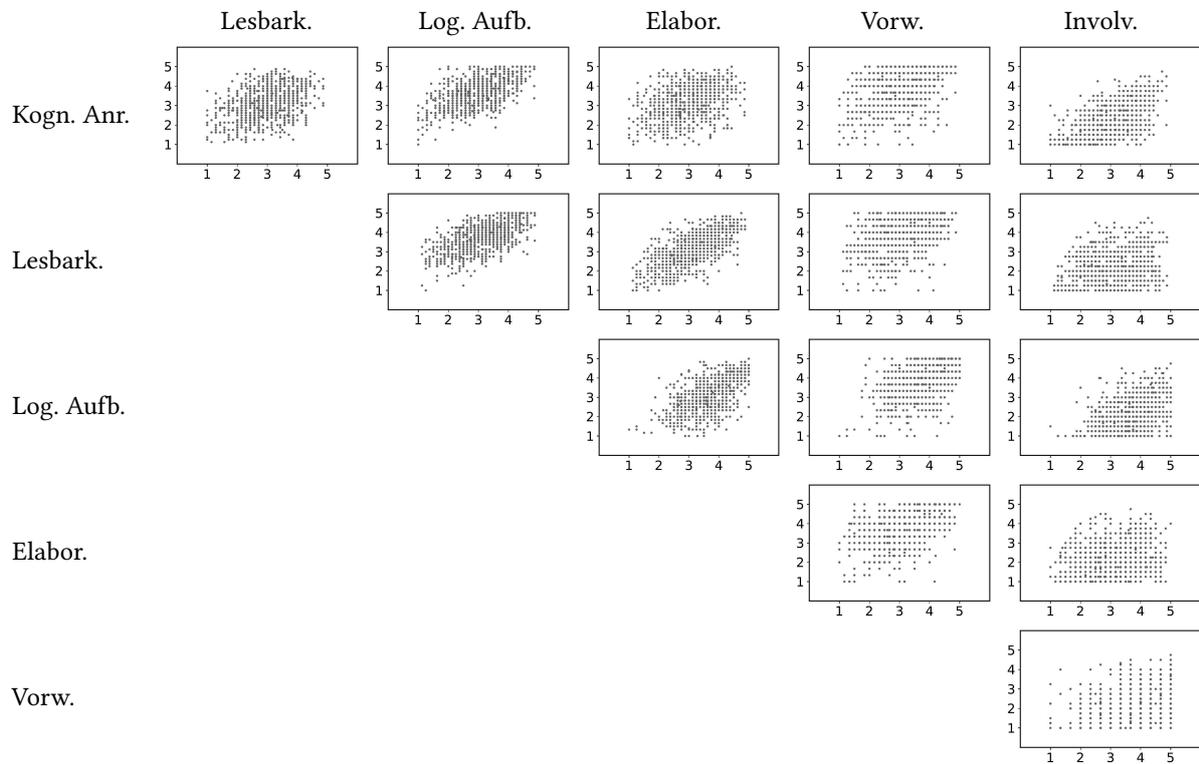


Abbildung 7.11.: Scatter-Plots für den Zusammenhang zwischen den Bewertungen zu jeweils zwei Facetten der affektiven Textwahrnehmung. Datengrundlage:  $n_{\text{HauptstudieBewFB}} = 781$ .

	Kogn. Anr.	Lesbark.	Log. Aufb.	Elabor.	Vorw.	Involv.
Kogn. Anr.	1.00	0.39	0.57	0.35	0.40	0.59
Lesbark.		1.00	0.60	0.69	0.41	0.23
Log. Aufb.			1.00	0.58	0.49	0.38
Elabor.				1.00	0.48	0.20
Vorw.					1.00	0.22
Involv.						1.00

Tabelle 7.11.: Korrelationsmatrix mit Koeffizienten nach Pearson (siehe Abschnitt 6.4.3) für die Zusammenhänge zwischen den Subfacetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* der affektiven Textwahrnehmung. Alle Korrelationen sind höchst signifikant. Datengrundlage:  $n_{\text{HauptstudieBewFB}} = 781$ .

Für die Zusammenhänge zwischen den einzelnen Subfacetten der affektiven Textwahrnehmung (siehe Tabelle 7.11) gilt, dass alle Korrelationen höchst signifikant ( $p < 0.001$ ) und positiv ( $r > 0$ ) sind.

Besonders starke, höchst signifikante Korrelationen ( $r > 0.5$ ; für die Richtwerte, die hier als Orientierung für die größenmäßige Einordnung dienen, siehe Cohen, 1988) bestehen dabei zwischen den folgenden Facetten:

- *Kognitive Anregung* und *Logischer Aufbau*.
- *Kognitive Anregung* und *Involviertheit*.
- *Lesbarkeit* und *Logischer Aufbau*.
- *Lesbarkeit* und *Elaboriertheit*.
- *Logischer Aufbau* und *Elaboriertheit*.

Zwischen allen übrigen Facetten bestehen schwache bis mittlere, höchst signifikante Korrelationen ( $0 < r < 0.5$ ).

In Abbildung 7.11 sind die eben beschriebenen quantitativen Korrelationswerte anhand von Scatter-Plots visualisiert. Je mehr sich die Punktwolken entlang einer Geraden von links unten nach rechts oben verteilen, desto stärker ist der Zusammenhang zwischen den beiden betrachteten Facetten. Die Abbildung lässt einen solchen Trend vor allem in denjenigen Plots, die zu einem Variablenpaar mit verhältnismäßig großem Pearson- $r$  gehören, erkennen (vergleiche Tabelle 7.11).

### **Zusammenhang zwischen globalem Verständlichkeitseindruck und den Facetten der affektiven Textwahrnehmung**

Analog zu den Korrelationsanalysen für die Facetten untereinander sind in Tabelle 7.12 die Pearson-Korrelationskoeffizienten (siehe Abschnitt 6.4.3) der einzelnen Subfacetten der Textwahrnehmung mit dem globalen Verständlichkeitseindruck dargestellt. Die Bewertungen einzelner Studienteilnehmender für die verschiedenen Facetten ergeben sich durch Mittelwertbildung (für eine genauere Beschreibung siehe Seite 164). Der globale Verständlichkeitseindruck wurde anhand des Items „Ich fand den Text insgesamt verständlich.“ im Bewertungsfragebogen erhoben. Alle Facetten korrelieren höchst signifikant und positiv mit dem Global-Item zur Verständlichkeit. Während die *Involviertheit* nur einen schwachen und der *Angemessene Bezug zum Vorwissen* einen mittleren Zusammenhang aufweisen, sind alle weiteren Zusammenhänge ( $r > 0.5$ ; für die Richtwerte zur vorsichtigen, größenmäßigen Einordnung siehe Cohen, 1988) als stark zu betrachten. Abbildung 7.12 ergänzt die mathematischen Resultate um einen visuellen Eindruck anhand von Scatter-Plots: Die positiven Zusammenhänge sind daran zu erkennen, dass die Punkteverteilung innerhalb der Plots einen Trend von links unten nach rechts oben aufweisen.

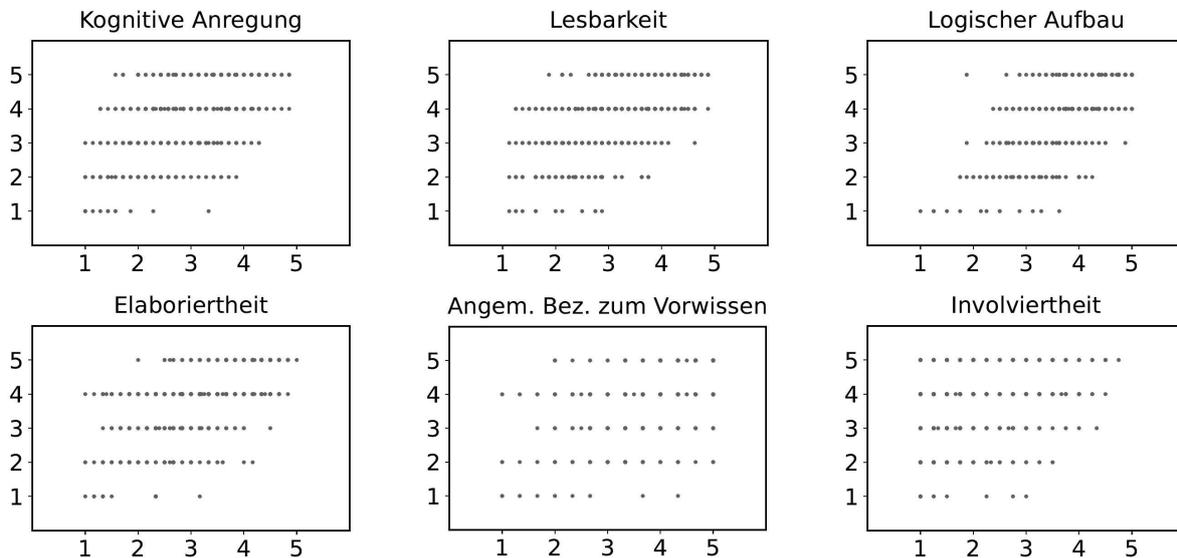


Abbildung 7.12.: Scatter-Plots für den Zusammenhang der Bewertung hinsichtlich einzelner Facetten der affektiven Textwahrnehmung (Rechtswertachse) mit der Bewertung im Item zum globalen Verständlichkeitseindruck (Hochwertachse). Datengrundlage:  $n = 755$  Datensätze mit nicht-fehlender Global-Bewertung.

	Kogn. Anr.	Lesbark.	Log. Aufb.	Elabor.	Vorw.	Involv.
glob.	0.51	0.56	0.61	0.59	0.46	0.29

Tabelle 7.12.: Pearson-Korrelationen (siehe Abschnitt 6.4.3) zwischen den einzelnen Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* der affektiven Textwahrnehmung mit dem globalen Eindruck (*glob.*) hinsichtlich der Verständlichkeit des Texts. Alle Korrelationen sind höchst signifikant. Datengrundlage:  $n = 755$  Datensätze mit nicht-fehlender Global-Bewertung.

### **Topikale Diskussion: Innere Struktur der affektiven Textwahrnehmung**

In diesem Kapitel wurde herausgearbeitet, inwiefern die affektive Textwahrnehmung auf Grundlage der empirischen Daten dieser Studie als ein aus sechs Facetten bestehendes Konstrukt aufgefasst werden kann. Die einzelnen Itembewertungen aus dem Fragebogen zur Textwahrnehmung werden in den sechs Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* bei geringem Informationsverlust zusammengefasst und zutreffend wiedergegeben. Die Facetten sind somit als Aspekte zu betrachten, hinsichtlich derer sich die Wirkung der Texte auf die Schüler\*innen beschreiben lässt: Inwiefern wirkt ein Text kognitiv anregend, einfach zu lesen, logisch im Aufbau, elaboriert (im Sinne der Ausführlichkeit der Erläuterungen), inwiefern erscheinen die Bezüge auf das Vorwissen angemessen und wie involvierend wirkt der Text. Es sei betont, dass diese ausgewählte Struktur für die affektive Textwahrnehmung (siehe Tabelle 7.10) ein Ergebnis von mehreren möglichen darstellt, welches die vorliegenden Textbewertungsdaten statistisch und inhaltlich angemessen repräsentiert und aussagekräftige weitere Analysen ermöglicht. Es wurden neben der Analyse, die zum hier berichteten Ergebnis führte, auch alternative PCA-Ansätze mit leicht unterschiedlichen Analysestrategien verfolgt (etwa: andere Anzahl an zu extrahierenden Hauptkomponenten, kein a-priori-Ausschluss der Items 11, 25 und 26, unterschiedliche Gewichtung der Item-Ausschluss-Kriterien *Kommunalität* und *Komplexität*). Ein Vergleich der Endergebnisse ergab, dass sich leicht veränderte, aber ähnliche Strukturen zeigen: Es ergäben sich leicht unterschiedliche Sets an zu eliminierenden Items, die extrahierten Hauptkomponenten könnten mit ähnlichen, gegebenenfalls verfeinerten oder vergrößerten Überbegriffen (je nach Anzahl der Hauptkomponenten) betitelt werden und die Hauptkomponenten würden eine geringfügig veränderte Itemzusammensetzung als in der hier ausgewählten Version aufweisen. Bei einer geringeren Anzahl extrahierter Hauptkomponenten würde jedoch weniger Gesamtvarianz erklärt werden und echte inhaltliche Information verloren gehen. Beispielshalber würden die beiden trennbaren Facetten „Involviertheit“ und „kognitive Anregung“ in eine allgemeinere Hauptkomponente, die mit „Aktivierung der Lesenden“ betitelt werden könnte, übergehen. Bei einer größeren Anzahl an extrahierten Hauptkomponenten würden solche Hauptkomponenten entstehen, die nur ein einzelnes Item umfassen und die inhaltliche Aussagekraft der Struktur würde sinken, da die Interpretierbarkeit und Unterscheidbarkeit der einzelnen Hauptkomponenten schwieriger und unklarer werden würde.

Im Hinblick auf die Möglichkeit, alternative Strukturen für die affektive Textwahrnehmung zu finden, ist auch folgende Einschränkung wichtig: Eine PCA ermittelt Hauptkomponenten, indem solche Items gruppiert werden, die viel gemeinsame Kovarianz aufweisen. Dabei gilt, dass einerseits *Personen* Varianz erzeugen, da verschiedene Personen einen Fragebogen zum gleichen Text unterschiedlich ausfüllen werden. Andererseits erzeugen auch *Texte* Varianz, da eine Person einen Fragebogen zu verschiedenen Texten unterschiedlich ausfüllen wird. In die PCA und alle weiteren Analysen geht damit die Unterschiedlichkeit der Personen in der Stichprobe ein und andererseits die Unterschiedlichkeit der Texte. Hinsichtlich der Stichprobe

dieser Studie ist davon auszugehen, dass sie repräsentativ für die Gesamtpopulation aller Mittelstufenschüler\*innen an bayerischen Realschulen und Gymnasien ist. Hinsichtlich der Texte jedoch ist zu beachten, dass die Texte hinsichtlich vieler Merkmale sehr ähnlich sind (z. B. Lerninhalte, grundlegend-inhaltlicher Aufbau, Layout). Daraus folgt, dass insbesondere die Ergebnisse zur Struktur der Textwahrnehmung, also die gefundenen Facetten und deren Zusammenhänge untereinander (siehe die später folgenden Analyseschritte) nicht ohne Weiteres auf allgemeines Textverstehen übertragbar sind. Für Texte, die mehr Varianz in der Wahrnehmung erzeugen, ist es durchaus denkbar, dass zusätzliche Varianz und damit Kovarianz in den Items entsteht und damit die Struktur von der hier gefundenen abweicht. Auch die gefundenen Korrelationen sind lediglich im Gültigkeitsrahmen der hier vorliegenden Studie mit gezielt variiertes Textgestaltung zu interpretieren. Die berichteten Zusammenhänge sind aus mathematischen Gründen stark davon abhängig, wie viel Varianz die verwendeten Texte in den einzelnen Facetten erzeugen. Es wird deutlich, dass großer Bedarf an Vergleichs- und Anschlussstudien besteht, welche die affektive Wahrnehmung naturwissenschaftlicher Texte in systematischer und differenzierter Art und Weise weiteruntersuchen.

Gleichwohl sind Interpretationen für die in dieser Studie gefundene Struktur der affektiven Textwahrnehmung interessant: Zunächst gilt, dass sich die theoretisch hergeleiteten Kategorien (siehe Seite 112) zu großen Teilen in den sechs Facetten wiederfinden lassen:

- Die meisten Items der theoretischen Kategorien *Lokale Kohärenz* und *Globale Kohärenz* finden sich gemeinsam mit den Items der *Kognitiven Gliederung* gebündelt in der Facette *Logischer Aufbau* wieder. Es zeigte sich hier also keine getrennte Wahrnehmung der Schüler\*innen der lokalen und der globalen Textkohärenz.
- Die *Informationsdichte* aus der Theorie geht direkt in die Facette *Elaboriertheit* über.
- Die empirische *Lesbarkeits*-Facette ist in ihrer Itemzusammensetzung zwar identisch mit der theoretischen Kategorie *Strukturelle Komplexität*. Da jedoch der Name *Lesbarkeit* greifbarer und intuitiv verständlicher erscheint, wird er im Folgenden für diese evidenzbasiert herausgearbeitete Facette verwendet. Im Vergleich mit weiteren Theorie-Facetten gilt, dass die empirische *Lesbarkeits*-Facette eine Erweiterung der Theorie-Kategorie *Lesbarkeit* um einige, auf Sätze bezogene Aspekte der theoretischen *Sprachlichen Einfachheit* darstellt. Die übrigen Aspekte der *Sprachlichen Einfachheit* aus der Theorie, die auf Wörter bezogen sind, sind der Empirie-Facette *Elaboriertheit* zuzuordnen.
- Die Theorie-Kategorie *Vorwissen* geht abzüglich eliminiertes Items nahezu direkt in die Facette *Angemessener Bezug zum Vorwissen* ein.
- Die in der Literatur genannten, weniger Items umfassenden Kategorien *Darstellungsin-tentionen* und *Einprägsamkeit* ordnen sich in die Facette *Kognitive Anregung* ein. Jene umfasst noch einige (kognitive) Aspekte der Theorie-Kategorien *Unterhaltsamkeit* und *Motivation*.
- Kognitive und emotionale Aspekte der Aktivierung erscheinen in den Bewertungsdaten getrennt. Denn die emotionsbezogenen Items aus den beiden Kategorien *Unterhaltsamkeit* und *Motivation* finden sich in einer eigenen Facette *Involviertheit* wieder.

Darüber hinaus sei mit Blick auf Tabelle 7.11 und die Auflistung der starken Korrelationen auf Seite 166 auf einige, spezielle Auffälligkeiten hingewiesen. Es fällt zunächst auf, dass die Facette *Logischer Aufbau* stark eingebunden erscheint, da sie mit drei von fünf anderen Facetten starke Korrelationen aufweist. Dies deutet darauf hin, dass (zumindest für die Texte dieser Studie) der logische Aufbau eine besonders zentrale Facette der affektiven Textwahrnehmung ist. Demgegenüber erweist sich der *Angemessene Bezug zum Vorwissen* auf der hier vorliegenden Datengrundlage als eher eigenständig und weniger verbunden mit anderen Facetten. Ebenso zeigt die *Involviertheit* keine starke Korrelationen mit anderen Facetten mit Ausnahme der *Kognitiven Anregung*. Eine mögliche Interpretation ist, dass sich im Zusammenhang zwischen *Involviertheit* und *kognitiver Anregung* zeigt, dass beide Facetten zu einer gemeinsamen *Aktivierung der Lesenden* beitragen und dabei die *Involviertheit* die emotionsbezogene und die *Kognitive Anregung* die kognitionsbezogene Seite ausmachen. Zuletzt seien noch die jeweils starken Zusammenhänge zwischen *Lesbarkeit*, *Logischem Aufbau* und *Elaboriertheit* herausgestellt. Zusammengenommen spiegelt sich in dieser Gruppe gewissermaßen die Textkohärenz wider, ohne dass in den Bewertungen der Lesenden eine Unterscheidung von lokaler und globaler Kohärenz zutage getreten wäre. Denn jede dieser drei Facetten birgt gewisse Aspekte der Kohärenz (die Kohärenz-Items aus Tabelle 6.7 wurden allesamt einer dieser drei Facetten zugeordnet, siehe Tabelle 7.10). Insgesamt lassen sich also die gefundenen Zusammenhänge zwischen den Facetten der affektiven Textwahrnehmung, die aus den Textbewertungsdaten dieser Studie gewonnen wurden, schlüssig interpretieren.

Abgesehen vom Fokus auf einzelne Facetten gilt ganz allgemein, dass alle Korrelationen aus Tabelle 7.11 positiv und dabei höchst signifikant sind. Dies kann als Indiz dafür gewertet werden, dass den Facetten ein gemeinsames Konstrukt in Form der affektiv wahrgenommenen Verständlichkeit des Texts zugrundeliegt. Dieser Gedanke wird weiter dadurch plausibilisiert, dass tendenziell starke und in jedem Fall positive und höchst signifikante Korrelationen zwischen den Facetten und dem globalen Verständlichkeitseindruck gefunden wurden. Der Globaleindruck zur Verständlichkeit scheint sich also einerseits zu einem gewissen Teil in jeder Subfacette auszudrücken und andererseits scheinen außerhalb der hier betrachteten Facetten der affektiven Textwahrnehmung keine weiteren Aspekte zu liegen, die unbeachtete, ausschlaggebende Einflüsse auf den Globaleindruck bergen. Dabei erwiesen sich auf Grundlage der hier ermittelten Korrelationen die *Elaboriertheit*, der *Logische Aufbau* und die *Lesbarkeit* – also die Kohärenz-Gruppe – gefolgt von der *Kognitiven Anregung* als am ausschlaggebendsten dafür, dass ein Text insgesamt verständlich wirkte. Die Korrelationen sind allerdings vor dem Hintergrund der in den Facetten erzeugten Varianz zu interpretieren. Der Größe nach geordnet gilt (vergleiche Tabelle 7.10 auf Seite 163): Die Varianz in den Itembewertungen für den *Logischen Aufbau* beträgt 0.52, für die *Kognitive Anregung* 0.65, für die *Lesbarkeit* 0.68, für die *Involviertheit* 0.70, für die *Elaboriertheit* 0.72 und für das *Vorwissen* 0.77. Da die Texte also beispielshalber weniger Varianz im *Logischen Aufbau* erzeugen als in den übrigen Facetten, erhält der *Logische Aufbau* durch die Beschaffenheit der Texte gewissermaßen weniger Gelegenheit, den Gesamteindruck zu beeinflussen. Im Allgemeinen, das

heißt für verschiedenartigere Texte, könnte sich die Stärke der Zusammenhänge verschieben. Womöglich wäre also beispielshalber der *Logische Aufbau* noch ausschlaggebender für den Globaleindruck, würde die Facette mehr Varianz aufweisen.

Einen interessanten Vergleich bietet eine Arbeit zur Qualität *mündlicher Erklärungen* von Heinze (2021). Als in jener Studie Schüler\*innen ihre Bewertung einer mündlichen Erklärung begründen sollten, wurden dabei in erster Linie Argumente, die mit dem strukturellen Aufbau, der Elaboriertheit, der Fachsprache (vor allem in Bezug auf die Häufung von Fachtermini und die Informationsdichte) und/oder dem vorausgesetzten Vorwissen zu tun haben, genannt (siehe Kapitel 6.2.2 in der Arbeit von Heinze). Zwar sind die Wahrnehmungskategorien bei Heinze (2021) nicht vollkommen deckungsgleich gefasst wie die Wahrnehmungsfacetten in dieser Arbeit, dennoch weisen beide Ergebnisse darauf hin, dass die Elaboriertheit (Ausführlichkeit versus hohe Informationsdichte) und eine nachvollziehbare, logische Gliederung besonders wichtig dafür sind, dass Schüler\*innen einen Text verständlich finden.

Darüber hinaus ist für eine Diskussion der Ergebnisse erwähnenswert, dass die Bedeutung des Global-Items für unterschiedliche Personen unterschiedlich gefärbt sein kann. Es ist denkbar, dass das Item „Ich fand den Text insgesamt verständlich“ je nach Person eine bestimmte Facette der affektiven Textwahrnehmung mehr oder weniger triggert und somit die Bewertung des Global-Items dadurch geprägt ist, wie die Person das Item interpretiert. Denkt eine Person bei dem Item beispielshalber mehr an gut erklärte Wörter als an den Zustand geistiger Aktivierung, so wird die Bewertung des Global-Items von dieser Person stark durch elaborationsbezogene Aspekte im Text beeinflusst sein. Dennoch ist anzunehmen, dass das Item als valide für den globalen (je nach persönlichen Neigungen individuell gewichteten) Eindruck hinsichtlich der Verständlichkeit des Text betrachtet werden kann.

Ein weiterer wichtiger Aspekt bei der Bewertung der Ergebnisse ist, dass die fünfstufigen Bewertungsskalen zu Decken- und Bodeneffekten führen können. Dadurch werden die Werte der berichteten Korrelationskoeffizienten verzerrt. Ein Blick zum Beispiel auf die Scatter-Plots (Abbildung 7.12 auf Seite 167) legt nahe, dass sich vor allem am oberen Rand der Skala Datenpunkte sammeln und womöglich eine erhöhte Stufenanzahl hier eine höhere Auflösung hätte bieten können. Die Verzerrungen mögen den Betrag der Korrelationskoeffizienten zwar beeinträchtigen, jedoch ist vermutlich der relative Vergleich einzelner Koeffizienten und vor allem die Richtung des Zusammenhangs robust gegen die Verzerrungen. Damit ist der Trend, den die Korrelationskoeffizienten anzeigen als zutreffend anzusehen.

Für diese Studie wurde von weiteren Analysen für den Zusammenhang zwischen Globaleindruck und Wahrnehmungsfacetten – z. B. in Form von ANOVAs oder multiplen, linearen Regressionsanalysen – abgesehen. In einer mehrfaktoriellen ANOVA, die die Effekte der Wahrnehmungsfacetten auf den Globaleindruck untersucht, wäre die Voraussetzung, dass die betrachteten Faktoren (also: die Facetten der Wahrnehmung) unabhängig sind, erheblich verletzt (siehe die Zusammenhangsanalyse der Wahrnehmungsfacetten rund um die Tabelle 7.11 auf Seite 165). Weiterführende Berechnungen wie eine multiple, lineare Regression beruhen

auf starken Annahmen wie der Linearität der Zusammenhänge zwischen Prädiktoren und abhängiger Variable, setzen ebenfalls Unabhängigkeit der Prädiktoren voraus (Problem der *Multikollinearität*) und benötigen zumindest annähernd kontinuierlich gestufte Prädiktorvariablen. Es ist davon auszugehen, dass derartige Analysen die hier vorliegende Datenstruktur überbeanspruchen würden. Sie sollten daher Gegenstand von Anschlussstudien sein, welche sich diese Arbeit als Startpunkt zunutze machen und mit intensiver und aufwendiger (weiter-)entwickelten Testinstrumenten daran anknüpfen können.

### 7.2.3 Zusammenhang zwischen sprachlicher Gestaltung und Globalurteil

Um Forschungsfrage 2b nach den Zusammenhängen zwischen sprachlicher Textgestaltung und affektiver Textwahrnehmung zu beantworten, wurden analog zu den Analysen für die gezeigte Behaltensleistung (Abschnitte 7.1.3 und 7.1.4) ANOVA-Modelle berechnet. Die unabhängigen Variablen bestehen in den sprachlichen Gestaltungsvariablen W, S und P, beziehungsweise in der Textversion<sup>21</sup>, die abhängige Variable ist die Bewertung einer Wahrnehmungsfacette beziehungsweise des Global-Items (Skala von 1 bis 5). Für die Voraussetzungsüberprüfungen (Normalverteilung und Varianzhomogenität der abhängigen Variablen) sei auf den Anhang D.4 verwiesen. Es zeigt sich, dass die Datengrundlage weitestgehend robust für ANOVAs ist, vor allem aufgrund der großen Stichprobe. Dennoch sind die vorausgesetzten Annahmen insbesondere hinsichtlich des Globaleindrucks möglicherweise verletzt und die Aussagekraft der Ergebnisse sollte nicht überbewertet werden.

Tabelle 7.13 enthält unter anderem die Bewertungsmittelwerte für das Global-Item je nach sprachlicher Textgestaltung. Die Mittelwerte liegen im Bereich von 3.5 bis 4.0. Die Texte wurden also alle insgesamt als tendenziell verständlich wahrgenommen. Darüber hinaus sind die Ergebnisse der beiden ANOVAs mit den unabhängigen Variablen W, S und P beziehungsweise mit der unabhängigen Variablen Textversion enthalten. Diese sagen aus:

---

<sup>21</sup>Die sprachliche Gestaltung der Textversionen A bis H ist ausführlich im Abschnitt 6.2.2 beschrieben. Für einen schnellen Überblick sei hier erneut kurz festgehalten:

---

A: wsp	B: wsP	C: wSp	D: wSP
E: Wsp	F: WsP	G: WSp	H: WSP

---

unabhängige Variable	F (7,754)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
W	7.73	0.017 0.006	0.01*	3.71 (W) 3.90 (w) +0.19 ( $\Delta$ )
S	5.97	0.030 0.015	0.03*	3.71 (S) 3.89 (s) +0.18 ( $\Delta$ )
P	2.73	0.099 0.099	0.00	3.85 (P) 3.74 (p) -0.11 ( $\Delta$ )
W:S	0.57	0.451	0.00	
W:P	0.23	0.633	0.00	
S:P	0.19	0.666	0.00	
W:S:P	0.43	0.512	0.00	
Textversion	2.64	0.011	0.02*	A: 3.92, B: 3.99, C: 3.79, D: 3.89 E: 3.71, F: 3.93, G: 3.56, H: 3.64

Tabelle 7.13.: ANOVA-Ergebnisse für die Abhängigkeit des globalen Verständlichkeitseindrucks von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen *Wortwahl*, *Satzstrukturen* und *Personalisierung* und deren Interaktionen („:“) sowie von der Textversion (Datengrundlage:  $n = 755$  nicht-fehlende Werte für das Global-Item). Es werden die Teststatistik F für die Modellparameter (7,754) und der p-Wert in unkorrigierter Form sowie in  $\alpha$ -korrigierter Form berichtet. Diejenigen Effekte, die nach einer Korrektur gemäß Bonferroni-Holm (siehe Abschnitt 6.4.3) signifikant sind, sind in der Effektstärke-Spalte ( $\omega^2$ ) mit \*\*\* für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$  markiert. Ergänzend sind in der letzten Spalte die Bewertungsmittelwerte für W-Texte gegenüber w-Texten, für S-Texte gegenüber s-Texten und für P-Texte gegenüber p-Texten sowie für die einzelnen Textversionen dargestellt. Mit  $\Delta$  ist der Unterschied „anspruchsgeminderte / schwach ausgeprägte Form“ minus „anspruchsgesteigerte / ausgeprägte Form“ bezeichnet.

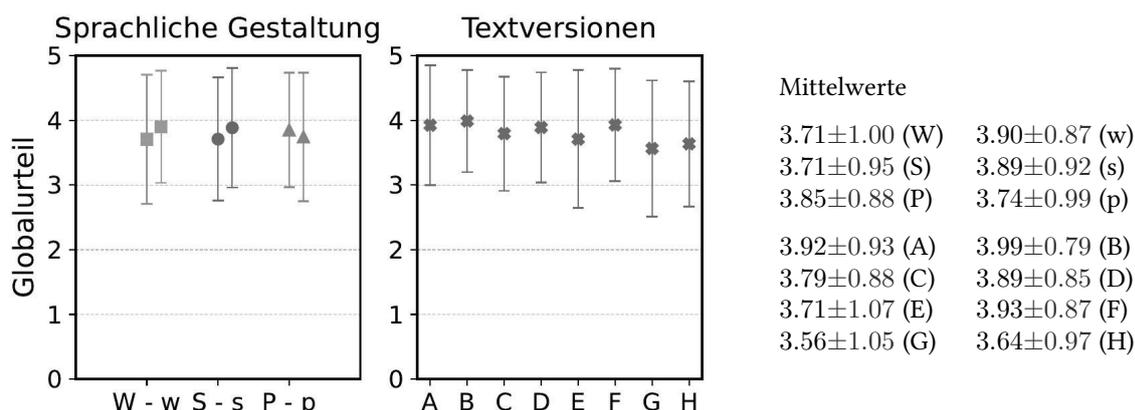


Abbildung 7.13.: Mittelwerte inklusive Fehlerbalken für das Global-Item in Abhängigkeit von der sprachlichen Gestaltung (links) beziehungsweise von der Textversion (rechts) (Datengrundlage:  $n = 755$  nicht-fehlende Werte für das Global-Item). Die Zahlenwerte der Mittelwerte und der Standardabweichungen sind nebenstehend angegeben.

Die ANOVA-Modelle zum Effekt der sprachlichen Gestaltung beziehungsweise der Textversion auf den Globaleindruck weisen auf einen signifikanten Einfluss der Wortwahl- und der Satzstruktur-Variable hin. Eine anspruchsgeminderte Ausprägung wirkte sich jeweils günstig auf den globalen Verständlichkeitseindruck aus. Es wurde eine  $\alpha$ -Fehler-Korrektur gemäß Bonferroni-Holm (für eine Erläuterung siehe Abschnitt 6.4.3; hier: drei paarweise Vergleiche W vs. w, S vs. s und P vs. p) berücksichtigt. Die Effektstärken liegen im Bereich kleiner bis mittelstarker Effekte (W:  $\omega^2 = 0.01$ , S:  $\omega^2 = 0.03$ ; siehe Seite 120 im Methoden-Abschnitt 6.4.3 für die Richtwerte). Interaktionseffekte erwiesen sich als nicht signifikant. Es wurde auch ein signifikanter Effekt kleiner Effektstärke der Textversion gefunden. Post-Hoc-Analysen deuten an, dass vor allem die Kombination anspruchsgesteigerter Wortwahl und anspruchsgesteigerter Satzstrukturen zu signifikant schlechteren Globalbewertungen führte.

In der Varianzanalyse für die globale Verständlichkeit erwies sich die Textversion als signifikante Einflussgröße. Es wurden daher paarweise Post-Hoc-Vergleiche in Form von t-Tests durchgeführt. Bei acht Textversionen und somit 28 paarweisen Vergleichen muss erheblich für eine  $\alpha$ -Fehler-Kumulierung korrigiert werden, sodass nach dieser Korrektur (Bonferroni-Holm, siehe Abschnitt 6.4.3) kein paarweiser Vergleich mehr signifikant ist. Es sei jedoch bemerkt, dass sich die deutlichsten Unterschiede jeweils gegenüber den Texten G (WSp) und H (WSP) ergaben (B vs. G:  $p_{\text{unkorr}} = 0.002$ ; B vs. H:  $p_{\text{unkorr}} = 0.006$ ; F vs. G:  $p_{\text{unkorr}} = 0.009$  und A vs. G:  $p_{\text{unkorr}} = 0.013$ ; um die vier niedrigsten, unkorrigierten p-Werte aufzuzählen). Diese beiden Textversionen G und H zeichnen sich durch die Kombination anspruchsgesteigerter Wortwahl und anspruchsgesteigerter Satzstrukturen aus und sie weisen die beiden schwächsten Mittelwerte für das Globalurteil auf (siehe Tabelle 7.13).

Zusätzlich zu den bis hierhin berichteten quantitativen Ergebnissen gibt Abbildung 7.13 visu-

ell einen Überblick, inwiefern unterschiedliche Texte zu unterschiedlichen Globaleindrücken führten. Es sind die Mittelwerte der Bewertungen des Global-Items für die verschiedenen sprachlichen Gestaltungen (W, S und P; linke Seite der Abbildung) sowie für die verschiedenen Textversionen (A bis H, rechte Seite der Abbildung) graphisch abgebildet. Dabei lässt sich erstens erkennen, dass eine anspruchsgeminderte Wortwahl zu einem besseren Globaleindruck führte. Sowohl liegt in der linksseitigen Grafik der *w*-Mittelwert über dem *W*-Mittelwert als auch liegen in der rechtsseitigen Grafik *alle* Mittelwerte der *w*-Textversionen höher als die der jeweils zugehörigen *W*-Textversion<sup>22</sup>. Analog zeigt sich zweitens sowohl in der linksseitigen Grafik als auch in den paarweisen Vergleichen der rechtsseitigen Grafik, dass anspruchsgeminderte Satzstrukturen günstiger für den globalen Verständlichkeitseindruck waren als anspruchsgesteigerte. Und drittens führte eine ausgeprägtere Personalisierung zu besseren Globalurteilen über die Verständlichkeit (ebenfalls ersichtlich anhand von paarweisen Vergleichen zugehöriger Textversionen). Mit Blick auf die Fehlerbalken ist dennoch zu beachten, dass sich alle Effekte *deutlich* innerhalb der Bereiche der Standardabweichungen abspielen, dass also die interpersonelle Varianz größer ist als die durch die Lektüre verschiedener Texte erzeugte Varianz.

### **Der Einfluss von Moderatorvariablen auf den Zusammenhang von sprachlicher Gestaltung und Globaleindruck**

In der Gesamtstichprobe wurde ein signifikanter Effekt der sprachlichen Gestaltungsvariablen *W* und *S* auf den globalen Verständlichkeitseindruck gefunden. Es wurden anschließend in analoger Vorgehensweise wie bei der Behaltensleistung (siehe Abschnitt 7.1.4) die Effekte der sprachlichen Gestaltung in einzelnen Subgruppen untersucht. Für jede der drei Gender-Subgruppen, für je zwei Subgruppen hinsichtlich Leseverstehen, Lesegeschwindigkeit und Lesegenauigkeit für die beiden Schulart-Subgruppen, für die beiden Jahrgangsstufen-Subgruppen sowie für Vorwissens-Subgruppen wurden also ANOVA-Modelle berechnet (alle einzelnen Tabellen sowie eine Übersicht über die Teilstichprobengrößen sind im Anhang D.8 einzusehen). Tabelle 7.14 stellt die Zahlenwerte für die Effektgrößen übersichtlich zusammen.

Für Personen mit einem geringeren Vorwissen (zahlenmäßig kleiner als die Gruppe der Personen mit höherem Vorwissen) zeigte die Wortwahl einen hoch signifikanten Effekt ( $p < 0.01$ ) bei kleiner bis mittelstarker Effektstärke.

Über alle übrigen Subgruppenvergleiche hinweg lässt sich zusammenfassen, dass in fast allen Fällen die Signifikanz der Wortwahl- und/oder der Satzstruktur-Variablen bei der Unterteilung der Gesamtstichprobe in Subgruppen entweder ganz verloren geht oder nur in der jeweils zahlenmäßig größten Subgruppe bestehen bleibt<sup>23</sup>.

<sup>22</sup>Mit „zugehörig“ ist gemeint, dass sich die zwei Textversionen ausschließlich in der Wortwahl-Variablen unterscheiden; also *wsp* gehört zu *Wsp*, *wSp* gehört zu *WSp* etc.

<sup>23</sup>Dies trifft in einem einzigen Fall nicht zu: Signifikante Effekte wurden in der achten Jahrgangsstufe ( $n = 147$ )

	$\omega^2$		
	W   S   P		
Gender			
	weiblich	männlich	divers
	.02*   .01*   .00	.00   .00   .00	-.04   .02   -.02
Lesefähigkeiten			
	niedrig		hoch
Verständnis	.00   .00   .00		.01*   .00   .00
Geschwindigkeit	.01   .00   .00		.01   .01   .00
Genauigkeit	.00   .02   -.01		.01*   .00   .00
Schulart			
	Realschule		Gymnasium
	.01   .00   .00		.01*   .01*   .01
Jahrgangsstufe			
	8. Klasse		9. Klasse
	.04*   .00   .00		.01   .00   .00
Vorwissen			
	niedrig		hoch
	.03**   .00   .00		.00   .01   .00

Tabelle 7.14.: Effektstärken ( $\omega^2$ ) aus ANOVA-Modellen für die Abhängigkeit des globalen Verständlichkeitseindrucks von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen *Wortwahl*, *Satzstrukturen* und *Personalisierung*. Es wurde für Teilstichproben bezüglich des Genders, bezüglich der Lesefähigkeiten (siehe dazu Abschnitt 6.3.1), bezüglich des Schulstatus und bezüglich des Vorwissens jeweils eine ANOVA berechnet. Signifikante Effekte sind mit Sternchen markiert (\*\* für  $p < 0.01$ , \* für  $p < 0.05$ ); dabei wurde eine  $\alpha$ -Fehler-Korrektur nach Bonferroni-Holm (siehe ebenfalls Abschnitt 6.4.3) angewandt. Datengrundlage:  $n = 781$ ; für die einzelnen Teilstichprobengrößen siehe Tabelle D.16 im Anhang.

Die Beobachtung, dass ein in der Gesamtstichprobe signifikanter Effekt in Teilstichproben nicht mehr signifikant ist, ist statistisch-mathematisch plausibel. Denn Unterschiede tendieren in zahlenmäßig kleiner werdenden Stichproben dazu, nicht mehr signifikant zu sein. Der Effekt der Personalisierung erwies sich in keiner Subgruppe als signifikant.

### Topikale Diskussion: *Sprachliche Gestaltung und globaler Verständlichkeitseindruck*

Die Ergebnisse dieses Unterabschnitts zur Forschungsfrage 2b sagen aus, dass sich mithilfe der sprachlichen Gestaltung signifikante, jedoch kleine Effekte auf den globalen Eindruck hinsichtlich der Verständlichkeit eines gelesenen Texts erzielen ließen. Die teilnehmenden Schüler\*innen bewerteten die wahrgenommene Verständlichkeit der Texte signifikant unterschiedlich, wenn die Texte unterschiedlich sprachlich gestaltet waren. Zu einer besseren Bewertung führten eine anspruchsgeminderte Wortwahl und die Verwendung anspruchsgeminderter Satzstrukturen. Es ist somit anzunehmen, dass die hier betrachteten Gestaltungsvariablen Wortwahl und Satzstrukturen zu den Einflussfaktoren auf den globalen Verständlichkeitseindruck, den ein Text erzeugt, zählen. Der Grad der Personalisierung erwies sich als

gefunden, obwohl die neunte Jahrgangsstufe in der Stichprobe deutlich stärker vertreten war ( $n = 640$ ). Für die Personen der achten Jahrgangsstufe bewirkte also eine anspruchsgeminderte Wortwahl einen verhältnismäßig starke Verbesserung des Verständlichkeitseindrucks.

nicht signifikant; in der Tendenz zeigte sich aber, dass eine ausgeprägtere Personalisierung zu einem besseren globalen Verständlichkeitseindruck führte.

Ein vergleichbares Resultat berichten Schmellentin, Dittmar, Gilg und Schneider (2017), in deren Studie sich das *Fachwortverstehen* als die Verständnishürde erwies, die Schüler\*innen am *bewusstesten* wird. Auch Hackemann, Heine und Höttecke (2022) kommen zu einem ähnlichen Ergebnis, indem für Texte auf dem anspruchsvollsten von drei Sprachniveaus eine geringere wahrgenommene Verständlichkeit bei Schüler\*innen gemessen wurde. Trotz der Übereinstimmung mit anderen empirischen Resultaten sollte nicht aus den Augen verloren werden, dass mit großer Sicherheit weitere Einflussfaktoren, die in der dieser Arbeit zugrundeliegenden Studie nicht Untersuchungsgegenstand waren (z. B. Interesse am Thema, alternative sprachliche Modifikationen an den Texten, Unterstützung der schriftlichen Informationen durch Bilder<sup>24</sup>), mit großer Wahrscheinlichkeit ebenfalls Einfluss auf den Globaleindruck nehmen. Anschlussuntersuchungen zu den hier berichteten Zusammenhängen könnten sich fokussiert damit auseinandersetzen, verschiedene Einflussfaktoren hinsichtlich deren Relevanz miteinander ins Verhältnis zu setzen.

Eine Betrachtung der Effekte der sprachlichen Gestaltung auf den Globaleindruck in verschiedenen Subgruppen je nach potenzieller Moderatorvariable (Gender, Lesefähigkeiten, Schulstatus und Vorwissen; siehe für Einzelheiten auch Abschnitt 7.1.4) führte zu der Erkenntnis, dass einerseits ein geringes Vorwissen den affektiven Effekt der sprachlichen Gestaltung besonders deutlich hervortreten lässt (hoch signifikanter Effekt der Wortwahl-Variablen in der *kleineren* Teilstichprobe). Andererseits wurde eine solche Beobachtung in nahezu keiner der übrigen Subgruppenvergleiche gemacht. Mit Ausnahme der Jahrgangsstufen-Unterteilung ging die Signifikanz der Effekte durch die Unterteilung der zahlenmäßig großen Gesamtstichprobe in kleinere Teilstichproben im Allgemeinen verloren oder blieb lediglich in der jeweils zahlenmäßig größten Subgruppe erhalten (vergleiche dazu Tabelle D.16 im Anhang D.8). Daraus lässt sich schließen, dass Schüler\*innen mit geringem Vorwissen beziehungsweise in der niedrigeren Jahrgangsstufe sensibler auf Textmodifikationen hinsichtlich der Wortwahl reagierten als Schüler\*innen mit einem größeren Vorwissen beziehungsweise in der höheren Jahrgangsstufe. Im Übrigen beeinflussten die betrachteten Moderatorvariablen den Effekt der sprachlichen Modifikationen auf den Globaleindruck nicht (Signifikanz und Effektstärke steigen bei Unterteilung in Subgruppen nicht an). Die sprachliche Gestaltung scheint sich weitestgehend unabhängig von Gender, Lesefähigkeiten und Schulart auf den globalen Eindruck ausgewirkt zu haben<sup>25</sup>.

Im Hinblick auf den Gültigkeitsbereich der berichteten Resultate ist erstens (wie in der Diskussion des Abschnitts 7.2.2 bereits erwähnt) zu beachten, dass sie sich auf physikalische Texte beziehen, welche sich lediglich hinsichtlich dreier, gezielt variiertes sprachlicher Gestal-

<sup>24</sup>Zu dieser Frage fand Starauschek (2006), dass Texte mithilfe von Bildern für Schüler\*innen verständlicher wirken.

<sup>25</sup>Es sei daran erinnert, dass eine ganz andere Beobachtung für das kognitive Gelingenskriterium des verstehenden Lesens auftrat: Effekte auf die Behaltensleistung wurden erst dadurch signifikant, dass einzelne Subgruppen betrachtet wurden. Vergleiche dazu Abschnitt 7.1.4.

tungsvariablen (Wortwahl, Satzstrukturen und Personalisierung) unterscheiden. Inwiefern die Resultate auf Texte mit anderem physikalischen oder nicht physikalischem Inhalt, mit anderer Vermittlungsabsicht (z. B. Instruktion zu prozeduralen Vorgängen in Form einer Versuchsanleitung oder ähnlichem) etc. übertragbar sind, kann hier nicht festgestellt werden. Zweitens liegen den Resultaten die Bewertungen hinsichtlich des Items „Ich fand den Text insgesamt verständlich“ zugrunde, welches kritisch hinsichtlich der Voraussetzungen für ANOVAs zu betrachten ist. Die große Stichprobe erlaubt zwar die Annahme, dass die Ergebnisse belastbar und aussagekräftig sind, dennoch sei auf die vermutliche Abweichung von der Normalverteilung, die Verletzung der Varianzhomogenität und das Auftreten von Boden- und Deckeneffekten hingewiesen (siehe den Anhang D.4).

Analog zum Globaleindruck wurden zur Beantwortung der Forschungsfrage 2b auch die unterschiedlichen Wahrnehmungsfacetten als abhängige Variable der sprachlichen Textgestaltung betrachtet. Die Ergebnisse dieser Analysen werden im folgenden Abschnitt dargestellt.

#### **7.2.4 Zusammenhang zwischen sprachlicher Gestaltung und affektiven Wahrnehmungsfacetten**

Einen weiteren Teil der Antwort auf Forschungsfrage 2b machen Analysen zum Zusammenhang zwischen sprachlicher Textgestaltung und den einzelnen Wahrnehmungsfacetten aus. Die Ergebnisse der dafür berechneten ANOVA-Modelle sind in Tabelle 7.15 dargestellt. Für jede der Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* wurde ein Modell mit den sprachlichen Gestaltungsvariablen W, S und P als unabhängige Variablen berechnet sowie eines mit der Textversion als unabhängige Variable. Ob die Verteilungen der Wahrnehmungsfacetten die Voraussetzungen für eine ANOVA erfüllen, wurde überprüft und ist detailliert im Anhang D.4 einsehbar. Es gilt zwar, dass die Voraussetzungen (Normalverteilung und Varianzhomogenität) insbesondere für die Facette *Angemessener Bezug zum Vorwissen* und *Involviertheit* möglicherweise nicht erfüllt sind, bei einer starken Stichprobengröße von ca.  $n = 100$  pro Textversion (siehe Tabelle 7.3 auf Seite 136) ist jedoch von einer Robustheit der Datengrundlage gegen Voraussetzungsverletzung auszugehen. Die Ergebnisse sollten jedoch in ihrer Aussagekraft nicht überschätzt werden.

7.2. Ergebnisse aus dem Bewertungsfragebogen

Unabhängige Variable	F(7,780)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
<b>Kognitive Anregung</b>				
W	4.41	0.108 0.04	0.00	2.92 (W) 3.04 (w) +0.12 ( $\Delta$ )
S	3.13	0.154 0.08	0.00	2.93 (S) 3.03 (s) +0.11 ( $\Delta$ )
P	1.35	0.245 0.24	0.00	3.01 (P) 2.95 (p) -0.06 ( $\Delta$ )
W:S	0.01	0.907	0.00	
W:P	0.04	0.839	0.00	
S:P	0.01	0.923	0.00	
W:S:P	0.00	0.997	0.00	
Textversion	1.29	0.254	0.00	A: 3.05, B: 3.13, C: 2.96, D: 3.03 E: 2.95, F: 3.01, G: 2.84, H: 2.89
<b>Lesbarkeit</b>				
W	6.45	0.034 0.01	0.01*	3.04 (W) 3.19 (w) +0.15 ( $\Delta$ )
S	5.94	0.034 0.01	0.01*	3.04 (S) 3.18 (s) +0.15 ( $\Delta$ )
P	0.06	0.807 0.81	0.00	3.11 (P) 3.10 (p) -0.01 ( $\Delta$ )
W:S	0.11	0.739	0.00	
W:P	0.41	0.522	0.00	
S:P	0.22	0.641	0.00	
W:S:P	0.02	0.885	0.00	
Textversion	1.94	0.061	0.01	A: 3.27, B: 3.23, C: 3.13, D: 3.13 E: 3.11, F: 3.12, G: 2.91, H: 3.00

7. Ergebnisse und Diskussion

Unabhängige Variable	F(7,780)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
<b>Logischer Aufbau</b>				
W	3.40	0.197 0.07	0.00	3.63 (W) 3.73 (w) +0.10 ( $\Delta$ )
S	0.05	0.825 0.82	0.00	3.67 (S) 3.69 (s) +0.02 ( $\Delta$ )
P	0.69	0.809 0.40	0.00	3.70 (P) 3.66 (p) -0.04 ( $\Delta$ )
W:S	1.53	0.216	0.00	
W:P	3.75	0.053	0.00	
S:P	0.02	0.898	0.00	
W:S:P	0.97	0.324	-0.00	
Textversion	1.51	0.160	0.00	A: 3.76, B: 3.65, C: 3.75, D: 3.76 E: 3.58, F: 3.76, G: 3.54, H: 3.65
<b>Elaboriertheit</b>				
W	9.85	0.005 0.00	0.01**	3.03 (W) 3.22 (w) +0.19 ( $\Delta$ )
S	2.54	0.223 0.11	0.00	3.06 (S) 3.17 (s) +0.11 ( $\Delta$ )
P	0.16	0.693 0.69	0.00	3.10 (P) 3.13 (p) +0.03 ( $\Delta$ )
W:S	1.95	0.163	0.00	
W:P	0.19	0.661	0.00	
S:P	6.67	0.010	0.01*	
W:S:P	0.43	0.513	0.00	
Textversion	3.13	0.003	0.02**	A: 3.35, B: 3.11, C: 3.15, D: 3.28 E: 3.18, F: 3.06, G: 2.87, H: 2.99

Unabhängige Variable	F (7,780)	p ( $\alpha$ -korr.) p (unkorr.)	$\omega^2$	Mittelwerte
<b>Angemessener Bezug zum Vorwissen</b>				
W	0.85	1.000 0.36	0.00	3.77 (W) 3.83 (w) +0.06 ( $\Delta$ )
S	0.00	1.000 0.97	0.00	3.79 (S) 3.80 (s) +0.01 ( $\Delta$ )
P	0.05	1.000 0.83	0.00	3.80 (P) 3.79 (p) -0.01 ( $\Delta$ )
W:S	0.13	0.723	0.00	
W:P	0.01	0.917	0.00	
S:P	0.11	0.736	0.00	
W:S:P	0.06	0.804	0.00	
Textversion	0.17	0.992	0.00	A: 3.83, B: 3.81, C: 3.81, D: 3.87 E: 3.78, F: 3.79, G: 3.75, H: 3.76
<b>Involviertheit</b>				
W	5.41	0.040 0.02	0.01*	2.12 (W) 2.25 (w) +0.12 ( $\Delta$ )
S	0.59	0.441 0.44	0.00	2.15 (S) 2.21 (s) +0.05 ( $\Delta$ )
P	53.13	0.000 0.00	0.06***	2.39 (P) 1.97 (p) -0.42 ( $\Delta$ )
W:S	1.14	0.286	0.00	
W:P	1.78	0.183	0.00	
S:P	0.20	0.653	0.00	
W:S:P	0.49	0.483	0.00	
Textversion	9.00	0.000	0.07***	A: 2.11, B: 2.38, C: 2.06, D: 2.48 E: 1.92, F: 2.43, G: 1.82, H: 2.30

Tabelle 7.15.: ANOVA-Ergebnisse für die Abhängigkeit der Wahrnehmungsfacetten von der sprachlichen Textgestaltung (**Wortwahl**, **Satzstrukturen**, **Personalisierung** und deren Interaktionen „:“; Textversion). Es werden die Teststatistik F für die Modellparameter (7,780) und der p-Wert (unkorrigiert sowie  $\alpha$ -korrigiert) berichtet. Diejenigen Effekte, die nach einer Bonferroni-Holm-Korrektur (siehe Abschnitt 6.4.3) signifikant sind, sind in der Effektstärke-Spalte ( $\omega^2$ ) mit \*\*\* für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$  markiert. Ergänzend sind rechts die mittleren Bewertungen angegeben. Dabei ist  $\Delta$  die Differenz „anspruchsgeminderte / schwach ausgeprägte Form“ minus „anspruchsgesteigerte / ausgeprägte Form“. Stichprobengröße  $n_{\text{HauptstudieBewFB}} = 781$ .

Ausgehend von den Ergebnissen in Tabelle 7.15 und somit unter Beachtung einer  $\alpha$ -Fehler-Korrektur nach Bonferroni-Holm (siehe Abschnitt 6.4.3) lässt sich für die einzelnen Facetten zusammenfassen:

*Kognitive Anregung*: die Mittelwerte liegen zwischen 2.8 und 3.2; die Texte wirkten im Mittel weder tendenziell wenig noch tendenziell stark kognitiv anregend. Die sprachliche Gestaltung zeigte weder in Form der Variablen W, S oder P noch in Form der Textversion einen signifikanten Einfluss auf diese Wahrnehmungsfacette.

Kurz: W (n.s.)<sup>26</sup>, S (n.s.), P (n.s.) und Textversion (n.s.)

*Lesbarkeit*: die Mittelwerte liegen zwischen 2.9 und 3.3; die Texte wirkten im Mittel weder tendenziell besonders anstrengend zu lesen noch besonders einfach zu lesen. Die sprachliche Gestaltung zeigte in Form der Wortwahl-Variablen und in Form der Satzstruktur-Variablen einen signifikanten Einfluss auf diese Wahrnehmungsfacette: eine anspruchsgeminderte Wortwahl und anspruchsgeminderte Satzstrukturen wirkten sich positiv auf die *Lesbarkeit* aus. Die Effektstärken sind jeweils klein. Die P-Variable und die Textversion zeigen keinen signifikanten Einfluss.

Kurz: W\*, S\*, P (n.s.) und Textversion (n.s.)

*Logischer Aufbau*: die Mittelwerte liegen zwischen 3.5 und 3.8; die Texte wirkten im Mittel tendenziell logisch im Aufbau. Die sprachliche Gestaltung zeigte weder in Form der Variablen W, S oder P noch in Form der Textversion einen signifikanten Einfluss auf diese Wahrnehmungsfacette.

Kurz: W (n.s.), S (n.s.), P (n.s.) und Textversion (n.s.)

*Elaboriertheit*: die Mittelwerte liegen zwischen 2.8 und 3.5; die Texte wirkten im Mittel weder besonders elaboriert und ausführlich erläutert noch besonders karg an Erläuterungen. Die sprachliche Gestaltung zeigte in Form der Wortwahl-Variablen einen hoch signifikanten Einfluss auf diese Wahrnehmungsfacette: Eine anspruchsgeminderte Wortwahl ließ den Text elaborierter wirken. Auch die Textversion zeigte einen hoch signifikanten Einfluss. Die Post-Hoc-Analyse aus 28 paarweisen t-Tests zeigte, dass sich vor allem der Text G (WSp, schwächster Mittelwert für Elaboriertheit) signifikant von den übrigen Texten unterscheidet. Die Effektstärken sind jeweils klein. Es wurde nach Bonferroni-Holm für  $\alpha$ -Fehler korrigiert (siehe Abschnitt 6.4.3). Die S- und die P-Variable zeigen keinen signifikanten Einfluss.

Kurz: W\*\*, S (n.s.), P (n.s.) und Textversion\*\*

---

<sup>26</sup>n.s.: nicht signifikant

*Angemessener Bezug zum Vorwissen:* Die Mittelwerte liegen zwischen 3.7 und 3.9; die Texte wirkten tendenziell angemessen auf das Vorwissen der Schüler\*innen bezogen. Die sprachliche Gestaltung zeigte weder in Form der Variablen W, S oder P noch in Form der Textversion einen signifikanten Einfluss auf diese Wahrnehmungsfacette.

Kurz: W (n.s.), S (n.s.), P (n.s.) und Textversion (n.s.)

*Involviertheit:* Die Mittelwerte liegen zwischen 1.8 und 2.5; die Texte erweckten tendenziell nicht den Eindruck von Involviertheit. Die sprachliche Gestaltung zeigte in Form der Wortwahl-Variablen einen signifikanten und in Form der Personalisierungs-Variablen einen höchst signifikanten Einfluss auf diese Wahrnehmungsfacette: Eine anspruchsgeminderte Wortwahl und eine ausgeprägtere Personalisierung ließen den Text jeweils personalisierter/erzählerischer wirken. Auch die Textversion zeigte einen höchst signifikanten Einfluss. Die Post-Hoc-Analyse aus 28 paarweisen t-Tests zeigte, dass ausschließlich Unterschiede zwischen solchen Textversionen signifikant sind, die sich in der Ausprägung der Personalisierung unterscheiden. Die Effekte sind klein für die Wortwahl und mittelgroß für die Personalisierung sowie für die Textversion. Es wurde nach Bonferroni-Holm für  $\alpha$ -Fehler korrigiert (siehe Abschnitt 6.4.3). Die S-Variable zeigte keinen signifikanten Einfluss.

Kurz: W\*, S (n.s.), P\*\*\* und Textversion\*\*\*

Die Abbildung 7.14 ergänzt diese Zusammenfassungen um eine visuelle Darstellung, indem die Mittelwerte der Bewertungen für die einzelnen Wahrnehmungsfacetten je nach sprachlicher Gestaltung der Texte (W, S und P; linke Seite der Abbildung) beziehungsweise je nach Textversion (A bis H, rechte Seite der Abbildung) geplottet werden. Sehr deutliche Auffälligkeiten sind in den Plots nicht erkennbar, jedoch weist Text G (WSp) in allen Subfacetten den schwächsten Mittelwert auf und für die *Involviertheit* wird der verhältnismäßig starke Einfluss der Personalisierungs-Variable augenfällig. Darüber hinaus sei bemerkt, dass sich alle Unterschiede *deutlich* innerhalb der Bereiche der Standardabweichungen abspielen. Im Verhältnis erzeugten also unterschiedliche Personen mehr Varianz als die Lektüre verschiedener Textversionen.

7. Ergebnisse und Diskussion

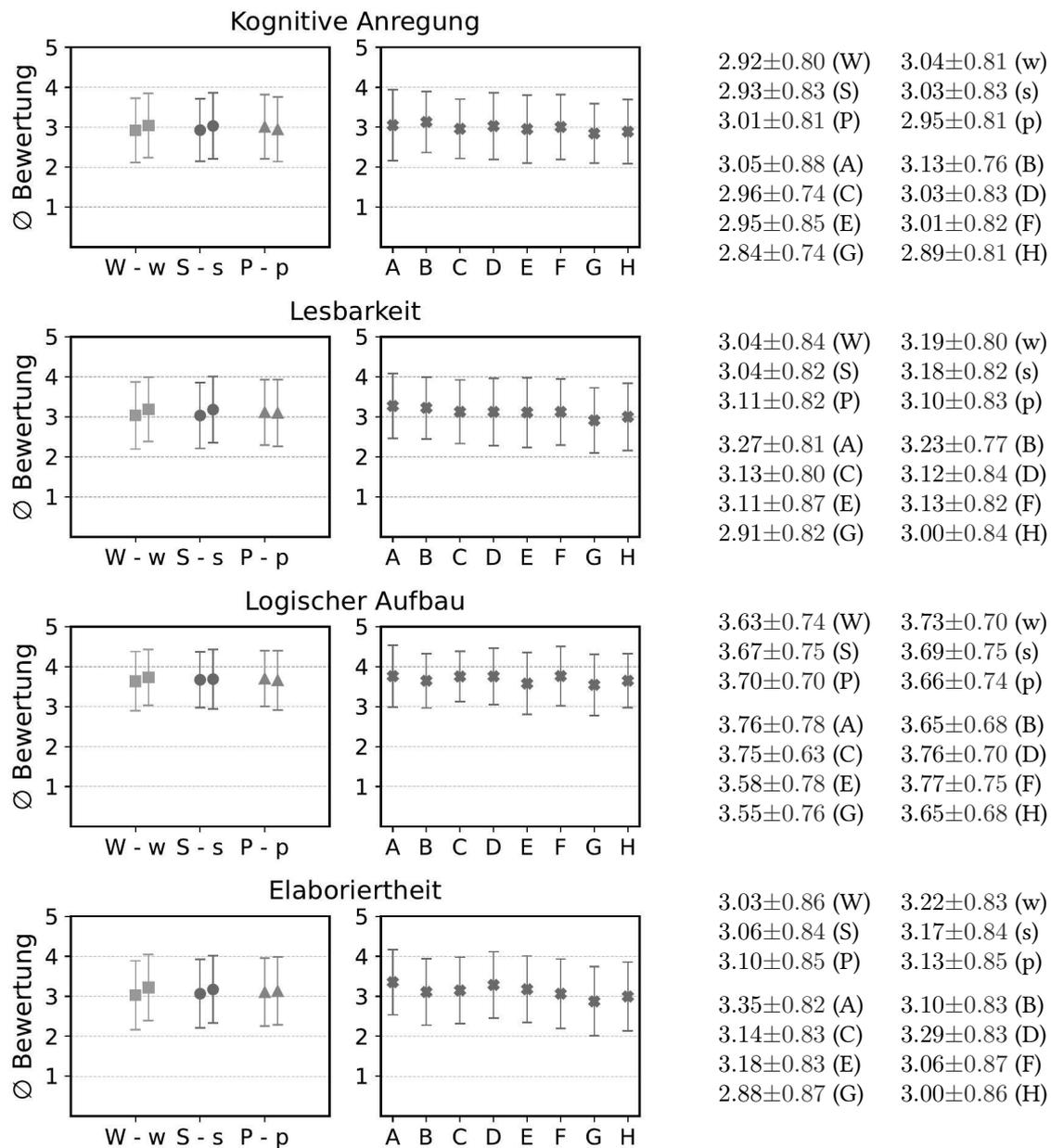


Abbildung 7.14.: Mittelwerte für die einzelnen Wahrnehmungsfacetten (Teil 1; Forts. n. Seite).

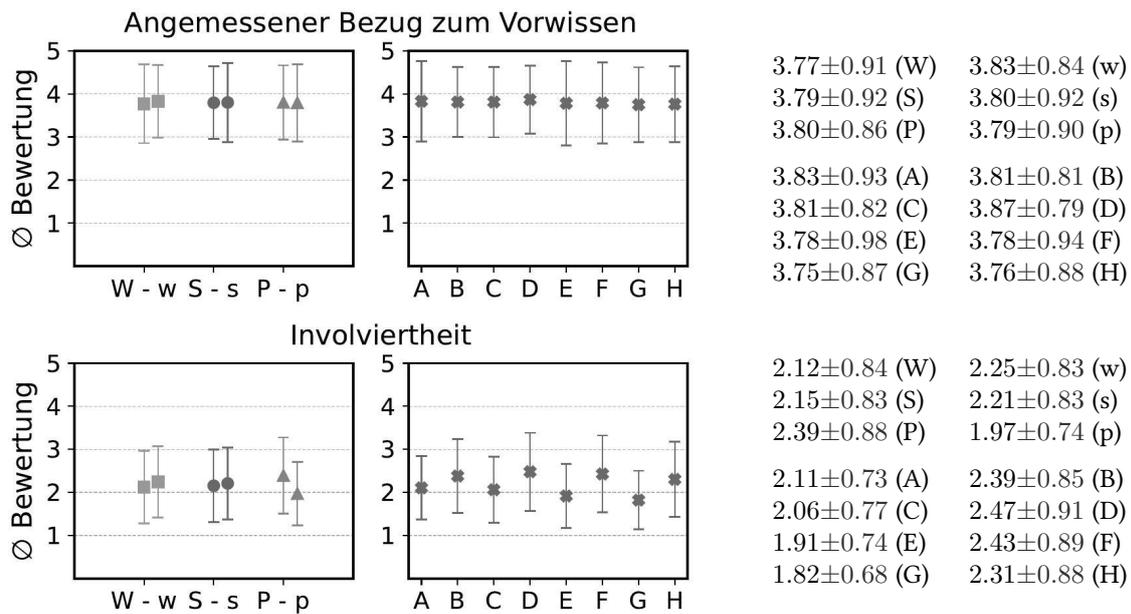


Abbildung 7.14.: Mittelwerte inklusive Fehlerbalken für die einzelnen Wahrnehmungsfacetten in Abhängigkeit von der sprachlichen Gestaltung (jeweils links) beziehungsweise von der Textversion (jeweils rechts). Datengrundlage  $n_{\text{HauptstudieBewFB}} = 781$ . Die Zahlenwerte für die Mittelwerte und Standardabweichungen sind nebenstehend angegeben.

### Der Einfluss von Moderatorvariablen auf den Zusammenhang von sprachlicher Gestaltung und Facetten der affektiven Textwahrnehmung

Hinsichtlich der einzelnen Subfacetten wurde auf eine zusätzliche Analyse von Moderatoreinflüssen bewusst verzichtet. Für die sechs Facetten eine Unterteilung der Stichprobe in Subgruppen vorzunehmen um zu Aussagen zu gelangen, inwiefern sich eine bestimmte sprachliche Gestaltungsvariable in bestimmten Subgruppen auf eine bestimmte Facette der Wahrnehmung stärker auswirkt als in anderen Subgruppen, bedeutet eine sehr große Anzahl an Kombinationen der Beobachtungsgrößen. Dies ist für diese Studie aufgrund einer übermäßigen Komplexität und Unübersichtlichkeit erstens unpraktikabel und geht zweitens über das Forschungsinteresse hinaus.

### Topikale Diskussion: Sprachliche Gestaltung und Facetten der affektiven Textwahrnehmung

Bevor die Ergebnisse dieses Abschnitts 7.2.4 diskutiert und interpretiert werden, sei hinsichtlich der Gültigkeit und Belastbarkeit an die vorhergehenden Abschnitte 7.2.2 und 7.2.3 erinnert. Die dort beschriebenen Limitationen setzen sich auf die im Folgenden zu diskutierenden und interpretierenden Resultate fort: Für ganz spezifisch sprachlich variierte Texte wurde Datenmaterial zur affektiven Textwahrnehmung von Schüler\*innen gesammelt. Aus diesem Material ließ sich eine sechsdimensionale Wahrnehmungsstruktur herausarbeiten und

dahingehend untersuchen, wie sie durch die sprachlichen Textmodifikationen beeinflussbar ist. Die Übertragung auf anderes Textmaterial ist nicht ohne Weiteres anzunehmen, sondern muss Gegenstand von Anschlussstudien bleiben.

Dieser Abschnitt 7.2.4 zu den Zusammenhängen zwischen der sprachlichen Gestaltung und den Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* vervollständigte nun die Datenanalysen zur Beantwortung der Forschungsfrage 2b. Es ließen sich für die Facetten *Lesbarkeit* ( $W^*$ ,  $S^*$ ), *Elaboriertheit* ( $W^{**}$ , Textversion<sup>\*\*</sup>) und *Involviertheit* ( $W^*$ ,  $P^{***}$ , Textversion<sup>\*\*\*</sup>) signifikante Effekte kleiner bis vereinzelt mittlerer Effektstärke nachweisen. In den übrigen Facetten ergaben sich keine signifikanten Mittelwertsunterschiede je nach sprachlicher Textgestaltung. Die Texte wirkten für die Schüler\*innen unkomplizierter zu lesen, wenn die Wortwahl und die Satzstrukturen anspruchsgemindert ausgeprägt waren. Die Bewertungen für *Elaboriertheit* fielen bei einer anspruchsgeminderten Wortwahl besser aus und bei Text G (WSp) besonders schlecht. Die *Involviertheit* schließlich reagierte am sensibelsten auf die hier vorgenommenen sprachlichen Variationen. Eine anspruchsgeminderte Wortwahl und vor allem eine ausgeprägte Personalisierung erhöhten die Bewertungen der Schüler\*innen hinsichtlich dieser Facette.

Aus Mangel an ähnlich konzipierten, dieser Studie vorangehenden Forschungsarbeiten zum Zusammenhang zwischen sprachlicher Textgestaltung und differenzierten Facetten der affektiven Textwahrnehmung lassen sich die Ergebnisse kaum vergleichend in den Stand der Forschung einbringen. Die Ergebnisse zu den Effekten der sprachlichen Textgestaltung auf einzelne Wahrnehmungsfacetten dieser Studie dienen vielmehr dazu, die bisherige Befundlage zu den Gelingenskriterien des verstehenden Lesens um eine affektive Komponente zu erweitern.

Für die Interpretation der Resultate dieses Abschnitts lässt sich mit Blick auf die Praxis sehr kurz fassen, dass sich offenbar anhand der sprachlichen Modifikationen, die in dieser Studie vorgenommen wurden, Auswirkungen auf die affektive Textwahrnehmung bei Schüler\*innen erzielen lassen. Genauer erwiesen sich die Facetten *Lesbarkeit*, *Elaboriertheit* und *Involviertheit* als beeinflussbar. Jedoch bleibt hinsichtlich der nicht-signifikant beeinflussten Facetten offen, ob die *Kognitive Anregung*, der *Logische Aufbau* und der *Angemessene Bezug zum Vorwissen* schon a priori kaum durch die sprachlichen Modifikationen betroffen waren oder ob die teilnehmenden Schüler\*innen hinsichtlich dieser Facetten weniger sensibel auf Änderungen reagieren. Hier wären weitere Studien mit anderem Textmaterial nötig.

Die Resultate dieses Abschnitts lassen sich auch in Zusammenhang mit den Erkenntnissen zum Globalurteil (siehe Abschnitt 7.2.3) bringen. Die Wortwahl und/oder die Satzstrukturen zeigten vor allem Einfluss auf die *Lesbarkeit* und die *Elaboriertheit*. Dies sind solche Facetten, die stark mit dem globalen Verständlichkeitseindruck korrelierten (siehe Abschnitt 7.2.2). Demgegenüber beeinflusste die Personalisierung ausschließlich die *Involviertheit*, die sich als nur schwach mit dem Globalurteil korreliert erwies. Es ist daher schlüssig, dass die Wortwahl und die Satzstrukturen signifikante Effekte auf den Globaleindruck zeigten und die Personalisierung – wenn auch ihr Einfluss auf die *Involviertheit* verhältnismäßig stark

war – nicht (siehe Abschnitt 7.2.3). Umso interessanter erscheint nun die Frage, inwiefern insbesondere eine veränderte Involviertheit für das Lernen (z. B. in kognitiver Hinsicht) relevant sein könnte.

Es soll daher im Anschluss an die folgende Zusammenfassung dieses Unterkapitels 7.2 der Bezug zur kognitiven Verständlichkeitskomponente, der Behaltensleistung, hergestellt werden: Im Vergleich der Ergebnisse zu Forschungsfrage 1 (sprachliche Gestaltung und Behaltensleistung; Unterkapitel 7.1) wurden anhand der sprachlichen Modifikationen deutlichere Effekte auf die affektive Textwahrnehmung erzielt – einerseits hinsichtlich des globalen Verständlichkeitseindrucks und andererseits hinsichtlich der Facetten *Lesbarkeit*, *Elaboriertheit* und *Involviertheit*. Gemäß der Grundidee dieser Studie, die in der Annahme eines mediatorischen Einflusses der affektiven Textwahrnehmung auf kognitionsbezogene Gelingenskriterien besteht (siehe Abbildung 4.1 auf Seite 64), bleibt nun noch zu prüfen, inwiefern die affektive Textwahrnehmung mit der gezeigten Behaltensleistung zusammenhängt.

### 7.2.5 Zusammenfassung und summarische Diskussion der Bewertungsfragebogen-Ergebnisse

In diesem Unterkapitel 7.2 wurden affektive Aspekte der Textverständlichkeit differenziert analysiert und in Verbindung mit der sprachlichen Textgestaltung gebracht. Dazu wurde zunächst die Datenstruktur anhand des faktoranalytischen Verfahrens der Hauptkomponentenanalyse (PCA) in ihren Dimensionen reduziert und in die sechs Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit* (im Sinne ausführlicher, expliziter Erläuterungen und Erklärungen im Text), *Angemessener Bezug zum Vorwissen* und *Involviertheit* eingeteilt (siehe Abschnitt 7.2.2). Alle Facetten korrelieren positiv und höchst signifikant mit dem globalen Verständlichkeitseindruck der Schüler\*innen und können somit als wesentliche Bestandteile der affektiv wahrgenommenen Textverständlichkeit gefasst werden. In Bezug auf Zusammenhänge zwischen den einzelnen Facetten und dem Globalurteil ist vor allem die verhältnismäßig schwache Korrelation der *Involviertheit* erwähnenswert: die *Involviertheit* scheint im Vergleich zu den anderen Facetten weniger Bedeutung für den globalen Verständlichkeitseindruck zu haben. Innerhalb der sechsdimensionalen Struktur zeigen sich die einzelnen Facetten nicht vollkommen eigenständig. Die Facetten *Lesbarkeit*, *Logischer Aufbau* und *Elaboriertheit* hängen beispielshalber stark miteinander zusammen und vor dem Hintergrund ihrer inhaltlichen Bezüge lässt sich in ihnen im Wesentlichen die Textkohärenz wiederfinden. Weitere Einzelheiten sind in Abschnitt 7.2.2, insbesondere ab Seite 164 und im Diskussionsteil ab Seite 168 beschrieben.

Inwiefern sich die sprachliche Textgestaltung auf die affektive Textwahrnehmung auswirkte, wurde in den Abschnitten 7.2.3 und 7.2.4 ausführlich beleuchtet. Mit der üblichen Notifikation für Signifikanz, das heißt \*\*\* für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$ , zeigte sich bei insgesamt überwiegend kleinen Effektstärken ( $0.01 \leq \omega^2 \leq 0.06$ , vergleiche Seite 120 im Methoden-Abschnitt 6.4.3 für die Richtwerte) in aller Kürze:

- Global-Item: W\*, S\*, P (n.s.)<sup>27</sup>, Textversion\*.
- Kognitive Anregung: W (n.s.), S (n.s.), P (n.s.), Textversion (n.s.).
- Lesbarkeit: W\*, S\*, P (n.s.), Textversion (n.s.).
- Logischer Aufbau: W (n.s.), S (n.s.), P (n.s.), Textversion (n.s.).
- Elaboriertheit: W\*\*, S (n.s.), P (n.s.), Textversion\*\*.
- Angemessener Bezug zum Vorwissen: W (n.s.), S (n.s.), P (n.s.), Textversion (n.s.).
- Involviertheit: W\*, S (n.s.), P\*\*\*, Textversion\*\*\*.

Die Wortwahl (und auch die Satzstrukturen) zeigten also ein etwas breiteres Wirkungsspektrum auf die Wahrnehmung. Die Personalisierung wirkte sich nur auf eine einzige Facette signifikant aus: Die Texte wirkten maßgeblich *involvierender* auf die Schüler\*innen, wenn sie mit einem hohen Grad an Personalisierung gestaltet waren.

Ausschließlich für das Global-Item wurde auch eine Betrachtung der Effekte in einzelnen Subgruppen vorgenommen. Dabei zeigte sich, dass die sprachlichen Textmodifikationen hinsichtlich der Wortwahl in besonderem Ausmaß für Schüler\*innen mit geringem Vorwissen-niveau und Personen der achten Jahrgangsstufe bedeutsam waren, damit die Schüler\*innen den Text als insgesamt verständlich beurteilten. In den übrigen Subgruppen wurden keine Moderationseffekte augenfällig. Es ist davon auszugehen, dass sich die Effekte aus der Gesamtstichprobe (W\*, S\*, P (n.s.), Textversion\*; siehe oben) etwa gleichverteilt in den Subgruppen wiederfinden lassen (siehe dafür ausführlich die Diskussion in Abschnitt 7.2.3 auf Seite 177).

### Summarische Diskussion

Der Gültigkeitsbereich der Ergebnisse beschränkt sich zunächst auf das konkrete Textmaterial, das in dieser Studie verwendet wurde. Hinsichtlich der Studienteilnehmenden kann dank der großen Anzahl davon ausgegangen werden, dass die Stichprobe repräsentativ für Mittelstufenschüler\*innen an Realschulen und Gymnasien ist. Für die physikalischen Sachtexte zum Thema *elektrische Spannung*, die hinsichtlich ihrer *Wortwahl / morphologischen Auffälligkeiten auf Wortebene*, ihren *Satzstrukturen / morphologischen Auffälligkeiten auf Phrasen- und Satzebene* und ihrem Grad an *erzählenden und personalisierenden Elemente* verändert wurden, konnten mit varianzanalytischen Verfahren sechs Facetten der Wahrnehmung herausgearbeitet werden und ihr Zusammenspiel mit dem globalen Verständlichkeitseindruck sowie der sprachlichen Textgestaltung untersucht werden. Inwiefern sich diese Struktur auch bei anderem Textmaterial (andere Inhalte, andere Textlänge, andere sprachliche Gestaltung) wiederfinden lässt, ist ein interessanter Forschungsgegenstand für weitere empirische Arbeiten.

Der Stand der Forschung zur affektiven Wahrnehmung von Texten hält nur wenige, gut vergleichbare empirische Ergebnisse für diese Studie bereit, da eine derart differenzierte Analyse eine in der Naturwissenschaftsdidaktik neuartige Herangehensweise ist. Insbesondere die Untersuchungen zur Auswirkung der sprachlichen Textgestaltung auf einzelne

---

<sup>27</sup>n.s.: nicht signifikant

Wahrnehmungsfacetten (Abschnitt 7.2.4) bieten neue, erste empirische Einblicke in das Wirkungsgefüge. Bezüge zu verwandten Studien wurden in den beiden Diskussions-Abschnitten auf Seite 168 für Abschnitt 7.2.2 und auf Seite 176 für Abschnitt 7.2.3 hergestellt. Die Passung der jeweiligen Ergebnisse ist gut: Verschiedene Facetten der Wahrnehmung scheinen für Schüler\*innen unterschiedlich wichtig für den globalen Verständlichkeitseindruck zu sein, den sie vom Text gewinnen – hier sind mit Blick auf die dieser Arbeit zugrundeliegenden Studie sowie auf die Studie von Heinze (2021) vor allem ausführliche, elaborierte Erläuterungen und ein logisch-nachvollziehbarer Aufbau zu nennen. „Von außen“ konnte mithilfe der sprachlichen Textgestaltung Einfluss auf den globalen Verständlichkeitseindruck genommen werden (vergleiche Abschnitt 7.2.3 in dieser Arbeit und Hackemann, Heine und Höttecke, 2022). Neben den Satzstrukturen zeigte sich dahingehend vor allem die Wortwahl als entscheidend. Dies deckt sich mit Schmellentin, Dittmar, Gilg und Schneider (2017).

Wie für das kognitive Gelingenskriterium – die Behaltensleistung – scheint also auch für den Verständlichkeitseindruck der Schüler\*innen vor allem die Wortwahl entscheidend zu sein. Es zeigte sich, dass eine anspruchsgesteigerte Wortwahl zu einem schlechteren Verständlichkeitsurteil führte. Indem also Textautor\*innen bei physikalischen Sachtexten darauf achten, Fremdwörter, lange Wörter, mehrdeutige Wörter und ungeläufige Wörter zu vermeiden, schaffen sie günstige Bedingungen dafür, dass der Text für Schüler\*innen insgesamt verständlich wirkt. Auch zu anspruchsvolle Satz- und Phrasenkonstruktionen (Passiv, Nominalisierungen Mehrwortkomplexe, Nebensatz- und Partizipialkonstruktionen) erwiesen sich als ungünstig. Einschränkend sei jedoch bemerkt, dass diese Studie nicht beantworten kann, ob es tatsächlich die Schüler\*innen sind, die sensibler auf die Wortwahl (und in zweiter Linie die Satzstrukturen) als auf andere sprachliche Gestaltungsvariablen reagiert haben, oder ob womöglich die Texte hinsichtlich dieser Gestaltungsvariablen besonders unterschiedlich gestaltet waren. Eine Offenlegung der Texte inklusive einer Übersicht über die variierten Textstellen je nach sprachlicher Gestaltungsvariable (siehe den digitalen Anhang unter <http://doi.org/10.5283/epub.58778> bzw. das Anhangskapitel A) soll daher für maximale Transparenz der Methode sorgen und eine gute Ausgangslage für Vergleichsstudien schaffen. Gegen einen derartigen Einwand, dass es die Texte sein könnten, die verhältnismäßig wenig Variation hinsichtlich der Personalisierung bieten würden und daher die Personalisierung nicht zu signifikanten Wahrnehmungs- (oder Behaltens-)Unterschieden führen würde, spricht der Befund, dass die Personalisierung durchaus einen höchst signifikanten Effekt auf die Wahrnehmungsfacetten *Involviertheit* zeigte. Lediglich schlug sich dieser Effekt nicht auf das Globalurteil oder die Behaltensleistung durch. Für die Praxis der Textproduktion oder -auswahl ist interessant, dass mithilfe von direkter Anrede, humorvollen Elementen und Füllwörtern zur Widergabe von Stimmungen ein Gefühl von Unterhaltsamkeit, emotionaler Angeregtheit und persönlicher Eingebundenheit bei den Schüler\*innen erzeugt wurde.

Nachdem in Unterkapitel 7.1 der Effekt sprachlicher Modifikationen auf kognitive Aspekte der Textverständlichkeit betrachtet wurde und in Unterkapitel 7.2 der Effekt auf affektive Aspekte der Textverständlichkeit, wird im folgenden Unterkapitel 7.3 nun die affektive

	Kogn. Anr.	Lesbark.	Log. Aufbau	Elabor.	Vorw.	Involv.	glob.
Korr.	0.16***	0.16***	0.20***	0.21***	0.13***	0.06	0.20***

Tabelle 7.16.: Pearson-Korrelationen (siehe Abschnitt 6.4.3) der einzelnen Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit* der affektiven Textwahrnehmung sowie des globalen Verständlichkeitseindrucks mit dem Gain-Score. Die Signifikanz wird mit \*\*\* für  $p < 0.001$ , \*\* für  $p < 0.01$  und \* für  $p < 0.05$  angezeigt. Datengrundlage:  $n_{\text{WT+BewFB}} = 751$ .

Textwahrnehmung der Schüler\*innen als möglicher Einflussfaktor für die Behaltensleistung im Zentrum des Interesses stehen.

### 7.3 Zusammenhang zwischen affektiver Textwahrnehmung und Behaltensleistung

Dieses Kapitel wird der Beantwortung der Forschungsfrage 3 dienen, also die Zusammenhänge zwischen dem globalen Verständlichkeitseindruck sowie den Facetten der affektiven Textwahrnehmung mit der Behaltensleistung beleuchten. Wie in Kapitel 4 erläutert wird, besteht die Grundidee dieser Studie darin, zu untersuchen, wie die sprachliche Gestaltung von Texten sowohl mit kognitiven als auch mit affektiven Gelingenskriterien des verstehenden Lesens zusammenspielt. Dabei soll insbesondere betrachtet werden, inwiefern die affektive Textwahrnehmung mediatorisch darauf Einfluss nimmt, wie sich die sprachliche Gestaltung auf die Behaltensleistung der Schüler\*innen auswirkt. Im Folgenden werden dazu die Korrelationen zwischen erstens dem Gain-Score als Maß für die Behaltensleistung (für Genaueres dazu siehe Seite 129 in Abschnitt 7.1.2) und zweitens den Bewertungen der Schüler\*innen zum globalen Verständlichkeits-Item sowie zu den sechs Wahrnehmungsfacetten analysiert. Die Wahrnehmungsfacetten sind (siehe Abschnitt 7.2.2): *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit*. Statistisch gesehen erlauben Korrelationen keine kausalen Schlüsse, dennoch wird hier aus *inhaltlich-logischen* Gründen die affektive Wahrnehmung als Regressor für das tatsächlich gezeigte Verständnis betrachtet (vergleiche dazu insbesondere Abbildung 4.1 auf Seite 64). Als Datengrundlage konnten dafür 751 Datensätze verwendet werden, die (a) für die Wissenstestauswertung verwendbar waren (siehe Abschnitt 7.1.1), (b) für die Fragebogenauswertung verwendbar waren (siehe Abschnitt 7.2.1) und (c) einen Wert für die Bewertung des Global-Items enthielten.

Tabelle 7.16 enthält die Pearson-Korrelationen (siehe Abschnitt 6.4.3) der einzelnen Wahrnehmungsfacetten sowie des globalen Verständlichkeitseindrucks mit dem Gain-Score. Alle Korrelationen sind positiv. Bis auf die Facette *Involviertheit* weisen alle Facetten sowie der Globaleindruck eine höchst signifikante Korrelation mit dem Gain-Score auf, welche jeweils

### 7.3. Zusammenhang zwischen affektiver Textwahrnehmung und Behaltensleistung

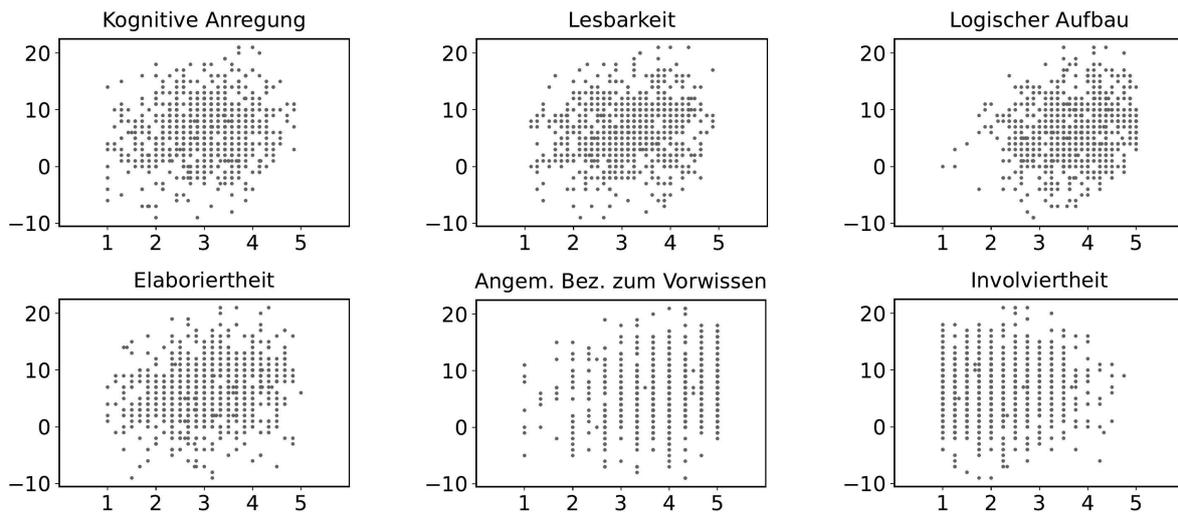


Abbildung 7.15.: Scatter-Plots für den Zusammenhang der Bewertung hinsichtlich einzelner Facetten der affektiven Textwahrnehmung (Rechtswertachse) mit dem Gain-Score (Hochwertachse). Datengrundlage:  $n_{WT+BewFB} = 751$ .

schwach bis mittelstark ausgeprägt ist (gemessen am Richtwert von  $r \geq 0.3$  für mittlere Effekte, siehe Cohen, 1988). Die stärksten Korrelationen weisen der globale Verständlichkeitsindruck, die Elaboriertheit und der Logische Aufbau auf. Die Korrelation der *Involviertheit* mit dem Gain-Score ist schwach und nicht signifikant.

Die Scatterplots in Abbildung 7.15 zeigen jedoch, dass die quantitativen Resultate und vor allem das hohe Signifikanzniveau, wofür die starke Stichprobengröße ein entscheidender Faktor ist, nicht überzubewerten sind. Es mag in den Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit* und *Angemessener Bezug zum Vorwissen* lediglich eine schwache Tendenz dazu erkennen sein, dass die Punktwolke in ihrer Ausdehnung von links unten nach rechts oben größer ist als von rechts unten nach links oben. Dies stützt das mathematische Resultat positiver Korrelationskoeffizienten und zeigt gleichzeitig an, dass das Ausmaß des Zusammenhangs nur gering ist. Vor allem hinsichtlich der *Involviertheit* lässt sich per Auge kein Zusammenhang mehr erkennen. Dies zeigte sich auch in der Pearson-Korrelation (sehr klein und nicht signifikant).

Von weiteren Analysen, die statistisch aussagekräftiger als die Betrachtung von Pearson-Korrelationskoeffizienten sind, jedoch dadurch auch stärkere Voraussetzungen an die Datengrundlage stellen, wurde an dieser Stelle abgesehen. Es wurde bereits auf Seite 171 in Abschnitt 7.2.2 argumentiert, inwiefern die aus dem Fragebogen extrahierten Subskalen (also: Wahrnehmungsfacetten) nicht mathematisch-statistisch überbeansprucht werden dürfen. Diese Studie liefert daher an dieser Stelle lediglich wertvolle, grundlegende Einblicke und erste aussagekräftige Eindrücke, die in Anschlussstudien weiterverfolgt werden sollten.

### **Diskussion: Zusammenhang affektive Textwahrnehmung – Behaltensleistung**

Die Ergebnisse dieses Abschnitts zeigen einen schwachen, positiv gerichteten Zusammenhang der affektiven Textwahrnehmung (Globaleindruck sowie Subfacetten) mit der Behaltensleistung beim verstehenden Lesen an. Mit Ausnahme der Facette *Involviertheit* korrelierten alle Facetten der affektiven Textwahrnehmung sowie der globale Verständlichkeitseindruck höchst signifikant mit der Behaltensleistung, welche anhand des Gain-Scores modelliert wurde. Vor allem für große Stichproben, bei denen auch kleine Unterschiede schnell signifikant sind, ist jedoch einschränkend zu bemerken, dass das Ausmaß der Korrelationen gering ist (siehe die Scatter-Plot-Abbildung 7.15 und beachte  $r < 0.3$  mit dem Richtwert für mittlere Effekte nach Cohen, 1988).

Diese Ergebnisse wurden auf Grundlage der spezifisch für diese Studie konstruierten Texte ermittelt. Es kann an dieser Stelle nicht beantwortet werden, wie sich die Korrelationen verändern würden, wenn andere Texte verglichen werden würden und somit mitunter mehr Varianz in der Wahrnehmung erzeugt werden würde. Denkbar wäre, andere sprachliche Modifikationen vorzunehmen, dem Text Bilder hinzuzufügen oder unterschiedliche Textinhalte zu betrachten. In einer Situation mit anderen Varianzen hinsichtlich der verschiedenen Wahrnehmungsfacetten würden sich vermutlich auch die Korrelationen der Wahrnehmungsfacetten mit dem Gain-Score ergeben. Es sei daran erinnert (siehe Tabelle 7.10 auf Seite 163 sowie die Argumentation auf Seite 168 in Abschnitt 7.2.2), dass die Stärke der berichteten Zusammenhänge von der Varianz in den Daten zu den betrachteten Variablen abhängt; dass also diejenigen Facetten mit einer geringeren Datenstreuung, wie z. B. der *Logische Aufbau* oder die *Kognitive Anregung*, a priori weniger Einfluss auf die abhängige Variable zeigen konnten.

Dennoch zeigt der gefundene positive Zusammenhang zwischen affektiver Textwahrnehmung und Behaltensleistung summarisch gesprochen an, dass verständlich empfundene Texte auch tendenziell zu einer besseren Behaltensleistung beim verstehenden Lesen führten und dass sich die teilnehmenden Schüler\*innen in dieser Studie mit keiner Verstehensillusion konfrontiert sahen (von einer Verstehensillusion bei Erklärvideos berichten Kulgemeyer und Wittwer, 2023; vergleiche auch Seite 67 im Kapitel zum Forschungsinteresse 4). Dieser Befund steht zunächst in einem scheinbaren Widerspruch zu den Ergebnissen von Hackemann, Heine und Höttecke (2022), die *keinen* Zusammenhang zwischen tatsächlichem Verständnis und wahrgenommener Verständlichkeit nachweisen. Die Autoren ziehen jedoch ihren Schluss daraus, dass die sprachliche Gestaltung der Texte sich ausschließlich auf die wahrgenommene Verständlichkeit auswirkt und das Verständnis unbeeinflusst lässt: „the perceived comprehensibility does not confound with students’ actual text comprehension because the latter does not vary significantly across levels of linguistic demand“ (Hackemann, Heine & Höttecke, 2022, S. 19). Diese Studie verfolgte demgegenüber einen anderen Ansatz, der die Wahrnehmung stärker betont und für eine generelle Aussage über den Zusammenhang von wahrgenommener Verständlichkeit und tatsächlichem Verständnis treffender ist: bei den Analysen zum Zusammenhang von affektiver und kognitiver Textwirkung wurde die sprach-

liche Gestaltung außer Acht gelassen und stattdessen „Person für Person“ betrachtet, wie die affektive Textwahrnehmung mit der gezeigten Behaltensleistung zusammenhängt. Dieser Ansatz führte offenbar zu dem anderen Ergebnis eines schwachen, aber höchst signifikanten Zusammenhangs.

Von einem Standpunkt, der die affektive Wahrnehmung als potenzielle Mediatorvariable zwischen Text und gezeigter Behaltensleistung beim verstehenden Lesen sieht (vergleiche Abbildung 4.1 auf Seite 64), lässt sich schließen: Insgesamt bewirkten *verständlich wirkende* Texte hinsichtlich des globalen Eindrucks sowie der Subfacetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit* und *Angemessener Bezug zum Vorwissen* tendenziell bessere Behaltensleistungen. Dennoch sind die Zusammenhänge schwach und *verständlich wirkende* Texte sind nicht ohne Weiteres als *kognitiv verständliche* Texte zu bewerten. Insbesondere die *Involviertheit* erwies sich als besonders unerheblich für die Behaltensleistung der Studienteilnehmenden. Dies lässt insgesamt den intuitiv plausiblen Schluss zu, dass die affektive und die kognitive Wirkung eines Texts getrennte Konstrukte sind, die teilweise durch gemeinsame Einflussfaktoren und teilweise durch spezifische Einflussfaktoren bestimmt werden.



Wichtig ist,  
dass man nicht aufhört zu fragen.

---

Albert Einstein

## 8 | RESÜMEE UND AUSBLICK

Am Ende der theoretischen und praktischen Vorbereitung, der Durchführung und schließlich der Auswertung dieser Studie lässt sich dem Kenntnisstand zur Textverständlichkeit in den naturwissenschaftlichen Fächern ein neuer Baustein hinzufügen. Zugleich markiert dies den Anfang weiterer Erkenntnisschritte. Denn es gilt, das Gesehene hinsichtlich seines Geltungsbereichs zu testen, sich andeutende Effekte mit genaueren Analysen weiterzuverfolgen und – um es kurz und mit Albert Einsteins Worten zu sagen – nicht aufzuhören, sich Fragen zu stellen. Bevor jedoch anhand eines schlaglichtartigen Resümees auf diese Arbeit zurückgeblickt wird und einige interessante Anschlussfragen formuliert werden, sei ein Hinweis auf eine Anwendung der theoretischen und empirischen Bestandteile dieser Arbeit in der universitären Lehre gegeben.

### **Anwendung der wissenschaftlichen Erkenntnisse in der Lehrkräftebildung**

Zusätzlich zum forschungsbezogenen Fortschritt bestand ein Ziel dieser Arbeit in einem Praxisbeitrag zur Verbesserung der Lehrkräftebildung. Das Forschungsprojekt war Teil des *Impuls+*-Projekts im Projektverbund *KOLEG 2*. *KOLEG 2* wurde vom Bundesministerium für Bildung und Forschung im Rahmen der *Qualitätsoffensive Lehrerbildung* gefördert. Ein zentrales Ziel der projektbezogenen Forschungsarbeiten war es, die Ergebnisse direkt für die Lehrkräftebildung nutzbar zu machen. In diesem Zug wurde begleitend zur Forschungsarbeit, welche ausführlich in dieser Dissertationsschrift dargelegt wurde, ein Seminar für angehende Lehrkräfte zum Thema *Sprache im Physikunterricht* entwickelt und mehrmals durchgeführt. Eine kurze Inhaltsbeschreibung sowie ein stichpunktartiger Semesterplan für den Seminarablauf im Anhang E bieten einen Einblick in die Lehrveranstaltung. Auf der Lernplattform *GRIPS* der Universität Regensburg sind die Seminarinhalte gemeinsam mit eigens für das Seminar entwickeltem Lehr-Lern-Material dokumentiert und für eine zukünftige Nutzung in der physikdidaktischen Lehre an der Universität Regensburg zur Verfügung gestellt<sup>1</sup>.

---

<sup>1</sup>Für UR-Mitglieder ist der Kurs einsehbar: <https://elearning.uni-regensburg.de/course/view.php?id=63619>.

### **Rück- und Ausblick in Bezug auf die wissenschaftlichen Erkenntnisse dieser Arbeit<sup>2</sup>**

Die empirischen Erkenntnisse dieser Arbeit basieren auf einer Erhebung an bayerischen Realschulen und Gymnasien, an der insgesamt 787 Schüler\*innen teilnahmen (siehe Abschnitt 6.5.2). Unter anderem dank dieser großen Anzahl an Studienteilnehmenden konnten aussagekräftige Feststellungen zu den Auswirkungen sprachlicher Veränderungen an physikalischen Sachtexten auf die Textverständlichkeit für Schüler\*innen getroffen werden. Es wurden dazu acht sprachlich unterschiedliche Textversionen zum Thema „Die elektrische Spannung“ erstellt (siehe Unterkapitel 6.2). Jede der acht Textversionen zeichnete sich dabei durch ein individuelles Muster im Anforderungs- beziehungsweise Ausprägungsgrad (jeweils hoch/niedrig) hinsichtlich der Wortwahl, der Satzstrukturen und des Grads an Personalisierung aus. Die teilnehmenden Schüler\*innen lasen jeweils eine ihnen per Zufall zugeteilte Textversion, bearbeiteten einen Wissenstest im Pre-Post-Design sowie einen Fragebogen zur Textbewertung direkt im Anschluss an die Lektüre. Darüber hinaus wurden die Lesefähigkeiten gemessen und das Gender, die Jahrgangsstufe und die Schulart für jede teilnehmende Person miterhoben (siehe die Unterkapitel 6.1 und 6.5.1).

Zusammenfassend zeigte sich, dass die sprachlich unterschiedlich gestalteten Texte zu unterschiedlichen Wirkungen auf unterschiedliche Schüler\*innengruppen führten. Anhand des folgenden, qualitativen Ergebnisresümeees wird deutlich gemacht, inwiefern eine differenzierte Vorgehensweise in mehrfacher Hinsicht entscheidend war, um Effekte nachzuweisen und präzise analysieren zu können.

Der Vergleich *einzelner, separat veränderter sprachlicher Gestaltungsvariablen* (Wortwahl, Satzstrukturen und Personalisierung) hinsichtlich deren Auswirkungen auf Lesende war aussagekräftiger als der Vergleich von Textversionen, also der Betrachtung aller Gestaltungsvariablen zugleich. Signifikante, kleine Unterschiede hinsichtlich der Behaltensleistung zeigten sich ausschließlich für Veränderungen im Anforderungsgrad der Wortwahl; und zwar erzielten bestimmte Subgruppen der Stichprobe bessere Resultate im Wissenstest für eine anspruchsgeminderte Wortwahl im Vergleich zu einer anspruchsgesteigerten. Vor allem die Wortwahl, teilweise auch die Satzstrukturen, führten je nach Anforderungsgrad auch zu signifikant unterschiedlichen affektiven Wahrnehmungen bei den Schüler\*innen – sowohl hinsichtlich des globalen Verständlichkeitseindrucks als auch hinsichtlich einiger Subfacetten der Wahrnehmung. Eine anspruchsgeminderte Gestaltung führte vor allem bei kleinen, vereinzelt auch mittleren Effektstärken zu günstigeren Textbewertungen. Als bemerkenswert erwies sich auch das Wirkungsprofil der Personalisierung: es wurde auf höchstem Signifikanzniveau ein großer Effekt für die Facette der *Involviertheit* nachgewiesen. Abgesehen von der Steigerung des Gefühls von *Involviertheit*, das sich bei ausgeprägter Personalisierung bei den Lesenden einstellte, erwies sich die Personalisierung jedoch als irrelevant für kognitive wie affektive Textwirkungsaspekte.

Um einzelne Effekte beobachten und herausarbeiten zu können, erschien es also wichtig,

---

<sup>2</sup>Eine ähnliche Ergebniszusammenfassung findet sich in Flieser und Rincke, 2024.

dass keine zu großen Sets an sprachlichen Modifikationen zugleich betrachtet wurden. Die Effekte traten im Allgemeinen weniger deutlich hervor, wenn die Textversionen, die gewissermaßen als Konglomerat aus allen drei sprachlichen Gestaltungsvariablen zu betrachten sind, miteinander verglichen wurden. Die der Studie vorangehende, möglichst umfassende Zusammenstellung und Sortierung der sprachlichen Gestaltungsmittel (Konstruktraum für die sprachliche Gestaltung, siehe Unterkapitel 5.1) erwies sich also als wertvolle Grundlage; insbesondere, da sie eine theoretisch begründete Einschätzung über die Trennbarkeit der acht rekonstruierten Gruppen erlaubte. Hier schließen sich interessante Überlegungen an: Wodurch kommt das Verschwinden der Effekte bei Betrachtung der Textversionen als Ganzes zustande? Denkbar wären rein statistische Effekte, da die Gruppengröße bei Einteilung der Studienteilnehmenden je nach gelesener Textversion geringer ist, als wenn die Studienteilnehmenden je nach sprachlicher Gestaltung ihres Textes hinsichtlich W beziehungsweise S beziehungsweise P in Gruppen eingeteilt und verglichen werden (vergleiche Tabelle 7.3 auf Seite 136). Und in kleineren Gruppen werden der Statistik des Zufalls folgend Effekte nicht so schnell signifikant. Denkbar ist jedoch auch, dass sich Effekte einzelner sprachlicher Gestaltungsmaßnahmen gegenseitig kompensieren. Weitere Studien, die den Fokus auf Interaktionseffekte legen und zum Beispiel auf Methoden der Multilevel-Analyse zurückgreifen (z. B. nach dem Vorbild von Hilbert, Stadler, Lindl, Naumann und Bühner, 2019) könnten hier ansetzen.

Eine *differenzierte Betrachtung einzelner Personengruppen je nach Gender, Schulstatus und Lesefähigkeiten* führte vor allem hinsichtlich der Auswirkungen sprachlicher Gestaltung auf die Behaltensleistung aber auch hinsichtlich der affektiven Textwahrnehmung zu aufschlussreichen Ergebnissen. Für die Behaltensleistung zeigte sich, dass Schüler\*innen mit tendenziell schwächeren Ausgangsbedingungen (geringeres Vorwissen, begrenztere Lesefähigkeiten, Realschüler\*innen im Vergleich zu Gymnasiast\*innen) stärker von einer anspruchsgeminderten Wortwahl profitieren als Schüler\*innen mit stärkeren Ausgangsbedingungen. Der globale Verständlichkeitseindruck schien weniger stark durch diese Personeneigenschaften moderiert zu sein. Lediglich für zwei Subgruppen (achte Jahrgangsstufe und geringeres Vorwissen) ist davon auszugehen, dass das globale Verständlichkeitsurteil der Schüler\*innen sensibler von der Wortwahl abhängt als das mittlere Urteil in der gesamten Gruppe. Die Schüler\*innen in diesen beiden Subgruppen beurteilten Texte mit anspruchsgeminderter Wortwahl als insgesamt verständlicher (signifikante bis hoch signifikante Effekte kleiner bis mittlerer Stärke).

Zusammengefasst scheint also vor allem das Vorwissen eine einflussstarke Moderatorvariable zu sein und zwar in dem Sinne, dass Schüler\*innen mit geringerem Vorwissen sensibler auf Textveränderungen – vor allem hinsichtlich der Wortwahl – reagieren. Da in keinem Fall ein negativer Effekt der anspruchsgeminderten Wortwahl gemessen wurde, deutet sich hier womöglich eine Option an, um einer größer werdende Schere zwischen leistungsstarken und leistungsschwachen Schüler\*innen entgegenzuwirken. Bei der Verwendung einer Wortwahl niedrigen Anspruchslevels ist jedoch zu beachten, dass eine Heranführung an fachsprachliche

Ausdrucksweisen auch beinhalten muss, den Umgang mit ungeläufigen, fremdsprachlichen Wörtern und Begriffen zu üben (siehe dazu die Seiten 27 und 90).

Es erwies sich darüber hinaus als sinnvoll und interessant, die *beiden Wirkungsebenen der kognitiven und der affektiven Textwirkung (Behaltensleistung und affektive Textwahrnehmung) zu untersuchen* und Unterschiede sowie Zusammenhänge festzustellen. Generell zeigte die sprachliche Gestaltung eher geringe Auswirkungen auf die Behaltensleistung. Dieses Ergebnis steht im Einklang zu den vorangehenden Arbeiten zur Verständlichkeit naturwissenschaftlicher Texte für Schüler\*innen (siehe Unterkapitel 3.4 sowie die Diskussion in Abschnitt 7.1.5). Für die Behaltensleistung deutete sich an, dass Text B (Ausprägungsmuster wsP) sich durch besonders gute Auswirkungen auf Lesende auszeichnete. Die affektive Textwahrnehmung der Schüler\*innen hing etwas deutlicher als die Behaltensleistung von der sprachlichen Gestaltung der Texte ab. Hier hoben sich zwei Texte *negativ* von den anderen ab: Die Texte G (WSp) und H (WSP) wurden signifikant als insgesamt schlechter verständlich beurteilt. Darüber hinaus stellte die differenzierte Messung der affektiven Textwahrnehmung einen für die Textverständlichkeitsforschung in den Naturwissenschaften neuen Forschungsansatz dar. Es wurde daher zunächst aus der Theorie abgeleitet, welche affektiven Aspekte mit der (wahrgenommenen) Verständlichkeit von Texten grundlegend in Verbindung stehen (Unterkapitel 5.2). Daraus wurden Items für einen Bewertungsfragebogen erstellt. Auf Grundlage der Fragebogendaten zeigte sich, dass die affektive Textwahrnehmung in sechs Dimensionen differenziert werden kann: sie besteht in der Einschätzung der Schüler\*innen, wie *kognitiv anregend* ein Text ist, wie *flüssig und unangestrengt er sich lesen lässt*, wie *logisch im Aufbau* er ist, wie *elaboriert* er ist, wie *angemessen die Bezugnahme des Texts auf das Vorwissen* ist und wie *involviert* sich die Schülerinnen fühlen. Die höchst signifikanten, überwiegend starken Korrelationen der Facetten miteinander sowie mit der Globalbewertung zur Verständlichkeit sprechen dafür, dass sich der Gesamteindruck hinsichtlich der Textverständlichkeit maßgeblich in diesen Facetten zeigt und fassen lässt. In diesem Zusammenhangsnetz hob sich die Facette der *Involviertheit* von den übrigen Facetten ab. Sie zeigte nur schwache Zusammenhänge mit dem globalen Verständlichkeitseindruck und der Behaltensleistung. Darüber hinaus war sie, wie oben bereits erwähnt, die einzige Facette, die von der sprachlichen Gestaltungsvariable *Personalisierung* maßgeblich beeinflusst war.

Diese ersten Schritte in Richtung einer intensiveren forschersischen Auseinandersetzung mit den affektiven Aspekten der Textwahrnehmung durch Schüler\*innen fügen sich in den allgemeinen Trend der Textverständlichkeitsforschung, der ausgehend von rein textstatistischen Merkmalen mehr und mehr die *aktive* Rolle der lesenden Person berücksichtigte (siehe Unterkapitel 3.3). Vor dem Hintergrund, dass ein Text damit je nach Lesezweck und Adressat\*innen *passend* gestaltet sein sollte und dass die Erwartungen, mit denen Schüler\*innen an Physiktexte herantreten, von fachspezifischen Vorerfahrungen geprägt sind (siehe vergleichend Seite 48 zur Relevanz des Vorwissens für das Textverstehen), erscheint es sinnvoll, Schüler\*innen direkt nach ihren Eindrücken vom Text beim Lesen zu fragen. Die Ergebnisse dieser Befragung legen eine wichtige Grundlage für den weiteren Erkenntnisfortschritt und werfen

anregende Anschlussfragen auf. In Bezug auf den Ausprägungsgrad an *Personalisierung* in den Texten blieb in dieser Studie unklar, inwiefern eine stärker personalisierte/erzählerische Textgestaltung über die deutliche Steigerung des Gefühls von *Involviertheit* hinaus weitere lernförderliche Konsequenzen haben kann. Womöglich fühlen sich Schüler\*innen durch die *Involviertheit* stärker geneigt, sich weiter mit dem Text auseinanderzusetzen oder die allgemeinen Lesefreude der Schüler\*innen lässt sich steigern. Zwar blieb in dieser Studie ein messbarer Effekt der Personalisierung auf das kognitive Erfolgskriterium aus, jedoch berichten Strohmaier, Ehmke, Härtig und Leiss (2023) von einer verhältnismäßig deutlichen Steigerung des Textverständnisses durch personalisiertere Texte. Es sollte daher ein erhebliches Forschungsinteresse für Anschlussstudien in der fokussierten Auseinandersetzung mit dem Wirkungspotenzial einer personalisierten Textgestaltung liegen. Jedoch bleibt auch ganz generell zu überprüfen, inwiefern sich eine ähnliche oder andere Struktur der affektiven Textwahrnehmung zeigt, wenn anderes Textmaterial verwendet wird. Da die Ergebnisse dieser Arbeit in weiten Teilen auf Varianzanalysen beruhen und Varianz zwar einerseits von den vielen teilnehmenden Personen erzeugt wird, andererseits jedoch auch durch die unterschiedlichen gelesenen Textversionen, sind Vergleichsstudien nötig. Diese sollten die Textwahrnehmung bei anders angelegten sprachlichen Modifikationen/Gestaltungsprinzipien (Bebilderung, unterschiedliche Gliederung, unterschiedliches Layout,...) oder bei alternativen Textinhalten in den Blick nehmen. In dieser Studie wurde bewusst ein *abstrakter*, physikalischer Textgegenstand gewählt, da anzunehmen ist, dass dadurch die Sprache entscheidend für den Aufbau von Vorstellungen ist (siehe Seite 20). Es bleibt zu überprüfen, inwiefern sich diese Vermutung womöglich darin widerspiegelt, wie sensibel sich Schüler\*innen auf sprachliche Texteigenschaften zeigen.

Mit diesem abschließenden Resümee wurde noch einmal herausgestellt, inwiefern diese Arbeit anhand empirischer Evidenz Antworten auf die Forschungsfragen finden konnte. Es wurden die Auswirkungen sprachlicher Gestaltungsmaßnahmen in Physiktexten auf die Behaltensleistung von Schüler\*innen untersucht, es wurden affektive Aspekte der Textwahrnehmung auf systematische Weise exploriert und analysiert und es wurden Zusammenhänge zwischen den kognitiven und affektiven Textwirkungen beleuchtet. Und dennoch stehen neben den Antworten vor allem wieder neue Fragen. Einige interessante Anknüpfungspunkte waren ebenfalls Teil dieses Resümees. Ganz im Sinne der Worte Albert Einsteins, die am Anfang dieses Kapitels stehen, sollen sie als Anstoß für weitere Forschungsunternehmen zur Verständlichkeit naturwissenschaftlicher Texte für Schüler\*innen dienen. Denn aus didaktischer Sicht ist es wichtig, die Prozesse des verstehenden Lesens wissenschaftlich weitergehend zu verstehen und Schüler\*innen in der Praxis beim Umgang mit schriftlichem Textmaterial besser unterstützen zu können. Das Lernen anhand von Texten ist ein wertvolles Bildungs- und Kulturgut, das es zu fördern gilt. Denn – um noch einmal den Einleitungsgedanken dieser Arbeit in Form der Verse 1966 f. aus Goethes *Faust I* (1808) zu zitieren – „was man schwarz auf weiß besitzt, kann man getrost nach Hause tragen“.



## LITERATURVERZEICHNIS

- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives* (P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths & M. C. Wittrock, Hrsg.). Longman.
- Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for Comprehending Discourse. *American Educational Research Journal*, 14(4), 367–381.
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., & Ring, K. (2005). *Expertise - Förderung von Lesekompetenz* (Unveränderter Nachdruck, Bd. 17). BMBF, Referat Publikationen; Internetredaktion.
- Bamberger, R., & Vanacek, E. (1984). *Lesen, Verstehen, Lernen, Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Jugend und Volk.
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220.
- Baurmann, J. (2006). Texte verstehen im Deutschunterricht. In H. Blühdorn, E. Breindl & U. H. Wassner (Hrsg.), *Text-Verstehen: Grammatik und darüber hinaus* (S. 239–253). De Gruyter.
- Berger, K., & Starauschek, E. (2009). Wie erkennt und erstellt man gute Bilder? Einige Regeln zur Gestaltung von Bildern für das Physiklernen. *Naturwissenschaften im Unterricht. Physik*, 20(109), 23–26.
- Bernholt, S., Härtig, H., & Retelsdorf, J. (2023). Reproduction Rather than Comprehension? Analysis of Gains in Students' Science Text Comprehension. *Research in Science Education*, 53(3), 493–506.
- Beyer, R., & Gerlach, R. (2018). *Sprache und Denken* (J. Kriz, Hrsg.; 2. Aufl.). Springer.
- Björnsson, C. H. (1968). *Lesbarkeit durch Lix*. Pedagogiskt Centrum.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Springer.
- Bortz, J., Lienert, G. A., & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Springer.

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Springer.
- Brainerd, C. J., & Reyna, V. F. (1988). Generic Resources, Reconstructive Processing, and Children's Mental Arithmetic. *Developmental Psychology*, 24(3), 324–334.
- Brandt, H. (2020). Exploratorische Faktorenanalyse. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 575–614). Springer.
- Brandt, H., & Moosbrugger, H. (2020). Planungsaspekte und Konstruktionsphasen von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 39–66). Springer.
- Brunner, M., Stallasch, S. E., & Lüdtke, O. (2023). Empirical Benchmarks to Interpret Intervention Effects on Student Achievement in Elementary and Secondary School: Meta-Analytic Results from Germany. *Journal of Research on Educational Effectiveness*, 1–39.
- Bryant, C. (2018). Researchpy. <https://github.com/researchpy/researchpy> [Zugriff: 01.12.2023].
- Busch, H., & Ralle, B. (2013). Diagnostik und Förderung fachsprachlicher Kompetenzen im Chemieunterricht. In *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 277–294). Waxmann.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. Aufl.). Routledge.
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(7).
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121–129.
- de Saussure, F. (1967). *Grundfragen Der Allgemeinen Sprachwissenschaft* (2. Aufl.). De Gruyter.
- Deutsches PISA-Konsortium. (2013). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich* (J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann & M. Weiß, Hrsg.). Springer-Verlag.
- Duden. (o.J.).
- Dürscheid, C. (2012). *Einführung in die Schriftlinguistik* (4. Aufl.). Vandenhoeck & Ruprecht.
- Ennemoser, M., Marx, P., Weber, J., & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 44(2), 53–67.
- Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education*, 28(5), 491–520.

- Fang, Z., Lamme, L., Pringle, R., Patrick, J., Sanders, J., Zmach, C., Charbonnet, S., & Henkel, M. (2008). Integrating Reading into Middle School Science: What we did, found and learned. *International Journal of Science Education*, 30(15), 2067–2089.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical Power Analyses Using G\* Power 3.1: Tests for Correlation and Regression Analyses. *Behavior research methods*, 41(4), 1149–1160.
- Feilke, H. (2013). Bildungssprache und Schulsprache am Beispiel literal-argumentativer Kompetenzen. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach: Sprachlichkeit und fachliches Lernen* (S. 113–130). Waxmann.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233.
- Flieser, K., & Rincke, K. (2024). Verständlichkeit physikalischer Texte für Schüler\*innen: Behaltensleistung und Textwahrnehmung im Fach Physik. In *Frühe naturwissenschaftliche Bildung – Tagungsband zur Jahrestagung der GDCP e.V. (im Druck)*.
- Frickel, D. (2018). Textpassung - Theoretische und empirische Ansätze zur Ermittlung der Gegenstandsadäquanz von (literarischen) Texten zwischen Einfachheit bzw. Verständlichkeit und Komplexität. In *Empirische Forschung in der Deutschdidaktik - Band 3: Forschungsfelder der Deutschdidaktik* (S. 185–198). Schneider.
- Gäde, J. C., Schermelleh-Engel, K., & Werner, C. S. (2020). Klassische Methoden der Reliabilitätschätzung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 305–334). Springer.
- Gademann, A., Guhn, M., & Zumbo, B. (2019). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, 17(1).
- Glück, H., & Rödel, M. (2016). *Metzler Lexikon Sprache* (5. Aufl.). J.B. Metzler Verlag.
- Gogolin, I., Dirim, I., Klinger, T., Lange, I., Lengyel, D., Michel, U., Neumann, U., Reich, H. H., Roth, H.-J., & Schwippert, K. (2011). *Förderung von Kindern und Jugendlichen mit Migrationshintergrund FörMig. Bilanz und Perspektiven eines Modellprogramms* (1. Aufl.). Waxmann.
- Gogolin, I., & Duarte, J. (2016). Bildungssprache. In J. Kilian, B. Brouer & D. Lüttenberg (Hrsg.), *Handbuch Sprache in der Bildung*. De Gruyter.
- Göpferich, S. (2002). *Textproduktion im Zeitalter der Globalisierung*. Stauffenburg.
- Gregory, M. (1967). Aspects of varieties differentiation. *Journal of Linguistics*, 3(2), 177–198.
- Groeben, N. (1978). *Die Verständlichkeit von Unterrichtstexten* (2., überarb. und erw. Aufl.). Aschendorff.
- Groeben, N. (1982). *Leserpsychologie: Textverständnis Textverständlichkeit*. Aschendorff.
- Habermas, J. (1977). Umgangssprache, Wissenschaftssprache, Bildungssprache. In Max-Planck-Gesellschaft (Hrsg.), *Jahrbuch 1977* (S. 36–51).

- Hackemann, T., Heine, L., & Höttecke, D. (2022). Challenging to Read, Easy to Comprehend? Effects of Linguistic Demands on Secondary Students' Text Comprehension in Physics. *International Journal of Science and Mathematics Education*, 20(1), 43–68.
- Halliday, M. A. K. (1966). *The Linguistic Sciences and Language Teaching*. Longmans.
- Halliday, M. A. K. (2004). Some Grammatical Problems In Scientific English. In *Applying English Grammar*. (S. 77–94). Routledge.
- Härtig, H., Bernholt, S., Precht, H., & Retelsdorf, J. (2015). Unterrichtssprache im Fachunterricht - Stand der Forschung und Forschungsperspektiven am Beispiel des Textverständnisses. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 55–67.
- Härtig, H., Fraser, N., Bernholt, S., & Retelsdorf, J. (2019). Kann man Sachtexte vereinfachen? – Ergebnisse einer Generalisierungsstudie zum Textverständnis. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 273–287.
- Heine, L., Domenech, M., Otto, L., Neumann, A., Krelle, M., Leiss, D., Höttecke, D., Ehmke, T., & Schwippert, K. (2018). Modellierung sprachlicher Anforderungen in Testaufgaben verschiedener Unterrichtsfächer: Theoretische und empirische Grundlagen. *Zeitschrift für Angewandte Linguistik*, 2018(69), 69–96.
- Heinze, J. (2021). *Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik* (Diss.). Universität Regensburg. Regensburg.
- Heiser, A. (2019). *Wirkstoff Werbetext*. Springer Gabler.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität - Diagnose, Evaluation und Verbesserung des Unterrichts* (6. Aufl.). Klett Kallmeyer.
- Helzel, A., & Rabe, T. (2023). Wie reflektieren Lehramtsstudierende ‚Sprache im Physikunterricht‘? In *Lernen, Lehren und Forschen in einer digital geprägten Welt – Tagungsband zur Jahrestagung der GDGP e.V.* (S. 458–461).
- Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM: Testing, Psychometrics, Methodology in Applied Psychology*, 26(1).
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2(3), 172–177.
- Holle, K. (2010). Psychologische Lesemodelle und ihre lesedidaktischen Implikationen. In *Texte lesen: Lesekompetenz - Textverstehen - Lesedidaktik - Lesesozialisation* (2. Aufl., S. 103–165). Schöningh.
- Holliday, W. G., & Cain, S. D. (2012). Teaching Science Reading Comprehension: A Realistic, Research-Based Approach. In B. J. Fraser, K. Tobin & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (S. 1405–1417). Springer Netherlands.
- Hopf, M., Schecker, H., Höttecke, D., & Wiesner, H. (2022). *Physikdidaktik kompakt* (1. vollständig neu bearbeitete Auflage). Aulis.

- Höttecke, D., Ehmke, T., Krieger, C., & Kulik, M. A. (2017). Vergleichende Messung fachsprachlicher Fähigkeiten in den Domänen Physik und Sport. *Zeitschrift für Didaktik der Naturwissenschaften*, 23(1), 53–69.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95.
- Johnson-Laird, P. N. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society of London.*, 295(1077), 353–374.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, 38(1), 23–31.
- Kattmann, U., Duit, R., Gropengießer, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18.
- Kehlmann, D. (2021). *Die Vermessung der Welt* (9. Aufl.). Rowohlt Taschenbuch Verlag.
- Kelava, A., & Moosbrugger, H. (2020). Deskriptivstatistische Itemanalyse und Testwertbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 143–157). Springer.
- Kernen, N., & Riss, M. (2012). *Textschwierigkeiten in Lehrmitteln für den naturwissenschaftlichen Unterricht in der Sekundarstufe I*. PH FHNW.
- Kintsch, W. (1998). *Comprehension*. Cambridge Univ. Press.
- Klafki, W. (1963). *Das pädagogische Problem des Elementaren und die Theorie der kategorialen Bildung*. Beltz.
- Kleinbub, I. (2018). Unterricht - Schlaglichter auf ein komplexes Forschungsfeld. In J. M. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik* (S. 11–26). Schneider.
- Klieme, E. (2006). Empirische Unterrichtsforschung - aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch*, 36(1), 15–43.
- Köller, O., Thiel, F., van Ackeren, I., Anders, Y., Becker-Mrotzek, M., Cress, U., Diehl, C., Kleickmann, T., Lütje-Klose, B., Prediger, S., Seeber, S., Ziegler, B., Kuper, H., Stanat, P., Maaz, K., & Lewalter, D. (2022). *Basale Kompetenzen vermitteln – Bildungschancen sichern. Perspektiven für die Grundschule. Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz* (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, Hrsg.). SWK.
- Köller, W. (1977). Der sprachtheoretische Wert des semiotischen Zeichenmodells. In *Zeichen, Text, Sinn: Zur Semiotik des literarischen Verstehens* (1. Aufl., S. 7–77). Vandenhoeck & Ruprecht.
- Krabbe, H., Rincke, K., & Aleksov, R. (2021). Language in Physics Instruction. In H. E. Fischer & R. Girwidz (Hrsg.), *Physics Education* (S. 361–382). Springer International Publishing.

- Krapp, A. (1993). Die Psychologie der Lernmotivation. Perspektiven der Forschung und Probleme ihrer pädagogischen Rezeption. *Zeitschrift für Pädagogik*, 39(2), 187–206.
- Krapp, A. (1999). Intrinsische Lernmotivation und Interesse. Forschungsansätze und konzeptuelle Überlegungen. *Zeitschrift für Pädagogik*, 45(3), 387–406.
- Kulgemeyer, C., & Schecker, H. (2009). Kommunikationskompetenz in der Physik: Zur Entwicklung eines domänenspezifischen Kommunikationsbegriffs. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 131–153.
- Kulgemeyer, C., & Wittwer, J. (2023). Misconceptions in Physics Explainer Videos and the Illusion of Understanding: An Experimental Study. *International Journal of Science and Mathematics Education*, 21(2), 417–437.
- Kultusministerkonferenz. (2019). Bildungssprachliche Kompetenzen in der deutschen Sprache stärken (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, Hrsg.).
- Kultusministerkonferenz. (2020). Bildungsstandards im Fach Physik für die allgemeine Hochschulreife - Beschluss der Kultusministerkonferenz vom 18.06.2020 (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, Hrsg.).
- Langer, I., Schulz von Thun, F., & Tausch, R. (2019). *Sich verständlich ausdrücken* (11. Aufl.). Ernst Reinhardt Verlag.
- Leisen, J. (2003). *Methoden-Handbuch. Deutschsprachiger Fachunterricht*. Varus.
- Leisen, J. (2005). Muss ich jetzt auch noch Sprache unterrichten?: Sprache und Physikunterricht. *Naturwissenschaften im Unterricht. Physik*, 16(87), 4–9.
- Leisen, J. (2022). *Sprachbildung und Sprachsensibler Fachunterricht in den Naturwissenschaften*. Kohlhammer.
- Lenhard, W. (2013). *Leseverständnis und Lesekompetenz* (1. Aufl.). Kohlhammer.
- Lexikon der Psychologie. (2021). *Dorsch - Lexikon der Psychologie* (20., überarbeitete Auflage). Hogrefe.
- Lilliefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399–402.
- Lutz, B. (2014). Von Hamburg über Karlsruhe nach Krems: Verständlichkeitsmodelle und deren praktische Anwendbarkeit. In *Tagungsband der tekomp-Frühjahrstagung* (S. 53–55).
- Lutz, B. (2015). *Verständlichkeitsforschung transdisziplinär: Plädoyer für eine anwenderfreundliche Wissensgesellschaft* (1. Aufl., Bd. 6). V&R Unipress.
- Mc Laughlin, G. H. (1969). SMOG Grading-a New Readability Formula. *Journal of Reading*, 12(8), 639–646.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning From Text. *Cognition and Instruction*, 14(1), 1–43.
- Merzyn, G. (1994). *Physikschulbücher, Physiklehrer und Physikunterricht* (Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel, Hrsg.). IPN.

- Möller, K. (2012). Konstruktion vs. Instruktion oder Konstruktion durch Instruktion? - Konstruktionsfördernde Unterstützungsmaßnahmen im Sachunterricht. In H. Giest, E. Heran-Dörr & C. Archie (Hrsg.), *Lernen und Lehren im Sachunterricht - Zum Verhältnis von Konstruktion und Instruktion* (S. 37–50). Julius Klinkhardt.
- Moosbrugger, H., & Brandt, H. (2020). Itemkonstruktion und Antwortverhalten. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3. Aufl., S. 67–89). Springer.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Aufl.). Springer.
- Moosbrugger, H., & Kelava, A. (Hrsg.). (2020). *Testtheorie und Fragebogenkonstruktion* (3. Aufl.). Springer.
- Morek, M., & Heller, V. (2012). Bildungssprache - Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *Zeitschrift für Angewandte Linguistik*, 2012(57), 67–101.
- Muckenfuß, H. (1995). *Lernen Im Sinnstiftenden Kontext* (1. Auflage). Cornelsen.
- Ogden, C. K., & Richards, I. A. (1930). *The Meaning of Meaning. A Study of the Influence of Language upon Thought and of The Science of Symbolism*. Kegan Paul, Trench, Trübner & Co. LTD.  
Deutsche Übersetzung der 10 Auflage von 1949. Original: The Meaning of Meaning.
- Patterson, A., Roman, D., Friend, M., Osborne, J., & Donovan, B. (2018). Reading for meaning: The foundational knowledge every teacher of science should have. *International Journal of Science Education*, 40(3), 291–307.
- Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and Science: Each in the Service of the Other. *Science*, 328(5977), 459–463.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Philipp, M., & Schilcher, A. (Hrsg.). (2012). *Selbstreguliertes Lesen - Ein Überblick über wirksame Leseförderansätze* (1. Aufl.). Kallmeyer in Verbindung mit Klett.
- Plath, J. (2020). Verstehensprozesse bei der Bearbeitung realitätsbezogener Mathematikaufgaben: Klassische Textaufgaben vs. Zeitungstexte. *Journal für Mathematik-Didaktik*, 41(2), 237–266.
- Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM*, 50(1), 159–171.
- Porst, R. (2014). *Fragebogen. Ein Arbeitsbuch* (4. Aufl.). Springer VS.
- Prediger, S. (2013). Darstellungen, Register und mentale Konstruktion von Bedeutungen und Beziehungen: Mathematikspezifische sprachliche Herausforderungen identifizieren und bearbeiten. In *Sprache im Fach: Sprachlichkeit und fachliches Lernen*. Waxmann.

- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014a). *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. (4. Aufl.). Springer.
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2014b). *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. (4. Aufl.). Springer.
- Rau, M. (2017). Conditions for the Effectiveness of Multiple Visual Representations in Enhancing STEM Learning. *Educational psychology review*, 29(4), 717–761.
- 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* (S. 635–655). Springer.
- Roelcke, T. D. (2020). *Fachsprachen* (4. Aufl.). Erich Schmidt Verlag.
- Rosebrock, C. (2020). *Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung* (9. Aufl.). Schneider Verlag Hohengehren GmbH.
- Rost, D. H. (2013). *Handbuch Intelligenz* (1. Aufl.). Beltz.
- Salzborn, S. (2016). John R. Searle: Speech Acts. In S. Salzborn (Hrsg.), *Klassiker der Sozialwissenschaften 100 Schlüsselwerke im Portrait* (2. Aufl., S. 268–271). Springer VS.
- Sander, J., & Rosebrock, C. (2016). Komplexität literarischer Texte als didaktische Kategorie. *Literatur im Unterricht*, 2, 143–157.
- Schecker, H., & Duit, R. (2018). Schülervorstellungen und Physiklernen. In *Schülervorstellungen und Physikunterricht: ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis* (S. 1–21). Springer Spektrum.
- Scheller, P. (2010). *Verständlichkeit im Physikschulbuch. Kriterien und Ergebnisse einer interdisziplinären Analyse* (Diss.). Universität Augsburg. Augsburg.
- Schilcher, A., & Wild, J. (2018). Lesen. Eine Schlüsselkompetenz im interdisziplinären Forschungsdiskurs. In J. M. Boelmann (Hrsg.), *Empirische Forschung in der Deutschdidaktik* (S. 163–184). Schneider.
- Schlepppegrell, M. J. (2019). Language and knowledge: How nouns contribute to knowledge construction across school subjects. *Research in Subject-matter Teaching and Learning (RISTAL)*, 2(1), 1–15.
- Schmellentin, C., Dittmar, M., Gilg, E., & Schneider, H. (2017). Sprachliche Anforderungen in Biologielehrmitteln. In B. Ahrenholz, B. Hövelbrinks & C. Schmellentin (Hrsg.), *Fachunterricht und Sprache in schulischen Lehr-/Lernprozessen* (S. 73–91). Narr Francke Attempto.
- Schneider, H., Gilg, E., Dittmar, M., & Schmellentin, C. (2019). Prinzipien der Verständlichkeit in Schulbüchern der Biologie auf der Sekundarstufe 1. In B. Ahrenholz, S. Jeuk, B. Lütke, J. Paetsch & H. Roll (Hrsg.), *Fachunterricht, Sprachbildung und Sprachkompetenzen* (S. 61–86). De Gruyter.
- Schneider, W., Schlagmüller, M., & Ennemoser, M. (2017). *LGVT 5-12+. Lesegeschwindigkeits- und -verständnistest für die Klassen 5-12*. (M. Hasselhorn, W. Schneider & U. Trautwein, Hrsg.; 2. Aufl.). Hogrefe.

- Schnotz, W. (1994). *Aufbau von Wissensstrukturen*. Beltz, Psychologie-Verl.-Union.
- Schulz-Gerlach, I. (2022). Bewertung von Mehrfachauswahlaufgaben.
- Schweller, K. G., Brewer, W. F., & Dahl, D. A. (1976). Memory for illocutionary forces and perlocutionary effects of utterances. *Journal of verbal learning and verbal behavior*, 15(3), 325–337.
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*.
- Shannon, C. L., & Weaver, W. (1964). *The mathematical theory of communication* (10. Aufl.). University of Illinois Press.
- Smolkowski, K. (2019). Gain Score Analysis. [https://homes.ori.org/keiths/Tips/Stats\\_Gain\\_Scores.html](https://homes.ori.org/keiths/Tips/Stats_Gain_Scores.html)  
[Zugriff: 24.11.2023].
- Solso, R. L. (2005). *Kognitive Psychologie* (6. Aufl.). Springer.
- Spitta, G. (Hrsg.). (2010). *Freies Schreiben - eigene Wege gehen* (3. Aufl.). Libelle-Verl.
- Staatsinstitut für Schulqualität und Bildungsforschung (ISB). (2022a). LehrplanPLUS Grundschule in Bayern - Onlinefassung.
- Staatsinstitut für Schulqualität und Bildungsforschung (ISB). (2022b). LehrplanPLUS Gymnasium in Bayern - Onlinefassung.
- Staatsinstitut für Schulqualität und Bildungsforschung (ISB). (2022c). LehrplanPLUS Realschule in Bayern - Onlinefassung.
- Starauschek, E. (2003). Ergebnisse einer Schülerbefragung über Physikschulbücher. *Zeitschrift für Didaktik der Naturwissenschaften*, 9, 135–146.
- Starauschek, E. (2006). Der Einfluss von Textkohäsion und gegenständlichen externen piktoralen Repräsentationen auf die Verständlichkeit von Texten zum Physiklernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 127–157.
- Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5. Aufl.). Routledge.
- Strobach, T. (2020). *Kognitive Psychologie* (M. Hasselhorn, W. Kunde & S. Schneider, Hrsg.; 1. Aufl.). Kohlhammer.
- Strohmaier, A. R., Ehmke, T., Härtig, H., & Leiss, D. (2023). On the role of linguistic features for comprehension and learning from STEM texts. A meta-analysis. *Educational Research Review*, 39, 100533.
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48, 1273–1296.
- Tergan, S. O. (1980). *Ist "Textverständlichkeit" gleich "Textverständlichkeit"? Überprüfung der Vergleichbarkeit zweier Verständlichkeitskonzepte* (Techn. Ber. Nr. 7). Deutsches Institut für Fernstudien an der Universität Tübingen. Tübingen.
- The pandas development team. (2020). Pandas-dev/pandas: Pandas.
- Urhahne, D., Dresel, M., & Fischer, F. (Hrsg.). (2019). *Psychologie für den Lehrberuf*. Springer.
- Vallat, R. (2018). Pingouin: Statistics in Python. *J. Open Source Software*, 3(31), 1026.

- Van Dijk, T. A. (1995). On Macrostructures, Mental Models, and Other Inventions: A Brief Personal History of the Kintsch-van Dijk Theory. In S. Mannes, C. R. Fletcher & C. A. Weaver III (Hrsg.), *Discourse Comprehension: Essays in Honor of Walter Kintsch* (1st Edition, S. 382–410). Routledge.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17, 261–272.
- 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). Narr Francke Attempto.
- von Goethe, J. W. (1808). *Faust - Der Tragödie Erster Teil* (durchgesehene Ausgabe aus dem Jahr 2000). Reclam.
- Wagenschein, M. (1970). *Ursprüngliches Verstehen und exaktes Denken I* (2. Aufl.). Klett.
- Wes McKinney. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & Jarrod Millman (Hrsg.), *Proceedings of the 9th Python in Science Conference* (S. 56–61).
- Wild, J., & Pissarek, M. (o.J.). Ratte. Regensburger Analysetool für Texte. <https://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/downloads/ratte/index.html> [Zugriff: 23.11.2023].
- Wilhelm, T., & Schecker, H. (2018). Strategien für den Umgang mit Schülervorstellungen. In *Schülervorstellungen und Physikunterricht: ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis* (S. 39–61). Springer Spektrum.
- Wygotski, L. S. (1979). *Denken Und Sprechen*. Fischer.

# **A | ZU DEN LESETEXTEN FÜR DIE EMPIRISCHE STUDIE**

## **A.1 Die Lesetexte**

Auf den folgenden Seiten ist exemplarisch die Version A des Lesetexts zu sehen, wie sie in der empirischen Studie verwendet wurde. Der Text und auch alle übrigen Versionen sind darüber hinaus im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> einsehbar.

## Lesetextversion A

### Von Gebirgsflüssen und Taschenlampen: Was ist elektrische Spannung?

Warum fließt ein Fluss in eine bestimmte Richtung?  
Ein Fluss fließt vom Gebirge ins Tal.  
Doch was hat das mit einer Taschenlampe zu tun?

#### Wovon handelt dieser Text?

Um die Gemeinsamkeiten von Gebirgsfluss und Taschenlampe zu verstehen, benötigt man den Begriff der **elektrischen Spannung**.

Dieser Text erklärt den Begriff und die Bedeutung der elektrischen Spannung. Dazu betrachtet man zunächst den Zusammenhang zum elektrischen Strom. Dann wird erklärt, was in der Physik genau unter der elektrischen Spannung verstanden wird. In der Physik hat man Worte und Zeichen für die elektrische Spannung. Diese werden am Ende des Texts erklärt.

#### Wozu dient die elektrische Spannung?

Die elektrische Spannung treibt den elektrischen Strom an. Folgender Vergleich macht dies klar: Wasser strömt vom Gebirge ins Tal. Es strömt also von der höheren Lage in eine tiefere Lage. Der Höhenunterschied bringt das Wasser zum Strömen. Somit wirkt der Höhenunterschied als Antrieb für den „Wasserstrom“.

Noch deutlicher wird das am Beispiel einer Taschenlampe. Deren Funktionsweise kann man sich ganz ähnlich vorstellen. Damit die Taschenlampe leuchtet, muss elektrische Ladung durch die Drähte in der Lampe fließen. Dazu ist eine Batterie nötig. Sie liefert die sogenannte elektrische Spannung. Strömende elektrische Ladung ist dabei elektrischer Strom. Und der braucht wie der Wasserstrom einen Antrieb. Beim Wasser treibt der Höhenunterschied den Wasserstrom an. Genauso ist die elektrische Spannung der **Antrieb des elektrischen Stroms**.

#### Was ist elektrische Spannung?

Eine Folge der elektrischen Spannung ist der elektrische Strom. Doch was ist elektrische Spannung genau?

Dazu erinnert man sich an den Wasserstrom und an

die Taschenlampe. Das Wasser fließt von einem höher liegenden Ort zu einem tiefer liegenden. Genauso strömt die elektrische Ladung von einem „*elektrisch höher*“ liegenden Ort zu einem „*elektrisch niedriger*“ liegenden. Den **Unterschied hinsichtlich der elektrischen Lage** nennt man in der Physik elektrische Spannung. Die Batterie in der Taschenlampe hat einen elektrisch hohen Anschluss und einen elektrisch niedrigen Anschluss. Dadurch liefert sie elektrische Spannung. Wenn man die beiden Anschlüsse elektrisch verbindet, dann verursacht die elektrische Spannung einen elektrischen Strom.

Man nennt den elektrisch hohen Anschluss den **Pluspol**. Der elektrisch niedrige Anschluss ist der **Minuspol** der Batterie.

#### Mit welchen Zeichen beschreibt man die elektrische Spannung?

Den Unterschied zwischen den Polen der Batterie kann man in Zahlen angeben. Dazu braucht man eine Einheit. Auf der Batterie finden sich Angaben wie 1,5 V oder 6 V. Das **V** steht für die Einheit **Volt**. Somit misst man die elektrische Spannung in Volt. Das Zeichen **U** bezeichnet in Formeln und Rechnungen die elektrische Spannung. Man schreibt zum Beispiel:  $U = 1,5 \text{ V}$ .

#### Wie spricht man über die elektrische Spannung?

Die elektrische Spannung ist der elektrische Lageunterschied der beiden Pole. Man misst die elektrische Spannung im Stromkreis immer zwischen *zwei* Punkten. Die elektrische Spannung herrscht nicht an *einem* Punkt. Ein Gegenstand *hat* keine elektrische Spannung. Deshalb sagt man: Die elektrische Spannung **liegt zwischen zwei Punkten an**.

#### Gibt es elektrische Spannung ohne elektrischen Strom?

Zwei Punkte können im Stromkreis elektrisch unterschiedlich hoch liegen. Das steht erst einmal nicht

in Verbindung mit elektrischem Strom. Wenn man die Batterie aus der Taschenlampe nimmt, dann liegen die beiden Pole immer noch elektrisch unterschiedlich hoch. Genauso wie der Berg ohne strömendes Wasser höher liegt als das Tal. Auch ohne elektrischen Strom bleibt die elektrische Spannung bestehen.

### Was muss man sich merken?

Wichtig zu wissen ist:

- Eine Batterie liefert eine elektrische Spannung. Zwischen ihren beiden Polen liegt eine elektrische Spannung an.
- Die elektrische Spannung vergleicht man mit dem Höhenunterschied beim Wasserstrom.
- Die elektrische Spannung kann elektrische Ladungen zum Strömen bringen. Die elektrische Spannung ist der Antrieb des elektrischen Stroms.
- Die elektrische Spannung ist der elektrische Lageunterschied zweier Orte. Man misst sie zwischen zwei Punkten.

- Man bezeichnet die elektrische Spannung mit dem Formelzeichen  $U$ .
- Man misst die elektrische Spannung in Volt. Man kürzt Volt mit dem Buchstaben V ab. Deshalb schreibt man zum Beispiel „1,5 V“ und spricht es als „1,5 Volt“ aus.

### Wie geht es weiter?

Die Gemeinsamkeiten eines Gebirgsflusses und einer Taschenlampe sind: Das fließende Wasser entspricht den fließenden Ladungen. Der Höhenunterschied zwischen Berg und Tal entspricht dem Unterschied in der elektrischen Lage. Dadurch gewinnt man eine bildliche Vorstellung von der elektrischen Spannung. Der Wasserstrom vom Berg ins Tal dient als **Modell** für einen elektrischen Strom zwischen zwei Punkten im Stromkreis. Das Modell hilft oftmals Vorgänge in Stromkreisen zu verstehen, zu erklären und vorherzusagen. Das Modell reicht nicht für alle richtigen Erklärungen oder Vorhersagen. Dafür gibt es noch weitere Modelle zur elektrischen Spannung.

## A.2 Zusatzmaterial zu den Lesetexten

### A.2.1 Outputs des Textanalyseprogramms RATTE

Im Fließtext wurde auf die Textanalyse mit dem Programm RATTE verwiesen (Seiten 90 und 100). Diese textstatistischen Analyseergebnisse sind im digitalen Anhang unter <http://doi.org/10.5283/epub.58778> einsehbar.

### A.2.2 Liste mit allen variierten Textstellen

Auf den folgenden Seiten sind alle variierten Textstellen zwischen den Versionen A bis H der Texte kompakt zusammengestellt.

#### Wortwahl / morphologische Auffälligkeiten auf Wortebene

Die folgende Tabelle A.1 listet Textstellen auf, die bezüglich der **Wortwahl-Variable** verändert wurden. Die Stellen sind nach ihrem Auftreten in den Lesetexten geordnet. Die Veränderungen wurden vorgenommen, indem die Wortwahl hinsichtlich der Fachsprachlichkeit, Mehrdeutigkeit, Geläufigkeit und Wortlänge variiert wurden, indem Fremdwörter verwendet oder vermieden wurden und indem einzelne Begriffe entweder durch sich wiederholende Wörter oder durch mehrere Synonyme im Text repräsentiert wurden.

Wortwahl / morphologische Auffälligkeiten auf Wortebene w	W
fließen, strömen	fließen, strömen, laufen, rauschen, sich bewegen
Gemeinsamkeiten	Affinitäten
verstehen	sich klarwerden
Begriff	Bezeichnung
genau	genau, exakt, explizit
Vergleich	Analogie
Unterschied	Differenz
Lage, Höhe, Ort	Lage, Höhe, Ort, Niveau, Punkt
ähnlich	vergleichbar
leuchten	Licht emittieren
Drähte in der Lampe	Lampendrähte
nötig	vonnöten
genauso	gleichermaßen, ebenso
Folge	Konsequenz
sich erinnern	sich entsinnen
Anschluss	Kontakt
verbinden	(miteinander) koppeln
verursachen	provozieren
nennen	betiteln
in Zahlen	quantitativ
brauchen	bedürfen
stehen für	kennzeichnen

w	W
messen	Messwerte erfassen, Messwerterfassung erfolgt
Zeichen	Symbol
in der Form ... schreiben	die Notation ... verwenden
Gegenstand	Objekt
liegen	positioniert sein
nehmen	entfernen
elektrisch unterschiedlich hoch liegen	eine elektrische Differenz aufweisen
bestehen	fortbestehen
wichtig	relevant
mit ... vergleichen	Parallelen zwischen ... ziehen
zum ... bringen	zum ... veranlassen
sprechen	artikulieren
entsprechen	entsprechen, gleichen
bildlich	sinnbildlich
dienen	fungieren
Modell	Funktionsmodell
helfen	Hilfe leisten
auf Erklärungen kommen	Erklärungen generieren
immer	ausschließlich
Vorhersage	Prognose
es gibt	es existiert/existieren

Tabelle A.1.: Sprachliche Variationen der Wortwahl-Variablen geordnet nach ihrer Position im Text. Auf die Wörter in der anspruchsteigernden Variante (W-Spalte) trifft im Vergleich zu den Wörtern in der w-Spalte einer oder mehrere der folgenden Punkte zu: die Wörter sind fachsprachlicher, mehrdeutig, Fremdwörter, weniger geläufig und/oder länger (umfassen mehr Silben). Darüber hinaus wurden für die W-Texte im Vergleich zu den w-Texte stellenweise mehrere Synonyme für einen Begriff anstelle von Wortwiederholungen gebraucht.

### **Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene**

Nahezu alle Sätze in der anspruchsteigernden Variante (S-Texte) sind Satzgefüge, die zwei oder mehrere Einzelsätze der anspruchsmindernden Variante (s-Sätze) verbinden. Dadurch sind die Sätze der S-Texte strukturell komplexer und länger (umfassen mehr Wörter). Dies kann hier nicht sinnvoll aufgelistet werden. Der folgende Vergleich einer Textstelle in der S- gegenüber der s-Variante macht die Anwendung von Satzgefügen daher exemplarisch deutlich:

A. Zu den Lesetexten für die empirische Studie

Damit die Taschenlampe leuchtet, muss elektrische Ladung durch die Drähte in der Lampe fließen. Dazu ist eine Batterie nötig. Sie liefert die sogenannte elektrische Spannung. Strömende elektrische Ladung ist dabei elektrischer Strom. Und der braucht wie der Wasserstrom einen Antrieb.

Aus Text A (*wsp*)

Damit die Taschenlampe leuchtet, ist durch die Drähte in der Lampe strömende elektrische Ladung nötig. Zum Ingangsetzen des Ladungsstroms in der Taschenlampe wird eine Batterie benötigt, durch welche die sogenannte elektrische Spannung zur Verfügung gestellt wird. Strömende elektrische Ladung ist dabei elektrischer Strom, welcher genauso wie der Wasserstrom einen Antrieb braucht.

Aus Text C (*wSp*)

Der folgenden Tabelle A.2 sind nun solche Sätze und Satzteile zu entnehmen, bei denen weitere sprachliche Variationen der Satzbau-Kategorie zum Tragen kommen (Passivkonstruktionen, Verbalklammern, Mehrwortkomplexe, Partizipialkonstruktionen, Regressionen, Nominalisierungen). Es sind aussagekräftige Fragmente aus den Sätzen in den Lesetexten dargestellt, um die Unterschiede in den Satzstrukturen möglichst deutlich zu machen und um die Übersichtlichkeit zu wahren.

Satzstrukturen / morphologische Auffälligkeiten auf Phrasen- und Satzebene s	S
Um ... zu verstehen	Um sich über ... klarzuwerden
Man benötigt ....	Es wird ... benötigt.
Dieser Text erklärt, was ...	In diesem Text wird erläutert, was ...
In der Physik hat man Wörter und Zeichen ... Diese werden am Ende des Texts erklärt.	Am Ende des Texts wird außerdem erklärt, mit welchen Worten sowie Zeichen in der Physik über die elektrische Spannung gesprochen wird.
Die elektrische Spannung treibt ... an.	Das Antreiben von ... schreibt man der elektrischen Spannung zu.
... macht dies klar.	Man kann sich dies mit ... klarmachen
Wasser strömt vom Gebirge ins Tal. Es strömt also von der höheren Lage in eine tiefere.	Vom Gebirge ins Tal, also von der höheren Lage in eine tiefere Lage strömendes Wasser.
Elektrische Ladung muss durch die Drähte in der Lampe strömen	Durch die Drähte in der Lampe strömende elektrische Ladung ist nötig.
Elektrische Ladung muss ... strömen. Dazu ist eine Batterie nötig.	Zum Ingangsetzen des Ladungsstroms in der Taschenlampe wird eine Batterie benötigt.
Dazu erinnert man sich an ....	Dazu kann man sich an ... erinnern.
Die Batterie hat einen elektrisch hohen Anschluss und einen .... Man nennt den elektrisch hohen Anschluss den Pluspol.	Die Batterie in der Taschenlampe hat einen Pluspol genannten elektrisch hohen Anschluss und einen ....
Man braucht eine Einheit.	Es wird eine Einheit gebraucht.

s	S
Man misst die elektrische Spannung in Volt.	Das Messen der elektrischen Spannung geschieht in Volt.
Das Zeichen U bezeichnet ....	... wird mit dem Zeichen U bezeichnet.
Man schreibt das in der Form ...:	Es wird in der folgenden Form geschrieben: ...
Die elektrische Spannung liegt zwischen zwei Punkten an.	Zwischen zwei Punkten anliegende elektrische Spannung.
Zwei Punkte liegen unterschiedlich hoch.	Zwei unterschiedlich hoch liegende Punkte.
Wenn man die Batterie aus der Taschenlampe nimmt, ...	Beim Herausnehmen der Batterie aus der Taschenlampe ....
Die elektrische Spannung bleibt bestehen.	Das Bestehenbleiben der elektrischen Spannung ist gegeben.
Eine Batterie liefert eine elektrische Spannung.	Eine Batterie dient dazu, elektrische Spannung bereitzustellen.
Die elektrische Spannung vergleicht man mit ....	Die elektrische Spannung kann man mit ... vergleichen.
Folgende Gemeinsamkeiten von Gebirgsfluss und Taschenlampe kennen wir nun:	In diesem unter anderem den Vergleich von Gebirgsfluss und Taschenlampe betreffenden Text, wurden folgende Gemeinsamkeiten enthüllt:
Der Wasserstrom vom Berg ins Tal dient als Modell.	Das Bild des vom Berg ins Tal strömenden Wasserstroms dient als Modell.

Tabelle A.2.: Sprachliche Variationen der Satzbau-Variablen geordnet nach ihrer Position im Text. Die anspruchsteigende Variante (S) zeichnet sich gegenüber der anspruchsmindernden Variante (s) jeweils durch eines oder mehrere der folgenden Merkmale aus: Die Sätze sind in passiver Form gehalten, enthalten Verbalklammern, Mehrwortkomplexe und/oder Partizipialkonstruktionen, sind regressiv formuliert und/oder es sind Nominalisierungen enthalten.

### Erzählende und personalisierende Elemente

Wo immer möglich, sind in den erzählenden und personalisierten Texten anstelle von unpersönlichen Passiv-, Nominal-, Reflexiv- und man-Formulierungen solche Formen zu finden, welche die Lesenden persönlich ansprechen (du-/wir-Formen) und darüber hinaus den Handlungsträger klar benennen (aktivischer Stil anstelle von Deagentivierung, siehe Kapitel 6.2.2). Weitere Unterschiede zwischen P- und p-Texten betreffen die Verwendung von werbenden Attributen, von humorvollen Elementen, von bedeutungsleeren Füllwörtern (oder andersherum: die Dichte bedeutungstragender Einheiten) und die Verwendung solcher Wörter und Phrasen, die Stimmungen wiedergeben. Die folgende Tabelle A.3 stellt die veränderten Sätze bzw. Satzteile – geordnet nach ihrem Auftreten im Lesetext – zum Vergleich gegenüber.

<b>Erzählende und personalisierende Elemente</b>	<b>P</b>
Warum fließt ein Fluss in eine bestimmte Richtung?	Warum fließt ein Fluss eigentlich in eine bestimmte Richtung?
Ein Fluss fließt vom Gebirge ins Tal.	Ein Fluss fließt nämlich vom Gebirge ins Tal.
Um die Gemeinsamkeiten von ... zu verstehen, benötigt man den Begriff der elektrischen Spannung. [X]	Wir wollen einige überraschende Gemeinsamkeiten von ... verstehen. Dazu benötigen wir den Begriff der elektrischen Spannung. Ohne eine elektrische Spannung würde deine Taschenlampe nicht leuchten.
Dieser Text erklärt ....	Dieser Text erklärt dir ....
Dazu betrachtet man ....	Dazu betrachten wir ....
was unter ... verstanden wird	was du dir unter ... vorstellen kannst
In der Physik hat man ....	In der Physik haben wir ....
Diese werden erklärt.	Diese werden dir erklärt.
Die elektrische Spannung treibt den elektrischen Strom an. Folgender Vergleich macht dies klar.	Stell dir die elektrische Spannung als „Tritt in den Hintern“ für den elektrischen Strom vor. Ein überraschender Vergleich erklärt diese Vorstellung.
Somit wirkt ... als ....	Somit kannst du ... als ... ansehen.
Noch deutlicher wird das am Beispiel ...	Noch schöner zeigt uns das ...
kann man sich ganz ähnlich vorstellen	kannst du dir ganz ähnlich vorstellen.
Damit die Taschenlampe leuchtet, ...	Damit unsere Taschenlampe leuchtet, ...
Dazu ist ... nötig.	Dazu benötigen wir ....
... braucht einen Antrieb. [X]	... braucht einen Antrieb. So wie für dich das Frühstück der Antrieb für deinen Schultag ist.
Die elektrische Spannung ist der Antrieb .... [X]	Die elektrische Spannung ist der Antrieb .... Und diesen Job macht sie richtig gut!
Eine Folge ist ....	Du kennst nun eine Folge ....
Dazu erinnert man sich an ....	Dazu erinnere dich an ....
nennt man in der Physik	nennen wir in der Physik
[X] Die Batterie hat einen elektrisch hohen Anschluss ...	Hier kommt die Batterie in der Taschenlampe ins Spiel: Sie hat einen elektrisch hohen Anschluss ....
Wenn man ... verbindet, ...	Wenn wir ... verbinden, ...
Man nennt ... den Pluspol. Der ... ist der Minuspol der Batterie. [X]	Wir nennen ... den Pluspol. Der ... ist der Minuspol. Und auch wenn es hier um Pole geht: Am Plus- oder Minuspol findest du sicherlich keine Eisbären oder Pinguine.
Mit welchen Zeichen beschreibt man ...?	Mit welchen Zeichen beschreiben wir ...?
kann man in Zahlen angeben	kannst du in Zahlen angeben
Dazu braucht man eine Einheit.	Dazu brauchst du eine Einheit.

<b>P</b>	<b>P</b>
Auf der Batterie finden sich Angaben wie ....	Auf einer Batterie findest du Angaben wie ....
Somit misst man ....	Somit messen wir ....
Man schreibt ....	Wir schreiben ....
Wie spricht man über ...?	Wie sprechen wir über ...?
... elektrische Lageunterschied der beiden Pole. [X]. Man misst ....	... der elektrische Lageunterschied der beiden Pole. Wir erinnern uns: keine Eisbären! Wir messen...
Deshalb sagt man:	Deshalb sagen wir:
Das steht erst einmal nicht in Verbindung mit elektrischem Strom.	Das hat erst einmal nichts mit elektrischem Strom zu tun.
Wenn man die Batterie aus der Taschenlampe nimmt, dann ...	Wenn du die Batterie aus der Taschenlampe nimmst, dann ...
Was muss man sich merken?	Was musst du dir merken?
Wichtig zu wissen ist:	Für dich ist wichtig:
Eine Batterie liefert eine elektrische Spannung.	Eine Batterie liefert unsere elektrische Spannung.
Die elektrische Spannung vergleicht man ...	Die elektrische Spannung vergleichen wir ...
Man misst sie zwischen zwei Punkten.	Wir messen sie zwischen zwei Punkten.
Man bezeichnet ....	Wir bezeichnen ....
Man misst ....	Wir messen ....
Man kürzt Volt mit ... ab.	Du kürzt Volt mit ... ab.
Deshalb schreibt man zum Beispiel ....	Deshalb schreibst du dann zum Beispiel ....
Die Gemeinsamkeiten eines Gebirgsflusses und einer Taschenlampe sind:	Folgende überraschende Gemeinsamkeiten eines Gebirgsflusses und einer Taschenlampe kennen wir nun:
Dadurch gewinnt man eine bildliche Vorstellung.	Dadurch gewinnst du eine bildliche Vorstellung.
Der Wasserstrom dient als Modell.	Der Wasserstrom dient uns als Modell.
Das Modell hilft oftmals ....	Das Modell hilft dir oftmals ....
Das Modell reicht nicht für ....	Das Modell reicht uns nicht für ....
Dafür gibt es noch weitere Modelle zur elektrischen Spannung. [X]	Dafür gibt es noch weitere „spannende“ Spannungsmodelle. Diese wirst du in Zukunft noch kennenlernen.

Tabelle A.3.: Sprachliche Variationen der Personalisierungs-Variablen geordnet nach ihrer Position im Text. Die P-Texte mit stark ausgeprägter Personalisierung / mit vielen erzählenden Elementen sind gegenüber den p-Texten durch häufige direkte Anrede, durch die Verwendung werbender Attribute und humorvoller Elemente, durch eine geringere Dichte an bedeutungstragenden Einheiten (erreichbar durch das Einfügen bedeutungsleerer Füllwörter und -phrasen) und durch die Wiedergabe von Stimmungen ausgezeichnet.



## **B | ZU DEN TESTINSTRUMENTEN DER EMPIRISCHEN STUDIE**

### **B.1 Zur Messung der Behaltensleistung: der Wissenstest**

Der Wissenstest, der in der Studie sowohl als Prä- als auch als Posttest diente, ist in seiner Version als Vorwissenstest auf den folgenden Seiten einzusehen. Die korrekten Lösungen sind mit einem Haken markiert. Zwischen dem ersten Durchführungszeitpunkt, an dem unter anderem das Vorwissen anhand dieses Tests erhoben wurde, und dem zweiten Zeitpunkt, an dem der Test als Behaltensleistungstest zum Einsatz kam, lagen mindestens sieben Tage.

## Vorwissen über die elektrische Spannung

---

### Was weißt du schon über die elektrische Spannung?

  Kreuze bitte in jeder Aufgabe **alle** Aussagen an, die du für richtig hältst.

#### Aufgabe 1

In den folgenden Aussagen geht es um den Vergleich zwischen einem Stromkreis und einem Gebirgsfluss.

- Die elektrische Spannung ist vergleichbar mit der Menge an Wasser, die pro Zeit strömt.
- Punkte im Stromkreis, die elektrisch hoch liegen, sind vergleichbar mit Orten, die hoch am Berg liegen.
- Die elektrische Spannung ist vergleichbar mit der Geschwindigkeit, mit der das Wasser strömt.
- Die elektrische Lage eines Punkts im Stromkreis ist vergleichbar mit dem Gefälle eines Orts am Fluss.
- Die elektrische Spannung ist vergleichbar mit dem Höhenunterschied von Berg und Tal.

#### Aufgabe 2

In den folgenden Aussagen geht es um einen (einfachen) Stromkreis mit geladener Batterie.

- Die Batterie stellt elektrische Spannung bereit.
- Die Batterie nutzt elektrischen Strom, um Spannung zu erzeugen.
- Der elektrische Strom im Stromkreis verursacht elektrische Spannung.
- Die Batterie wandelt elektrische Spannung in elektrischen Strom um.
- In der Batterie ist elektrischer Strom gespeichert.

#### Aufgabe 3

In den folgenden Aussagen geht es um die elektrische Spannung als physikalische Größe.

- Wir können die elektrische Spannung als Antrieb des elektrischen Stroms verstehen.
- Elektrische Spannung kann elektrischen Strom verursachen.
- Nur wenn es elektrischen Strom gibt, kann es auch eine elektrische Spannung geben.
- Eine elektrische Spannung kann Ladungsträger zum Strömen bringen.
- In einem einfachen Stromkreis (Batterie, Draht, ein elektrisches Bauteil) gibt es keinen elektrischen Strom ohne elektrische Spannung.

#### Aufgabe 4

In den folgenden Aussagen geht es um einen (einfachen) Stromkreis mit Batterie.

- Wenn ein Punkt im Stromkreis elektrisch höher liegt als ein anderer, können wir zwischen den beiden Punkten eine elektrische Spannung messen.
- Jeder Punkt im Stromkreis hat eine elektrische Spannung.
- Jedem Punkt im Stromkreis können wir eine elektrische Lage zuschreiben.
- Zwischen zwei Punkten mit gleicher elektrischer Lage messen wir eine elektrische Spannung  $U = 0$  V.
- Wir müssen die elektrische Spannung immer entlang ihrer Flussrichtung (von elektrisch hoch nach elektrisch tief) messen.

### Aufgabe 5

In den folgenden Aussagen geht es um einen (einfachen) Stromkreis mit geladener Batterie.

- Zwischen den beiden Polen der Batterie liegt elektrische Spannung an.
- Der Pluspol liegt elektrisch höher als der Minuspol.
- Die elektrische Spannung fließt im Stromkreis von einem Pol der Batterie zum anderen.
- Die elektrische Spannung liegt am Pluspol der Batterie an.
- Die elektrische Spannung liegt am Minuspol der Batterie an.

### Aufgabe 6

In den folgenden Aussagen geht es um die elektrische Spannung als physikalische Größe.

- Die elektrische Spannung liegt zwischen zwei Punkten im Stromkreis an.
- Ein Draht im Stromkreis transportiert elektrische Spannung.
- Die elektrische Spannung fließt im Stromkreis vom Pluspol zum Minuspol.
- Zwischen den beiden Anschlüssen einer Batterie liegt Spannung an.
- Die elektrische Spannung ist vor einer Glühlampe höher als dahinter.



In den letzten beiden Aufgaben ist **genau eine** Aussage richtig.  
Kreuze bitte diejenige Aussage an, die du für richtig hältst.

### Aufgabe 7

In den folgenden Aussagen geht es um die Einheit der elektrischen Spannung.

- Die Einheit der elektrischen Spannung ist „C“ für „Coulomb“.
- Die Einheit der elektrischen Spannung ist „A“ für „Ampère“.
- Die Einheit der elektrischen Spannung ist „V“ für „Volt“.
- Die Einheit der elektrischen Spannung ist „Ω“ für „Ohm“.
- Da die elektrische Spannung nur einen Unterschied zweier Punkte beschreibt, hat sie keine Einheit.

### Aufgabe 8

In den folgenden Aussagen geht es um die mathematische Darstellung der elektrischen Spannung.

- Das Formelzeichen der elektrischen Spannung ist ein „S“.
- Das Formelzeichen der elektrischen Spannung ist ein „I“.
- Das Formelzeichen der elektrischen Spannung ist ein „Ω“.
- Das Formelzeichen der elektrischen Spannung ist ein „U“.
- Das Formelzeichen der elektrischen Spannung ist ein „V“.

## B.2 Zur Messung der Textwahrnehmung: der Bewertungsfragebogen

In der Hauptstudie kam der Bewertungsfragebogen in vier Versionen mit unterschiedlichen Itemreihenfolgen zum Einsatz. Der Fragebogen wurde unmittelbar nach der Textlektüre bearbeitet. In dieser Arbeit beziehen sich alle Angaben auf die hier abgebildete Nummerierung.

**Welche Textversion hast du gelesen?**

- Version A       Version B       Version C       Version D  
 Version E       Version F       Version G       Version H

**Welchen Gesamteindruck vom Text hattest du?**

	Stimme voll zu	Stimme eher zu	Habe keine Tendenz	Stimme eher nicht zu	Stimme gar nicht zu
Ich fand den Text insgesamt verständlich.	<input type="checkbox"/>				

**Und wie war dein Eindruck zu den folgenden Details?**

	Stimme voll zu	Stimme eher zu	Habe keine Tendenz	Stimme eher nicht zu	Stimme gar nicht zu
1.) Ich fand den Text humorvoll gestaltet.	<input type="checkbox"/>				
2.) Ich fand die Sätze im Text kompliziert.	<input type="checkbox"/>				
3.) Der Text hat mich emotional angesprochen.	<input type="checkbox"/>				
4.) Ich fand die Reihenfolge der einzelnen Textabschnitte sinnvoll.	<input type="checkbox"/>				
5.) Es hat mir Freude gemacht, den Text zu lesen.	<input type="checkbox"/>				
6.) Der Text hat mich zum Nachdenken angeregt.	<input type="checkbox"/>				
7.) Ich war fasziniert von den Textinhalten.	<input type="checkbox"/>				
8.) Ich fand den Text unterhaltsam.	<input type="checkbox"/>				
9.) Ich fand den Aufbau der einzelnen Sätze verwirrend.	<input type="checkbox"/>				
10.) Ich konnte dem Satzbau beim Lesen problemlos vom Satzanfang bis zum Satzende folgen.	<input type="checkbox"/>				
11.) Ich fand die Gliederung des Texts typisch für Physiktex-te.	<input type="checkbox"/>				
12.) Ich fand die Informationen im Text wirr.	<input type="checkbox"/>				
13.) Ich konnte gut nachvollziehen, wie die Textinhalte aufeinander aufbauen.	<input type="checkbox"/>				
14.) Der Text hat mich an mein Vorwissen zum Thema Elektrizitätslehre erinnert.	<input type="checkbox"/>				

*Bitte umblättern*

B.2. Zur Messung der Textwahrnehmung: der Bewertungsfragebogen

	Stimme voll zu	Stimme eher zu	Habe keine Tendenz	Stimme eher nicht zu	Stimme gar nicht zu
15.) Ich wusste am Satzende nicht mehr, wie der Satz begonnen hatte.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16.) Der Text hat Interesse für das Thema Elektrische Spannung bei mir geweckt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17.) Ich fand die Einteilung des Texts in seine Abschnitte passend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18.) Der Text erschien mir sehr voll an Informationen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19.) Ich hätte bei schwierigen Aspekten noch eine Erklärung aus einer anderen Sicht gebraucht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20.) Für mich war beim Lesen eines Satzes stets klar, wie er sich auf den vorherigen Satz bezieht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21.) Ich musste mit den Augen zwischen einzelnen Textstellen vor und zurück hüpfen anstatt den Text flüssig zu lesen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22.) Ich fand den Text auf den ersten Blick übersichtlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23.) Ich musste meine volle Aufmerksamkeit auf den Text richten, um ihm Information zu entnehmen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24.) Mir erschienen aufeinanderfolgende Sätze inhaltlich logisch miteinander verknüpft.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25.) Der Text rief gedankliche Konflikte bei mir hervor.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26.) Es gab Stellen im Text, die überraschend für mich waren.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27.) Um den Text zu verstehen, hätte ich mehr Erklärung zu den einzelnen Aspekten der elektrischen Spannung gebraucht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28.) Mir war klar, warum es sich lohnt, den Text zu lesen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29.) Für mich war beim Lesen eines Textabschnitts stets klar, wie er sich auf die übrigen Absätze bezieht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30.) Ich konnte mein Wissen über Elektrizitätslehre um neue Informationen aus dem Text erweitern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31.) Beim Lesen brachten mich einzelne Textstellen zum Stocken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32.) Ich konnte mir die Informationen aus dem Text merken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33.) Ich habe mein Wissen, das nichts mit Elektrizitätslehre zu tun hat, eingesetzt um den Text zu verstehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34.) Schwierige Begriffe im Text haben mir Verständnisprobleme bereitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35.) Ich konnte im Text durchgehend einen roten Faden erkennen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Bitte umblättern

B. Zu den Testinstrumenten der empirischen Studie

	Stimme voll zu	Stimme eher zu	Habe keine Tendenz	Stimme eher nicht zu	Stimme gar nicht zu
36.) Die einzelnen Abschnitte setzen sich für mich logisch zu einem Gesamtbild der elektrischen Spannung zusammen.	<input type="checkbox"/>				
37.) Mir war in jedem Textabschnitt klar, was das Thema dieses Abschnitts ist.	<input type="checkbox"/>				
38.) Ich hätte mehr Erklärung zu den Fachbegriffen im Text gebraucht.	<input type="checkbox"/>				
39.) Mir war klar, inwiefern der Text mir Erfahrungen aus meiner Lebenswelt erklärt.	<input type="checkbox"/>				
40.) Ich habe mich beim Lesen persönlich angesprochen gefühlt.	<input type="checkbox"/>				
41.) Ich musste den Text in einem langsamen Tempo lesen.	<input type="checkbox"/>				
42.) Die Informationen im Text knüpften an mein bisheriges Wissen an.	<input type="checkbox"/>				
43.) Mein Vorwissen über Elektrizitätslehre reichte aus, um den Text zu verstehen.	<input type="checkbox"/>				
44.) Mir war klar, wie die Wörter im Satz miteinander in Beziehung stehen.	<input type="checkbox"/>				

# C | ZUM KONSTRUKTRAUM FÜR DIE SPRACHLICHEN GESTALTUNGSMITTEL

Auf den folgenden Seiten werden die acht Gruppen für die sprachlichen Gestaltungsmittel (siehe Kapitel 5.1) mit allen zugeordneten, konkreten Mitteln aufgelistet. In Klammern ist jeweils angegeben, in welchen Quellen das sprachliche Mittel erwähnt wird. Dabei wurden die folgenden Abkürzungen verwendet:

- Artelt et al., 2005; Abkürzung im Folgenden: *Art05*.
- Fang, 2006; Abkürzung im Folgenden: *Fan06*.
- Frickel, 2018; Abkürzung im Folgenden: *Fri18*.
- Göpferich, 2002; Abkürzung im Folgenden: *Goe02*
- Heine et al., 2018; Abkürzung im Folgenden: *Hei18*.
- Höttecke, Ehmke, Krieger und Kulik, 2017; Abkürzung im Folgenden: *Hoe17*.
- Kernen und Riss, 2012; Abkürzung im Folgenden: *Ker12*.
- Lenhard, 2013; Abkürzung im Folgenden: *Len13*.
- Lutz, 2015; Abkürzung im Folgenden: *Lut15*.
- Starauschek, 2006; Abkürzung im Folgenden: *Sta06*.
- Patterson, Roman, Friend, Osborne und Donovan, 2018; Abkürzung im Folgenden: *Pat18*.
- Rincke, 2010; Abkürzung im Folgenden: *Rin10*.

## **Wortwahl / morphologische Auffälligkeiten auf Wortebene**

### **Einfach-Zuordnungen**

- Fremdwörter  
(Hei18, S. 72 | Goe02, S. 177)
- Wortlänge (= Silbenzahl pro Wort)  
(Sta06, S. 128 | Ker12, S. 5 | Hei18, S. 77)
- Zusammengesetzte Wörter (Komposita, Präfix-/Suffixkonstruktionen, ...)  
(Art05, S. 18 | Rin10, S. 238 | Ker12, S.10 | Hei18, S. 72 u. S. 87 | Hoe17, S. 55)

### C. Zum Konstrukttraum für die sprachlichen Gestaltungsmittel

- Geläufigkeit (= Häufigkeit) der Wörter; Anteil nicht bekannter Wörter  
(Art05, S. 15, 18 | Len13, S. 31 | Ker12, S. 4 | Hei18, S. 79, 82)
- Fachlicher Wortschatz (= Fachbegriffe)  
(Sta06, S. 141 | Hei18, S. 86, 87 | Hoe17, S. 55 | Lut15, S. 241 | Goe02, S. 157, 176)
- Technischer, akademischer Wortschatz  
(Pat18, S. 296 f. | Fan06, S. 494, 506)
- fachspezifische Akronyme  
(Rin10, S. 238)
- Alltagswortschatz mit fachspezifischer Bedeutung  
(Pat18, S. 296 | Ker12, S. 11 | Hoe17, S. 59 | Fan06, S. 494)
- Mehrdeutigkeiten (allgemein, insbes. grammatikalisch)  
(Ker12, S. 11 | Hoe17, S. 55 | Goe02, S. 183 | Fan06, S. 494)
- Neologismen  
(Hoe17, S. 55)
- Tautologien  
(Goe02, S. 165)
- Präzision (vs. Abstraktheit) der Wörter  
(Goe02, S. 177)

#### **Zweifach-Zuordnungen**

- Nominalisierungen (substantivierte Infinitive, nominalisierte Adjektive)  
(Hei18, S. 72, 77 | Hoe17, S. 54, 55 | Goe02, S. 179 | Rin10, S. 238 | Pat18, S. 296 | Fan06, S. 500)
- Idiome, standardisierte sprachliche *Fertigstücke* (= Phraseologismen)  
(Hei18, S. 78, 84 | Goe02, S. 160 | Hoe17, S. 54, 59)
- Funktionsverbgefüge (= Verb-Sinnentleerung)  
(Rin10, S. 238 | Hoe17, S. 55 | Goe02, S. 160)
- Wortwiederholungen (nominale vs. synonymiale Verbindungen)  
(Art05, S. 15, 24, 134 | Sta06, S. 134, 135 | Goe02, S. 185 | Hei18, S. 83)
- Pronomen, pronomiale Verbindungen  
(Sta06, S. 134 | Goe02, S. 183 | Art05, S. 24 | Fri18, S. 190 | Fan06, S. 496)
- Explizite Definitionen (anstelle von Fachtermini)  
(Goe02, S. 165)
- Jugendslang  
(Goe02, S. 170)

#### **Drei- und Mehrfach-Zuordnungen**

- Ellipsen (vs. Redundanz)  
(Hei18, S. 83 | Fan06, S. 497, 506 | Goe02, S. 175)
- Metaphorische Ausdrücke  
(Goe02, S. 171)

## Satzstrukturen / morphologische Auffälligkeiten auf Satzebene

### Einfach-Zuordnungen

- Satzlänge (= Wörter pro Satz)  
(Sta06, S. 128 | Fri18, S. 186 | Len13, S. 31 | Ker12, S. 4, 10 | Hei18, S. 74 | Hoe17, S. 55 | Lut15, S. 246)
- Mehrwortkomplexe  
(Rin10, S. 238)
- Nominalisierungsgruppen, Nominalphrasen  
(Rin10, S. 238 | Hei18, S. 77 | Hoe17, S. 55 | Fan06, S. 501, 512)
- Satzglieder vs. Gliedsätze  
(Rin10, S. 238 | Ker12, S. 11 | Hoe17, S. 54 f. | Fan06, S. 502 | Goe02, S. 181)
- Hypotaxe vs. Nebenordnung  
(Len13, S. 31 | Ker12, S. 11 | Hei18, S. 79 | Hoe17, S. 55 | Fan06, S. 502 | Goe02, S. 177]
- Verbalklammern (= Satzklammern)  
(Goe02, S. 182 | Hei18, S. 79)
- Nebensatzlänge  
(Hei18, S. 87)
- Phrasenlänge  
(Hei18, S. 80)
- Nominativkonstruktionen vs. Verwendung alternativer Kasus  
(Hei18, S. 72 | Hoe17, S. 55 | Goe02, S. 179)
- Einschübe  
(Hei18, S. 83 | Goe02, S. 182 | Fan06, S. 503)
- Partizipialkonstruktionen  
(Fan06, S. 498)
- Regressivitätsgrad  
(Goe02, S. 180)
- Affirmative Ausdrucksweise  
(Goe02, S. 179)
- Fachsprachliche, spezifische Präpositionen  
(Fan06, S. 495 | Rin10, S. 247)

### Zweifach-Zuordnungen

- Passivische/aktivische Konstruktionen  
(Sta06, S. 128 | Rin10, S. 238, 247 | Hei18, S. 72, 77, 79 | Hoe17, S. 54 | Goe02, S. 170, 178)
- Füllwörter/-phrasen (= Propositionsdichte pro Satz)  
(Hei18, S. 74 | Goe02, S. 160)
- Idiome, standardisierte sprachliche *Fertigstücke* (= Phraseologismen)  
(Hei18, S. 78, 84 | Goe02, S. 160 | Hoe17, S. 54, 59)
- Nominalisierungen (substantivierte Infinitive, nominalisierte Adjektive)  
(Hei18, S. 72, 77 | Hoe17, S. 54, 55 | Goe02, S. 179 | Rin10, S. 238 | Pat18, S. 296 | Fan06, S. 500)

### C. Zum Konstruktionsraum für die sprachlichen Gestaltungsmittel

- Distanz von Relativsatz und Bezugswort  
(Hei18, S. 76, 78)
- Funktionsverbgefüge (= Verb-Sinentleerung)  
(Rin10, S. 238 | Hoe17, S. 55 | Goe02, S. 160)
- Konjunktiv- vs. Indikativkonstruktionen  
(Hei18, S. 84 | Hoe17, S. 54)
- Deagentivierung  
(Hoe17, S. 56)
- Inversionen (= Topikalisierungen)  
(Hei18, S. 79 | Goe02, S. 175)

#### **Drei- und Mehrfach-Zuordnungen**

- Ellipsen (vs. Redundanz)  
(Hei18, S. 83 | Fan06, S. 497, 506 | Goe02, S. 175)
- Konjunktionen  
(Hei18, S. 84 | Goe02, S. 174, 180 | Art05 S. 14, 24 | Sta06, S. 134, 139 | Pat18, S. 297 | Fri18, S. 190 | Len13, S. 29 | Fan06, S. 514)
- Fragen/Aufgaben zum Text  
(Hei18, S. 87 | Ker12, S. 4, 6, 12 | Goe02, S. 171)

### **Erzählende und personalisierende Textelemente**

#### **Einfach-Zuordnungen**

- Unpersönliche Wendungen („man“) vs. persönliche Anrede („du“/„wir“)  
(Rin10, S. 238 | Hoe17, S. 54 | Goe02, S. 170)
- Dialoge  
(Pat18, S. 492)
- Lebensweltnähe im Themeneinstieg  
(Ker12, S. 10)
- Alltagsnähe in den Beispielen  
(Goe02, S. 170)
- Alltagsnähe in den Analogien/Parallelen  
(Ker12, S. 10)
- Werbende Attribute  
(Goe02, S. 167)
- Anekdoten  
(Goe02, S. 167)
- Witze  
(Goe02, S. 171)
- Wiedergabe von Stimmungen  
(Ker12, S. 10)

### **Zweifach-Zuordnungen**

- Passivische/aktivische Konstruktionen  
(Sta06, S. 128 | Rin10, S. 238, 247 | Hei18, S. 72, 77, 79 | Hoe17, S. 54 | Goe02, S. 170, 178)
- Füllwörter/-phrasen (= Propositionsdichte pro Satz)  
(Hei18, S. 74 | Goe02, S. 160)
- Symbole (i. S. d. Stilmittels)  
(Art05, S. 16)
- Allegorien  
(Art05, S. 16)
- Konjunktiv- vs. Indikativkonstruktionen  
(Hei18, S. 84 | Hoe17, S. 54)
- Deagentivierung  
(Hoe17, S. 56)
- Jugendslang  
(Goe02, S. 170)

### **Drei- und Mehrfach-Zuordnungen**

- Metaphorische Ausdrücke  
(Art05, S. 16, 28 | Rin10, S. 238 | Len13, S. 30 | Goe02, S. 171)

## **Hilfestellungen zur Textstruktur und Aufmerksamkeitssteuerung**

### **Einfach-Zuordnungen**

- (Verbalisierte/nicht-verbalisierte) Gedankensprünge/Themenwechsel  
(Art05, S. 25)
- Relationen zwischen Abstraktem und Konkretem  
(Art05, S. 25)
- (Typographische) Topic-Indikatoren  
(Art05, S. 25 | Sta06, S. 133 | Len13, S. 30)
- Verbalisierte Hervorhebungen  
(Lut15, S. 252)
- Verbalisierung des Lesezwecks  
(Len13, S. 12 | Goe02, S. 156)

### **Zweifach-Zuordnungen**

- Rück- und Vorverweise  
(Art05, S. 24 | Sta06, S. 134 | Hei18, S. 72)
- Abstraktionen (= Verallgemeinerungen)  
(Art05, S. 25)
- Spezifizierungen, Inklusionen  
(Art05, S. 28, Sta06, S. 139)
- (Zwischen-)Überschriften  
(Art05, S. 25 | Len13, S. 30 | Hei18, S. 72)

### C. Zum Konstruktivraum für die sprachlichen Gestaltungsmittel

- Unterstreichungen  
(Art05, S. 25)
- Typengrößenvariation  
(Art05, S. 25 | Lut15, S. 252)
- Fett-/Kursivdruck  
(Sta06, S. 139 | Ker12, S. 10 | Lut15, S. 252)
- Advance Organizers  
(Art05, S. 28 | Len13, S. 30 | Ker12, S. 5 | Goe02, S. 172)

#### **Drei- und Mehrfach-Zuordnungen**

- Analogien  
(Art05, S. 28 | Sta06, S. 149 | Len13, S. 30 | Goe02, S. 171)
- Fragen/Aufgaben zum Text  
(Hei18, S. 87 | Ker12, S. 4, 6, 12 | Goe02, S. 171)
- Konjunktionen  
(Hei18, S. 84 | Goe02, S. 174, 180 | Art05 S. 14, 24 | Sta06, S. 134, 139 | Pat18, S. 297 | Fri18, S. 190 | Len13, S. 29 | Fan06, S. 514)

#### **Inhaltliche Hilfestellungen**

##### **Einfach-Zuordnungen**

- Redundanz bedeutungsrelevanter Informationen  
(Art05, S. 15 | Ker12, S. 5 | Hei18, S. 80 | Goe02, S. 167)
- Beispiele (zur inhaltlichen Elaboration)  
(Art05, S. 25 | Len13, S. 30)
- Inhaltliche Vergleiche  
(Art05, S. 25 | Goe02, S. 171)
- Inhaltliches Fazit/Zusammenfassung  
(Ker12, S. 11)
- Gegenüberstellungen (= Kontrastierung)  
(Pat18, S. 171 | Goe02, S. 171)
- Begriffserklärungen  
(Art05, S. 28 | Ker12, S. 11 | Fan06, S. 494 | Goe02, S. 178)
- Explizite Bezeichnungen („das heißt“ etc.)  
(Sta06, S. 139)

##### **Zweifach-Zuordnungen**

- Explizite Definitionen (anstelle von Fachtermini)  
(Goe02, S. 165)
- Abstraktionen (= Verallgemeinerungen)  
(Art05, S. 25)
- Spezifizierungen, Inklusionen  
(Art05, S. 28, Sta06, S. 139)

- Advance Organizers  
(Art05, S. 28 | Len13, S. 30 | Ker12, S. 5 | Goe02, S. 172)

### **Drei- und Mehrfach-Zuordnungen**

- Analogien  
(Art05, S. 28 | Sta06, S. 149 | Len13, S. 30 | Goe02, S. 171)
- Metaphorische Ausdrücke  
(Art05, S. 16, 28 | Rin10, S. 238 | Len13, S. 30 | Goe02, S. 171)
- Konjunktionen  
(Hei18, S. 84 | Goe02, S. 174, 180 | Art05 S. 14, 24 | Sta06, S. 134, 139 | Pat18, S. 297 | Fri18, S. 190 | Len13, S. 29 | Fan06, S. 514)
- Fragen/Aufgaben zum Text  
(Hei18, S. 87 | Ker12, S. 4, 6, 12 | Goe02, S. 171)

## **Typographische Markierungen**

### **Einfach-Zuordnungen**

- Schriftfarbe  
(Goe02, S. 157, 183)

### **Zweifach-Zuordnungen**

- (Zwischen-)Überschriften  
(Art05, S. 25 | Len13, S. 30 | Hei18, S. 72)
- Unterstreichungen  
(Art05, S. 25)
- Typengrößenvariation  
(Art05, S. 25 | Lut15, S. 252)
- Fett-/Kursivdruck  
(Sta06, S. 139 | Ker12, S. 10 | Lut15, S. 252)
- (Typographische) Topic-Indikatoren  
(Art05, S. 25 | Sta06, S. 133 | Len13, S. 30)

### **Drei- und Mehrfach-Zuordnungen**

–

## **Kohäsionsmittel**

### **Einfach-Zuordnungen**

- Wiederaufnahme von Satzteilen  
(Art05, S. 24)
- Adverbien (als verbindendes Element)  
(Goe02, S. 174)
- Explizite Bezüge zwischen Textteilen  
(Len13, S. 29)

### **Zweifach-Zuordnungen**

- Wortwiederholungen (nominale vs. synonymiale Verbindungen)  
(Art05, S. 15, 24 | Sta06, S. 134, 135 | Goe02, S. 185 | Hei18, S. 83)
- Pronomen, pronomiale Verbindungen  
(Sta06, S. 134 | Goe02, S. 183 | Art05, S. 24 | Fri18, S. 190 | Fan06, S. 496)
- Rück- und Vorverweise  
(Art05, S. 24 | Sta06, S. 134 | Hei18, S. 72)
- Wortwiederholungen (nominale vs. synonymiale Verbindungen)  
(Art05, S. 15, 24, 134 | Sta06, S. 134, 135 | Goe02, S. 185 | Hei18, S. 83)
- fachsprachliche, spezifische Präpositionen  
(Fan06, S. 495 | Rin10, S. 247)
- Inversionen (= Topikalisierungen)  
(Hei18, S. 79 | Goe02, S. 175)
- Distanz von Relativsatz und Bezugswort  
(Hei18, S. 76, 78)

### **Drei- und Mehrfach-Zuordnungen**

- Ellipsen (vs. Redundanz)  
(Hei18, S. 83 | Fan06, S. 497, 506 | Goe02, S. 175)
- Konjunktionen  
(Hei18, S. 84 | Goe02, S. 174, 180 | Art05 S. 14, 24 | Sta06, S. 134, 139 | Pat18, S. 297 | Fri18, S. 190 | Len13, S. 29 | Fan06, S. 514)

## **Darstellungsmodi und Zeichensysteme**

### **Einfach-Zuordnungen**

- Tabellen  
(Pat18, S. 193 | Ker12, S. 10 | Fan06, S. 492)
- Diagramme/Graphen  
(Pat18, S. 193 | Ker12, S. 10 | Fan06, S. 492)
- Listen  
(Goe02, S. 186)

### **Zweifach-Zuordnungen**

- Symbole (i. S. d. Stilmittels)  
(Art05, S. 16)
- Allegorien  
(Art05, S. 16)

### **Drei- und Mehrfach-Zuordnungen**

- Analogien  
(Art05, S. 28 | Sta06, S. 149 | Len13, S. 30 | Goe02, S. 171)
- Metaphorische Ausdrücke  
(Art05, S. 16, 28 | Rin10, S. 238 | Len13, S. 30 | Goe02, S. 171)

## D | STATISTISCHER ANHANG

### D.1 Testkennwerte aus der Pilotierungsstudie

An der Pilotierung des Bewertungsfragebogens nahmen insgesamt  $n_{\text{PilotierungFB}} = 99$  Schüler\*innen aus vier achten Klassen verschiedener bayerischer Gymnasien teil; mit drei dieser Klassen wurde auch der Wissenstests pilotiert ( $n_{\text{PilotierungWT}} = 71$ ). Die Daten wurden für die Testevaluation sowohl des Wissenstests als auch des Bewertungsfragebogens genutzt.

Die Kennwerte aus der Pilotierungsstudie des Wissenstests sind in Tabelle D.1 dargestellt. Die Tabelle D.2 listet ferner die einzelnen Distraktoren-Schwierigkeiten der sechs Multiple-Choice-Aufgaben auf. Auf Grundlage dieser Kennwerte wurden geringfügige Anpassungen des Tests für die Hauptstudie vorgenommen. Diese sind im Einzelnen im Hauptteil dieser Arbeit beschrieben.

Für den Bewertungsfragebogen sind die berechneten Testkennwerten in Tabelle D.3 dargestellt. Die Itemselektion wurde anhand einer Hauptkomponentenanalyse vorgenommen. Der Fragebogen wurde von den Teilnehmenden direkt nach der Textlektüre durchgeführt.

Kennwert	gesamter Test							
Anzahl	71							
Max. Score	34							
Mittelwert	17.66							
Schwierigkeit	0.52							
Varianz	27.00							

Kennwert	A1	A2	A3	A4	A5	A6	A7	A8
Anzahl	71	71	71	68	71	71	71	71
Max. Task-Score	5	5	5	5	5	5	2	2
Mittelwert	3.85	2.72	1.46	1.60	2.75	2.03	1.66	1.66
Schwierigkeit	0.77	0.54	0.29	0.32	0.55	0.41	0.83	0.83
Varianz	1.85	1.81	2.05	1.24	1.74	2.06	0.56	0.56
Trennschärfe	0.20	0.32	0.27	0.50	0.43	0.28	0.28	0.47

Tabelle D.1.: Itemkennwerte aus der Pilotierungsstudie für den Wissenstest. Der Test wurde im Anschluss an die Textlektüre durchgeführt. Die Aufgabennummern beziehen sich auf den Test, wie er in Anhang B zu finden ist. Mit „Anzahl“ ist die Anzahl der nicht-fehlenden Werte bezeichnet. Unter „Max. (Task-)Score“ ist die maximal erreichbare Punktzahl angegeben. „Mittelwert“ bezeichnet das arithmetische Mittel. Die Schwierigkeit setzt Mittelwert und maximal erreichbare Punktzahl ins Verhältnis. Die Varianz wird aus der Quadratsumme der Mittelwertabweichungen gebildet. Die Trennschärfe wurde als *part-whole-korrigierte* Korrelation mit dem Test-Score berechnet.

Schwierigkeitsindex für die einzelnen Distraktoren						
Distraktor-Nr.	A1	A2	A3	A4	A5	A6
i	.83	.61	.58	.59	.68	.41
ii	.79	.77	.28	.65	.76	.66
iii	.99	.79	.61	.22	.46	.45
iv	.79	.76	.34	.44	.83	.51
v	.90	.59	.46	.65	.82	.87

Tabelle D.2.: Distraktor-Schwierigkeitsindizes auf Basis der Pilotierungsdaten für den Wissenstest. Die Schwierigkeit entspricht hier dem Mittelwert, der erreichten Punkte, da jede korrekte Entscheidung mit 1 Punkt bewertet wird. Für Aufgabe vier lagen  $n = 68$  bearbeitete Aufgaben vor, alle anderen Aufgaben wurden von  $n = 71$  Teilnehmenden bearbeitet. Die Aufgaben- und Distraktorennummern beziehen sich auf den Test, wie er in Anhang B zu finden ist.

Kennwert	global	1	*2	3	4	5	6	7	8
Anzahl	96	92	97	92	90	95	96	94	96
Mittelwert	2.06	4.04	3.65	4.34	1.93	2.82	2.95	3.15	3.1
Korrelation mit <i>global</i>	1.0	0.29	-0.61	0.36	0.41	0.52	0.26	0.37	0.31
erwartete Korrelation	pos.	pos.	neg.	pos.	pos.	pos.	pos.	pos.	pos.
Schwierigkeit	0.27	0.76	0.66	0.83	0.23	0.46	0.49	0.54	0.53
Varianz	0.79	1.41	1.3	1.03	0.93	1.45	1.63	1.64	1.61
Kennwert	*9	10	11	*12	13	14	*15	16	17
Anzahl	95	95	90	95	91	91	95	95	94
Mittelwert	4.04	2.31	2.23	4.03	1.89	1.91	3.58	2.84	2.0
Korrelation mit <i>global</i>	-0.62	0.65	0.04	-0.61	0.53	0.51	-0.52	0.43	0.29
erwartete Korrelation	neg.	pos.	pos.	neg.	pos.	pos.	neg.	pos.	pos.
Schwierigkeit	0.76	0.33	0.31	0.76	0.22	0.23	0.64	0.46	0.25
Varianz	1.09	1.08	1.2	1.27	0.78	1.03	1.76	1.4	0.89
Kennwert	*18	*19	20	*21	22	*23	24	25	26
Anzahl	97	87	92	92	95	99	95	79	91
Mittelwert	2.07	3.23	2.11	3.26	1.84	2.41	2.09	3.87	3.79
Korrelation mit <i>global</i>	-0.13	-0.26	0.56	-0.31	0.43	-0.33	0.6	-0.07	-0.12
erwartete Korrelation	neg.	neg.	pos.	neg.	pos.	neg.	pos.	pos.	pos.
Schwierigkeit	0.27	0.56	0.28	0.57	0.21	0.35	0.27	0.72	0.7
Varianz	1.06	1.51	0.9	1.71	0.98	1.33	0.82	1.35	1.44
Kennwert	*27	28	29	30	*31	32	33	*34	35
Anzahl	96	90	94	97	94	94	85	96	83
Mittelwert	3.6	2.44	2.05	2.38	3.56	2.35	3.86	3.91	2.4
Korrelation mit <i>global</i>	-0.6	0.53	0.6	0.28	-0.42	0.5	-0.26	-0.42	0.41
erwartete Korrelation	neg.	pos.	pos.	pos.	neg.	pos.	pos.	neg.	pos.
Schwierigkeit	0.65	0.36	0.26	0.35	0.64	0.34	0.71	0.73	0.35
Varianz	1.41	1.11	0.82	1.64	1.39	1.4	1.39	1.46	1.44
Kennwert	36	37	*38	39	40	*41	42	43	44
Anzahl	93	96	97	89	88	97	92	98	93
Mittelwert	1.96	1.89	3.76	2.22	3.8	3.55	1.99	1.77	1.99
Korrelation mit <i>global</i>	0.52	0.49	-0.33	0.26	0.26	-0.27	0.43	0.52	0.63
erwartete Korrelation	pos.	pos.	neg.	pos.	pos.	neg.	pos.	pos.	pos.
Schwierigkeit	0.24	0.22	0.69	0.31	0.7	0.64	0.25	0.19	0.25
Varianz	0.97	1.02	1.48	1.21	1.82	1.53	1.21	1.2	0.81

Tabelle D.3.: Itemkennwerte aus der Pilotierungsstudie für den Bewertungsfragebogen. Es nahmen insgesamt  $n_{\text{PilotierungFB}}=99$  Personen teil. Die Itemnummern beziehen sich auf den Fragebogen, wie er in Anhang B zu finden ist. Mit „Anzahl“ ist die Anzahl der nicht-fehlenden Werte bezeichnet. Die Korrelation wurde gemäß der Berechnungsformel für die *Pearson-Produkt-Moment-Korrelation* ermittelt. Die Schwierigkeit setzt Mittelwert und maximale Bewertung ins Verhältnis. Die Varianz wird aus der Quadratsumme der Mittelwertabweichungen gebildet. Invertierte formulierte Items sind mit einem \* gekennzeichnet.

## D.2 Testkennwerte aus der Hauptstudie

An der Hauptuntersuchung dieser Studie nahmen insgesamt  $n_{\text{Hauptstudie}} = 787$  Schüler\*innen aus der achten und neunten Jahrgangsstufe bayerischer Gymnasien und Realschulen teil.

Zur Auswertung des Wissenstests konnten  $n_{\text{HauptstudieWT}} = 782$  Datensätze genutzt werden. Die ermittelten Testkennwerte sind in Tabelle D.4 dargestellt.

Für die Datenanalyse des Fragebogens zur Testbewertung konnten  $n_{\text{HauptstudieFB}} = 781$  Datensätze genutzt werden. Die Daten wurden in erster Linie für die Beantwortung der Forschungsfragen genutzt, es wurde aber zu Kontrollzwecken auch ein Vergleich zwischen den Hauptstudien-Testkennwerte und denen der Pilotierung angestellt. Die Werte der Hauptstudie sind im Folgenden in Tabelle D.5 dargestellt; für die Kennwerte der Pilotierung siehe Tabelle D.3 auf Seite 240.

Kennwert	ges. Prä-Test	ges. Post-Test						
Anzahl	782	782						
Max. Score	34	34						
Mittelwert	11.5	17.7						
Schwierigkeit	0.34	0.52						
Varianz	17.2	27.4						

Kennwert	A1 <sup>Prä</sup>	A2 <sup>Prä</sup>	A3 <sup>Prä</sup>	A4 <sup>Prä</sup>	A5 <sup>Prä</sup>	A6 <sup>Prä</sup>	A7 <sup>Prä</sup>	A8 <sup>Prä</sup>
Anzahl	782	782	782	782	782	782	782	782
Max. Task-Score	5	5	5	5	5	5	2	2
Mittelwert	1.47	1.95	1.05	1.24	1.67	1.71	1.41	1.01
Schwierigkeit	0.29	0.39	0.21	0.25	0.33	0.34	0.70	0.50
Varianz	1.59	1.39	1.21	1.39	1.04	1.90	0.84	1.00
Trennschärfe	0.25	0.23	0.21	0.14	0.19	0.22	0.15	0.23

Kennwert	A1 <sup>Post</sup>	A2 <sup>Post</sup>	A3 <sup>Post</sup>	A4 <sup>Post</sup>	A5 <sup>Post</sup>	A6 <sup>Post</sup>	A7 <sup>Post</sup>	A8 <sup>Post</sup>
Anzahl	782	782	782	782	782	782	782	782
Max. Task-Score	5	5	5	5	5	5	2	2
Mittelwert	3.11	2.67	1.73	1.96	2.55	2.25	1.77	1.62
Schwierigkeit	0.62	0.53	0.35	0.39	0.51	0.45	0.88	0.81
Varianz	2.08	2.07	1.97	2.01	1.72	2.04	0.42	0.61
Trennschärfe	0.29	0.24	0.38	0.37	0.39	0.32	0.21	0.27

Tabelle D.4.: Itemkennwerte aus der Hauptstudie für den Wissenstest. Es wurden insgesamt  $n_{\text{HauptstudieWT}} = 782$  Datensätze für die Ermittlung der dargestellten Kennwerte sowie für die Ergebnisauswertung hinsichtlich der Forschungsfragen verwendet. Der Test wurde als Prä-Test sieben bis achtundzwanzig Tage vor der Textlektüre und als Post-Test direkt im Anschluss an die Textlektüre durchgeführt. Die Aufgabennummern beziehen sich auf den Test, wie er in Anhang B.1 zu finden ist. Mit „Anzahl“ ist die Anzahl der nicht-fehlenden Werte bezeichnet (es wurden alle Aufgaben von allen Teilnehmenden als bearbeitet gewertet – siehe dazu Abschnitt 7.1). Unter „Max. (Task-)Score“ ist die maximal erreichbare Punktzahl angegeben. *Mittelwert* bezeichnet das arithmetische Mittel. Die Schwierigkeit setzt Mittelwert und maximal erreichbar Punktzahl ins Verhältnis. Die Varianz wird aus der Quadratsumme der Mittelwertabweichungen gebildet. Die Trennschärfe wurde als *part-whole-korrigierte* Korrelation mit dem Test-Score berechnet.

D. Statistischer Anhang

Kennwert	global	1	*2	3	4	5	6	7	8
Anzahl	755	777	778	775	777	778	779	775	778
Mittelwert	3.80	2.28	2.60	1.62	3.83	2.70	2.70	2.54	2.59
Korrelation mit <i>global</i>	1.00	0.20	-0.51	0.09	0.30	0.44	0.23	0.29	0.29
erwartete Korrelation	pos.	pos.	neg.	pos.	pos.	pos.	pos.	pos.	pos.
Schwierigkeit	0.70	0.32	0.40	0.15	0.71	0.42	0.42	0.38	0.40
Varianz	0.89	1.48	1.33	0.90	0.91	1.24	1.48	1.12	1.43
Kennwert	*9	10	11	*12	13	14	*15	16	17
Anzahl	773	779	777	778	777	776	778	777	777
Mittelwert	2.44	3.50	3.27	2.38	3.71	3.90	2.86	2.70	3.90
Korrelation mit <i>global</i>	-0.49	0.47	-0.12	-0.55	0.50	0.30	-0.45	0.34	0.30
erwartete Korrelation	neg.	pos.	pos.	neg.	pos.	pos.	neg.	pos.	pos.
Schwierigkeit	0.36	0.63	0.57	0.35	0.68	0.73	0.47	0.43	0.72
Varianz	1.24	1.34	1.22	1.25	0.94	1.15	1.66	1.36	0.95
Kennwert	*18	*19	20	*21	22	*23	24	25	26
Anzahl	777	777	775	781	779	776	778	778	773
Mittelwert	3.98	2.98	3.40	2.75	3.87	3.82	3.74	2.33	2.47
Korrelation mit <i>global</i>	-0.13	-0.46	0.48	-0.31	0.28	-0.27	0.47	-0.26	-0.02
erwartete Korrelation	neg.	neg.	pos.	neg.	pos.	neg.	pos.	pos.	pos.
Schwierigkeit	0.74	0.49	0.60	0.44	0.72	0.71	0.68	0.33	0.37
Varianz	0.94	1.53	1.14	1.70	1.35	1.19	0.88	1.30	1.24
Kennwert	*27	28	29	30	*31	32	33	*34	35
Anzahl	779	776	775	778	777	775	779	776	774
Mittelwert	2.68	3.18	3.42	3.57	2.93	3.45	2.97	2.61	3.30
Korrelation mit <i>global</i>	-0.46	0.41	0.49	0.32	-0.32	0.53	0.07	-0.46	0.48
erwartete Korrelation	neg.	pos.	pos.	pos.	neg.	pos.	pos.	neg.	pos.
Schwierigkeit	0.42	0.55	0.60	0.64	0.48	0.61	0.49	0.40	0.58
Varianz	1.47	1.43	1.20	1.42	1.56	1.15	1.23	1.50	1.39
Kennwert	36	37	*38	39	40	*41	42	43	44
Anzahl	776	775	775	776	775	780	775	776	778
Mittelwert	3.74	3.89	2.67	3.38	2.24	3.23	3.74	3.75	3.79
Korrelation mit <i>global</i>	0.45	0.43	-0.46	0.33	0.25	-0.33	0.33	0.49	0.48
erwartete Korrelation	pos.	pos.	neg.	pos.	pos.	neg.	pos.	pos.	pos.
Schwierigkeit	0.69	0.72	0.42	0.60	0.31	0.56	0.68	0.69	0.70
Varianz	1.07	1.14	1.53	1.34	1.38	1.51	1.09	1.32	0.98

Tabelle D.5.: Itemkennwerte aus der Hauptstudie für den Bewertungsfragebogen. Die Itemnummern beziehen sich auf den Fragebogen, wie er in Anhang B zu finden ist. Mit „Anzahl“ ist die Anzahl der nicht-fehlenden Werte bezeichnet. Die Korrelation wurde gemäß der Berechnungsformel für die *Pearson-Produkt-Moment-Korrelation* ermittelt. Die Schwierigkeit setzt Mittelwert und maximale Bewertung ins Verhältnis. Die Varianz wird aus der Quadratsumme der Mittelwertabweichungen gebildet. Invertiert formulierte Items sind mit einem \* gekennzeichnet.

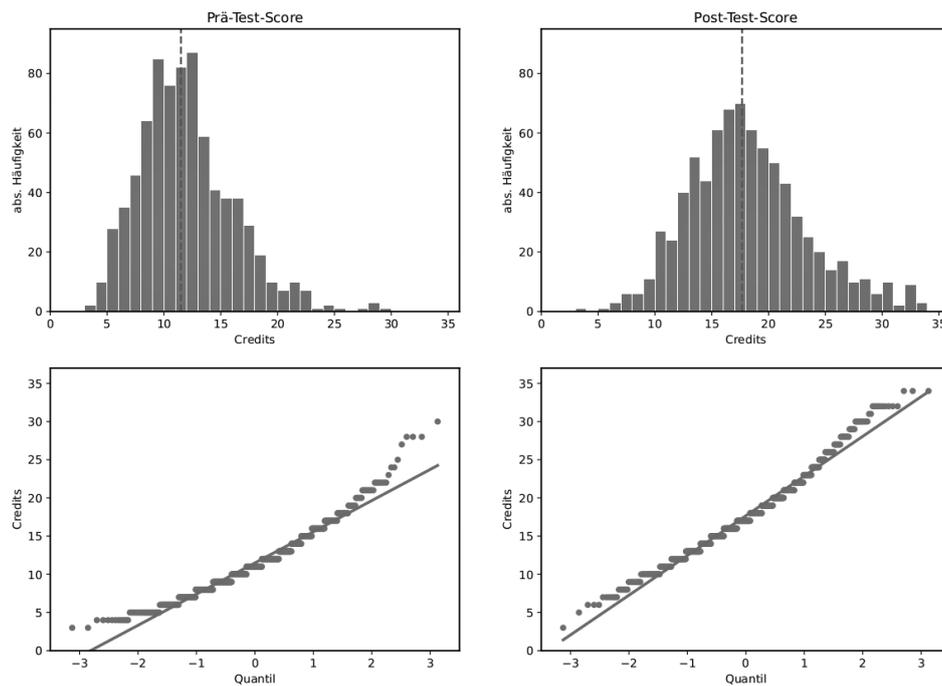


Abbildung D.1.: Graphische Überprüfung der Normalverteilungsannahme für die Prä-Post-Vergleiche. Oben links ist die Häufigkeitsverteilung der Prä-Test-Scores abgebildet, rechts oben diejenige für die Post-Test-Scores; die gestrichelte Linie markiert jeweils den Mittelwert. Darunter sind die entsprechenden Q-Q-Plot zu sehen.

### D.3 Robustheitsanalyse des Wissenstests

Im Abschnitt 6.4.3 wird erläutert, dass die Datenanalyse anhand von t-Tests und ANOVAs robust ist, solange keine erheblichen Voraussetzungsverletzungen vorliegen. Konkret bedeutet das für Analysen zur Behaltensleistung, dass zu überprüfen bleibt, ob die Creditverteilung in der Population normalverteilt ist und ob die Varianzen in den zu vergleichenden Gruppen homogen sind. Dies wird zum Zweck der Prä-Post-Vergleiche für den Prä-Test-Score und den Post-Test-Score durchgeführt. Zum Zweck der Zusammenhangsanalysen zwischen sprachlicher Gestaltung und Behaltensleistung wird die Stichprobe gemäß der unabhängigen Variable (*sprachliche Gestaltung*) in Gruppen geteilt und in diesen zu vergleichenden Gruppen werden die Voraussetzungen für die jeweilige Gain-Score-Verteilung überprüft. Die folgenden statistischen Testergebnisse zusammen mit den Abbildungen D.1, D.2, D.3 und D.4 zeigen, dass die Aussagen der t-Tests und ANOVAs der Behaltensleistung belastbar sind.

#### Zur Normalverteilungsannahme

Die Normalverteilungsannahme wird anhand eines Lilliefors-Tests statistisch überprüft. Der Lilliefors-Test ist eine Weiterentwicklung des Kolmogorov-Smirnov-Tests für Normalverteilung für unbekannte Populationsparameter (Lilliefors, 1967). Das Ergebnis für die Teststatistik und den p-Wert der Prä-Test-Score-Verteilung ist  $D = 0.11$  und  $p = 0.001$ , für den Post-Test ergeben sich  $D = 0.08$  und  $p = 0.001$ . Für die Gain-Score-Verteilungen in den Subgruppen je

## D. Statistischer Anhang

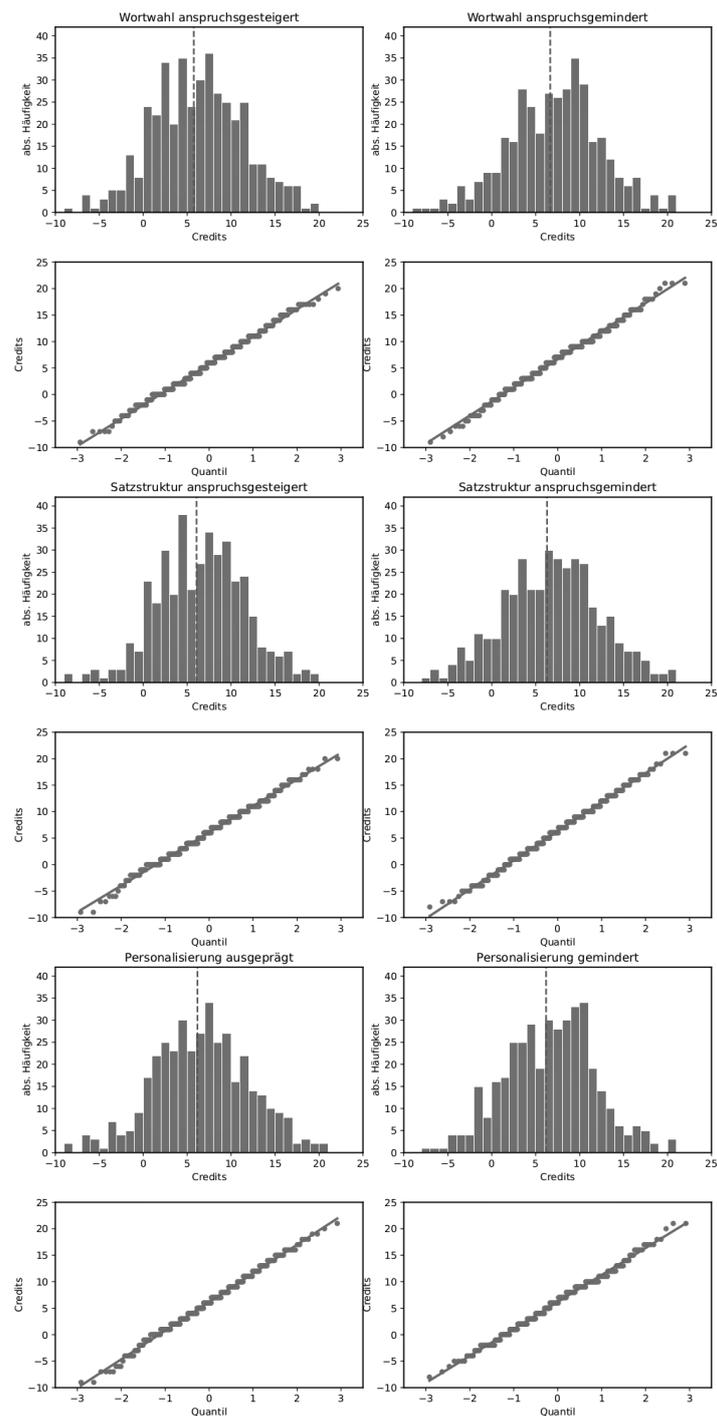


Abbildung D.2.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen Behaltensleistung und sprachlicher Gestaltung. Es sind die Gain-Score-Häufigkeitsverteilungen in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

nach sprachlicher Gestaltung findet man:

- Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.0560, p = 0.0037$
- Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.0610, p = 0.0032$
- Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.0587, p = 0.0024$
- Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.0507, p = 0.0226$
- Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0470, p = 0.0430$
- Subgruppe „Personalisierung gemindert“:  $D = 0.0622, p = 0.0010$

Damit ist die Nullhypothese in allen Fällen abzulehnen ( $p < 0.05$ ) und der Test spricht *nicht* für das Vorliegen einer Normalverteilung. Allerdings tendieren statistische Tests mit wachsender Stichprobengröße dazu, auch bei sehr kleinen Unterschieden bereits signifikant zu werden – es lohnt daher im vorliegenden Fall ein prüfender Blick auf die Verteilungsgrafiken. Auf Grundlage der Plots für die Prä-Post-Vergleiche in Abbildung D.1 ist die Annahme einer Normalverteilung weder eindeutig gerechtfertigt noch widerlegt. Eine Gaußsche Glockenform deutet sich in beiden Verteilungen an, die Verteilung ist jedoch vor allem für den Prä-Test-Score leicht rechtsschief. Die Q-Q-Plots, bei denen die Datenquantile mit den Quantilen einer Normalverteilung verglichen werden, zeigen, dass vor allem im zentralen Bereich, in dem sich die meisten Datenpunkte sammeln, die gemessenen Verteilungen beide stark einer Normalverteilung gleichen – lediglich am Rand sind Abweichungen erkennbar, da die Datenpunkte nicht mehr auf der Normalverteilungslinie liegen. Hinsichtlich der Gain-Score-Verteilungen in den Subgruppen lassen sich ähnliche Schlüsse ziehen: Die Grafiken weisen allesamt starke Ähnlichkeiten zur Glockenkurve auf, wenn auch leichte Verzerrungen erkennbar sind. Die Q-Q-Plots zeigen auch hier, dass sich die Abweichung von der Normalverteilung vor allem an den Rändern manifestiert.

Zusammengefasst lässt sich festhalten, dass die Annahme einer Normalverteilung in der Population fraglich ist. Die Grafiken zeigen jedoch, dass keine *gravierenden* Unterschiede zwischen gemessener Häufigkeitsverteilung und Normalverteilung vorliegen, sodass dank der großen Stichprobe in allen Fällen von Robustheit gegenüber der (vermutlich geringfügigen) Verletzung der Normalverteilungsannahme auszugehen ist.

### Zur Annahme der Varianzhomogenität

Die Annahme der Varianzhomogenität wird mithilfe des Levene-Tests (Rasch, Friese, Hofmann & Naumann, 2014a, S. 44) statistisch überprüft. Das Ergebnis im Prä-Post-Vergleich (Teststatistik  $L = 33.52$  und p-Wert  $p < 0.0001$ ) spricht gegen das Vorliegen der Nullhypothese homogener Varianzen. Eine grafische Auswertung (Abbildung D.3) stützt diesen Schluss. Für die Vergleiche der Gain-Scores zwischen Personen, die unterschiedlich gestaltete Texte gelesen haben, gilt hingegen, dass der Levene-Test die Annahme der Varianzhomogenität stützt. Der Test wird nicht signifikant: Teststatistik  $L = 0.76$  und p-Wert  $p = 0.58 \gg 0.05$ . Auch der Boxplot in Abbildung D.4 lässt auf der Grundlage von Augenscheinargumenten diese Annahme zu.

Trotz der offensichtlichen Verletzung der Varianzhomogenitätsvoraussetzung, sind die t-

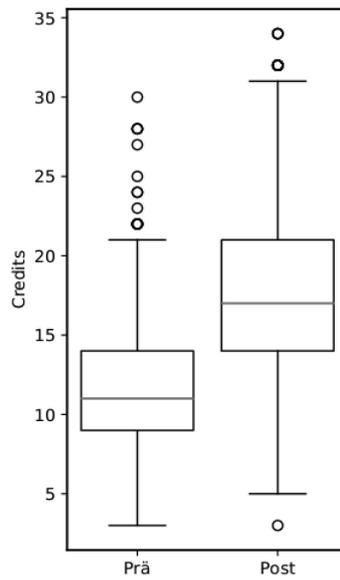


Abbildung D.3.: Boxplot der Prä-Test-Score-Verteilung (links) und der Post-Test-Score-Verteilung (rechts) zur Veranschaulichung der Varianzen in beiden Verteilungen.

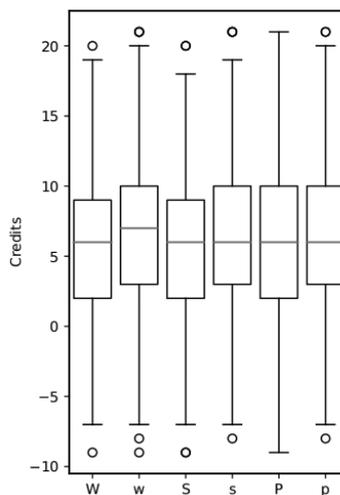


Abbildung D.4.: Boxplot der Gain-Score-Verteilungen in den einzelnen Subgruppen je nach sprachlicher Textgestaltung. Der Plot veranschaulicht, inwiefern die Varianzen in den Subgruppen vergleichbar groß sind.

Test-Vergleiche zwischen Prä-Test-Score und Post-Test-Score als belastbar zu betrachten. Denn bei großen Stichproben und insbesondere gleicher Gruppengröße (liegt hier vor, da mit jedem Prä-Test-Score genau der zugehörige Post-Test-Score verglichen wird) zeigt sich der t-Test robust gegen nicht erfüllte Voraussetzungen (Bortz & Schuster, 2010, S. 122). Für die Gain-Score-Verteilungen in den Subgruppen je nach sprachlicher Gestaltung der Texte kann von homogenen Varianzen ausgegangen werden, sodass anhand von t-Tests und ANOVAS belastbare Aussagen getroffen werden können.

Für alle weiteren Untersuchungen – etwa die spezifische Betrachtung einzelner Task-Scores oder die Unterteilung in Subgruppen je nach Textversion anstelle der sprachlichen Gestaltungsvariablen W, S und P – wurden keine weiteren, eigenen Robustheitsanalysen vorgenommen. Die entsprechenden Verteilungen sind mit den oben analysierten direkt verbunden, sodass sich kein stark unterschiedliches Bild ergeben wird und die Argumentation von oben sich ohne weiteres auch auf die weiteren Vergleiche und Zusammenhangsanalysen beziehen lässt.

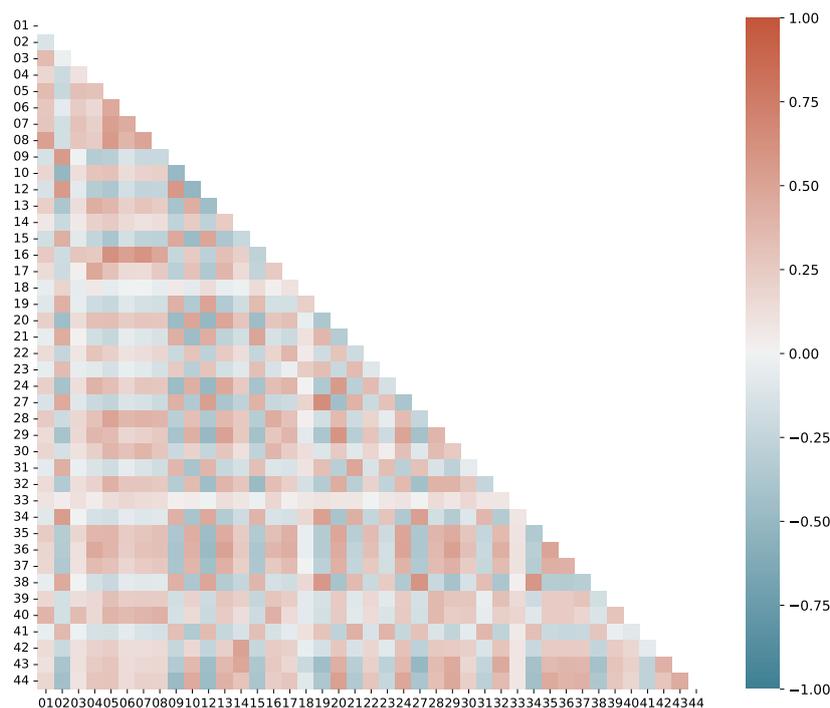


Abbildung D.5.: Heatmap für Item-Item-Korrelationen. Je nach Größe des Pearson-Korrelationskoeffizienten in der Korrelationsmatrix werden die Felder der Heatmap tiefer rot (stark positive Korrelation), farblos (nahezu keine Korrelation) oder tiefer blau (stark negative Korrelation) eingefärbt.

## D.4 Robustheitsanalyse des Bewertungsfragebogens

Die Grundlage für die Arbeit mit den Bewertungsfragebogen-Daten stellte eine PCA auf dem Datensatz ( $n_{\text{HauptstudieBewFB}} = 781$  mit fehlenden Werten durchsetzte Schüler\*innenbewertungen zu 41 Items und dem Global-Item) dar.

Generell gilt, dass eine Stichprobengröße von über 300 Personen beziehungsweise von mehr als zehnmal so vielen Personen wie Items (hier:  $10 \times 41 = 410$ ) eine solide Grundlage für eine PCA darstellt. Dies ist hier zutreffend ( $n_{\text{HauptstudieBewFB}} = 781$ ).

Darüber hinaus müssen gewisse Anforderungen an die Datengrundlage, genauer: an die Korrelationen zwischen den Items, erfüllt sein, damit eine PCA zu aussagekräftigen Ergebnissen führt. Es sind daher Tests nötig, die überprüfen, ob die Itemkorrelationen zu schwach für eine belastbare PCA sind<sup>1</sup>. Auf dem hier vorliegenden Datensatz wurden dazu unter Anwendung der pairwise-deletion-Methode (siehe Seite 115 im Abschnitt 6.3.3) Item-Item-Korrelationen berechnet. Der Bartlett's Test<sup>2</sup> (Field, Miles & Field, 2012, S. 770) weist mit den Werten  $\chi^2 = 13539$  für die Teststatistik und  $p < 0.001$  für die Fehlerwahrscheinlichkeit darauf hin, dass sich die Korrelationsmatrix signifikant von der Einheitsmatrix unterscheidet und

<sup>1</sup>Zu starke Korrelationen zwischen Items wären andererseits ein Anzeichen für Multikollinearität in den Daten. Dies wird anhand der Determinante der Item-Item-Korrelationsmatrix überprüft. Jedoch ist Multikollinearität lediglich problematisch für die Faktorenanalyse und nicht für die PCA (Field, Miles & Field, 2012, S. 771).

<sup>2</sup>`cortest.bartlett()`-Methode aus dem `psych`-Paket der Statistiksoftware R; siehe auch Abschnitt 6.4.3

somit ausreichend starke Item-Item-Korrelationen vorliegen. Auch der Kaiser-Meyer-Olkin-Kennwert<sup>3</sup>, der das Verhältnis zwischen der Summe der quadrierten Item-Item-Korrelationen zur Summe der quadrierten partiellen Item-Korrelation bezeichnet, spricht mit einem Wert von 0.95 für eine einwandfreie Eignung der Daten für eine PCA (Richtwerte nach Field, Miles und Field, 2012, S. 769 f. ab 0.8 sehr gut und ab 0.9 exzellent). Die beiden eben genannten quantitativen Tests wurden noch um eine Augenscheinanalyse anhand einer Heatmap für Korrelationen ergänzt (siehe Abbildung D.5). Die Items erscheinen dabei weitestgehend geeignete Korrelationen miteinander aufzuweisen. Dennoch fallen drei Items mit einer durchgehend sehr blass eingefärbten Zeile beziehungsweise Spalte auf. Die Items 03, 18 und 33 scheinen insgesamt sehr schwach mit den übrigen Items korreliert zu sein. Die entsprechenden Mittelwerte über die gesamte Spalte beziehungsweise Zeile der Korrelationsmatrix hinweg liegen jedoch auch für diese Items jeweils bei  $\bar{r} > 0.1$ , sodass keine weiteren itemselektiven Maßnahmen auf Grundlage der Heatmap-Analyse vorgenommen wurden.

Analog zu den ANOVA-Analysen zur Behaltenstest bleiben auch für den Bewertungsfragebogen zur affektiven Textwahrnehmung die Voraussetzungen, ob die abhängige Variable (Globalbewertung beziehungsweise Bewertung hinsichtlich einer Wahrnehmungsfacette) jeweils normalverteilt ist und ob die Varianzen in den zu vergleichenden Gruppen je nach sprachlicher Gestaltung des gelesenen Texts homogen sind. Im Folgenden werden die Ergebnisse entsprechender statistischer Tests (Lilliefors-Test für Normalverteilung sowie Levene-Test für Varianzhomogenität) berichtet sowie die Merkmalsverteilungen graphisch abgebildet.

### Zur Normalverteilungsannahme

Die Normalverteilungsannahmen werden jeweils anhand eines Lilliefors-Tests statistisch überprüft (siehe auch Abschnitt D.3 sowie Lilliefors, 1967). Das Ergebnis für die Teststatistik und den p-Wert der Prä-Test-Score-Verteilung ist

- Globaleindruck. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.2894, p = 0.0010$
- Globaleindruck. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.3010, p = 0.0010$
- Globaleindruck. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.2973, p = 0.0010$
- Globaleindruck. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.2955, p = 0.0010$
- Globaleindruck. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.2972, p = 0.0010$
- Globaleindruck. Subgruppe „Personalisierung gemindert“:  $D = 0.2951, p = 0.0010$
- Kognitive Anregung. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.2894, p = 0.0010$
- Kognitive Anregung. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.0792, p = 0.0010$
- Kognitive Anregung. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.0592, p = 0.0024$
- Kognitive Anregung. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.0728, p = 0.0010$
- Kognitive Anregung. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0621, p = 0.0014$
- Kognitive Anregung. Subgruppe „Personalisierung gemindert“:  $D = 0.0488, p = 0.0269$
- Lesbarkeit. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.0530, p = 0.0081$
- Lesbarkeit. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.0815, p = 0.0010$

<sup>3</sup>KMO()-Funktion des *EFAtools*-Paket der Statistiksoftware R; siehe auch Abschnitt 6.4.3

#### D. Statistischer Anhang

- Lesbarkeit. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.0550, p = 0.0065$
- Lesbarkeit. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.0736, p = 0.0010$
- Lesbarkeit. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0597, p = 0.0025$
- Lesbarkeit. Subgruppe „Personalisierung gemindert“:  $D = 0.0722, p = 0.0010$
- Logischer Aufbau. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.0601, p = 0.0014$
- Logischer Aufbau. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.0769, p = 0.0010$
- Logischer Aufbau. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.0747, p = 0.0010$
- Logischer Aufbau. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.0583, p = 0.0036$
- Logischer Aufbau. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0636, p = 0.0010$
- Logischer Aufbau. Subgruppe „Personalisierung gemindert“:  $D = 0.0818, p = 0.0010$
- Elaboriertheit. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.0582, p = 0.0023$
- Elaboriertheit. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.0723, p = 0.0010$
- Elaboriertheit. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.0603, p = 0.0019$
- Elaboriertheit. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.0771, p = 0.0010$
- Elaboriertheit. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0655, p = 0.0010$
- Elaboriertheit. Subgruppe „Personalisierung gemindert“:  $D = 0.0688, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.1215, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.1541, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.1361, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.1386, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.1247, p = 0.0010$
- Angem. Bez. z. Vorwissen. Subgruppe „Personalisierung gemindert“:  $D = 0.1494, p = 0.0010$
- Involviertheit. Subgruppe „Wortwahl anspruchsgesteigert“:  $D = 0.1165, p = 0.0010$
- Involviertheit. Subgruppe „Wortwahl anspruchsgemindert“:  $D = 0.1126, p = 0.0010$
- Involviertheit. Subgruppe „Satzstruktur anspruchsgesteigert“:  $D = 0.1299, p = 0.0010$
- Involviertheit. Subgruppe „Satzstruktur anspruchsgemindert“:  $D = 0.1027, p = 0.0010$
- Involviertheit. Subgruppe „Personalisierung ausgeprägt“:  $D = 0.0954, p = 0.0010$
- Involviertheit. Subgruppe „Personalisierung gemindert“:  $D = 0.1182, p = 0.0010$

Damit ist die Nullhypothese im Allgemeinen abzulehnen ( $p < 0.05$ ) und der Test spricht *nicht* für das Vorliegen einer Normalverteilung. Mit einer analogen Argumentation zu den Ausführungen in Abschnitt D.3 gilt jedoch, dass vor allem ein prüfender Blick auf die Verteilungsgrafiken lohnt. Auf Grundlage der Plots in den Abbildungen D.6, D.7, D.8, D.9, D.10, D.11 und D.12 ist die Annahme einer Normalverteilung ebenfalls eher abzulehnen. Die Grafiken D.6 bis D.12 rufen zunächst in einigen Fällen den Eindruck hervor, dass die Skala Bodenbeziehungsweise Deckeneffekte verursacht: es häufen sich Bewertungen bei den Polen 1 beziehungsweise 5 der Skala oder das Maximum der Skala liegt so weit rechts beziehungsweise links, dass Ausläufer der Verteilung nicht mehr umfasst werden können. Neben diesen Effekten macht vor allem der Globaleindruck, der sich allein an einem Item festmacht und somit nur in 5 diskreten Stufen erhoben wurde, eine Entscheidung für oder gegen das Vorliegen einer Normalverteilung schwer. Auch die Facetten *Angemessener Bezug zum Vorwissen* und *Involviertheit* weisen deutliche, augenscheinliche Unterschiede zur Normalverteilung auf. In den übrigen Verteilungen deutet sich tendenziell eine Gaußsche Glockenform an. Die

Q-Q-Plots stützen die Einschätzung, dass lediglich die Verteilungen der Facetten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau* und *Elaboriertheit* und auch des Globaleindruck einer Normalverteilung ähneln.

Zusammengefasst lässt sich festhalten, dass die Annahme einer Normalverteilung in der Population recht fraglich ist. Die Grafiken sprechen nur für einige der betrachteten Merkmale dafür, dass keine schwerwiegenden Abweichungen zur Normalverteilung in den Häufigkeitsverteilungen vorliegen. Dennoch ist aufgrund der großen Stichprobe in allen Fällen von Robustheit gegenüber der (in den meisten Fällen geringfügigen) Verletzung der Normalverteilungsannahme auszugehen.

## D. Statistischer Anhang

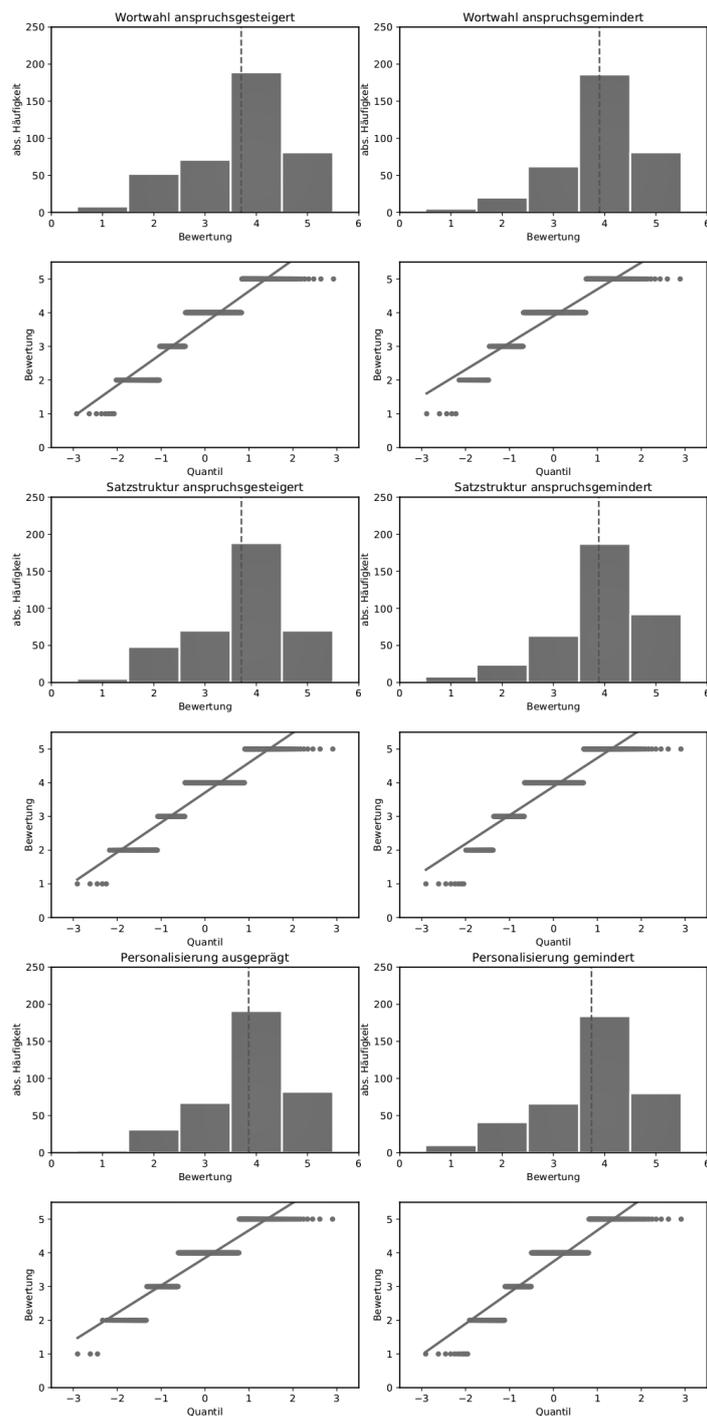


Abbildung D.6.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen Globaleindruck und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen zum Global-Item in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

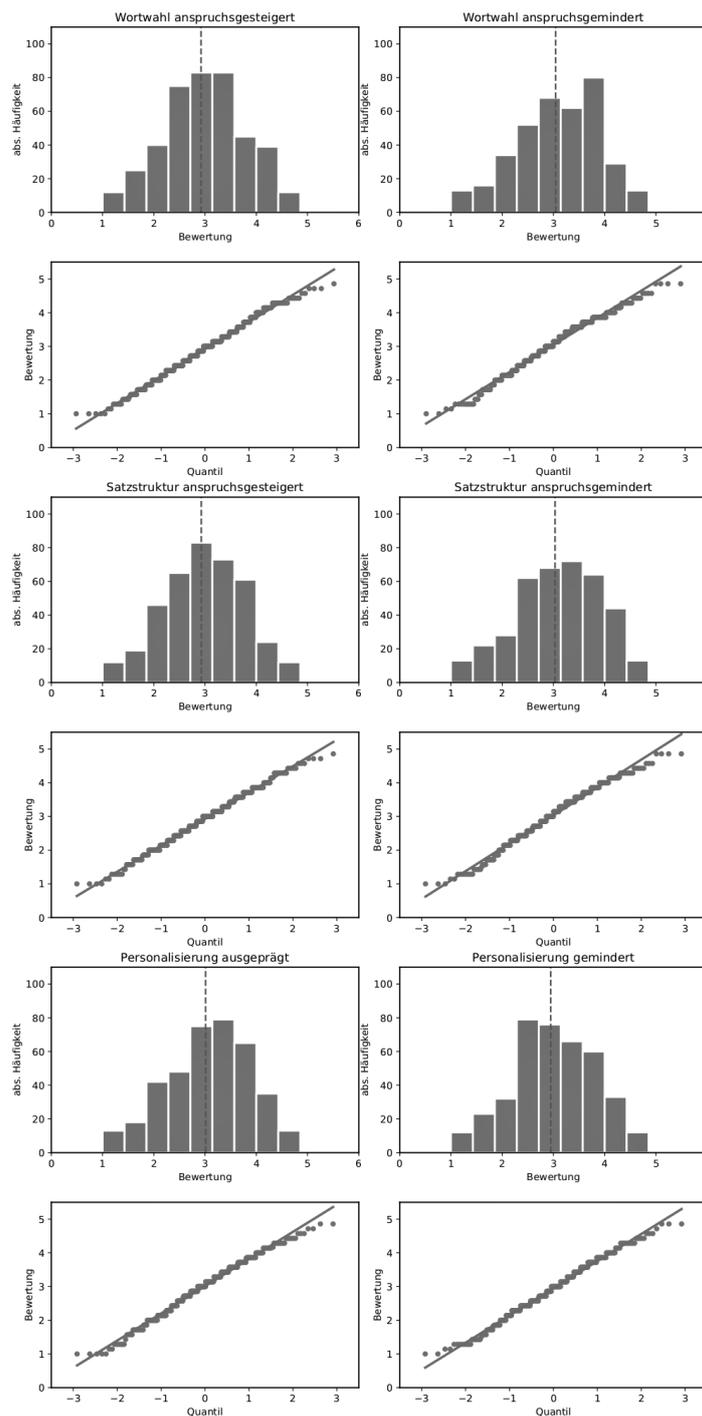


Abbildung D.7.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen *Kognitiver Anregung* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Kognitiver Anregung* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

## D. Statistischer Anhang

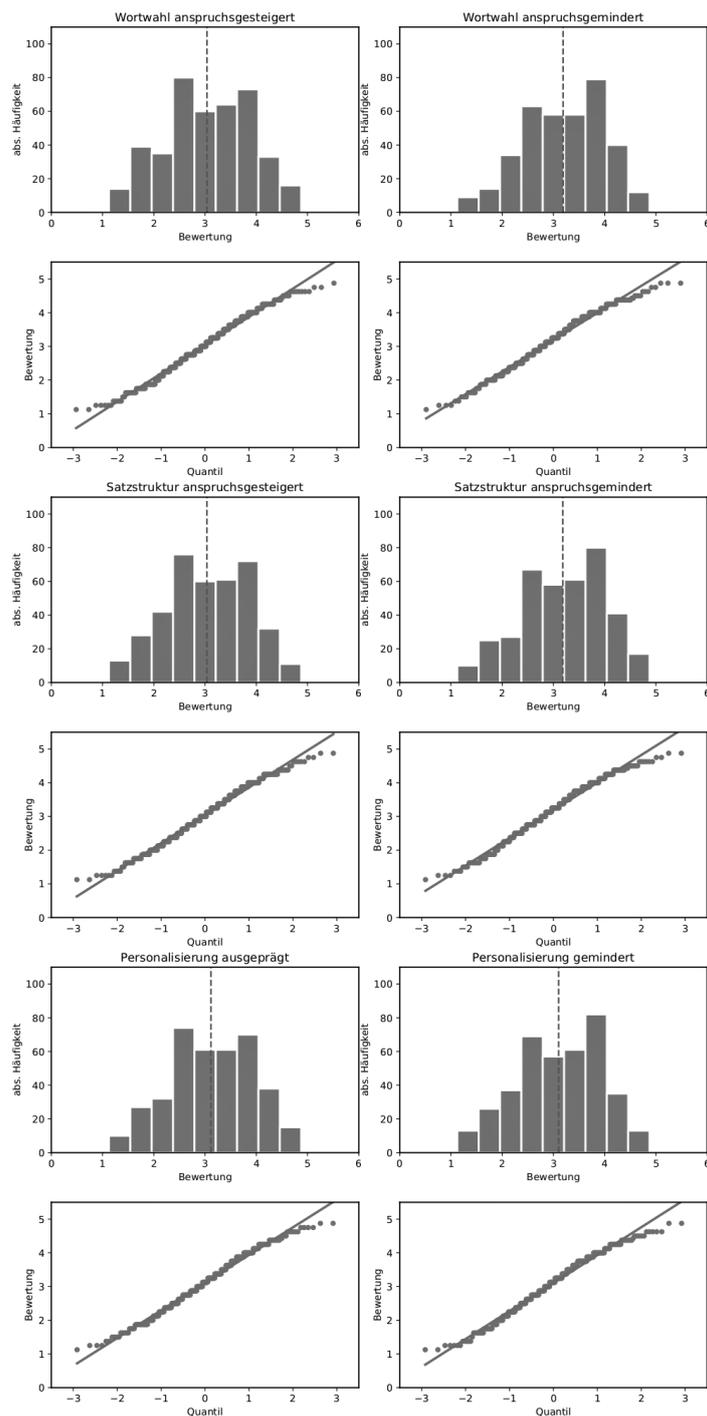


Abbildung D.8.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen *Lesbarkeit* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Lesbarkeit* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

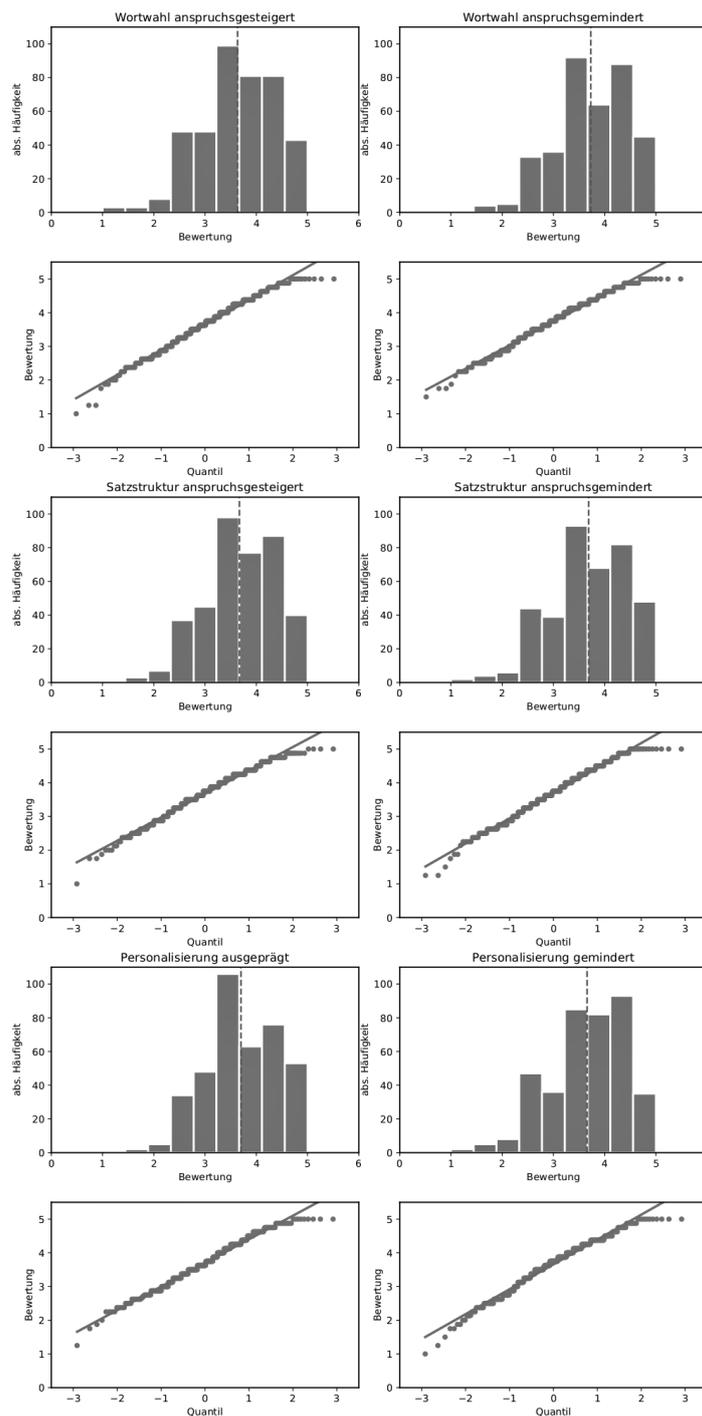


Abbildung D.9.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhanganalysen zwischen *Logischem Aufbau* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Logischem Aufbau* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

## D. Statistischer Anhang

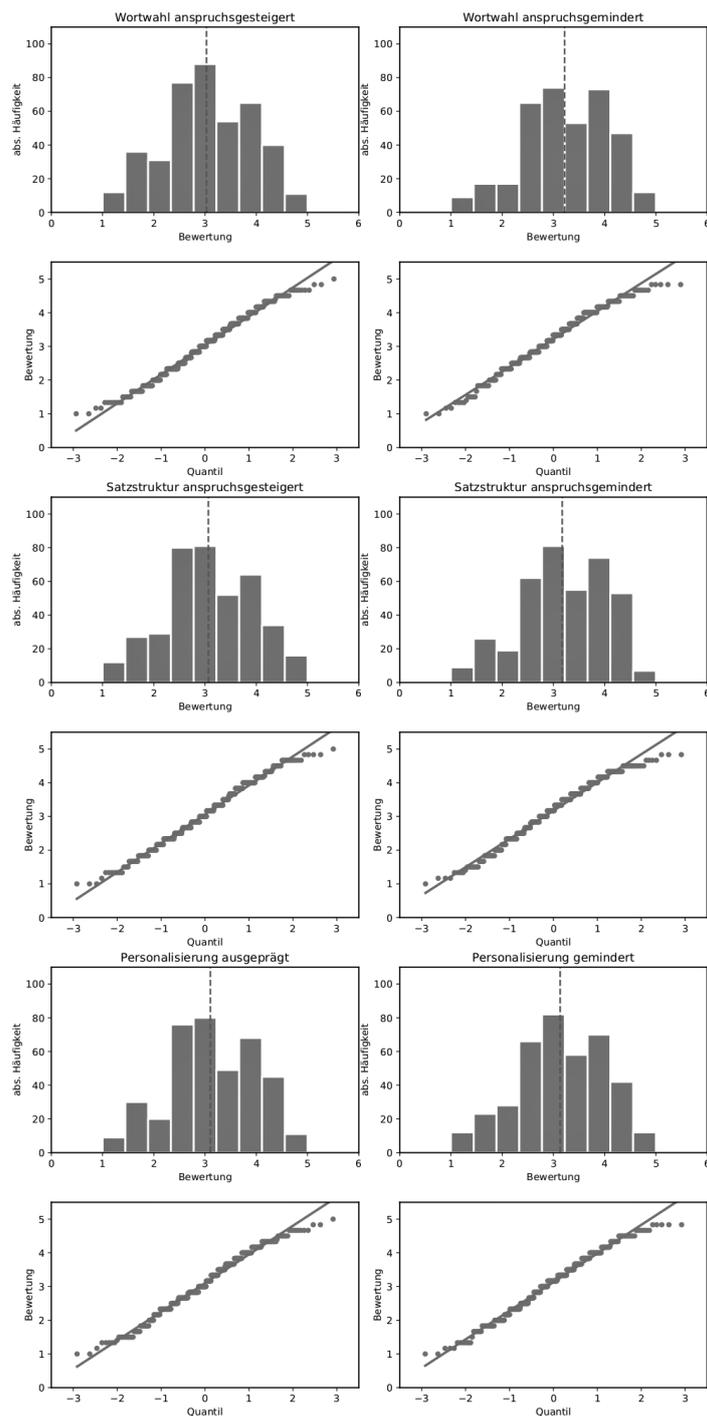


Abbildung D.10.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen *Elaboriertheit* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Elaboriertheit* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

#### D.4. Robustheitsanalyse des Bewertungsfragebogens

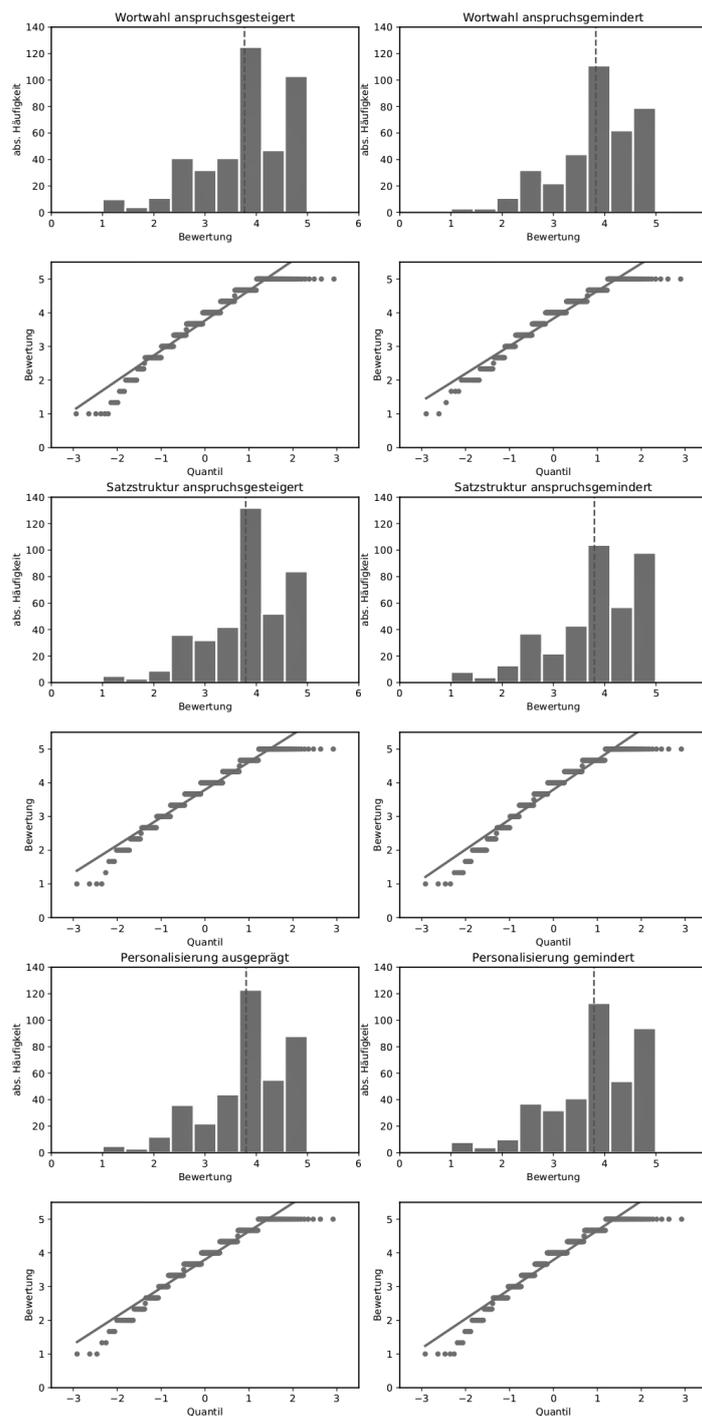


Abbildung D.11.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen *Angemessenem Bezug zum Vorwissen* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Angemessenem Bezug zum Vorwissen* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

## D. Statistischer Anhang

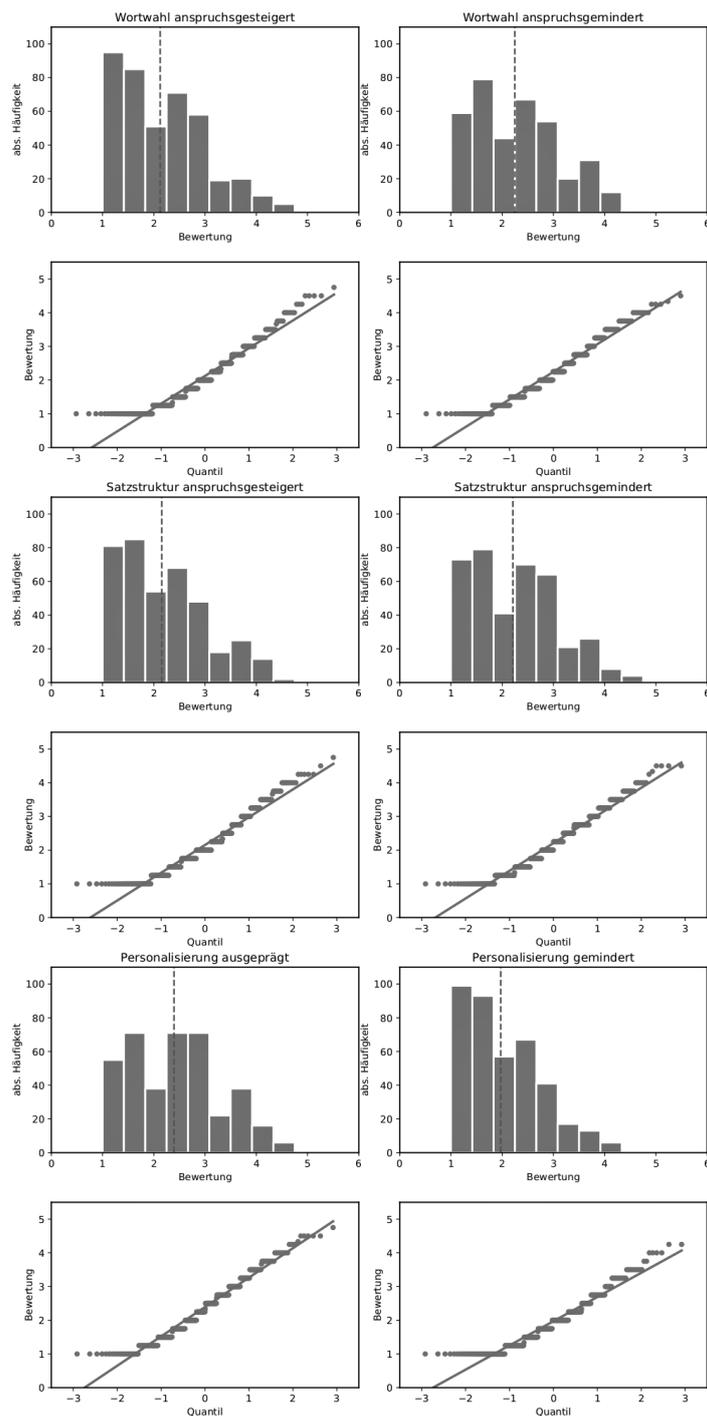


Abbildung D.12.: Graphische Überprüfung der Normalverteilungsannahme für die Zusammenhangsanalysen zwischen *Involviertheit* und sprachlicher Gestaltung. Es sind die Häufigkeitsverteilungen für die Bewertungen hinsichtlich *Involviertheit* in verschiedenen Stichproben-Subgruppen abgebildet. Die gestrichelte Linie markiert jeweils den Mittelwert. Die Subgruppen wurden je nach gelesenen Text gebildet. Dementsprechend fließen in die einzelnen Plots die Daten genau derjenigen Personen ein, die einen Text mit entsprechender Gestaltung gelesen haben. Unter den Verteilungshistogrammen sind jeweils die zugehörigen Q-Q-Plots abgebildet.

### Zur Annahme der Varianzhomogenität

Die Annahme der Varianzhomogenität wird mithilfe des Levene-Tests (Rasch, Frieze, Hofmann & Naumann, 2014a, S. 44) statistisch überprüft. Es gilt für die einzelnen Aspekte der affektiven Textwahrnehmung:

- Globaleindruck:  $L = 6.1704$ ,  $p \approx 0$ . Der Test stützt die Homogenitätsannahme nicht.
- Kognitive Anregung:  $L = 0.4805$ ,  $p = 0.7911$ . Der Test stützt die Homogenitätsannahme.
- Lesbarkeit:  $L = 0.1768$ ,  $p = 0.9714$ . Der Test stützt die Homogenitätsannahme.
- Logischer Aufbau:  $L = 0.7309$ ,  $p = 0.6002$ . Der Test stützt die Homogenitätsannahme.
- Elaboriertheit:  $L = 0.2000$ ,  $p = 0.9625$ . Der Test stützt die Homogenitätsannahme.
- Angemessener Bezug zum Vorwissen:  $L = 1.3568$ ,  $p = 0.2376$ . Der Test stützt die Homogenitätsannahme.
- Involviertheit:  $L = 4.0455$ ,  $p = 0.0012$ . Der Test stützt die Homogenitätsannahme nicht.

In zwei Fällen scheint die Varianzhomogenitätsannahme verletzt; in den übrigen fünf Fällen wird der Levene-Test nicht signifikant. Die quantitativ-statistischen Kennwerte wurden durch keine zusätzliche graphische Analyse in Form von Boxplots ergänzt. Vor dem Hintergrund sehr großer Gruppengrößen (jeweils ca. 400 Datenpunkte) wird dennoch von belastbaren Aussagen durch die ANOVAs ausgegangen.

Variable	Teststatistik F	p-Wert	$\omega^2$	Mittelwerte	
Wortwahl	7.86	.01	.01**	W	w
Satzstruktur	0.23	.63	.00	17.7	17.6
Personalisierung	0.02	.88	.00	S	s
WTpre_tot_credits	134.07	.00	.15***	17.4	17.9
WTpre_tot_credits:Wortwahl	6.51	.01	.01**	P	p
WTpre_tot_credits:Satzstruktur	0.72	.40	.00	17.6	17.8
WTpre_tot_credits:Personalisierung	0.07	.79	.00		

Tabelle D.6.: ANCOVA mit der abhängigen Variable *Post-Test-Score* und der unabhängigen Variable *sprachliche Gestaltung* in Form der Wortwahl, der Satzstruktur und der Personalisierung. Es wurde hinsichtlich des Vorwissens in Form des *Prä-Test-Scores* kontrolliert. In der letzten Spalte, in der die Effektstärke  $\omega^2$  angegeben ist, sind signifikante Effekte mit Sternchen markiert. Die p-Werte sind nicht für eine Alpha-Fehler-Kumulierung korrigiert. Die Modellparameter für die F-Statistik sind (7, 781). Nebenstehend sind die *Post-Test-Score*-Mittelwerte für die Textversionen je nach Ausprägung der sprachlichen Gestaltungsvariablen Wortwahl, Satzstruktur und Personalisierung angegeben.

## D.5 Zusammenhangsanalysen

### sprachliche Gestaltung – Post-Test-Score

In der Tabelle D.6 wird das Ergebnis einer ANCOVA mit der abhängigen Variable *Post-Test-Score*, der unabhängigen Variable *sprachliche Gestaltung* und der Kontrollvariable *Prä-Test-Score* gezeigt. Nebenstehend sind die entsprechenden Mittelwerte angegeben. Entscheidend sind die Effekte „WTpre\_tot\_credits:Wortwahl“, „WTpre\_tot\_credits:Satzstruktur“ und „WTpre\_tot\_credits:Personalisierung“, die den Haupteffekt der sprachlichen Gestaltungsvariablen unter Kontrolle des *Prä-Test-Scores* wiedergeben.

Das Ergebnis der ANOVA ist in großen Teilen im Einklang mit den im Hauptteil berichteten Ergebnissen. Während das Vorwissen erwartungsgemäß einen höchst signifikanten, sehr starken Effekt auf das *Post-Test-Ergebnis* zeigt, ist hinsichtlich der sprachlichen Gestaltung kaum ein einflussreicher Faktor zu identifizieren. Allenfalls zeigt sich die Wortwahl mit  $p_{\text{WTpre\_tot\_credits:Wortwahl}} = 0.01$  signifikant mit geringer Effektstärke  $\omega^2 = 0.001$  für das *Post-Test-Ergebnis*. Man bemerke jedoch, dass hier keine Alpha-Fehler-Korrektur für multiple Mittelwertvergleiche vorgenommen wurde. Man bemerke mit Blick auf die Mittelwerte neben Tabelle D.6 darüber hinaus, dass die Richtung des Effekts sich *scheinbar* umkehrt. Während wie im Hauptteil berichtet der *Gain-Score* bei w-Texten höher ist als bei W-Texten, ist das *Post-Test-Ergebnis* für W-Texte höher als für w-Texte. Dies ändert jedoch nichts an dem Befund, dass das *Dazulernen* in w-Texten größer war in W-Texten.

Die hier dargestellte ANCOVA verfolgt die Frage, ob sich unterschiedliche *Post-Test-Ergebnisse* erwarten lassen, je nachdem, welche sprachliche Gestaltung vorliegt. Indem dabei das Vorwissen in Form des *Pre-Test-Scores* kontrolliert wird, liegt der Analyse eine Situation zugrunde, in der alle Testpersonen den gleichen „Startpunkt“ hinsichtlich des Vor-

wissens haben. Eine solche ANCOVA ist aus diesem Grund nur dann geeignet, wenn davon auszugehen ist, dass das Vorwissen in allen Textversions-Gruppen gleichverteilt ist und wenn das Vorwissen ohne Fehler gemessen wurde (Smolkowski, 2019). Das Ergebnis der Review-Analyse von Smolkowski ist entsprechend, dass eine ANCOVA nur verwendet werden sollte, wenn das Design randomisiert ist und die einzelnen Gruppen gleiche Startbedingungen haben (außerdem können nur Haupteffekte zuverlässig abgebildet werden). Dass die Bedingungen zutreffen kann hier aufgrund der randomisiert verteilten Textversionen erwartet werden.

Gender	Teilstichprobengrößen		
	weiblich 471	männlich 281	divers 30
Lesefähigkeiten	niedrig		hoch
Verständnis	351		429
Geschwindigkeit	424		358
Genauigkeit	207		573
Schulart	Realschule 293		Gymnasium 489
Jahrgangsstufe	8. Klasse 147		9. Klasse 635
Vorwissen	niedrig 346		hoch 436

Tabelle D.7.: Übersicht über die Verteilung der Gesamtstichprobe in die einzelnen Subgruppen für die Betrachtung des Gain Scores in Abhängigkeit von der sprachlichen Textgestaltung.

## D.6 Moderierte Zusammenhangsanalysen: sprachliche Gestaltung – Gain-Score

Im Folgenden werden alle ANOVA-Tabellen für die Betrachtung der Auswirkungen der sprachlichen Gestaltung auf die Behaltensleistung in unterschiedlichen Subgruppen der Gesamtstichprobe. Die Subgruppen wurden auf der Grundlage kategorialer Moderatorvariablen beziehungsweise auf der Grundlage eines Mediansplits oder Normwerttabellen bei kontinuierlichen Moderatorvariablen gebildet. Die jeweiligen Subgruppengrößen sind in Tabelle D.7 zusammengestellt.

- In Tabelle D.8 sind die Ergebnisse der ANOVA-Modelle je nach **Gender** aufgelistet.
- In den Tabellen D.9, D.10 und D.11 sind die Ergebnisse der ANOVA-Modelle je nach **Lesefähigkeiten** aufgelistet.
- In den Tabellen D.12 und D.13 sind die Ergebnisse der ANOVA-Modelle je nach **Schulstatus** aufgelistet.
- In Tabelle D.14 sind die Ergebnisse der ANOVA-Modelle je nach Niveau des **Vorwissens** aufgelistet.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Gender: weiblich					
W	3.85	0.152, 0.051	0.01	6.07 (W)	6.95 (w)
S	2.67	0.206, 0.103	0.00	6.13 (S)	6.89 (s)
P	0.02	0.890, 0.890	0.00	6.53 (P)	6.41 (p)
W : S	0.80	0.370	0.00		
W : P	0.00	0.958	0.00		
S : P	0.79	0.375	0.00		
W : S : P	0.02	0.878	0.00		
Gender: männlich					
W	2.02	0.470, 0.157	0.00	5.43 (W)	6.34 (w)
S	0.08	1.000, 0.775	0.00	5.99 (S)	5.79 (s)
P	0.04	1.000, 0.851	0.00	5.83 (P)	5.93 (p)
W : S	0.00	0.998	0.00		
W : P	0.10	0.752	0.00		
S : P	0.88	0.349	0.00		
W : S : P	1.05	0.306	0.00		
Gender: divers					
W	0.54	0.938, 0.469	-0.02	3.50 (W)	5.21 (w)
S	1.21	0.849, 0.283	0.01	5.60 (S)	3.00 (s)
P	0.10	0.938, 0.757	-0.03	3.62 (P)	4.82 (p)
W : S	0.22	0.647	-0.03		
W : P	0.42	0.525	-0.02		
S : P	0.05	0.826	-0.03		
W : S : P	1.30	0.266	0.01		

Tabelle D.8.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung je nach Gender ergänzend zur Abbildung 7.6 im Abschnitt 7.1.4 (Modellparameter (7,463) bzw. (7,273) bzw. (7,22) für weiblich bzw. männlich bzw. divers). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Leseverständnis: niedrig					
W	6.50	0.034, 0.011	0.02*	4.75 (W)	6.25 (w)
S	0.59	0.678, 0.442	0.00	5.20 (S)	5.78 (s)
P	0.92	0.678, 0.339	0.00	5.67 (P)	5.26 (p)
W : S	0.01	0.931	0.00		
W : P	0.11	0.743	0.00		
S : P	0.03	0.852	0.00		
W : S : P	0.37	0.542	0.00		
Leseverständnis: hoch					
W	0.72	1.000, 0.398	0.00	6.57 (W)	7.03 (w)
S	0.00	1.000, 0.946	0.00	6.81 (S)	6.75 (s)
P	0.26	1.000, 0.610	0.00	6.64 (P)	6.91 (p)
W : S	1.28	0.260	0.00		
W : P	0.00	0.985	0.00		
S : P	0.01	0.904	0.00		
W : S : P	0.70	0.402	0.00		

Tabelle D.9.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung je nach Leseverständnis-Niveau ergänzend zur Abbildung 7.7 im Abschnitt 7.1.4 (Modellparameter (7,343) bzw. (7,421) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Lesegeschwindigkeit: niedrig					
W	11.46	0.002, 0.001	0.02**	5.44 (W)	7.10 (w)
S	0.57	0.900, 0.450	0.00	5.99 (S)	6.45 (s)
P	0.06	0.900, 0.800	0.00	6.26 (P)	6.17 (p)
W : S	0.04	0.836	0.00		
W : P	0.04	0.843	0.00		
S : P	1.33	0.250	0.00		
W : S : P	0.00	0.949	0.00		
Lesegeschwindigkeit: hoch					
W	0.00	1.000, 0.998	0.00	6.13 (W)	6.12 (w)
S	0.00	1.000, 0.998	0.00	6.16 (S)	6.09 (s)
P	0.03	1.000, 0.871	0.00	6.05 (P)	6.19 (p)
W : S	0.25	0.617	0.00		
W : P	0.06	0.813	0.00		
S : P	1.12	0.291	0.00		
W : S : P	0.14	0.710	0.00		

Tabelle D.10.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung je nach Lesegeschwindigkeit-Niveau ergänzend zur Abbildung 7.7 im Abschnitt 7.1.4 (Modellparameter (7,416) bzw. (7,350) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Lesegenauigkeit: niedrig					
W	2.85	0.280, 0.093	0.01	3.70 (W)	4.97 (w)
S	0.45	0.503, 0.503	0.00	3.98 (S)	4.54 (s)
P	1.61	0.412, 0.206	0.00	4.67 (P)	3.85 (p)
W : S	0.00	0.950	0.00		
W : P	1.27	0.261	0.00		
S : P	0.18	0.670	0.00		
W : S : P	2.40	0.123	0.01		
Lesegenauigkeit: hoch					
W	2.78	0.287, 0.096	0.00	6.55 (W)	7.26 (w)
S	0.78	0.752, 0.376	0.00	6.74 (S)	7.06 (s)
P	0.16	0.752, 0.690	0.00	6.78 (P)	6.99 (p)
W : S	1.06	0.303	0.00		
W : P	0.61	0.433	0.00		
S : P	0.25	0.620	0.00		
W : S : P	1.97	0.161	0.00		

Tabelle D.11.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung je nach Lesegenauigkeit-Niveau ergänzend zur Abbildung 7.7 im Abschnitt 7.1.4 (Modellparameter (7,199) bzw. (7,565) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Realschule					
W	10.23	0.005, 0.002	0.03**	4.67 (W)	6.61 (w)
S	0.70	0.665, 0.405	0.00	5.48 (S)	6.08 (s)
P	0.94	0.665, 0.332	0.00	6.06 (P)	5.49 (p)
W : S	0.65	0.422	0.00		
W : P	0.87	0.351	0.00		
S : P	0.31	0.578	0.00		
W : S : P	1.74	0.188	0.00		
Gymnasium					
W	0.83	1.000, 0.363	0.00	6.23 (W)	6.68 (w)
S	0.00	1.000, 0.981	0.00	6.42 (S)	6.42 (s)
P	0.36	1.000, 0.551	0.00	6.24 (P)	6.60 (p)
W : S	0.01	0.930	0.00		
W : P	0.31	0.575	0.00		
S : P	0.00	0.968	0.00		
W : S : P	0.20	0.652	0.00		

Tabelle D.12.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung ergänzend zur Abbildung 7.8 im Abschnitt 7.1.4 (Modellparameter (7,285) bzw. (7,481) für Realschule bzw. Gymnasium). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
<b>8. Jahrgangsstufe</b>					
W	1.21	0.821, 0.274	0.00	5.07 (W)	6.05 (w)
S	0.01	1.000, 0.932	-0.01	5.48 (S)	5.69 (s)
P	0.39	1.000, 0.532	0.00	5.86 (P)	5.30 (p)
W : S	0.98	0.323	0.00		
W : P	0.01	0.929	-0.01		
S : P	0.24	0.626	-0.01		
W : S : P	0.48	0.491	0.00		
<b>9. Jahrgangsstufe</b>					
W	4.76	0.088, 0.029	0.01	5.89 (W)	6.81 (w)
S	0.29	1.000, 0.589	0.00	6.20 (S)	6.43 (s)
P	0.08	1.000, 0.778	0.00	6.24 (P)	6.38 (p)
W : S	0.04	0.835	0.00		
W : P	0.00	0.957	0.00		
S : P	0.04	0.851	0.00		
W : S : P	0.00	0.986	0.00		

Tabelle D.13.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung je nach Jahrgangsstufe ergänzend zur Abbildung 7.8 im Abschnitt 7.1.4 (Modellparameter (7,139) bzw. (7,627) für Jahrgangsstufe 8 bzw. 9). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.14.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Vorwissensniveau: niedrig					
W	9.23	0.008, 0.003	0.02**	7.38 (W)	8.86 (w)
S	0.01	0.946, 0.928	0.00	8.08 (S)	8.19 (s)
P	0.52	0.946, 0.473	0.00	8.29 (P)	7.98 (p)
W : S	4.56	0.033	0.01*		
W : P	0.37	0.542	0.00		
S : P	0.53	0.468	0.00		
W : S : P	0.37	0.543	0.00		
Vorwissensniveau: hoch					
W	0.01	1.000, 0.916	0.00	4.62 (W)	4.61 (w)
S	0.24	1.000, 0.627	0.00	4.48 (S)	4.77 (s)
P	0.48	1.000, 0.486	0.00	4.45 (P)	4.78 (p)
W : S	1.19	0.277	0.00		
W : P	0.38	0.539	0.00		
S : P	0.22	0.642	0.00		
W : S : P	0.06	0.804	0.00		

Tabelle D.14.: ANOVA-Ergebnisse für die Abhängigkeit des Gain-Scores von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen **Wortwahl**, **Satzstrukturen** und **Personalisierung** in Teilstichproben je nach Vorwissensniveau ergänzend zur Abbildung 7.9 im Abschnitt 7.1.4. Es werden die Teststatistik F für die Modellparameter (7,338) (niedrig) und (7,428) (hoch) berichtet sowie der p-Wert in unkorrigierter Form und in  $\alpha$ -korrigierter Form (Bonferroni-Holm; siehe 6.4.3). In der Spalte für die Effektstärke  $\omega^2$  sind signifikante Werte mit Sternchen markiert. Ganz rechts sind Gain-Score-Mittelwerte je nach sprachlicher Gestaltung der Texte eingetragen.

## **D.7 PCA für den Bewertungsfragebogen**

In Tabelle D.15 werden die Faktorladungen der 36 in die Analyse einbezogenen Items des Bewertungsfragebogens auf die sechs Hauptkomponenten dargestellt. Bei Faktorladungen wird üblicherweise ab einem Richtwert von 0.3 von einer merklichen Ladung auf eine bestimmte Komponente gesprochen (vergleiche Stevens, 2009).

Item	Kogn. Anr.	Lesbark.	Log. Aufb.	Elabor.	Vorw.	Involv.	Komm.	Kompl.
07	0.78	-0.07	-0.05	0.02	-0.10	0.09	0.62	1.1
16	0.77	-0.03	-0.02	0.02	0.07	0.05	0.65	1.0
06	0.74	0.08	-0.11	-0.02	0.01	0.07	0.52	1.1
30	0.66	0.08	0.18	0.01	0.00	-0.26	0.49	1.5
05	0.65	-0.09	0.06	-0.01	0.08	0.16	0.64	1.2
28	0.51	0.01	0.18	-0.01	0.21	-0.01	0.49	1.6
08	0.45	-0.16	0.11	0.10	-0.06	0.44	0.60	2.5
32	0.41	-0.10	-0.04	-0.35	0.21	-0.12	0.50	2.9
31	0.03	0.75	0.01	-0.06	-0.01	0.00	0.52	1.0
41	-0.04	0.73	0.10	-0.04	0.01	0.01	0.48	1.0
21	-0.02	0.73	-0.04	0.02	-0.01	0.10	0.58	1.0
23	0.04	0.64	0.22	0.11	0.00	-0.17	0.47	1.5
10	0.00	-0.56	0.32	0.00	0.03	0.04	0.56	1.6
02	0.02	0.49	-0.13	0.29	-0.03	0.00	0.55	1.8
15	-0.26	0.45	-0.09	0.13	-0.07	0.07	0.50	2.1
09	-0.05	0.44	-0.32	0.11	-0.07	0.08	0.54	2.2
17	0.00	0.03	0.77	0.06	-0.02	-0.01	0.56	1.0
04	-0.06	0.00	0.74	0.09	0.06	0.12	0.54	1.1
36	0.18	0.00	0.55	-0.22	0.03	0.01	0.56	1.6
24	0.07	-0.12	0.54	-0.16	0.09	0.04	0.54	1.4
13	0.10	-0.03	0.53	-0.24	0.02	0.06	0.52	1.5
35	0.07	-0.09	0.50	-0.17	0.05	0.15	0.51	1.6
29	0.03	-0.05	0.47	-0.32	0.10	0.06	0.53	1.9
44	-0.03	-0.24	0.33	-0.19	0.22	0.05	0.47	3.4
19	-0.02	0.04	-0.07	0.79	0.09	-0.03	0.65	1.1
38	-0.04	0.03	-0.06	0.74	-0.05	0.14	0.64	1.1
27	-0.02	0.03	-0.06	0.73	-0.10	-0.03	0.66	1.1
34	0.04	0.23	-0.05	0.59	-0.02	0.07	0.57	1.4
18	0.14	0.01	0.32	0.53	0.00	-0.33	0.41	2.6
12	-0.13	0.29	-0.29	0.35	-0.04	0.06	0.61	3.4
14	-0.01	-0.06	-0.02	0.10	0.88	-0.02	0.73	1.0
42	0.01	0.08	0.02	0.01	0.83	0.03	0.67	1.0
43	0.02	-0.05	-0.01	-0.40	0.54	0.05	0.65	1.9
01	0.06	-0.02	0.19	0.06	0.00	0.71	0.59	1.2
03	0.09	0.07	0.01	0.01	0.08	0.66	0.49	1.1
40	0.36	0.08	0.03	-0.11	0.02	0.47	0.49	2.1

Tabelle D.15.: Faktorladungen der Items des Bewertungsfragebogens auf die sechs extrahierten Komponenten *Kognitive Anregung*, *Lesbarkeit*, *Logischer Aufbau*, *Elaboriertheit*, *Angemessener Bezug zum Vorwissen* und *Involviertheit*. In der Spalte „Komm.“ sind die Kommunalitäten, in der Spalte „Kompl.“ sind die Komplexitäten der Items eingetragen. Die Items sind so angeordnet, dass zunächst alle Items mit höchster Faktorladung auf die Facette *Kognitive Anregung* aufgelistet sind (Items 7, 16, ..., 32), gefolgt von denen mit höchster Faktorladung auf *Lesbarkeit* (Items 31, 41, ..., 9) und so weiter.

Gender	Teilstichprobengrößen		
	weiblich 457	männlich 271	divers 27
Lesefähigkeiten	niedrig		hoch
Verständnis	336		417
Geschwindigkeit	409		346
Genauigkeit	195		558
Schulart	Realschule 286		Gymnasium 469
Jahrgangsstufe	8. Klasse 145		9. Klasse 610
Vorwissen	niedrig 331		hoch 420

Tabelle D.16.: Übersicht über die Verteilung der Gesamtstichprobe in die einzelnen Subgruppen für die Betrachtung des globalen Verständlichkeitseindrucks in Abhängigkeit von der sprachlichen Textgestaltung.

## D.8 Moderierte Zusammenhangsanalysen: sprachliche Gestaltung – Globaleindruck

Analog zum Zusammenhang zwischen der sprachlichen Gestaltung und der Behaltensleistung wurde auch der Zusammenhang zwischen sprachlicher Gestaltung und globalem Verständlichkeitseindruck in verschiedenen Subgruppen analysiert. Die Stichprobengrößen sind in Tabelle D.16 zusammengefasst.

- In Tabelle D.17 sind die Ergebnisse der ANOVA-Modelle je nach **Gender** aufgelistet.
- In den Tabellen D.18, D.19 und D.20 sind die Ergebnisse der ANOVA-Modelle je nach **Lesefähigkeiten** aufgelistet.
- In den Tabellen D.21 und D.22 sind die Ergebnisse der ANOVA-Modelle je nach **Schulstatus** aufgelistet.
- In Tabelle D.23 sind die Ergebnisse der ANOVA-Modelle je nach Niveau des **Vorwissens** aufgelistet.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Gender: weiblich					
W	7.99	0.015, 0.005	0.02*	3.66 (W)	3.91 (w)
S	5.21	0.046, 0.023	0.01*	3.68 (S)	3.89 (s)
P	2.78	0.096, 0.096	0.00	3.84 (P)	3.71 (p)
W : S	0.26	0.607	0.00		
W : P	0.02	0.889	0.00		
S : P	1.19	0.276	0.00		
W : S : P	4.63	0.032	0.01*		
Gender: männlich					
W	1.12	0.872, 0.291	0.00	3.83 (W)	3.94 (w)
S	0.10	1.000, 0.746	0.00	3.87 (S)	3.90 (s)
P	0.21	1.000, 0.648	0.00	3.91 (P)	3.86 (p)
W : S	0.00	0.989	0.00		
W : P	1.38	0.241	0.00		
S : P	0.05	0.818	0.00		
W : S : P	4.37	0.037	0.01*		
Gender: divers					
W	0.00	1.000, 1.000	-0.04	3.29 (W)	3.23 (w)
S	1.46	0.723, 0.241	0.02	2.85 (S)	3.64 (s)
P	0.53	0.953, 0.477	-0.02	3.30 (P)	3.24 (p)
W : S	1.46	0.241	0.02		
W : P	0.06	0.811	-0.04		
S : P	0.94	0.345	0.00		
W : S : P	0.23	0.634	-0.03		

Tabelle D.17.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Gender (Modellparameter (7,455) bzw. (7,270) bzw. (7,26) für weiblich bzw. männlich bzw. divers). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

D. Statistischer Anhang

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Leseverständnis: niedrig					
W	1.87	0.518, 0.173	0.00	3.65 (W)	3.80 (w)
S	1.15	0.567, 0.283	0.00	3.66 (S)	3.78 (s)
P	0.07	0.790, 0.790	0.00	3.72 (P)	3.72 (p)
W : S	0.11	0.737	0.00		
W : P	0.45	0.501	0.00		
S : P	0.05	0.818	0.00		
W : S : P	0.35	0.553	0.00		
Leseverständnis: hoch					
W	6.93	0.026, 0.009	0.01*	3.74 (W)	3.98 (w)
S	4.22	0.081, 0.041	0.01	3.75 (S)	3.96 (s)
P	3.51	0.081, 0.062	0.01	3.95 (P)	3.76 (p)
W : S	1.17	0.279	0.00		
W : P	0.01	0.938	0.00		
S : P	0.03	0.873	0.00		
W : S : P	0.14	0.706	0.00		

Tabelle D.18.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Leseverständnis-Niveau (Modellparameter (7,335) bzw. (7,416) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Lesegeschwindigkeit: niedrig					
W	4.38	0.111, 0.037	0.01	3.69 (W)	3.88 (w)
S	2.70	0.202, 0.101	0.00	3.70 (S)	3.85 (s)
P	1.18	0.279, 0.279	0.00	3.82 (P)	3.73 (p)
W : S	0.01	0.941	0.00		
W : P	2.17	0.141	0.00		
S : P	0.00	0.975	0.00		
W : S : P	1.70	0.193	0.00		
Lesegeschwindigkeit: hoch					
W	3.79	0.157, 0.052	0.01	3.72 (W)	3.93 (w)
S	3.55	0.157, 0.060	0.01	3.72 (S)	3.92 (s)
P	1.91	0.168, 0.168	0.00	3.89 (P)	3.75 (p)
W : S	0.94	0.332	0.00		
W : P	0.70	0.403	0.00		
S : P	0.60	0.440	0.00		
W : S : P	0.14	0.713	0.00		

Tabelle D.19.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Lesegeschwindigkeit-Niveau (Modellparameter (7,408) bzw. (7,345) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Lesegenauigkeit: niedrig					
W	0.70	0.810, 0.405	0.00	3.57 (W)	3.74 (w)
S	3.99	0.142, 0.047	0.02	3.50 (S)	3.77 (s)
P	0.00	0.997, 0.997	-0.01	3.64 (P)	3.65 (p)
W : S	1.44	0.231	0.00		
W : P	0.01	0.909	-0.01		
S : P	0.02	0.876	-0.01		
W : S : P	0.07	0.790	0.00		
Lesegenauigkeit: hoch					
W	6.43	0.035, 0.011	0.01*	3.75 (W)	3.95 (w)
S	3.32	0.109, 0.069	0.00	3.77 (S)	3.93 (s)
P	3.72	0.109, 0.054	0.00	3.92 (P)	3.77 (p)
W : S	1.59	0.208	0.00		
W : P	0.16	0.693	0.00		
S : P	0.10	0.750	0.00		
W : S : P	0.22	0.641	0.00		

Tabelle D.20.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Lesegenauigkeit-Niveau (Modellparameter (7,194) bzw. (7,557) für das niedrige bzw. hohe Niveau). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Realschule					
W	5.04	0.077, 0.026	0.01	3.46 (W)	3.71 (w)
S	0.83	0.724, 0.362	0.00	3.55 (S)	3.65 (s)
P	0.16	0.724, 0.687	0.00	3.62 (P)	3.59 (p)
W : S	0.87	0.351	0.00		
W : P	0.02	0.897	0.00		
S : P	0.44	0.509	0.00		
W : S : P	0.14	0.713	0.00		
Gymnasium					
W	7.64	0.018, 0.006	0.01*	3.82 (W)	4.06 (w)
S	7.11	0.018, 0.008	0.01*	3.80 (S)	4.03 (s)
P	3.23	0.073, 0.073	0.00	3.99 (P)	3.84 (p)
W : S	0.02	0.897	0.00		
W : P	0.11	0.745	0.00		
S : P	0.00	0.968	0.00		
W : S : P	0.78	0.377	0.00		

Tabelle D.21.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Schulart (Modellparameter (7,468) bzw. (7,285) für Gymnasium bzw. Realschule). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
<b>8. Jahrgangsstufe</b>					
W	6.94	0.028, 0.009	0.04*	3.26 (W)	3.68 (w)
S	0.56	0.912, 0.456	0.00	3.41 (S)	3.56 (s)
P	0.01	0.913, 0.913	-0.01	3.49 (P)	3.48 (p)
W : S	2.00	0.159	0.01		
W : P	2.06	0.153	0.01		
S : P	1.22	0.272	0.00		
W : S : P	0.26	0.610	-0.01		
<b>9. Jahrgangsstufe</b>					
W	4.34	0.075, 0.038	0.01	3.80 (W)	3.96 (w)
S	5.73	0.051, 0.017	0.01	3.78 (S)	3.96 (s)
P	3.09	0.079, 0.079	0.00	3.94 (P)	3.81 (p)
W : S	0.00	0.949	0.00		
W : P	1.40	0.237	0.00		
S : P	1.17	0.281	0.00		
W : S : P	0.71	0.401	0.00		

Tabelle D.22.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung je nach Jahrgangsstufe (Modellparameter (7,144) bzw. (7,609) für Jahrgangsstufe 8 bzw. 9). Für weitere Einzelheiten siehe die Unterschrift der Tabelle D.23.

unabhängige Variable	F	p ( $\alpha$ -korr.), p (unkorr.)	$\omega^2$	Mittelwerte	
Vorwissensniveau: niedrig					
W	10.82	0.003, 0.001	0.03**	3.52 (W)	3.87 (w)
S	1.78	0.366, 0.183	0.00	3.62 (S)	3.78 (s)
P	0.36	0.549, 0.549	0.00	3.73 (P)	3.67 (p)
W : S	0.96	0.329	0.00		
W : P	0.21	0.647	0.00		
S : P	0.49	0.487	0.00		
W : S : P	0.20	0.659	0.00		
Vorwissensniveau: hoch					
W	1.25	0.265, 0.265	0.00	3.83 (W)	3.92 (w)
S	4.00	0.138, 0.046	0.01	3.77 (S)	3.97 (s)
P	2.81	0.189, 0.095	0.00	3.95 (P)	3.79 (p)
W : S	0.00	0.971	0.00		
W : P	0.04	0.833	0.00		
S : P	0.01	0.933	0.00		
W : S : P	1.89	0.170	0.00		

Tabelle D.23.: ANOVA-Ergebnisse für die Abhängigkeit der Bewertung zum Global-Item von der sprachlichen Gestaltung der Texte hinsichtlich der Variablen **Wortwahl**, **Satzstrukturen** und **Personalisierung** in Teilstichproben je nach Vorwissen ergänzend zu Tabelle 7.14 im Abschnitt 7.2.3. Es werden die Teststatistik F für die Modellparameter (7,330) (niedrig) und (7,439) (hoch) berichtet sowie der p-Wert in unkorrigierter Form und in  $\alpha$ -korrigierter Form (Bonferroni-Holm; siehe 6.4.3). In der Spalte für die Effektstärke  $\omega^2$  sind signifikante Werte mit Sternchen markiert. Ganz rechts sind Mittelwerte der Global-Bewertung je nach sprachlicher Gestaltung der Texte eingetragen.



# E | ZUM SEMINAR „SPRACHE IM PHYSIKUNTERRICHT“

## **Seminartitel**

Idee, Medien, Gestaltung – Kreativer Physikunterricht.

*Thema:* Sprache im Fach Physik – Verständliche Textgestaltung.

## **Seminarbeschreibung**

### **Wie muss ein Text aussehen, damit man ihn versteht?**

In diesem Seminar werden wir uns damit beschäftigen, an welchen Stellschrauben Sie als Lehrkraft drehen können, um Schüler\*innen beim verstehenden Lesen in den Naturwissenschaften zu unterstützen.

Anhand von Textbeispielen und eigenen Textproduktionen wollen wir uns mit der Verständlichkeit von physikalischen Sachtexten beschäftigen. Dazu betrachten wir auf der Basis von Forschungsergebnissen Themen rund um das Leseverstehen: Wie läuft der Leseprozess ab? Welche Rolle spielt (insbes. geschriebene, fachspezifische) Sprache im Unterricht der Naturwissenschaften? Welche stilistischen Merkmale auf der Textoberfläche erzeugen Schwierigkeit, welche erleichtern das verstehende Lesen? Welche unterstützenden Möglichkeiten bieten sich auf der Ebene der Textgliederung und -struktur?

## Seminarablauf

Parallel zum Theorie- und Übungsteil des Seminars erhalten die Studierenden am Ende jeder Sitzung Gelegenheit, in Zweier- bis Dreiergruppen an einer eigenen Textproduktion zu arbeiten. Zusätzlich sind die Studierenden aufgefordert, auch zuhause an diesen Texten zu arbeiten. Die entstandenen Texte werden am Ende des Seminars vor dem Hintergrund der erarbeiteten Gütekriterien für verständliche Texte im Plenum besprochen.

### Semesterphase 01

1. Organisatorisches inklusive gegenseitigem (interaktiv-spielerischem) Kennenlernen und inklusive Vorstellung der Arbeitsmedien (GRIPS<sup>1</sup>, Miro<sup>2</sup>)
2. Die Bedeutsamkeit und Umsetzung sprachbewussten Unterrichts: Wissensvermittlung anhand von Texten, Teilprozesse des Leseverstehens, Verstehen/Nicht-Verstehen/Missverstehen, Methodenswerkzeuge für sprachexpliziten Unterricht und für Leseunterricht
3. Inhaltliche Erarbeitung und fachliche Analyse des Textthemas für die eigenständigen Textproduktionen

### Semesterphase 02

1. Theorie zur Textverständlichkeit: Modelle für verständliche Texte, fachsprachliche Charakteristika, Adressat\*innenorientierung
2. Eigenständige Textproduktionen ausgehend von inhaltlichen Mindmaps

### Semesterphase 03

1. Spezielle sprachliche Mittel: Personifizierung, (un)persönliche Wendungen, ...
2. Unterrichtsbeobachtungen mit Fokus auf die Wirkungen bestimmter sprachlicher Gestaltungsmittel (mithilfe der Videoplattform <https://unterrichtonline.org/>)
3. Fertigstellung der eigenständigen Textproduktionen und Evaluation im Plenum (unter Verwendung etablierter Regeln zum Feedback-Geben und Feedback-Empfangen)

---

<sup>1</sup><https://elearning.uni-regensburg.de/>

<sup>2</sup><https://miro.com/product-overview/>

Bisher erschienene Bände der Reihe  
*Studien zum Physik- und Chemielernen*

ISSN 1614-8967

**Vollständige Übersicht auf unserer Website**



<https://www.logos-verlag.de/spcl>

**Aktuelle Bände**

- 300 Amany Annaggar (2020): A Design Framework for Video Game-Based Gamification Elements to Assess Problem-solving Competence in Chemistry Education  
ISBN 978-3-8325-5150-6 52.00 EUR
- 301 Alexander Engl (2020): CHEMIE PUR – Unterrichten in der Natur. Entwicklung und Evaluation eines kontextorientierten Unterrichtskonzepts im Bereich Outdoor Education zur Änderung der Einstellung zu „Chemie und Natur“  
ISBN 978-3-8325-5174-2 59.00 EUR (open access)
- 302 Christin Marie Sajons (2020): Kognitive und motivationale Dynamik in Schülerlaboren. Kontextualisierung, Problemorientierung und Autonomieunterstützung der didaktischen Struktur analysieren und weiterentwickeln  
ISBN 978-3-8325-5155-1 56.00 EUR (open access)
- 303 Philipp Bitzenbauer (2020): Quantenoptik an Schulen. Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik  
ISBN 978-3-8325-5123-0 59.00 EUR (open access)
- 304 Malte Ubben (2020): Typisierung des Verständnisses mentaler Modelle mittels empirischer Datenerhebung am Beispiel der Quantenphysik  
ISBN 978-3-8325-5181-0 43.50 EUR (open access)
- 305 Wiebke Hinrike Kuske-Janßen (2020): Sprachlicher Umgang mit Formeln von LehrerInnen im Physikunterricht am Beispiel des elektrischen Widerstandes in Klassenstufe 8  
ISBN 978-3-8325-5183-4 47.50 EUR (open access)
- 306 Kai Bliesmer (2020): Physik der Küste für außerschulische Lernorte. Eine Didaktische Rekonstruktion  
ISBN 978-3-8325-5190-2 58.00 EUR (open access)
- 307 Nikola Schild (2021): Eignung von domänenspezifischen Studieneingangsvariablen als Prädiktoren für Studienerfolg im Fach und Lehramt Physik  
ISBN 978-3-8325-5226-8 42.00 EUR (open access)

- 308 Daniel Aeverbeck (2021): Zum Studienerfolg in der Studieneingangsphase des Chemie-  
studiums. Der Einfluss kognitiver und affektiv-motivationaler Variablen  
ISBN 978-3-8325-5227-5 51.00 EUR
- 309 Martina Strübe (2021): Modelle und Experimente im Chemieunterricht. Eine Video-  
studie zum fachspezifischen Lehrerwissen und -handeln  
ISBN 978-3-8325-5245-9 45.50 EUR
- 310 Wolfgang Becker (2021): Auswirkungen unterschiedlicher experimenteller Repräsen-  
tationen auf den Kenntnisstand bei Grundschulkindern  
ISBN 978-3-8325-5255-8 50.00 EUR
- 311 Marvin Rost (2021): Modelle als Mittel der Erkenntnisgewinnung im Chemieunter-  
richt der Sekundarstufe I. Entwicklung und quantitative Dimensionalitätsanalyse eines  
Testinstruments aus epistemologischer Perspektive  
ISBN 978-3-8325-5256-5 44.00 EUR (open access)
- 312 Christina Kobl (2021): Förderung und Erfassung der Reflexionskompetenz im Fach  
Chemie  
ISBN 978-3-8325-5259-6 41.00 EUR (open access)
- 313 Ann-Kathrin Beretz (2021): Diagnostische Prozesse von Studierenden des Lehramts.  
eine Videostudie in den Fächern Physik und Mathematik  
ISBN 978-3-8325-5288-6 45.00 EUR (open access)
- 314 Judith Breuer (2021): Implementierung fachdidaktischer Innovationen durch das An-  
gebot materialgestützter Unterrichtskonzeptionen. Fallanalysen zum Nutzungsverhalten  
von Lehrkräften am Beispiel des Münchener Lehrgangs zur Quantenmechanik  
ISBN 978-3-8325-5293-0 50.50 EUR (open access)
- 315 Michaela Oettle (2021): Modellierung des Fachwissens von Lehrkräften in der Teil-  
chenphysik. Eine Delphi-Studie  
ISBN 978-3-8325-5305-0 57.50 EUR (open access)
- 316 Volker Brüggemann (2021): Entwicklung und Pilotierung eines adaptiven Multistage-  
Tests zur Kompetenzerfassung im Bereich naturwissenschaftlichen Denkens  
ISBN 978-3-8325-5331-9 40.00 EUR (open access)
- 317 Stefan Müller (2021): Die Vorläufigkeit und soziokulturelle Eingebundenheit natur-  
wissenschaftlicher Erkenntnisse. Kritische Reflexion, empirische Befunde und fachdi-  
daktische Konsequenzen für die Chemielehrer\*innenbildung  
ISBN 978-3-8325-5343-2 63.00 EUR
- 318 Laurence Müller (2021): Alltagsentscheidungen für den Chemieunterricht erkennen  
und Entscheidungsprozesse explorativ begleiten  
ISBN 978-3-8325-5379-1 59.00 EUR
- 319 Lars Ehlert (2021): Entwicklung und Evaluation einer Lehrkräftefortbildung zur Pla-  
nung von selbstgesteuerten Experimenten  
ISBN 978-3-8325-5393-7 41.50 EUR (open access)

- 320 Florian Seiler (2021): Entwicklung und Evaluation eines Seminarkonzepts zur Förderung der experimentellen Planungskompetenz von Lehramtsstudierenden im Fach Chemie  
ISBN 978-3-8325-5397-5 47.50 EUR (open access)
- 321 Nadine Boele (2021): Entwicklung eines Messinstruments zur Erfassung der professionellen Unterrichtswahrnehmung von (angehenden) Chemielehrkräften hinsichtlich der Lernunterstützung  
ISBN 978-3-8325-5402-6 46.50 EUR
- 322 Franziska Zimmermann (2022): Entwicklung und Evaluation digitalisierungsbezogener Kompetenzen von angehenden Chemielehrkräften  
ISBN 978-3-8325-5410-1 49.50 EUR
- 323 Lars-Frederik Weiß (2021): Der Flipped Classroom in der Physik-Lehre. Empirische Untersuchungen in Schule und Hochschule  
ISBN 978-3-8325-5418-7 51.00 EUR
- 324 Tilmann Steinmetz (2021): Kumulatives Lehren und Lernen im Lehramtsstudium Physik. Theorie und Evaluation eines Lehrkonzepts  
ISBN 978-3-8325-5421-7 51.50 EUR
- 325 Kübra Nur Celik (2022): Entwicklung von chemischem Fachwissen in der Sekundarstufe I. Validierung einer Learning Progression für die Basiskonzepte „Struktur der Materie“, „Chemische Reaktion“ und „Energie“ im Kompetenzbereich „Umgang mit Fachwissen“  
ISBN 978-3-8325-5431-6 55.00 EUR
- 326 Matthias Ungermann (2022): Förderung des Verständnisses von Nature of Science und der experimentellen Kompetenz im Schüler\*innen-Labor Physik in Abgrenzung zum Regelunterricht  
ISBN 978-3-8325-5442-2 55.50 EUR
- 327 Christoph Hoyer (2022): Multimedial unterstütztes Experimentieren im webbasierten Labor zur Messung, Visualisierung und Analyse des Feldes eines Permanentmagneten  
ISBN 978-3-8325-5453-8 45.00 EUR
- 328 Tobias Schüttler (2022): Schülerlabore als interessefördernde authentische Lernorte für den naturwissenschaftlichen Unterricht nutzen  
ISBN 978-3-8325-5454-5 50.50 EUR
- 329 Christopher Kurth (2022): Die Kompetenz von Studierenden, Schülerschwierigkeiten beim eigenständigen Experimentieren zu diagnostizieren  
ISBN 978-3-8325-5457-6 58.50 EUR
- 330 Dagmar Michna (2022): Inklusiver Anfangsunterricht Chemie. Entwicklung und Evaluation einer Unterrichtseinheit zur Einführung der chemischen Reaktion  
ISBN 978-3-8325-5463-7 49.50 EUR
- 331 Marco Seiter (2022): Die Bedeutung der Elementarisierung für den Erfolg von Mechanikunterricht in der Sekundarstufe I  
ISBN 978-3-8325-5471-2 66.00 EUR

- 332 Jörn Hägele (2022): Kompetenzaufbau zum experimentbezogenen Denken und Arbeiten. Videobasierte Analysen zu Aktivitäten und Vorstellungen von Schülerinnen und Schülern der gymnasialen Oberstufe bei der Bearbeitung von fachmethodischer Instruktion  
ISBN 978-3-8325-5476-7      56.50 EUR (open access)
- 333 Erik Heine (2022): Wissenschaftliche Kontroversen im Physikunterricht. Explorationsstudie zum Umgang von Physiklehrkräften und Physiklehramtsstudierenden mit einer wissenschaftlichen Kontroverse am Beispiel der Masse in der Speziellen Relativitätstheorie  
ISBN 978-3-8325-5478-1      48.50 EUR (open access)
- 334 Simon Goertz (2022): Module und Lernzirkel der Plattform FLexKom zur Förderung experimenteller Kompetenzen in der Schulpraxis. Verlauf und Ergebnisse einer Design-Based Research Studie  
ISBN 978-3-8325-5494-1      66.50 EUR
- 335 Christina Toschka (2022): Lernen mit Modellexperimenten. Empirische Untersuchung der Wahrnehmung und des Denkens in Analogien beim Umgang mit Modellexperimenten  
ISBN 978-3-8325-5495-8      50.00 EUR (open access)
- 336 Alina Behrendt (2022): Chemiebezogene Kompetenzen in der Übergangsphase zwischen dem Sachunterricht der Primarstufe und dem Chemieunterricht der Sekundarstufe I  
ISBN 978-3-8325-5498-9      40.50 EUR (open access)
- 337 Manuel Daiber (2022): Entwicklung eines Lehrkonzepts für eine elementare Quantenmechanik. Formuliert mit In-Out Symbolen  
ISBN 978-3-8325-5507-8      48.50 EUR
- 338 Felix Pawlak (2022): Das Gemeinsame Experimentieren (an-)leiten. Eine qualitative Studie zum chemiespezifischen Classroom-Management  
ISBN 978-3-8325-5508-5      46.50 EUR
- 339 Liza Dopatka (2022): Konzeption und Evaluation eines kontextstrukturierten Unterrichtskonzeptes für den Anfangs-Elektrizitätslehreunterricht  
ISBN 978-3-8325-5514-6      69.50 EUR
- 340 Arne Bewersdorff (2022): Untersuchung der Effektivität zweier Fortbildungsformate zum Experimentieren mit dem Fokus auf das Unterrichtshandeln  
ISBN 978-3-8325-5522-1      39.00 EUR (open access)
- 341 Thomas Christoph Münster (2022): Wie diagnostizieren Studierende des Lehramtes physikbezogene Lernprozesse von Schüler\*innen?. Eine Videostudie zur Mechanik  
ISBN 978-3-8325-5534-4      44.50 EUR (open access)
- 342 Ines Komor (2022): Förderung des symbolisch-mathematischen Modellverständnisses in der Physikalischen Chemie  
ISBN 978-3-8325-5546-7      46.50 EUR

- 343 Verena Petermann (2022): Überzeugungen von Lehrkräften zum Lehren und Lernen von Fachinhalten und Fachmethoden und deren Beziehung zu unterrichtsnahem Handeln  
ISBN 978-3-8325-5545-0 47.00 EUR (open access)
- 344 Jana Heinze (2022): Einfluss der sprachlichen Konzeption auf die Einschätzung der Qualität instruktionaler Unterrichtserklärungen im Fach Physik  
ISBN 978-3-8325-5553-5 42.00 EUR (open access)
- 345 Jannis Weber (2022): Mathematische Modellbildung und Videoanalyse zum Lernen der Newtonschen Dynamik im Vergleich  
ISBN 978-3-8325-5566-5 68.00 EUR (open access)
- 346 Fabian Sterzing (2022): Zur Lernwirksamkeit von Erklärvideos in der Physik. Eine Untersuchung in Abhängigkeit von ihrer fachdidaktischen Qualität und ihrem Einbettungsformat  
ISBN 978-3-8325-5576-4 52.00 EUR (open access)
- 347 Lars Greitemann (2022): Wirkung des Tablet-Einsatzes im Chemieunterricht der Sekundarstufe I unter besonderer Berücksichtigung von Wissensvermittlung und Wissenssicherung  
ISBN 978-3-8325-5580-1 50.00 EUR
- 348 Fabian Poengen (2022): Diagnose experimenteller Kompetenzen in der laborpraktischen Chemielehrer\*innenbildung  
ISBN 978-3-8325-5587-0 48.00 EUR
- 349 William Lindlahr (2023): Virtual-Reality-Experimente. Entwicklung und Evaluation eines Konzepts für den forschend-entwickelnden Physikunterricht mit digitalen Medien  
ISBN 978-3-8325-5595-5 49.00 EUR
- 350 Bert Schlüter (2023): Teilnahmemotivation und situationales Interesse von Kindern und Eltern im experimentellen Lernsetting KEMIE  
ISBN 978-3-8325-5598-6 43.00 EUR
- 351 Katharina Nave (2023): Charakterisierung situativer mentaler Modellkomponenten in der Chemie und die Bildung von Hypothesen. Eine qualitative Studie zur Operationalisierung mentaler Modellkomponenten für den Fachbereich Chemie  
ISBN 978-3-8325-5599-3 43.00 EUR
- 352 Anna B. Bauer (2023): Experimentelle Kompetenz Physikstudierender. Entwicklung und erste Erprobung eines performanzorientierten Kompetenzstrukturmodells unter Nutzung qualitativer Methoden  
ISBN 978-3-8325-5625-9 47.00 EUR (open access)
- 353 Jan Schröder (2023): Entwicklung eines Performanztests zur Messung der Fähigkeit zur Unterrichtsplanung bei Lehramtsstudierenden im Fach Physik  
ISBN 978-3-8325-5655-6 46.50 EUR
- 354 Susanne Gerlach (2023): Aspekte einer Fachdidaktik Körperpflege. Ein Beitrag zur Standardentwicklung  
ISBN 978-3-8325-5659-4 45.00 EUR

- 355 Livia Murer (2023): Diagnose experimenteller Kompetenzen beim praktisch-naturwissenschaftlichen Arbeiten. Vergleich verschiedener Methoden und kognitive Validierung eines Testverfahrens  
ISBN 978-3-8325-5657-0 41.50 EUR (open access)
- 356 Andrea Maria Schmid (2023): Authentische Kontexte für MINT-Lernumgebungen. Eine zweiteilige Interventionsstudie in den Fachdidaktiken Physik und Technik  
ISBN 978-3-8325-5605-1 57.00 EUR (open access)
- 357 Julia Ortman (2023): Bedeutung und Förderung von Kompetenzen zum naturwissenschaftlichen Denken und Arbeiten in universitären Praktika  
ISBN 978-3-8325-5670-9 37.00 EUR (open access)
- 358 Axel-Thilo Prokop (2023): Entwicklung eines Lehr-Lern-Labors zum Thema Radioaktivität. Eine didaktische Rekonstruktion  
ISBN 978-3-8325-5671-6 49.50 EUR
- 359 Timo Hackemann (2023): Textverständlichkeit sprachlich variiertes physikbezogener Sachtexte  
ISBN 978-3-8325-5675-4 41.50 EUR (open access)
- 360 Dennis Dietz (2023): Vernetztes Lernen im fächerdifferenzierten und integrierten naturwissenschaftlichen Unterricht aufgezeigt am Basiskonzept Energie. Eine Studie zur Analyse der Wirksamkeit der Konzeption und Implementation eines schulinternen Curriculums für das Unterrichtsfach „Integrierte Naturwissenschaften 7/8“  
ISBN 978-3-8325-5676-1 49.50 EUR
- 361 Ann-Katrin Krebs (2023): Vielfalt im Physikunterricht. Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten  
ISBN 978-3-8325-5672-3 65.50 EUR (open access)
- 362 Simon Kaulhausen (2023): Strukturelle Ursachen für Klausurmisserfolg in Allgemeiner Chemie an der Universität  
ISBN 978-3-8325-5699-0 37.50 EUR (open access)
- 363 Julia Eckoldt (2023): Den (Sach-)Unterricht öffnen. Selbstkompetenzen und motivationale Orientierungen von Lehrkräften bei der Implementation einer Innovation untersucht am Beispiel des Freien Explorierens und Experimentierens  
ISBN 978-3-8325-5663-1 48.50 EUR (open access)
- 364 Albert Teichrow (2023): Physikalische Modellbildung mit dynamischen Modellen  
ISBN 978-3-8325-5710-2 58.50 EUR (open access)
- 365 Sascha Neff (2023): Transfer digitaler Innovationen in die Schulpraxis. Eine explorative Untersuchung zur Förderung der Implementation  
ISBN 978-3-8325-5687-7 59.00 EUR (open access)
- 366 Rahel Schmid (2023): Verständnis von Nature of Science-Aspekten und Umgang mit Fehlern von Schüler\*innen der Sekundarstufe I. Am Beispiel von digital-basierten Lernprozessen im informellen Lernsetting Smartfeld  
ISBN 978-3-8325-5722-5 53.50 EUR (open access)

- 367 Dennis Kirstein (2023): Individuelle Bedingungs- und Risikofaktoren für erfolgreiche Lernprozesse mit kooperativen Experimentieraufgaben im Chemieunterricht. Eine Untersuchung zum Zusammenhang von Lernvoraussetzungen, Lerntätigkeiten, Schwierigkeiten und Lernerfolg beim Experimentieren in Kleingruppen der Sekundarstufe I  
ISBN 978-3-8325-5729-4 50.50 EUR (open access)
- 368 Frauke Düwel (2024): Argumentationslinien in Lehr-Lernkontexten. Potenziale englischer Fachtexte zur Chromatografie und deren hochschuldidaktische Einbindung  
ISBN 978-3-8325-5731-7 63.00 EUR (open access)
- 369 Fabien Güth (2023): Interessenbasierte Differenzierung mithilfe systematisch variiertes Kontextaufgaben im Fach Chemie  
ISBN 978-3-8325-5737-9 48.00 EUR (open access)
- 370 Oliver Grewe (2023): Förderung der professionellen Unterrichtswahrnehmung und Selbstwirksamkeitsüberzeugungen hinsichtlich sprachsensibler Maßnahmen im naturwissenschaftlichen Sachunterricht. Konzeption und Evaluation einer video- und praxisbasierten Lehrveranstaltung im Masterstudium  
ISBN 978-3-8325-5738-6 44.50 EUR (open access)
- 371 Anna Nowak (2023): Untersuchung der Qualität von Selbstreflexionstexten zum Physikunterricht. Entwicklung des Reflexionsmodells REIZ  
ISBN 978-3-8325-5739-3 59.00 EUR (open access)
- 372 Dominique Angela Holland (2023): Bildung für nachhaltige Entwicklung (BNE) kooperativ gestalten. Vergleich monodisziplinärer und interdisziplinärer Kooperation von Lehramtsstudierenden bei der Planung, Durchführung und Reflexion von Online-BNE-Unterricht  
ISBN 978-3-8325-5760-7 47.00 EUR (open access)
- 373 Renan Marcello Vairo Nunes (2024): MINT-Personal an Schulen. Eine Untersuchung der Arbeitssituation und professionellen Kompetenzen von MINT-Lehrkräften verschiedener Ausbildungswege  
ISBN 978-3-8325-5778-2 51.00 EUR (open access)
- 374 Mats Kieserling (2024): Digitalisierung im Chemieunterricht. Entwicklung und Evaluation einer experimentellen digitalen Lernumgebung mit universeller Zugänglichkeit  
ISBN 978-3-8325-5786-7 45.50 EUR
- 375 Cem Aydin Salim (2024): Die Untersuchung adaptiver Lernsettings im Themenbereich „Schwimmen und Sinken“ im naturwissenschaftlichen Unterricht  
ISBN 978-3-8325-5787-4 49.00 EUR (open access)
- 376 Novid Ghassemi (2024): Evaluation eines Lehramtsmasterstudiengangs mit dem Profil Quereinstieg im Fach Physik. Erkenntnisse zu Eingangsbedingungen, professionellen Kompetenzen und Aspekten individueller Angebotsnutzung  
ISBN 978-3-8325-5789-8 41.50 EUR (open access)

- 377 Martina Flurina Cavelti (2024): Entwicklung und Validierung eines Messinstruments zur Erfassung der Schülerkompetenzen im Bereich des wissenschaftlichen Skizzierens im Fach Chemie in der Sekundarstufe I  
ISBN 978-3-8325-5829-1 45.00 EUR (open access)
- 378 Tom Bleckmann (2024): Formatives Assessment auf Basis von maschinellem Lernen. Eine Studie über automatisiertes Feedback zu Concept Maps aus dem Bereich Mechanik  
ISBN 978-3-8325-5842-0 46.50 EUR (open access)
- 379 Jana Marlies Rehberg (2024): Das physikspezifische Mindset zum Studienbeginn. Fragebogenentwicklung und Aufbau einer Online-Intervention  
ISBN 978-3-8325-5850-5 59.50 EUR (open access)
- 380 Florian Trauten (2024): Entwicklung und Evaluation von automatisierten Feedbackschleifen in Online-Aufgaben im Fach Chemie  
ISBN 978-3-8325-5859-8 46.00 EUR (open access)
- 381 Johanna Dejanovikj (2024): Reflexion als Lerngegenstand im Unterricht zur Förderung von Bewertungskompetenz  
ISBN 978-3-8325-5860-4 41.00 EUR (open access)
- 382 Katharina Flieser (2024): Verständlichkeit physikalischer Sachtexte. Untersuchungen zum Wirkungsgefüge zwischen sprachlicher Textgestaltung und der Behaltensleistung sowie der Textwahrnehmung im Schulfach Physik  
ISBN 978-3-8325-5858-1 44.00 EUR (open access)
- 383 Stephanie Neppl (2024): Perspektivenübernahme im Physikunterricht. Explorative Interviewstudie zu einer Seminarkonzeption mit dem Schwerpunkt Perspektivenübernahme bei der Planung von Physikunterricht  
ISBN 978-3-8325-5865-9 48.00 EUR (open access)
- 384 Katja Plicht (2024): Ein Physikübungskonzept zur Förderung der Problemlösekompetenz. Entwicklung und empirische Evaluation eines Strategietrainings auf der Basis von Expertisemerkmalen  
ISBN 978-3-8325-5875-8 45.00 EUR (open access)

Vollständige Übersicht unter: <https://www.logos-verlag.de/spcl>

Alle erschienenen Bücher können unter der angegebenen ISBN direkt online (<http://www.logos-verlag.de>) oder telefonisch (030 - 42 85 10 90) beim Logos Verlag Berlin bestellt werden.

# Studien zum Physik- und Chemielernen

Herausgegeben von Martin Hopf und Mathias Ropohl

Die Reihe umfasst inzwischen eine große Zahl von wissenschaftlichen Arbeiten aus vielen Arbeitsgruppen der Physik- und Chemiedidaktik und zeichnet damit ein gültiges Bild der empirischen physik- und chemiedidaktischen Forschung im deutschsprachigen Raum.

Die Herausgeber laden daher Interessenten zu neuen Beiträgen ein und bitten sie, sich im Bedarfsfall an den Logos-Verlag oder an ein Mitglied des Herausgeberteams zu wenden.

## **Kontaktadressen:**

Univ.-Prof. Dr. Martin Hopf  
Universität Wien,  
Österreichisches Kompetenzzentrum  
für Didaktik der Physik,  
Porzellangasse 4, Stiege 2,  
1090 Wien, Österreich,  
Tel. +43-1-4277-60330,  
e-mail: martin.hopf@univie.ac.at

Prof. Dr. Mathias Ropohl  
Didaktik der Chemie,  
Fakultät für Chemie,  
Universität Duisburg-Essen,  
Schützenbahn 70, 45127 Essen,  
Tel. 0201-183 2704,  
e-mail: mathias.ropohl@uni-due.de

Das verstehende Lesen naturwissenschaftlicher Texte gilt als besonders anspruchsvoll für Lernende. Zur Förderung des Leseverständnisses ist daher geeignetes Textmaterial nötig.

Dieses Forschungsvorhaben untersuchte, inwiefern sprachliche Gestaltungsmittel die Textwirkung auf Lernende beeinflussen. Es wurde ein theoretisches Fundament in Form einer Strukturierung sprachlicher Gestaltungsmittel (z.B. Wortwahl) sowie der Erfassung affektiver Textwirkungsaspekte (z.B. Logik im Aufbau) erarbeitet. In der darauf gründenden empirischen Studie an bayerischen Gymnasien und Realschulen lasen Lernende der 8. und 9. Jahrgangsstufe sprachlich unterschiedlich gestaltete Texte zur elektrischen Spannung. Es wurden die Behaltensleistungen sowie die Textbewertungen der  $N = 787$  Teilnehmenden analysiert.

Die Datenauswertung ergab, dass in erster Linie eine anspruchsgeminderte Wortwahl und darüber hinaus anspruchsgeminderte Satzstrukturen sowie ein ausgeprägter Grad an Personalisierung zu leicht besseren Behaltensleistungen und Textbewertungen führten. Dabei wurde die affektive Textwahrnehmung – allgemein sowie differenziert in einzelne Aspekte – stärker beeinflusst als die Behaltensleistung. Tendenziell profitierten Lernende mit schwächeren Ausgangsbedingungen stärker von einer anspruchsgeminderten Textgestaltung.

Die Studie liefert Erkenntnisse zur Textverständlichkeit in den Naturwissenschaften und schafft eine Grundlage für weitere Forschung insbesondere zu affektiven Textwirkungen und deren lernförderlichen Konsequenzen.

Logos Verlag Berlin

ISBN 978-3-8325-5858-1