
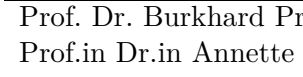


Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Physik

Der Einfluss von Diversitätsdarstellungen in einer digitalen Lernumgebung auf die intrinsische Motivation von Schülerinnen und Schülern ab der 9. Klassenstufe

Masterarbeit

zur Erlangung des akademischen Grades
Master of Education (M. Ed.)

Eingereicht von: Passlida Saila
Geboren am: 
Geboren in: 
Erstgutachter: Prof. Dr. Burkhard Priemer
Zweitgutachterin: Prof.in Dr.in Annette Upmeyer zu Belzen
Eingereicht am: 1. Dezember 2025

Danksagung

An dieser Stelle möchte ich all den Menschen danken, die zum Gelingen dieser Masterarbeit beigetragen und mich auf meinem Weg begleitet haben.

Mein besonderer Dank gilt Prof. Priemer, der mir die Möglichkeit gegeben hat, diese Arbeit im Rahmen des *Humboldt-Explorers*-Projekt zu schreiben. Ebenso danke ich Prof.in Upmeyer zu Belzen, die sich schnell als Zweitprüferin zur Verfügung gestellt hat.

Ein herzliches Dankeschön richte ich an die Schulleitung, Lehrkräfte und Schüler:innen der Untersuchungsschule. Ohne ihre Offenheit, Bereitschaft und Unterstützung wäre die Durchführung dieser Untersuchung nicht möglich gewesen.

Mein Dank gilt außerdem der Arbeitsgruppe *Humboldt-Explorers*, die für mich nicht nur Kolleg:innen, sondern auch Freund:innen geworden sind. Besonders hervorheben möchte ich [REDACTED] der die Animationen für die Erklärvideos erstellte. Sowie [REDACTED] die ihre Stimme für das Voice-over zur Verfügung gestellt hat. Ein besonderer Dank gilt an [REDACTED] die mich in jeder Phase dieser Arbeit begleitet, mir konstruktive Rückmeldungen gegeben, mich stets ermutigt und unterstützt hat.

Auch möchte ich mich bei meinen Freund:innen bedanken, die sich nicht nur als Akteur:innen für die Erklärvideos zur Verfügung gestellt haben, sondern mir mit ihrem Zuspruch und ihrer Geduld sehr geholfen haben.

Von Herzen möchte ich mich auch bei [REDACTED] bedanken, der nicht nur die Arbeit sorgfältig korrektur gelesen, sondern mich auch kontinuierlich motiviert, unterstützt und in herausfordernden Momenten gestärkt hat.

Abschließend möchte ich allen genannten und ungenannten Personen für ihre Unterstützung, ihre Zeit und ihre Hilfsbereitschaft danken. Ohne sie wäre diese Arbeit in dieser Form nicht möglich gewesen.

Inhaltsverzeichnis

Abbildungsverzeichnis	A
Tabellenverzeichnis	B
Zusammenfassung	1
1 Einleitung	2
2 Theoretischer Hintergrund	5
2.1 Konzept der Motivation	5
2.1.1 Motivation	5
2.1.2 Extrinsische und intrinsische Motivation	6
2.1.3 Selbstbestimmungstheorie (Self-Determination Theory, SDT)	7
2.1.4 Intrinsic Motivation Inventory (IMI)	8
2.2 Digitale Lernumgebung	12
2.2.1 Erklärvideo	12
2.2.2 Interaktive digitale Quizaufgaben	14
2.3 Diversität	15
2.3.1 Definition der Diversität	16
2.3.2 Intersektionalität	17
2.3.3 Diversität im Physikunterricht	18
2.3.4 Diversitätsdarstellung in Erklärvideos und Quizaufgaben	20
2.3.5 Diversität und Motivation	21
2.3.6 Altersabhängige Wahrnehmung und Wirkung von Diversität	21
3 Forschungsfragen	22
4 Methode	23
4.1 Forschungsdesign	23
4.2 Settings: Humboldt Explorers	25
4.3 Zielgruppe	25
4.4 Intervention	27
4.4.1 Erklärvideo	27
4.4.2 Interaktive digitale Quizaufgaben	35
4.5 Messinstrument	38
4.5.1 Auswahl der Subskala	38
4.5.2 Eingesetzte Items	39
4.5.3 Gütekriterien der IMI-Items	42
4.5.4 Offene Fragen	44
4.6 Durchführung	44
4.6.1 Ablaufplan der Durchführung	45
4.6.2 Technische Probleme bei der Durchführung	46
4.7 Datenaufbereitung und -auswertung	46
4.7.1 Quantitativ	46
4.7.2 Qualitativ	48
5 Darstellung der Ergebnisse	49
5.1 Reliabilität und Validität der IMI-Subskalen	49
5.2 Quantitative Ergebnisse	52
5.2.1 Über alle Jahrgangsstufen hinweg:	53
5.2.2 Gruppenvergleich über alle Klassenstufen je Subskala:	54
5.2.3 Median der Gruppenunterschiede nach Jahrgangsstufe	57

5.3	Qualitative Ergebnisse	57
6	Diskussion der Ergebnisse	62
6.1	(F1) Wie wirken diversitätsberücksichtigende digitale Lernumgebungen auf die intrinsische Motivation von Schüler:innen?	62
6.2	(F2) Inwiefern unterscheiden sich die Wirkungen diversitätsberücksichtigender digitale Lernumgebungen auf die Motivation von Schüler:innen je nach Jahrgangsstufe?	65
6.3	Limitierung	66
7	Ausblick	68
	Literaturverzeichnis	71
A	Intervention	75
A.1	Qualitätskriterien für ein gutes Erklärvideos nach Kulgemeyer	75
A.2	Voice-over-Skript	76
A.3	Storyboard	79
B	Messinstrumenten: Fragebogen	86
C	Ergebnisse der quantitativen Analyse	89
C.1	Validitäts- und Reliabilitätsprüfung	89
C.2	Levene-Test	91
C.3	Mittelwertaggregation und deskriptiven Statistik	91
D	Kodierung der offenen Fragen	92
D.1	Frage: Was hat euch besonders gut gefallen?	92
D.2	Frage: Was hat euch nicht gefallen?	95
	Selbstständigkeitserklärung	98

Abbildungsverzeichnis

1	Übersicht der Regulationsstile nach der Organismic Integration Theory, übersetzt von Fröhlich (2020, S. 16)	7
2	Four Layers of Diversity nach Gardenwartz und Rowe 2003, erstellt von Krebs (2023, S. 38)	16
3	Zeitlicher Ablauf der Untersuchung pro Durchführung	24
4	Animationsszenen	29
5	Quizaufgabe - Einfluss auf die Gesundheit	36
6	Beschreibung der Bezirke: Wedding (links) und Grunewald (rechts)	36
7	Quizaufgabe - Statische Kennzahlen des Messstation Wedding	37
8	Quizaufgabe: Wedding und Grunewald vergleichen	37
9	Scree-Plot der Eigenwerte-Faktoren-Diagramm	50
10	Faktorenladungen der Items für eine dreifaktorielle Faktorstruktur	51
11	Vergleich der Ausprägungen der drei IMI-Subskalen zwischen Diversitäts- und Kontrollgruppe (inkl. Signifikanzniveaus)	53
12	Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala <i>Interesse</i> .	55
13	Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala <i>Nützlichkeit</i>	56
14	Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala <i>Anstrengung</i> .	56
15	Prozentuale Verteilung der Antwortkategorien beider Gruppen auf die Frage: „Was hat euch besonders gut gefallen?“	59
16	Prozentuale Verteilung der Antwortkategorien beider Gruppen auf die Frage: „Was hat euch nicht gefallen?“	61
17	Voice-over Skript	78
18	Storyboard	85
19	Erhebungsmessinstrument: Fragebogen	88

Tabellenverzeichnis

1	Übersicht der Teilnehmenden nach Klassenstufe, Klassenstärke und Gruppenzugehörigkeit	26
2	Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 1/3	31
3	Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 2/3	32
4	Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 3/3	33
5	Szenen mit KI-generierten Bilder von Personen - Teil 1/2	34
6	Szenen mit KI-generierten Bilder von Personen- Teil 2/2	35
7	Alle Items der Subskala <i>Interesse</i>	40
8	Alle Items der Subskala <i>Anstrengung</i>	41
9	Alle Items der Subskala <i>Nützlichkeit</i>	42
10	Übersicht der ausgewertete Stichprobe nach Klassenstufe und Gruppenzugehörigkeit	46
11	Cronbach's α und McDonald's ω der drei IMI-Subskalen	50
12	Stichproben nach Gruppen und Klassenstufe	54
13	Gruppe mit dem höheren Median in der jeweiligen Jahrgangsstufe, sowie die Medianwerte	57
14	Kriterien für gute Erklärvideos nach Kulgemeyer (2020, S. 423)	75
15	Faktorladung einzelner Items der dreifaktoriellen Lösung	89
16	Rotierte Ladungsmatrix der dreifaktoriellen Lösung	89
17	Varianzaufklärung der dreifaktoriellen Lösung	89
18	Faktorladung einzelnen Items der vierfaktoriellen Lösung	90
19	Rotierte Ladungsmatrix der vierfaktoriellen Lösung	90
20	Varianzaufklärung der vierfaktoriellen Lösung	90
21	Ergebnisse des Levene-Tests zur Überprüfung der Varianzhomogenität	91
22	Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala <i>Interesse</i>	91
23	Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala <i>Anstrengung</i>	91
24	Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala <i>Nützlichkeit</i>	91
25	Kodierung der Antworten der Ohne-Diversität-Gruppe auf die Frage „Was hat euch besonders gut gefallen?“	92
26	Kodierung der Antworten der Diversität-Gruppe auf die Frage „Was hat euch besonders gut gefallen?“	93
27	Kategorien der Antworten beider Gruppe auf die Frage „Was hat euch besonders gut gefallen?“	94
28	Kodierung der Antworten der Ohne-Diversität-Gruppe auf die Frage „Was hat euch nicht gefallen?“	95
29	Kodierung der Antworten der Diversität-Gruppe auf die Frage „Was hat euch nicht gefallen?“	96
30	Kategorien der Antworten der beiden Gruppe auf die Frage „Was hat euch nicht gefallen?“	97

Zusammenfassung

Diese Masterarbeit untersucht im Rahmen eines Mixed-Methods-Designs den Einfluss von Diversitätsdarstellungen in einer digitalen Lernumgebung auf die intrinsische Motivation von Schüler:innen der Klassenstufen 9 bis 12 an einer Schule in Thailand, [REDACTED] im Rahmen des Humboldt-Explorers-Projekts. Die Untersuchung folgt einem experimentellen Design mit zwei Gruppen: einer Lernumgebung mit Diversitätsdarstellung und einer Kontrollversion ohne entsprechende Darstellung. Beide Versionen setzten sich aus einem Erklärvideo zum Thema Feinstaub sowie interaktiven Übungen zusammen. Zur Erfassung der Motivation wurde eine deutschsprachige Übersetzung dreier Subskalen des Intrinsic Motivation Inventory (*Interesse*, *Anstrengung*, *Nützlichkeit*) eingesetzt. Die quantitative Auswertung erfolgte mittels Mittelwertvergleichen und t-Tests, während offene Antworten qualitativ inhaltsanalytisch ausgewertet wurden, um die quantitativen Ergebnisse zu ergänzen und zu differenzieren. Die Ergebnisse zeigen, dass die Diversitätsdarstellung keinen systematischen Einfluss auf die intrinsische Motivation ausübte. Für die Subskalen *Interesse* und *Nützlichkeit* fanden sich keine signifikanten Unterschiede zwischen den Gruppen. Die Subskala *Anstrengung* wies zwar einzelne Unterschiede auf, diese sind jedoch aufgrund kleiner Stichprobenumfänge und eingeschränkter psychometrischer Eigenschaften der Skala nicht als belastbare Effekte zu interpretieren. Auch zwischen den Jahrgangsstufen ergaben sich lediglich geringe und nicht systematische Differenzen. Die qualitativen Daten stützen diese Befunde.

Insgesamt deutet die Untersuchung darauf hin, dass eine einmalige, subtil gestaltete Diversitätsdarstellung in einem kurzen Erklärvideo und einige interaktive Übungen (Quizaufgaben) nicht ausreichen, um motivationale Prozesse im Sinne der Selbstbestimmungstheorie zu beeinflussen. Im Ausblick werden Ansätze zur Steigerung der Wirksamkeit digitaler Lernumgebungen diskutiert, insbesondere durch höhere Sichtbarkeit, wiederholte Exposition und kontextbezogene Repräsentation von Diversität.

1 Einleitung

Die heutige Gesellschaften sind weltweit durch eine vielschichtige Vielfalt geprägt, die sich in sozialen, kulturellen, sprachlichen, religiösen, geschlechtlichen, sexuellen und körperlichen Dimensionen manifestiert. Die UNESCO (2009) hebt in ihrem Weltbericht *cultural diversity* hervor, dass kulturelle und sprachliche Vielfalt global zunimmt und als zentrale Ressource für gesellschaftliche Teilhabe und Innovation gilt (vgl. UNESCO, 2009). Migration ist eine dieser Diversitätsdimensionen. Weltweit leben rund 281 Millionen internationale Migrant:innen, etwa 3,6 % der Weltbevölkerung, und zusätzlich waren 117 Millionen Menschen im Jahr 2022 durch Konflikte, Instabilität oder Katastrophen vertrieben (vgl. Pope, 2022). Diese Menschen bringen nicht nur ihre Kultur und Sprache mit, sondern auch ihren religiösen Glauben und weltanschauliche Zugehörigkeit. Neben dieser Dimensionen der Diversität gibt es unter anderem auch die körperliche Beeinträchtigungen. Daten der Weltgesundheitsorganisation (WHO) im Jahr 2023 zeigen, dass etwa 16 % der Weltbevölkerung mit einer körperliche Beeinträchtigung lebt (vgl. WHO, 2023).

Deutsche Schulen spiegeln diese gesellschaftliche Diversität wider (vgl. KMK, 2025). Die Schüler:innenschaft ist kulturell, sprachlich und biografisch zunehmend heterogen, insbesondere mit Blick auf Migrationserfahrungen, Mehrsprachigkeit und diverse familiäre Hintergründe (vgl. KMK, 2025). Gesellschaftliche Ungleichheiten und Exklusionsmechanismen entstehen nicht erst im Erwachsenenalter, sondern wirken bereits im Kindes- und Jugendalter (vgl. OECD, 2025). Eine diversitätsbewusste Bildung ist hierfür nicht bloß ein ergänzender Beitrag, sondern eine strukturelle Voraussetzung um sozialer Teilhabe, Anerkennung und Chancengerechtigkeit zu fördern (vgl. OECD, 2025; KMK, 2025). Vor diesem Hintergrund soll Schule ein Ort sein, an dem gesellschaftliche Vielfalt sichtbar wird und an dem Schüler:innen lernen können, Unterschiede nicht nur zu tolerieren, sondern als wertvolle Ressource zu verstehen (vgl. KMK, 2025). Diese Anforderungen gelten für alle Fächer, einschließlich der Naturwissenschaften (vgl. OECD, 2025; KMK, 2025). Für den Physikunterricht ist dies besonders relevant, da er als stark maskulin konnotiertes Fach gilt, traditionell als Fach mit geringer Diversitätssensibilität wahrgenommen wird und vielfach eher stereotype Vorstellungen westlich-weißer Männer reproduziert (vgl. Tajmel, 2009). Das spiegelt sich auch in den Unterrichtsmaterialien wider. Schulbuchanalyse zeigen, dass Physikschulbücher häufig nur geringe Sensibilität für Diversität aufweisen und eher weiterhin stereotype oder homogene Darstellungen reproduzieren (vgl. Tolksdorf, 2017). Aus physikdidaktischer Perspektive ist diese Diskrepanz nicht folgenlos. Tajmel (2009) argumentiert, dass fehlende oder verzerrte Repräsentation in naturwissenschaftlichen Kontexten dazu führen kann, dass sich bestimmte Lerngruppen weniger mit dem Fach identifizieren, sich weniger zugehörig fühlen und seltener akademische Selbstwirksamkeit entwickeln. Untersuchung aus der sozialpsychologischen und MINT-Bildungsforschung (z.B. Cheryan et al., 2009) bestätigen, dass mangelnde Repräsentation negative Auswirkungen auf die Interesse an dessen Umfeld hat. Eine diversitätsbewusste Gestaltung des Physikunterrichts, wie durch kontextorientierte und lebensweltnahe Zugänge, kann das Interesse der Schüler:innen fördern (vgl. Schnirch, 2006).

Intrinsische Motivation spielt eine entscheidende Rolle für ein nachhaltiges Lernen. Sie entsteht, wenn Lernende Inhalte als interessant, persönlich relevant oder bedeutungsvoll erleben und trägt maßgeblich zu vertieftem Verständnis, ausdauerndem Lernen und selbstständiger Weiterarbeit bei (vgl. Schiefele & Schaffner, 2015; Deci & Ryan, 1985). Entsprechend der Selbstbestimmungstheorie (Self-Determination Theory, SDT) von Deci und Ryan (1993) hängt die Qualität der Motivation davon ab, inwieweit die drei grundlegende psychologische Bedürfnisse erfüllt sind: Autonomie, Kompetenz und sozialer Eingebundenheit.

Gerade digitale Lernumgebungen bieten hier neue Möglichkeiten. Sie eröffnen neue Wege, um Lernprozesse flexibel, adaptiv und beteiligungsorientiert zu gestalten. Digitale Lernumgebungen bieten ein zeitgemäßes Format, das nicht nur fachliche Inhalte zugänglich macht, sondern auch digitale Kompetenzen fördert, die im Rahmen der 21st Century Skills als zentral gelten (vgl. Hosang, 2021). Die Kultusministerkonferenz (KMK) fordert die systematische Integration digitaler Medien in Unterrichts- und Lernprozesse (vgl. Bogedan, 2017). Zusätzlich bieten sie vielfältige Möglichkeiten, Lernprozesse zu unterstützen, Autonomie und Motivation der Lernenden zu fördern (vgl. Bogedan, 2017). Wie stark Lernende sich in digitalen Umgebungen motivieren, hängt jedoch nicht allein von der mediendidaktischen Qualität ab, sondern auch davon, ob Inhalte als persönlich bedeutsam wahrgenommen werden und die Lernumgebung Zugehörigkeitserleben unterstützt (vgl. Findeisen et al., 2019; Cheryan et al., 2009). Damit wird deutlich, dass Lernumgebungen dann besonders motivierend wirken können, wenn sie an die Lebenswelt der Lernenden anknüpfen und ihnen ermöglichen, Inhalte in einem authentischen Kontext zu erfahren. Ein solches authentisches Lernsetting bietet sich im vorliegenden Forschungsvorhaben in besonderer Weise an, da der Untersuchungsort selbst durch ein reales, für die Schüler:innen unmittelbar erfahrbares Umweltphänomen geprägt ist.

Der Untersuchungsort dieser Arbeit ist eine deutsche Auslandsschule in [REDACTED], eine Stadt im Norden Thailands. Die Schüler:innenschaft ist hier aufgrund der internationalen Einbettung und der Mischung aus deutschsprachigen, thailändischen und weiteren kulturellen Hintergründen in besonderem Maße von Diversität geprägt. Die Stadt [REDACTED] ist jährlich von der sogenannten *Burning Season*¹ betroffen. Aufgrund der geografischen Tallage können Schadstoffe nur begrenzt abziehen, wodurch sich die Luftqualität deutlich verschlechtert. Vor diesem Hintergrund bietet sich Feinstaub als persönliches und alltagsrelevantes Thema für diese Schüler:innen an. Die Schule hat neben Maßnahmen bei sehr schlechter Luftqualität auch mehrere Luftqualitätssensoren, die auf dem Schulgelände installiert sind und die Feinstaubkonzentration messen, um ihre Schüler:innenschaft vor den schlechten Luftverschmutzung und Feinstaubbelastung zu schützen. Die Messdaten der Sensoren sind digital zugänglich. Die Lernenden sind somit regelmäßig mit realen Umweltdaten konfrontiert.

Die wiederkehrende Auseinandersetzung der Schüler:innen mit lokal erhobenen Messdaten schafft damit nicht nur einen kontextualisierten Zugang zum Thema, sondern erfüllt zugleich zentrale Bedingungen wirksamen naturwissenschaftlichen Lernens. Vor diesem Hintergrund lässt sich das

¹Die Burning Season in [REDACTED] ist eine jährliche Periode, meist von Februar bis April, in der die Luftqualität aufgrund von landwirtschaftlichen Bränden und Waldbränden zu extrem erhöhten Feinstaubwerten führt.

Thema Feinstaub in besonderer Weise zur Förderung physikalischer Denk- und Arbeitsweisen nutzen. Die Arbeit mit realen Messwerten und deren Auswertung fördert bei Schüler:innen zentrale Kompetenzen der Kommunikation, da die Lernenden physikalische Phänomene aus ihrer eigenen Lebenswelt kritisch analysieren und die gewonnenen Erkenntnisse sachgerecht darstellen und begründen (vgl. SenBJF, 2015). Die Auseinandersetzung mit authentischen Datensätzen unterstützt damit nicht nur den Erwerb fachlicher Fähigkeiten, sondern stärkt auch grundlegende Aspekte des wissenschaftlichen Denkens und Arbeitens. Digitale Medien bieten hierfür besondere Chancen: Sie ermöglichen einen Zugang zu Messdaten, visualisieren komplexe Zusammenhänge und unterstützen die strukturierte Analyse physikalischer Phänomene (vgl. Bernstein & Wilhelm, 2023). Damit bietet das Thema Feinstaub nicht nur inhaltlich, sondern auch methodisch und entwicklungspsychologisch ein Potenzial für den Physikunterricht.

Während theoretische Modelle diese Aspekte als bedeutsam für motivationale Prozesse beschreiben, ist bislang wenig darüber bekannt, inwiefern sie durch diversitätssensible Darstellungen in digitalen Lernumgebungen tatsächlich beeinflusst werden. Vor diesem Hintergrund untersucht die vorliegende Arbeit, inwiefern Diversitätsdarstellungen in einer digitalen Lernumgebung zum Thema Feinstaub die intrinsische Motivation von Lernenden der Sekundarstufe beeinflussen. Die Untersuchung verbindet damit mehrere Perspektiven moderner Bildungsforschung: Umweltbildung und Nachhaltigkeit, datenbasiertes Lernen im Physikunterricht, digital unterstützte Lernprozesse sowie die Bedeutung von Diversität und Repräsentation. Für diese Untersuchung werden zwei digitale Lerneinheiten zum Thema Feinstaub erstellt, die inhaltlich identisch sind und sich nur dadurch unterscheiden, dass die Lernumgebung der Experimentalgruppe Diversitätsdarstellungen enthält und die Lernumgebung der Kontrollgruppe keine Diversitätsdarstellungen enthält.

Die vorliegende Arbeit ist wie folgt aufgebaut:

Kapitel 2 stellt den theoretischen Rahmen und behandelt grundlegende motivationspsychologische Konzepte, insbesondere die Selbstbestimmungstheorie nach Deci und Ryan (1993) sowie das Intrinsic Motivation Inventory (IMI) als Messinstrument zur Erfassung intrinsischer Motivation. Darüber hinaus werden Merkmale digitaler Lernumgebungen sowie Diversität und ihre zentralen Aspekte im physikdidaktischen Kontext erläutert. Auf dieser theoretischen Grundlage werden in Kapitel 3 die Forschungsfragen der Untersuchung formuliert. Kapitel 4 beschreibt das methodische Vorgehen und stellt das experimentelle Forschungsdesign der Untersuchung vor. Dazu werden das Untersuchungssetting, die Zielgruppe sowie die Intervention erläutert. Im Rahmen der Intervention wird insbesondere die Darstellung von Personen in der digitalen Lernumgebung der beiden Untersuchungsgruppen beschrieben. Darüber hinaus werden die eingesetzten Messinstrumente vorgestellt und das Vorgehen zur Prüfung von Reliabilität und Validität dargelegt. Anschließend werden die angewandten quantitativen und qualitativen Analyseverfahren erläutert. Kapitel 5 präsentiert die Ergebnisse der Untersuchung. Kapitel 6 diskutiert diese Befunde vor dem Hintergrund des theoretischen Rahmens und beantwortet die Forschungsfragen. Darüber hinaus werden die Limitierungen der Untersuchung dargestellt und kritisch eingeordnet. Das abschließende Kapitel 7 bietet einen Ausblick auf weiterführende Forschungsperspektiven

und Implikationen für die Gestaltung digitaler Lernumgebungen im Zusammenhang mit diversitätssensibler Darstellung.

2 Theoretischer Hintergrund

In diesem Kapitel werden die für diese Arbeit zentralen Begriffe eingeführt und erläutert.

Im ersten Abschnitt 2.1 wird das Konstrukt der Motivation erläutert. Dabei wird der Unterschied zwischen intrinsischer und extrinsischer Motivation vorgestellt, bevor im Anschluss auf die Selbstbestimmungstheorie nach Deci und Ryan (1993) eingegangen wird.

Im zweiten Abschnitt 2.2 wird der Einsatz von Erklärvideos im Physikunterricht thematisiert. Hier werden die Potenziale und Herausforderungen audiovisueller Lehr-Lernmedien beschrieben, Qualitätsmerkmale nach Kulgemeyer vorgestellt und Zusammenhänge zwischen Erklärvideos und der Lern- bzw. Motivationsförderung der Schüler:innen dargestellt. Anschließend wird der Einsatz interaktiver digitaler Quizaufgaben als eine weitere Methode zur Förderung der Motivation von Lernenden thematisiert.

Im letzten Abschnitt 2.3 wird das Thema Diversität behandelt. Dabei werden zentrale soziale Diversitätsdimensionen und ihre intersektionalen Verflechtungen eingeführt und insbesondere ihre Bedeutung für den Physikunterricht erläutert. Zudem werden theoretische Ansätze und empirische Befunde zu den Auswirkungen der Diversitätsdarstellungen in digitalen Lernumgebung vorgestellt. Abschließend folgt ein Überblick über Studien zu allgemeinen Diversitätsdarstellungen, auch außerhalb digitaler Lernumgebungen, und deren Einfluss auf Motivation und Zugehörigkeitsgefühl.

2.1 Konzept der Motivation

Um das Konzept der Motivation einzuordnen, wird im Folgenden (Abschnitt 2.1.1) zunächst die Definition von Motivation erläutert. Anschließend erfolgt in Abschnitt 2.1.2 eine Unterscheidung zwischen extrinsischer und intrinsischer Motivation, bevor auf die Selbstbestimmungstheorie nach Deci und Ryan (1993) (Abschnitt 2.1.3) eingegangen wird, die ein differenziertes Verständnis motivationaler Prozesse ermöglicht und den theoretischen Rahmen dieser Arbeit bildet. Darauf aufbauend wird das Intrinsic Motivation Inventory (IMI) im Abschnitt 2.1.4 als Messinstrument zur Erfassung intrinsischer Motivation vorgestellt.

2.1.1 Motivation

In der Psychologie wird Motivation als Bereitschaft verstanden, ein Ziel mit Ausdauer, Intensität und Anstrengung zu verfolgen (vgl. Schiefele & Schaffner, 2015). Sie stellt keinen direkt beobachtbaren Gegenstand dar, sondern ein zentrales Konstrukt der Verhaltensklärung. Rheinberg und Vollmeyer (2019) Motivation als *„eine aktivierende Ausrichtung des momentanen Lebensvollzugs auf einen positiv bewerteten Zielzustand“* (siehe Rheinberg & Vollmeyer, 2019, S. 17). Die Entstehung von Motivation lässt sich sowohl auf innere Faktoren, wie Bedürfnisse oder Interesse, als auch auf äußere Einflüsse, etwa Belohnungen oder soziale Aufforderungen, zurück-

führen. Während erstere die intrinsische Motivation kennzeichnen, also Handlungen um ihrer selbst willen auslösen, beruhen letztere auf extrinsischer Motivation, die durch äußere Anreize gesteuert wird (vgl. Schiefele & Schaffner, 2015). Der folgende Abschnitt erläutert diese beiden grundlegenden Motivationsarten und beschreibt, wodurch sie sich hinsichtlich ihrer Entstehung und Wirkung unterscheiden.

2.1.2 Extrinsische und intrinsische Motivation

Ein Verhalten wird als extrinsisch motiviert verstanden, wenn es nicht um seiner selbst willen ausgeführt wird, sondern durch äußere Anreize, um bestimmte Ziele zu erreichen oder negative Konsequenzen zu vermeiden. Typische extrinsische Anreize sind materielle Belohnungen, soziale Anerkennung oder das Streben nach guten Noten, aber auch der Einfluss von Druck oder die Vermeidung von Strafen (vgl. Hennecke & Brandstätter, 2016). Die Handlung wird in diesem Fall als Mittel zum Zweck betrachtet.

Die intrinsische Motivation beschreibt eine Form des Handelns, die aus eigenem Interesse und innerem Antrieb erfolgt. Die Anreize liegen in der Tätigkeit selbst und zeigen sich durch positive Lebenszustände, die bereits während der Handlung auftreten (vgl. Deci & Ryan, 1993). Sie entstehen insbesondere dann, wenn Personen herausfordernde Aufgaben bewältigen, ihre eigene Wirksamkeit erfahren oder das Gefühl haben, selbstbestimmt handeln zu können (vgl. Deci & Ryan, 1993). Intrinsisch motivierte Personen führen Aktivitäten also nicht aufgrund äußerer Belohnungen aus, sondern weil sie den Gegenstand oder die Tätigkeit selbst als interessant und lohnend empfinden (vgl. Hennecke & Brandstätter, 2016). Sie gilt als die qualitativ hochwertigste und nachhaltigste Form der Motivation (vgl. Deci & Ryan, 1993).

Dennoch sind extrinsische und intrinsische Motivation nicht als Gegensätze zu verstehen, sondern vielmehr als unterschiedliche Formen motivationaler Regulation. Beide können sich wechselseitig beeinflussen, teils abschwächend, teils stabilisierend (vgl. Deci & Ryan, 1993). So kann sich beispielsweise die intrinsische Motivation einer Schülerin am Lösen mathematischer Aufgaben verringern, wenn das Verhalten durch äußere Anreize wie Geld verstärkt wird. Erfolgt die externe Rückmeldung hingegen in Form sozialer Anerkennung, etwa durch das Lob einer Lehrkraft, kann dies die intrinsische Motivation unterstützen und aufrechterhalten.

Die Betrachtung extrinsischer und intrinsischer Motivation verdeutlicht, dass menschliches Verhalten nicht ausschließlich durch äußere oder innere Anreize erklärbar ist. Vielmehr bewegen sich Handlungen auf einem Spektrum unterschiedlicher Regulationsformen, die je nach Kontext mehr oder weniger selbstbestimmt ausfallen können (vgl. Deci & Ryan, 1993). Motiviertes Handeln lässt sich daher nicht nur entlang der Unterscheidung zwischen extrinsischer und intrinsischer Motivation betrachten, sondern auch hinsichtlich des Grades an Selbstbestimmung (vgl. Deci & Ryan, 1993). Die Selbstbestimmungstheorie von Deci und Ryan (1993) beschreibt, wie sich Motivation in Abhängigkeit vom Grad der Selbstbestimmung entwickelt und unter welchen Bedingungen selbstbestimmte, und damit intrinsisch motivierte, Lernprozesse gefördert werden können (vgl. Frühwirth, 2020).

2.1.3 Selbstbestimmungstheorie (Self-Determination Theory, SDT)

Die Selbstbestimmungstheorie (Self-Determination Theory, SDT) von Deci und Ryan (1993) zählt zu den zentralen motivationstheoretischen Ansätzen der modernen Psychologie. Sie postuliert, dass die Qualität der Motivation – ob intrinsisch oder extrinsisch – maßgeblich davon abhängt, inwieweit drei grundlegende psychologische Bedürfnisse erfüllt sind: das Bedürfnis nach Autonomie, das Bedürfnis nach Kompetenz und das Bedürfnis nach sozialer Eingebundenheit (vgl. Fröhlich, 2020).

Das Bedürfnis nach Autonomie beschreibt das Streben, das eigene Handeln als selbstbestimmt und im Einklang mit den eigenen Werten und Interessen zu erleben. Das Bedürfnis nach Kompetenz bezieht sich auf das Erleben von Wirksamkeit im eigenen Tun – also darauf, sich Herausforderungen gewachsen zu fühlen und Erfolge zu erzielen. Das Bedürfnis nach sozialer Eingebundenheit schließlich meint das Gefühl, sich mit anderen verbunden zu fühlen und Teil einer Gemeinschaft zu sein (vgl. Deci & Ryan, 1993). Werden diese Bedürfnisse erfüllt, begünstigt dies die Entstehung intrinsischer Motivation, also einer Form des Handelns, die aus eigenem Interesse und innerem Antrieb erfolgt.

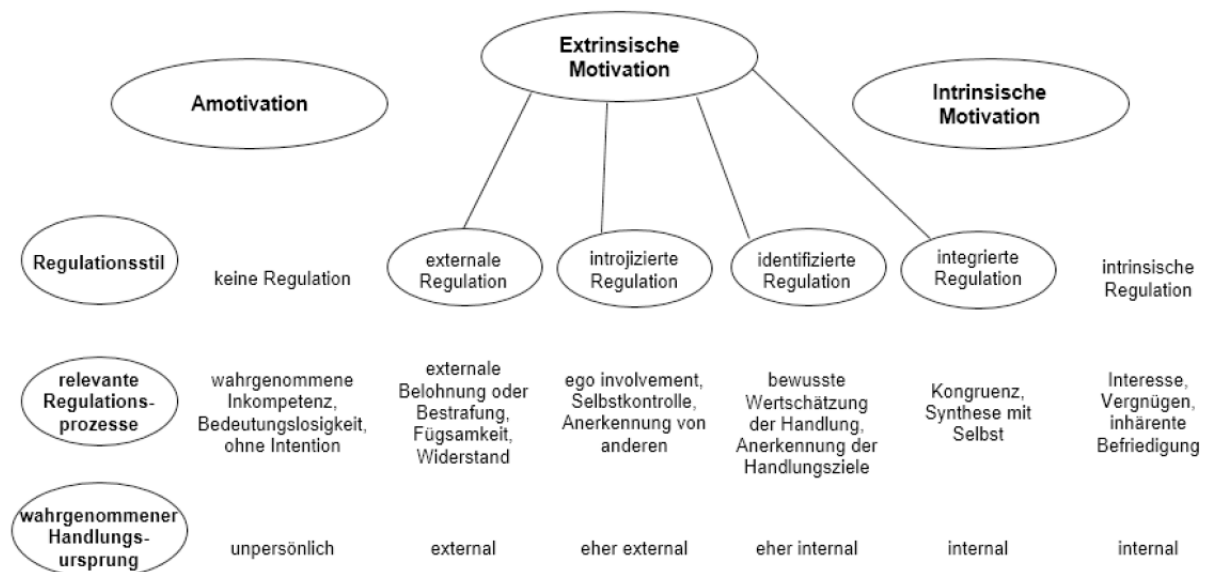


Abbildung 1: Übersicht der Regulationsstile nach der Organismic Integration Theory, übersetzt von Fröhlich (2020, S. 16)

Zur Veranschaulichung der unterschiedlichen Stufen der Motivation zeigt Abbildung 1 das von Fröhlich (2020, S. 16) übersetzte Motivationskontinuum² nach Deci und Ryan (1993). Es verdeutlicht, dass Motivation nicht als dichotomes Konstrukt verstanden wird, sondern in Abstufungen des wahrgenommenen Grades an Selbstbestimmung verläuft.

Am einen Ende steht die Amotivation, also das Fehlen einer Handlungsabsicht, während am anderen Ende die intrinsische Motivation verortet ist, bei der eine Tätigkeit aus eigenem Interesse und Freude ausgeführt wird. Dazwischen liegen verschiedene Formen extrinsischer Motivation,

²Motivationskontinuum ist ein Konzept der Selbstbestimmungstheorie von Deci und Ryan (1993), das verschiedene Formen der Motivation auf einem Spektrum einordnet.

die sich danach unterscheiden, wie stark äußere Gründe bereits in das Selbst integriert wurden. Dazu zählen die externale Regulation (Handeln aufgrund äußerer Anreize, wie Belohnung oder Bestrafung), die introjizierte Regulation (Handeln aus innerem Druck oder Pflichtgefühl) sowie die identifizierte (Handeln, weil der persönliche Wert oder Nutzen erkannt wird) und integrierte Regulation (Handeln, bei dem äußere Ziele vollständig in das eigene Selbstkonzept übernommen werden) (vgl. Fröhlich, 2020).

Für die vorliegende Arbeit ist die Selbstbestimmungstheorie von besonderer Bedeutung, da sie den theoretischen Rahmen für das Verständnis der intrinsischen Motivation bildet.

2.1.4 Intrinsic Motivation Inventory (IMI)

Die Erfassung intrinsischer Motivation kann über verschiedene Ansätze erfolgen. Neben Verhaltensmaßen, wie der Dauer freiwilliger Beschäftigung mit einer Aufgabe, gelten Selbstberichtungsverfahren als besonders praktikabel (vgl. Henneke & Brandstätter, 2016). Ein bekanntes und theoretisch fundiertes Instrument ist das Intrinsic Motivation Inventory (IMI), das im Rahmen der Selbstbestimmungstheorie entwickelt wurde (vgl. Wilde et al., 2009). Es ist ein mehrdimensionales Instrument zur Erfassung unterschiedlicher subjektiver Erfahrungen im Zusammenhang mit einer Tätigkeit (vgl. Wilde et al., 2009).

Das ursprünglich in englischer Sprache veröffentlichte IMI von *Center for Self-Determination Theory* (CSDT) umfasst 45 Items in sieben Subskalen und kann je nach Anwendungskontext in gekürzter Form eingesetzt werden (vgl. CSDT, 2022). In der Forschung werden häufig Varianten mit 9 bis 29 Items und unterschiedlichen Strukturen der Subskalen verwendet (vgl. IMI-english, 2022). Die Subskalen des IMI berücksichtigen sowohl die positiven Facetten der intrinsischen Motivation als auch hemmende Faktoren. Die sieben Subskalen sind:

1. **interest/enjoyment (Interesse/Vergnügen)**: misst die Freude und das Interesse an der Aufgabe und gilt als direktester Indikator intrinsischer Motivation.

Innerhalb der Selbstbestimmungstheorie wird Interesse als Ausdruck selbstbestimmten Verhaltens interpretiert, das durch die Befriedigung der drei psychologischen Grundbedürfnisse unterstützt wird (vgl. Deci & Ryan, 1993). Interesse kann somit als emotions- und wertebezogene Ausprägung intrinsischer Motivation verstanden werden: Lernende wenden sich einem Thema aus eigenem Antrieb zu, weil sie es als bedeutsam, herausfordernd oder persönlich relevant erleben (vgl. Deci & Ryan, 1993).

2. **perceived competence (wahrgenommene Kompetenz)**: erfasst, inwieweit Lernende sich in der Tätigkeit als fähig und wirksam erleben.

Innerhalb der Selbstbestimmungstheorie entspricht diese Subskala dem psychologischen Grundbedürfnis nach Kompetenz, das eine zentrale Voraussetzung für intrinsische Motivation darstellt. Wenn Lernende das Gefühl haben, eine Aufgabe erfolgreich bewältigen zu können, fördert dies ihre Bereitschaft, sich aktiv und selbstbestimmt mit dem Lerngegenstand auseinanderzusetzen. Wahrgenommene Kompetenz wirkt somit motivierend, indem sie positive Leistungsrückmeldungen, Fortschrittserleben und das Gefühl der eigenen Wirk-

samkeit unterstützt (vgl. Deci & Ryan, 1993).

3. **effort (Anstrengung)**: spiegelt wider, wie viel eigenen Einsatz Lernende einer Tätigkeit beimessen.

Sie beschreibt das Ausmaß an investierter Anstrengung und persönlicher Bedeutsamkeit, die eine Person einer Aktivität beimisst. Sie stellt eine eigenständige Variable dar, die in Studien eingesetzt wird, wenn das Interesse an motivationalem Engagement oder der Aufrechterhaltung von Anstrengung relevant ist (vgl. Deci & Ryan, 1993). In der Selbstbestimmungstheorie wird diese Komponente als Ausdruck einer internalisierten Motivation verstanden, also einer Form, bei der Lernende Anstrengung nicht aufgrund äußerer Kontrolle, sondern aus einem inneren Verantwortungsgefühl oder Bedeutsamkeit der Tätigkeit zeigen (vgl. Deci & Ryan, 1993). Ein höheres Maß an selbst initiiertem Anstrengung weist darauf hin, dass Lernende die Aufgaben als herausfordernd, sinnvoll und selbstbestimmt wahrnehmen.

4. **felt pressure/tension (Druck/Anspannung)**: erfasst das Ausmaß an empfundenem Druck oder innerer Anspannung während der Tätigkeit.

Sie beeinflusst die intrinsische Motivation negativ, da ein höheres Druckerleben mit geringerer Selbstbestimmung einhergeht.

5. **perceived choice (wahrgenommene Wahlfreiheit)**: beschreibt das Ausmaß der empfundenen Selbstbestimmung bei der Ausführung einer Aufgabe.

Innerhalb der Selbstbestimmungstheorie entspricht diese Subskala dem psychologischen Grundbedürfnis nach Autonomie (vgl. Wilde et al., 2009). Sie erfasst, inwiefern Lernende den Eindruck haben, ihre Handlungen aus eigenem Antrieb zu wählen und nicht durch äußeren Druck oder Kontrolle gesteuert zu sein. Ein hohes Maß an wahrgenommener Wahlfreiheit stärkt das Gefühl der persönlichen Entscheidungs- und Handlungsfreiheit und fördert damit intrinsische Motivation. Lernende engagieren sich intensiver, wenn sie erleben, dass ihre Entscheidungen, Ziele und Vorgehensweisen selbstbestimmt sind (vgl. Deci & Ryan, 1993).

6. **value/usefulness (Wert/Nützlichkeit)**: erfasst, ob Lernende die Aufgabe als persönlich bedeutsam und für ihr Leben oder ihre Zukunft relevant wahrnehmen.

Sie dient dazu, den Grad zu erfassen, in dem Personen eine Aktivität als nützlich, wertvoll oder persönlich bedeutsam wahrnehmen. Im theoretischen Rahmen der Selbstbestimmungstheorie entspricht dies einer identifizierten Regulation, einer autonomen Form extrinsischer Motivation, bei der Lernende den Wert einer Tätigkeit verinnerlicht haben und sie deshalb aus Überzeugung ausführen (vgl. Deci & Ryan, 1993). Eine hohe wahrgenommene Nützlichkeit weist darauf hin, dass die Lernumgebung Lernende dabei unterstützt, inhaltliche Relevanz und Wertbezug zu erkennen – ein zentraler Aspekt nachhaltiger Motivation.

7. **relatedness (soziale Eingebundenheit)**: beschreibt das Gefühl, sich im sozialen Lernumfeld zugehörig und unterstützt zu fühlen.

In der Selbstbestimmungstheorie entspricht diese Subskala dem psychologischen Grundbedürfnis nach sozialer Eingebundenheit (vgl. Wilde et al., 2009). Sie erfasst, inwiefern Lernende positive, vertrauensvolle Beziehungen zu anderen Personen erleben und sich als Teil einer Gemeinschaft wahrnehmen. Ein hohes Maß an sozialer Eingebundenheit fördert intrinsische Motivation, da Lernende sich sicher, akzeptiert und wertgeschätzt fühlen – Bedingungen, die selbstbestimmtes Lernverhalten begünstigen (vgl. Deci & Ryan, 1993).

Die Subskala *interest/enjoyment* ist die einzige Dimension von den sieben, die intrinsische Motivation selbst abbildet (vgl. CSDT, 2022). Allerdings repräsentieren die Subskalen *perceived competence*, *perceived choice* und *relatedness* die drei psychologischen Grundbedürfnisse des Menschen – nach Kompetenz, Autonomie und sozialer Eingebundenheit der Selbstbestimmungstheorie von Deci und Ryan (1993). Ihre Befriedigung fördert das Erleben von Selbstbestimmung und damit intrinsische Motivation. Die Subskalen *effort* und *value/usefulness* nehmen im Rahmen des IMI eine besondere Stellung ein, da sie nicht direkt die intrinsische Motivation erfassen, sondern in spezifischen theoretischen Kontexten herangezogen werden. Die Subskala *pressure/tension* hingegen erfasst erlebten Druck und Anspannung und wirkt sich negativ auf die intrinsische Motivation aus.

Laut *Center for Self-Determination theory* (CSDT, 2022) zeigen frühere Studien, dass die Reihenfolge der Fragen nur einen vernachlässigbaren Einfluss hat und die Auswahl einzelner Subskalen die übrigen Skalen nicht beeinträchtigt. Die Verwendung aller Items in einer Untersuchung werden nur selten angewandt. Daher werden in empirischen Untersuchungen meist nur jene Subskalen eingesetzt, die für die jeweilige Forschungsfrage von Bedeutung sind (vgl. CSDT, 2022).

In der ursprünglichen Beschreibung des Instruments finden sich keine Angaben zur Messgenauigkeit. CSDT (2022) weist lediglich darauf hin, dass das IMI inhaltlich schlüssig erscheine. Entsprechend empfehlen sie, bei einer Anwendung des Fragebogens die Faktorenstruktur empirisch zu überprüfen (vgl. CSDT, 2022). In der Folge wurde die englische Version des IMI in zahlreichen empirischen Studien in unterschiedlichen Kontexten erfolgreich eingesetzt, etwa im Sport- und Bewegungsbereich, in technologiegestützten Lehr-Lern-Umgebungen sowie im schulischen Lernen auf ihre Validität und Reliabilität geprüft (vgl. McAuley et al., 1989; Ostrow & Heffernan, 2018; Bosch, 2024).

In einer aktuellen Studie im schulischen Lernen überprüfte Bosch (2024) die Validität und Reliabilität des englischsprachigen IMI in einem Blended-Learning-Kontext. In dieser Untersuchung mit über 1.200 Studierenden eines universitären Wirtschaftskurses wurde mithilfe einer konfirmatorischen Faktorenanalyse gezeigt, dass die Faktorstruktur des IMI weitgehend bestätigt werden konnte. Alle Subskalen zeigten signifikante Faktorladungen und erwartungsgemäße Zusammenhänge, wobei insbesondere der Faktor *Druck/Anspannung* negativ mit den übrigen Skalen korrelierte. Die interne Konsistenz erwies sich (Cronbachs $\alpha = 0,75\text{--}0,92$) als sehr zufriedenstellend, lediglich die Skala *soziale Eingebundenheit* zeigte geringere Werte (vgl. Bosch, 2024). Damit unterstreicht die Studie, dass das IMI auch in technologiegestützten Lernumgebungen ein zuverlässiges und weitgehend valides Instrument zur Erfassung der intrinsischen Motivation darstellt

und somit für den Einsatz in modernen digitalen Lernkontexten geeignet ist (vgl. Bosch, 2024).

Neben der englischsprachigen Originalversion ist auf der Homepage des *Center for Self-Determination theory* (CSDT) auch eine deutschsprachige Fassung des IMI verfügbar (vgl. Homepage der Center for Self-Determination Theory, o.D.). Diese enthält jedoch nicht alle Subskalen der englischen Version, insbesondere fehlen *Nützlichkeit* und *soziale Eingebundenheit*. Auch hier finden sich keine Angaben zur Messgenauigkeit.

Eine der ersten empirisch geprüften deutschsprachigen Adaptionen stammt von Wilde et al. (2009), die die *Kurzskala Intrinsischer Motivation* (KIM) entwickelten. Ziel ihrer Untersuchung war es, eine an den deutschsprachigen Kontext angepasste und zugleich zeitökonomische Kurzversion des IMI hinsichtlich ihrer internen Konsistenz und Reliabilität zu überprüfen (vgl. Wilde et al., 2009). Die KIM umfasst vier Subskalen (*Interesse/Vergnügen*, *wahrgenommene Kompetenz*, *wahrgenommene Wahlfreiheit* und *Druck/Anspannung*), mit jeweils drei Items, die auf einer fünfstufigen Likert-Skala beantwortet werden. Die Erhebung erfolgte im Rahmen einer Untersuchung zur Motivation von Schüler:innen beim selbstgesteuerten Lernen im Berliner Naturkundemuseum. Die Teilnehmenden füllten den Fragebogen unmittelbar nach dem Museumsbesuch sowie vier Wochen später erneut aus.

Ein Vergleich der Items der KIM mit der deutschen Version der CSDT zeigt, dass sich die beiden Instrumente nicht nur in der Anzahl der Items pro Subskala unterscheiden, sondern auch in der sprachlichen Ausgestaltung der jeweiligen Items (vgl. Wilde et al., 2009; Self-Determination Theory, 2022). So lautet etwa ein Item der Subskala *Interesse/Vergnügen* in der KIM: „Ich fand die Tätigkeit in der Ausstellung sehr interessant.“ (vgl. Wilde et al., 2009, S. 45). Die entsprechende Formulierung in der CSDT-Version ist stärker abstrahiert und allgemeiner gehalten: „Ich würde diese Tätigkeit als sehr interessant bezeichnen.“ (vgl. Self-Determination Theory, 2022, S. 1). Zum Vergleich sei die englische Originalversion angeführt: „I would describe this activity as very interesting.“ (vgl. IMI-english, 2022, S. 3)

Zur Prüfung der Validität führten Wilde et al. (2009) eine konfirmatorische Faktorenanalyse (CFA) durch, um zu überprüfen, ob die KIM vierfaktorielle Faktoranalyse klar voneinander abgrenzbare Indikatoren intrinsisch motivierten Verhaltens mit hoher Messzuverlässigkeit erfasst. Die Ergebnisse bestätigten eine vierfaktorielle Struktur. Drei Subskalen wiesen hohe Trennschärfen eines Items und Eigenwerte auf (Vergnügen: 21,8%, $\lambda = 2,620$; Wahlfreiheit: 18,2%, $\lambda = 2,19$; Kompetenz: 17,8%, $\lambda = 2,13$), während die Subskala *Druck/Anspannung* (13,3 %, $\lambda = 1,60$) nur geringe Trennschärfen zeigte. Zusätzlich spiegeln ihre internen Konsistenzen dieses Muster wider. Für drei Subskalen wurden hohe Reliabilitäten erreicht ($\alpha = 0,75\text{--}0,89$), während *Druck/Anspannung* eine geringere Reliabilität aufwies ($\approx 0,53$) (vgl. Wilde et al., 2009). Dies könnte daran liegen, dass der außerschulische Besuch nicht den Druck und die Anspannung erzeugt wie im schulischen Umfeld (vgl. Wilde et al., 2009).

Obwohl für alle sieben IMI-Subskalen deutsche Übersetzungen existieren, wurden in der Studie von Wilde et al. (2009) ausschließlich vier Subskalen empirisch geprüft; die Subskalen *Anstrengung/Wichtigkeit*, *Wert/Nützlichkeit* und *soziale Eingebundenheit* blieben unberücksichtigt.

2.2 Digitale Lernumgebung

Digitale Lernumgebungen bieten vielfältige Möglichkeiten, Lernprozesse zu unterstützen und Autonomie und Motivation der Lernenden zu fördern (vgl. Bogedan, 2017). In diesem Abschnitt werden zentrale Elemente solcher Umgebungen behandelt: Erklärvideos und interaktive digitale Quizaufgaben. Zunächst wird in 2.2.1 auf den didaktischen Einsatz und die Qualitätsmerkmale von Erklärvideos nach Kulgemeyer (2020) eingegangen (Abschnitt 2.2.1), bevor der Zusammenhang zwischen Erklärvideos und Lernmotivation erläutert wird. Anschließend wird der Einfluss interaktiver Quizaufgaben auf Motivation und Lernengagement betrachtet (Abschnitt 2.2.2).

2.2.1 Erklärvideo

Eine einheitliche Definition des Begriffs **Erklärvideo** existiert nicht. In der Literatur finden sich Bezeichnungen wie Lernvideo oder Lehrfilm, die jedoch unterschiedliche Konnotationen haben (vgl. Meller, 2024). Häufig ist die Definition von Wolf (2015b) verbreitet, der **Erklärvideo** als *„eigenproduzierte Filme, in denen erläutert wird, wie man etwas macht oder wie etwas funktioniert bzw. in denen abstrakte Konzepte erklärt werden“* beschreibt (vgl. Wolf, 2015b, S. 123). Dabei ist zu unterstreichen, dass Erklärvideos im Gegensatz zu Video-Tutorials nicht primär auf das Nachahmen einer Handlung, sondern auf das Verständnis von Konzepten abzielen, und sich zugleich von Performanzvideos unterscheiden, bei denen die reine Darstellung einer Fertigkeit im Vordergrund steht (vgl. Wolf, 2015b). Damit hebt Wolf (2015b) insbesondere die Tatsache hervor, dass die Mehrheit frei verfügbarer Erklärvideos nicht von professionellen Medienschaffenden, sondern von Laien produziert werden können. Allgemein sind Erklärvideos im Vergleich zu klassischen Filmen kürzer, einfacher gestaltet und stärker auf Anschaulichkeit ausgerichtet. Damit zeichnen sie sich durch einen niedrigschwelligen Zugang aus, der Lernenden insbesondere bei komplexen Inhalten eine verständliche Aufbereitung ermöglicht.

Für die vorliegende Arbeit werden Erklärvideos als kurze audiovisuelle Formate verstanden, die komplexe Sachverhalte anschaulich und didaktisch aufbereitet darstellen, um Verständnis zu fördern.

Einsatz von Erklärvideos im Unterricht

In den letzten Jahren haben Erklärvideos, nicht zuletzt durch Plattformen wie YouTube, schul.cloud oder andere digitale Bildungsmedienportale, erheblich an Bedeutung gewonnen. Sie dienen Schüler:innen nicht nur außerhalb der Unterrichtszeit zur individuellen Wiederholung und Nachbereitung, sondern werden zunehmend auch im schulischen Kontext eingesetzt (vgl. Obermoser, 2018). So finden Erklärvideos Anwendung bei der Einführung neuer Inhalte, der Festigung von Wissen oder als Bestandteil selbstgesteuerten Lernens im Rahmen von Blended-Learning- und Flipped-Classroom-Konzepten (vgl. Obermoser, 2018). Ein wesentlicher Vorteil für Schüler:innen liegt dabei in der zeitlichen und räumlichen Flexibilität: Lernende können Videos beliebig oft ansehen, pausieren und in ihrem eigenen Tempo bearbeiten und es ist nicht ortsgebunden (vgl. Kulgemeyer, 2020). Gerade in heterogenen Lerngruppen eröffnet dies Möglichkeiten zur individuellen Förderung. Besonders im Physikunterricht, der häufig durch abstrakte und schwer zugängli-

che Konzepte geprägt ist, übernehmen Erklärvideos eine zentrale Brückenfunktion. Komplexe Phänomene wie elektromagnetische Wellen, der Aufbau von Atomen oder Bewegungsabläufe in der Mechanik lassen sich mithilfe von Animationen und Visualisierungen deutlich anschaulicher darstellen als durch statische Abbildungen. Empirische Studien zeigen, dass Lernende bei animierten Visualisierungen systematischer mentale Modelle bilden und dadurch signifikant bessere Transferleistungen erzielen als bei rein textbasierten Darstellungen (vgl. Mayer, 2002). Auf diese Weise werden nicht nur grundlegende Verständlichkeit unterstützt, sondern zugleich die Anwendung des Gelernten auf neue Problemsituationen erleichtert. Darüber hinaus können Erklärvideos Lernenden Einblicke in Prozesse ermöglichen, die im Klassenzimmer experimentell nicht oder nur eingeschränkt zugänglich sind – beispielsweise mikroskopische Vorgänge oder sicherheitskritische Experimente mit Hochspannung.

Neben der Rezeption bietet auch die eigenständige Produktion von Erklärvideos durch Schüler:innen didaktisches Potenzial. Die Erstellung erfordert eine vertiefte Auseinandersetzung mit den Inhalten und fördert zugleich kommunikative sowie mediale Kompetenzen, wodurch Erklärvideos nicht nur als Lernmedium, sondern auch als Lernprodukt einen Mehrwert im Unterricht darstellen (vgl. Kulgemeyer, 2020).

Trotz ihres Potenzials stehen Erklärvideos auch in der Kritik. Die Wahrnehmung über audiovisuelle Medien wird im Vergleich zu unmittelbaren, körperlich-sinnlichen Erfahrungen häufig als reduzierter wahrgenommen (vgl. Meller, 2024). Die Produktion erfordert zudem Zeit und Kosten, deren Nutzen nicht immer im Verhältnis zum Lernertrag steht. Darüber hinaus ist ihre ökologische Nachhaltigkeit fraglich (vgl. Meller, 2024). Frei verfügbare Videos weisen häufig Qualitätsprobleme auf, etwa fachliche Fehler oder ungeeignete Vermittlungskonzepte, und unterliegen der algorithmischen Steuerung von Plattformen wie YouTube (vgl. Wolf, 2015a). Gerade im Physikunterricht kann der Vorrang von Unterhaltung vor Erklärung problematisch sein, da exakte Begriffe und konsistente Modelle unverzichtbar sind. Aus diesem Grund sollten Erklärvideos nach sorgfältiger Prüfung ihrer fachlichen Richtigkeit, didaktischer Reduktion und Lernwirksamkeit eingesetzt werden (vgl. Wolf & Kulgemeyer, 2022). Besonders wirksam ist ihre Integration, wenn sie durch Aufgaben, Diskussionen oder Experimente ergänzt werden, die ein aktives Verarbeiten fördern.

Qualität von Erklärvideos

Kulgemeyer (2020) formuliert sieben Kernideen, die die Wirksamkeit von Erklärvideos bestimmen. Diese umfassen die Adaption an das Vorwissen, die Interessen und mögliche Fehlvorstellungen der Lernenden, den gezielten Einsatz von Visuellen Werkzeugen (wie Beispielen, Modellen oder geeigneten Sprachebenen) sowie die Verdeutlichung der Relevanz des Lerninhalts für die Zielgruppe. Darüber hinaus betont Kulgemeyer (2020) die Bedeutung einer klaren Struktur, beispielsweise durch den Wechsel von Regel- und Beispielphasen sowie durch prägnante Zusammenfassungen. Eine präzise und kohärente Erklärung trägt dazu bei, den kognitiven Aufwand gering zu halten und die Kernideen verständlich zu vermitteln. Schließlich sollte ein Erklärvideo ein neues oder komplexes Prinzip behandeln, das sich Lernende nicht ohne Unterstützung selbst erschließen können, und es sollte idealerweise in einen unterrichtlichen Kontext eingebettet sein,

der durch passende Lernaufgaben vertieft wird. Eine ausführliche Übersicht der sieben Kernideen und zugehörigen Kriterien befindet sich im Anhang (siehe Tabelle 14, S. 16).

Erklärvideo und Motivation

Erklärvideos stellen eine zunehmend genutzte Lernform im schulischen und universitären Kontext dar. Sie ermöglichen durch ihre multimediale Aufbereitung und flexible Nutzung eine anschauliche Vermittlung komplexer Inhalte und können Motivation der Lernenden fördern (vgl. Conradt, 2009).

Studien zeigen, dass insbesondere Lernende mit niedrigem Vorwissen von Erklärvideos profitieren und durch deren Einsatz Interesse und Lernfreude gesteigert werden können (vgl. Sterzig, 2022). Sterzig (2022) konnte in einer empirischen Untersuchung mit Studierenden nachweisen, dass das generelle Interesse an Erklärvideos einen positiven Einfluss auf den Lernzuwachs hat. Auch wenn in dieser Studie die Qualität der Videos keinen signifikanten Unterschied im Lernerfolg bewirkte, deutete sich an, dass Lernende motiviert bleiben, wenn sie sich mit einem Medium identifizieren und eigenständig relevante Inhalte erkennen (vgl. Sterzig, 2022).

Für die vorliegende Untersuchung bedeutet dies, dass Erklärvideos nicht nur als Informationsquelle, sondern auch als potenzieller Motivationsfaktor verstanden werden. Entscheidend ist dabei die Art der Gestaltung. Wie Kulgemeyer (2020) betont, entfalten Erklärvideos ihr lernförderliches Potenzial nur dann, wenn sie adressatengerecht strukturiert und didaktisch reflektiert konzipiert sind. Aufbauend auf diesem Verständnis untersucht die vorliegende Arbeit, inwiefern eine diversitätssensible Gestaltung solcher Videos die intrinsische Motivation beeinflusst.

2.2.2 Interaktive digitale Quizaufgaben

Ein weiteres zentrales Element digitaler Lernumgebungen stellen interaktive digitale Quizaufgaben dar, die Lernende aktiv in den Lernprozess einbeziehen und unmittelbares Feedback ermöglichen.

Während der Covid-19-Pandemie zeigte sich beispielsweise, dass der Einsatz der digitalen Lernplattform *Quizizz*,³ in der Lernende alleine oder in Echtzeit oder zeitversetzt spielen, um Wissen spielerisch zu festigen, im Grundlagenphysikunterricht nicht nur den Lernerfolg, sondern auch das Interesse und die Motivation der Studierenden deutlich steigerte (vgl. Henrich et al., 2022). Darüber hinaus belegen empirische Befunde, dass die konkrete Gestaltung digitaler Aufgaben einen erheblichen Einfluss auf die kognitive Belastung und die Bearbeitungsstrategien von Lernenden hat. So konnten Arslan et al. (2020) nachweisen, dass bereits geringfügige Änderungen im Aufgabenlayout, etwa bei Drag-and-Drop-Aufgaben, die Effizienz der Bearbeitung beeinflussen. Die Art der digitalen Interaktion spielt somit eine zentrale Rolle für die Qualität des Lernprozesses. Einen weiteren Zugang bieten gamifizierte Quizformate. Eine Masterarbeit von Rose (2015) zeigt, dass die Gamifizierung von Online-Physikübungen die Motivation und das Engagement der Lernenden steigern kann, jedoch nicht zwangsläufig zu besseren Lernergebnissen führt. Die

³Quizizz ist eine digitale Lernplattform, auf der Lehrkräfte interaktive, gamifizierte Quizze und Übungen erstellen können.

Ergebnisse betonen dennoch den motivationsfördernden Effekt spielerischer Elemente in digitalen Lernumgebungen. Diese Befunde werden von Özdemir (2024) durch eine Meta-Analyse zu *Kahoot!*⁴ gestützt. Die Ergebnisse zeigen, dass eine gamifizierte Lernplattform ein wirksames digitales Werkzeug ist, das den Unterricht interaktiver und motivierender gestaltet (vgl. Özdemir, 2024). Es steigert sowohl Lernerfolg und Behaltensleistung als auch Motivation und positive Einstellungen der Lernenden und kann Angstgefühle leicht reduzieren (vgl. Özdemir, 2024).

Neben geschlossenen Formaten wie Drag-and-Drop-, oder Multiple-Choice-Aufgaben können insbesondere offene Aufgabenformen zu einer tieferen inhaltlichen Auseinandersetzung führen. Offene Aufgabenformen unterstützen in interaktiven Tests mit unmittelbarem Feedback ein reflektierteres und kognitiv anspruchsvolleres Bearbeiten als Multiple-Choice-Formate und führen dadurch zu einer tieferen Auseinandersetzung mit dem Inhalt (vgl. Attali et al., 2015). Auch die Nutzung von Online-Übungen in Lernmanagementsystemen, wie *Moodle* hat sich als förderlich erwiesen. Solche e-Assessment-Formate unterstützen nicht nur den Wissenserwerb, sondern auch selbstgesteuertes Lernen, indem sie direkte Rückmeldungen ermöglichen und den Lernfortschritt transparent machen (vgl. Cohen & Sasson, 2016). Cohen und Sasson (2016) betonen, dass regelmäßiges Feedback durch digitale Quizze die Lernmotivation und die Entwicklung positiver Lernhaltungen wesentlich unterstützt. Dadurch wird e-Assessment zu einem integralen Bestandteil moderner Lehr-Lern-Konzepte, der sowohl Motivation als auch nachhaltigen Lernerfolg begünstigt (vgl. Cohen & Sasson, 2016).

Für die vorliegende Untersuchung lässt sich daraus ableiten, dass interaktive und gamifizierte Quizformate wie *Quizizz* oder *Kahoot!* in besonderem Maße geeignet sind, Lernende aktiv in digitale Lernprozesse einzubeziehen und ihre Motivation zu fördern. Entscheidend ist dabei, wie die Gestaltung der Aufgaben, die Lernmotivation der Lernenden beeinflusst. Motivationale Impulse können beispielsweise durch spielerische Elemente, unmittelbares Feedback und ein adressatengerechtes Design dargestellt werden. Interaktive Quizformate bilden daher einen zentralen Bestandteil der untersuchten Lernumgebung.

2.3 Diversität

Neben der fachlichen und technischen Gestaltung der Lernumgebung spielte auch die visuelle und inhaltliche Darstellung eine zentrale Rolle im Design der Einheit. Im folgenden Abschnitt wird das Thema Diversität im Kontext schulischen Lernens und insbesondere des Physikunterrichts behandelt. Zunächst wird der Begriff **Diversität** in seinen zentralen Dimensionen (Abschnitt 2.3.1) definiert, sowie die Intersektionalität (Abschnitt 2.3.2) beschrieben. Anschließend wird auf die Bedeutung von Diversität im Physikunterricht (Abschnitt 2.3.3) eingegangen, wobei insbesondere die Repräsentation unterschiedlicher gesellschaftlicher Gruppen in Unterrichtsmaterialien thematisiert wird. Darauf folgt eine Betrachtung der Darstellung von Diversität in audiovisuellen Medien, insbesondere in Erklärvideos und interaktiven Lernumgebungen (Abschnitt 2.3.4), die für die Konzeption dieser Untersuchung relevant sind. Der darauffolgenden Abschnitt 2.3.5

⁴*Kahoot!* ist eine digitale Lernplattform, auf der interaktive Quizspiele erstellt und in Echtzeit gespielt werden können, um Wissen auf spielerische Weise zu überprüfen.

erläutert den Zusammenhang zwischen Diversität und Motivation, um die theoretische Grundlage für die anschließende empirische Untersuchung zu schaffen. Abschließend wird in 2.3.6 die Bedeutung des Alters für die Wahrnehmung und Einordnung von Diversität in Lernkontexten betrachtet.

2.3.1 Definition der Diversität

Diversität (engl. diversity) bezeichnet im Deutschen allgemein Vielfalt und wird in verschiedenen Kontexten unterschiedlich verstanden. In der Soziologie beschreibt Diversität die Vielfalt und Heterogenität von Menschen aufgrund ihrer Gemeinsamkeiten und Unterschiede, die im gesellschaftlichen Kontext zu Privilegien oder Diskriminierung führen können (vgl. Krebs, 2023). Das bedeutet, dass jede Person Diversitätsfaktoren in sich trägt. Gardenswartz und Rowe⁵ gliedern hierfür die Persönlichkeit in vier Ebenen, die sogenannten *Four Layers of Diversity*: (1) Persönlichkeit, (2) innere Ebene, (3) äußere Ebene und (4) organisationale Ebene. Abbildung 2 veranschaulicht das Zusammenspiel dieser vier Ebenen, die als drei konzentrische Kreise die Persönlichkeit im Zentrum umgeben.

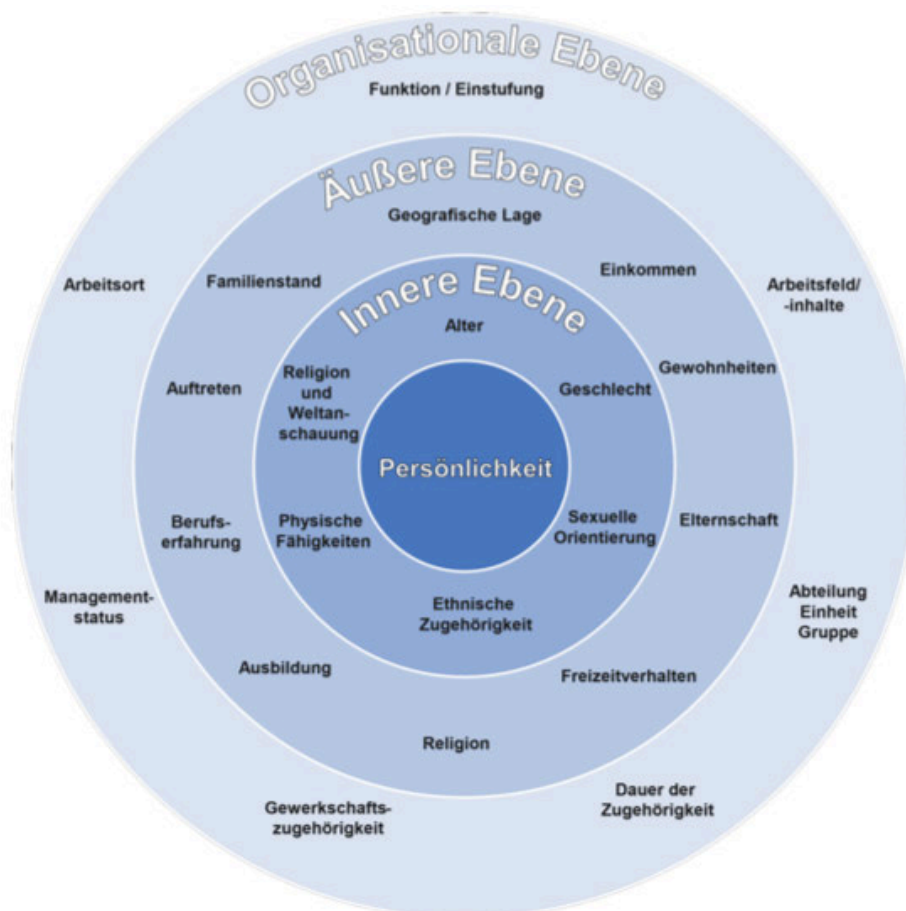


Abbildung 2: Four Layers of Diversity nach Gardenswartz und Rowe 2003, erstellt von Krebs (2023, S. 38)

Die konzentrischen Kreise verdeutlichen, dass die Persönlichkeit von unterschiedlichen Diversitätsfaktoren umgeben und beeinflusst wird (vgl. Wedl, 2018). Die innere Ebene umfasst weitge-

⁵zitiert nach Krebs (2023, S. 38)

hend unveränderbare Merkmale, die meist von Geburt an festgelegt sind, wie Alter, Geschlecht, sexuelle Orientierung, ethnische Zugehörigkeit, physische Fähigkeiten sowie Religion oder Weltanschauung. Die Merkmale der inneren Ebene sind eng mit der individuellen Identität verbunden und prägen maßgeblich die Wahrnehmung sowie das Erleben sozialer Zugehörigkeit oder Differenzierung. Die äußere Ebene beinhaltet Merkmale, die in gewissem Maße veränderbar oder frei wählbar sind, beispielsweise Einkommen, Familienstand, Elternschaft, Religion oder persönliche Gewohnheiten. Die organisationale Ebene schließlich beschreibt Diversitätsfaktoren im beruflichen und institutionellen Kontext, etwa Position, Abteilung, Arbeitsort oder Gewerkschaftszugehörigkeit.

Erkennbar ist, dass der Faktor Religion sowohl der inneren als auch der äußeren Ebene zugeordnet ist. Diese Doppelzuordnung ergibt sich daraus, dass religiöse Zugehörigkeit einerseits durch Sozialisation und kulturelle Prägung früh im Leben verankert und damit relativ stabil ist (innere Ebene), andererseits jedoch durch persönliche Entscheidungen oder gesellschaftliche Einflüsse verändert werden kann (äußere Ebene).

Im Rahmen dieser Arbeit wird der Begriff **Diversität** auf jene Merkmale der inneren Ebene bezogen, die als weitgehend unveränderbar gelten. Dazu zählen Alter, Geschlecht, sexuelle Orientierung, ethnische Zugehörigkeit, physische Fähigkeiten sowie Religion.

2.3.2 Intersektionalität

Für eine differenzierte Betrachtung von Diversität reicht es nicht aus, einzelne Dimensionen der Diversität isoliert zu betrachten. Genau diesen blinden Fleck adressiert das Konzept der Intersektionalität, das von Crenshaw (2020) entwickelt wurde. Sie zeigt, dass Diskriminierungserfahrungen Schwarzer Frauen nicht verstanden werden können, wenn Kategorien wie „Race“ und Geschlecht isoliert betrachtet werden. Erst die gemeinsame Betrachtung beider Kategorien macht spezifische Benachteiligungen sichtbar, die in eindimensionalen Ansätzen unsichtbar bleiben (vgl. Crenshaw, 2020). Ihr Ansatz macht deutlich, dass Menschen an den Schnittstellen mehrerer sozialer Positionierungen spezifischen Barrieren begegnen können, die über additive Erklärungen hinausgehen.

Intersektionalität beschreibt damit das Zusammenspiel verschiedener Dimensionen, wie Geschlecht, Ethnizität, Klasse, körperliche Verfasstheit oder Alter, und fragt, wie diese miteinander verflochten soziale Erfahrungen prägen (vgl. Winker & Degele, 2020). Für die Analyse solcher Überlagerungen schlagen Winker und Degele (2020) ein Mehrebenenmodell vor, das soziale Ungleichheiten auf drei miteinander verbundenen Ebenen beschreibt *Strukturebene* (gesellschaftliche Macht- und Herrschaftsverhältnisse), *Repräsentationsebene* (kulturelle Normen, Diskurse und symbolische Bedeutungen) und *Identitätsebene* (alltägliche Interaktionen und Prozesse der Identitätskonstruktion). Dieses Modell zeigt, dass Diskriminierungen nicht isoliert entstehen, sondern aus der Verknüpfung verschiedener sozialer Positionierungen hervorgehen können (vgl. Winker & Degele, 2020).

Das Verständnis von Intersektionalität kann somit zusammengefasst werden: Es beschreibt, wie

unterschiedliche Kategorien sozialer Ungleichheit zusammenwirken, welche Auswirkungen ihre Überschneidungen auf struktureller, kultureller und individueller Ebene haben und warum eindimensionale Erklärungsansätze komplexe Formen von Diskriminierungen nicht erfassen können (vgl. Winker & Degele, 2020). Hinweise darauf liefert auch die Masterarbeit von Tolksdorf (2017), die Physik-Schulbücher hinsichtlich unterschiedlicher sozialer Merkmale untersuchte und dabei aufzeigte, dass Kombinationen wie ethnische Zugehörigkeit, Geschlecht und Körpnormen bestimmte stereotype Darstellungen begünstigen können.

Im schulischen Kontext bedeutet eine intersektionale Perspektive, Diversität nicht nur entlang einzelner Merkmale zu betrachten. Vielmehr ist zu reflektieren, wie vielfältige Darstellungen in ihrer Gesamtheit Lernenden unterschiedliche Identifikationsmöglichkeiten eröffnen. Damit liefert Intersektionalität einen theoretischen Rahmen, um Diversität als mehrdimensionale Ressource für Motivation und Teilhabe zu verstehen.

2.3.3 Diversität im Physikunterricht

Der Bildungsauftrag von Schulen umfasst zunehmend die Aufgabe, die gesellschaftliche Vielfalt der Lernenden zu reflektieren und in der Unterrichtsgestaltung pädagogisch wirksam zu berücksichtigen (vgl. Bogedan, 2017), dies gilt auch für den Physikunterricht.

Die Physik gilt als stark maskulin konnotiertes Fach (vgl. Hannover, 2011). Dadurch werden stereotype Rollenerwartungen und implizite Vorurteile fortgeführt, die die Wahrnehmung von Kompetenz und Zugehörigkeit beeinflussen. Die Studie von Schnirch (2006) untersucht, wie Physikunterricht geschlechtergerecht gestaltet werden kann, um Interesse und Motivation von Schüler:innen zu fördern. Grundlage bildet die Selbstbestimmungstheorie, operationalisiert durch das Intrinsic Motivation Inventory (IMI).⁶ Der Unterricht wurde kontextorientiert und lebensweltbezogen gestaltet, um sowohl Mädchen als auch Jungen anzusprechen (vgl. Schnirch, 2006). Physikalische Themen wurden in alltagsnahe und soziale Kontexte eingebettet und durch offene, handlungsorientierte Lernformen ergänzt, die eigenständiges Arbeiten und Kooperation ermöglichten (vgl. Schnirch, 2006). In entwickelten Lernumgebungen wurden Realschulklassen untersucht, wobei beide Geschlechter positive Werte in den Subskalen *Interesse*, *wahrgenommene Kompetenz*, *wahrgenommene Wahlfreiheit* und *Nützlichkeit* zeigten (vgl. Schnirch, 2006). Damit adressiert die Lernumgebung zentrale motivationale Bedingungen, die für die Entstehung intrinsischer Motivation relevant sind. Zwischen Mädchen und Jungen ergaben sich keine signifikanten Unterschiede in der Motivation, was auf eine erfolgreiche gendergerechte Gestaltung hinweist (vgl. Schnirch, 2006). Mädchen an reinen Mädchenschulen erzielten jedoch tendenziell höhere Werte, insbesondere im *Interesse* und in der *wahrgenommenen Kompetenz* (vgl. Schnirch, 2006). Die Ergebnisse verdeutlichen, dass ein kontextualisierter, lebensweltbezogener und differenzierender Physikunterricht geeignet ist, geschlechtsspezifische Unterschiede in der Motivation zu reduzieren und die Lernmotivation insgesamt zu stärken.

⁶IMI ist ein Messinstrument zur Erfassung intrinsischer Motivation. Im Abschnitt 2.1.4 (S. 8) wird es näher erläutert.

Das Forschungsfeld der Physikdidaktik zeigt insgesamt ein deutliches Ungleichgewicht in der Thematisierung verschiedener Diversitätsdimensionen. Eine Metaanalyse der Zeitschrift *Physical Review Physics Education Research* (PRPER) ergab, dass „gender“ in 131 Artikeln thematisiert wurde, während Begriffe wie „sexual orientation“ nur in neun und „gender identity“ lediglich in zwei Artikeln vorkamen (vgl. Frazer, 2017). Ebenso selten werden physische Fähigkeiten und Beeinträchtigungen behandelt: „Disability“ wurde nur in zehn Artikeln erwähnt, „learning disability“ in zwei und „physical disability“ in einem einzigen Beitrag (vgl. Frazer, 2017). Die Autor:innen führen dies darauf zurück, dass viele Beeinträchtigungen unsichtbar sind (z. B. Aufmerksamkeitsdefizite) und daher selten explizit berücksichtigt werden, obwohl sie das Lernen in Physik erheblich beeinflussen können.

Diversität im Physikunterricht geht aber weit über Geschlechterfragen hinaus und umfasst verschiedene Dimensionen – etwa Leistungsvoraussetzungen, sprachliche und kulturelle Hintergründe, soziale Rollen sowie individuelle Lernwege (vgl. Krebs, 2023, S. 47–48). Tajmel (2009) beschreibt, dass die sprachliche und kulturelle Diversität der Gesellschaft den Physikunterricht vor neue Herausforderungen stellt. Ziel ist es, Chancengleichheit als Qualitätsmerkmal des Unterrichts zu etablieren (vgl. Tajmel, 2009). Benachteiligungen entstehen vor allem für Schüler:innen mit Migrationshintergrund, die in naturwissenschaftlichen Fächern oft geringere Leistungen zeigen (vgl. Tajmel, 2009). Ursachen sind laut Tajmel (2009) nicht Defizite der Lernenden, sondern strukturelle und didaktische Schwächen des Unterrichts. Vor allem sprachliche, kulturelle und institutionelle Barrieren erschweren den Zugang zu naturwissenschaftlicher Bildung (vgl. Tajmel, 2009). Zur Überwindung dieser Barrieren fordert Tajmel (2009): Sprachförderung als festen Bestandteil des Fachunterrichts, Einbeziehung unterschiedlicher Kulturen und Lebenswelten, geschlechtssensible Unterrichtsgestaltung, sowie gezielte Lehrer:innenfortbildung. Dadurch können Schüler:innen stärker mit dem Unterricht identifizieren, was Motivation und Interesse am naturwissenschaftlichen Lernen fördert (vgl. Tajmel, 2009).

Nicht nur die diversitätssensible Gestaltung des Physikunterrichts fördert die Identifizierung der Schüler:innen, sondern auch die visuelle und inhaltliche Gestaltung von Lehrmaterialien. Die Schulbuchanalyse von Tolksdorf (2017) zeigt, dass deutsche Physikschulbücher bisher nur geringe Sensibilität für Diversität aufweisen. Abgebildete Personen sind überwiegend weiß, männlich, jung und ohne erkennbare Behinderung, während Frauen, People of Color, ältere Menschen oder Personen mit körperlichen Einschränkungen kaum vorkommen (vgl. Tolksdorf, 2017). Diese eingeschränkte Repräsentation und die daraus resultierenden geringen Identifikationsmöglichkeiten der Lernenden könnten eine Ursache für die in den naturwissenschaftlichen Fächern beobachteten Leistungsunterschiede zwischen Schüler:innen sein (vgl. Tolksdorf, 2017). Sie hat weitreichende Folgen, da sie Lernenden unbewusst vermitteln kann, wer als „typisch“ für Physik gilt und wer dazugehört. Solche impliziten Botschaften können sich negativ auf Identifikation, Interesse und Selbstwirksamkeitserwartung von Schüler:innen auswirken, insbesondere bei jenen, die sich in den Darstellungen nicht wiederfinden.

Daher sollte eine diversitätssensible Physikdidaktik die Vielfalt der Lernenden stärker berück-

sichtigen und Lernumgebungen gestalten, die allen Schüler:innen Identifikationsmöglichkeiten bieten. Sichtbare Diversität, etwa durch inklusive Sprache (vgl. Tajmel, 2024) oder durch MINT-Vorbilderinnen, können Geschlechterstereotype abbauen und Freude Fach stärken (vgl. González-Pérez et al., 2020).

2.3.4 Diversitätsdarstellung in Erklärvideos und Quizaufgaben

Erklärvideos auf YouTube führen trotz des offenen und vielfältigen Charakters der Plattform häufig traditionelle Geschlechterstereotype fort (vgl. Honkomp-Wilkens & Wolf, 2024). In schulischen und naturwissenschaftlichen Themen dominieren vor allem weiße, männliche Personen, während Frauen und People of Color überwiegend in freizeitorientierten Bereichen auftreten (vgl. Honkomp-Wilkens & Wolf, 2024). Laut Honkomp-Wilkens und Wolf (2024) hinterfragen Jugendliche diese Geschlechterordnung kaum und gehen häufig davon aus, dass Männer besser erklären können (vgl. Honkomp-Wilkens & Wolf, 2024). Daher sollte Medienbildung stärker geschlechtersensibler gestaltet werden.

Diese Erkenntnisse verdeutlichen, dass die Repräsentation von Personen in audiovisuellen Lernformaten nicht nur Fragen der Sichtbarkeit, sondern auch der Wahrnehmung und Identifikation berührt. Dabei spielt insbesondere die Darstellungsform eine zentrale Rolle: Sollten die Inhalte animiert vermittelt werden oder durch reale Personen in authentischen Umgebungen (Live-Action-Darstellungen)?

Die Soziale Identitätstheorie nach Tajfel und Turner (1997) beschreibt, dass Menschen ihr Selbstbild wesentlich aus der Zugehörigkeit zu sozialen Gruppen ableiten. Individuen streben nach einer positiven sozialen Identität, die durch günstige Vergleiche der eigenen Gruppe mit relevanten Fremdgruppen gestützt wird. Wird die eigene Gruppe hingegen negativ bewertet, entstehen Strategien, um den Selbstwert zu stabilisieren, etwa durch eine Neubewertung von Gruppenmerkmalen oder eine verstärkte Identifikation mit der eigenen ethnischen Gruppe (In-Group). Aufbauend auf der Sozialen Identitätstheorie zeigen empirische Befunde, dass positive mediale Darstellungen der eigenen ethnischen Gruppe in Medien das Selbstwertgefühl und die „ingroup“-Bewertung von Angehörigen dieser Gruppe stärken können (vgl. McKinley et al., 2014). Insgesamt deuten die Befunde darauf hin, dass positive Repräsentationen vor allem innerhalb der eigenen Gruppe wirksam sind, während sie die Wahrnehmung oder Bewertung anderer Gruppen kaum beeinflussen.

In der Dissertation von Sato (2022) zeigte sich in einer Studie zu Lernvideos zum Vokabellernen kein Unterschied im Lernerfolg zwischen animierten und Live-Action-Videos. Allerdings wurden Videos mit menschlichen Stimmen als authentischer und lebendiger wahrgenommen als solche mit synthetischen Stimmen, wobei die Stimme einen stärkeren Einfluss auf das Engagement hatte als die visuelle Darstellungsform (vgl. Sato, 2022).

Für Lernumgebungen bedeutet dies, dass Rezipient:innen bzw. Lernende stärkeres Engagement zeigen können, wenn sie in medialen Darstellungen Personen wahrnehmen, die ihnen in zentralen Merkmalen ähneln (z. B. Ethnie, Geschlecht oder körperliche Fähigkeit). Solche Repräsentatio-

nen fördern das Gefühl von Zugehörigkeit, sozialer Eingebundenheit und Selbstwirksamkeit, was im Sinne der Selbstbestimmungstheorie von Ryan und Deci (2000) eine zentrale Voraussetzung für intrinsische Motivation bildet.

2.3.5 Diversität und Motivation

Nach Wlodkowski (2002) ist Motivation untrennbar mit kulturellen Erfahrungen und gesellschaftlichen Zugehörigkeiten verbunden. Lernprozesse, die Vielfalt sichtbar machen und an die kulturellen Hintergründe der Lernenden anknüpfen, fördern das Gefühl von Zugehörigkeit, Relevanz und Kompetenz – zentrale Voraussetzungen intrinsischer Motivation (vgl. Wlodkowski, 2002). Studien zum Stereotype Threat belegen, dass Lernleistungen insbesondere dann stabil bleiben oder sich verbessern, wenn Lernende ein Umfeld erleben, das Zugehörigkeit, Sicherheit und positive Identifikationsmöglichkeiten bietet (vgl. Eckert, 2014). Solche Umgebungen verhindern die kognitive Belastung durch stereotype Erwartungen und unterstützen damit die Verarbeitung und den Lernerfolg.

Die Studie von Cheryan et al. (2009) untersuchte in vier Experimenten, wie stereotypisch gestaltete Umgebungen in der Informatik die Wahrnehmung von Zugehörigkeit und das Interesse potenzieller Studierender beeinflussen. Dabei wurde getestet, ob visuelle Objekte – wie Poster, Zeitschriften oder Dekorationen – Geschlechterstereotype über Informatiker:innen aktivieren und so die Fachwahl beeinflussen (vgl. Cheryan et al., 2009). Die Ergebnisse zeigten, dass Frauen in stereotypisch „männlich“ geprägten Umgebungen (z. B. mit Science-Fiction-Postern und Videospielen) signifikant weniger Interesse an Informatik zeigten als Männer (vgl. Cheryan et al., 2009). In neutral oder nicht-stereotypischen Umgebungen (z. B. mit Naturbildern oder Kunstpostern) verschwanden diese Geschlechterunterschiede. Die Befunde wurden durch das Gefühl der Zugehörigkeit vermittelt: Frauen fühlten sich in den stereotypischen Umgebungen weniger zugehörig und identifizierten sich weniger mit dem Fach (vgl. Cheryan et al., 2009). Cheryan et al. (2009) schließen daraus, dass stereotype visuelle Reize in Lern- und Arbeitsumgebungen signalisieren „wer dazugehört“ (vgl. Cheryan et al., 2009). Eine bewusst vielfältige und weniger stereotype Gestaltung kann daher das Zugehörigkeitsgefühl und das Interesse – insbesondere von Frauen – erhöhen. Darüber hinaus könnte eine breitere Repräsentation von Identitäten auch Männern zugutekommen, die sich mit bestimmten Aspekten des Informatikstereotyps (z. B. sozialer Isolation oder exzessiver Technikfokussierung) nicht identifizieren (vgl. Cheryan et al., 2009).

2.3.6 Altersabhängige Wahrnehmung und Wirkung von Diversität

Um die theoretische Grundlage der vorliegenden Untersuchung zu vervollständigen, wird im Folgenden der Frage nachgegangen, welche Rolle das Alter für die Wahrnehmung und Bedeutung von Diversität in Lernkontexten spielt.

Mulvey et al. (2022) zeigen, dass ein soziales Klima in naturwissenschaftlichen Unterrichtskontexten, zentrale Bedeutung für das Engagement von Jugendlichen besitzt. Dabei sind wahrge-

nommene Inklusion, erlebte Diskriminierung und Zugehörigkeit zentrale Prädiktoren des Engagements. Auffällig ist, dass ältere Schülerinnen und Schüler in diesen Bereichen niedrigere Werte berichten und häufiger Diskriminierung wahrnehmen, was auf eine zunehmende Sensibilität gegenüber Ungleichheiten im Verlauf der Adoleszenz hinweist (Mulvey et al., 2022). Die Studie unterstreicht damit die Bedeutung eines wertschätzenden und diversitätssensiblen Unterrichtsklimas für motivationale Prozesse in naturwissenschaftlichen Lernumgebungen.

Auch die Analyse internationaler Bildungsdatensätze durch Kleemola et al. (2025) macht es deutlich, dass Zugehörigkeitsgefühl altersabhängig variiert. Während in der Primarstufe noch vergleichsweise geringe Differenzen bestehen, nehmen sie in der Sekundarstufe deutlich zu und fallen in der Tertiärbildung am stärksten aus. Besonders Lernende aus sozioökonomisch benachteiligten oder migrierten Gruppen berichten mit zunehmendem Alter ein sinkendes Zugehörigkeitsgefühl, was auf eine wachsende Sensibilität gegenüber Ungleichheiten hindeutet (vgl. Kleemola et al., 2025).

Vor diesem Hintergrund erscheint es plausibel, dass die Wirkungen einer diversitätssensiblen Darstellung in digitalen Lernumgebung auf die Motivation von Schüler:innen je nach Jahrgangsstufe unterschiedlich ausfallen, da sie auf altersabhängig unterschiedliche motivationalen Ausgangslagen und Zugehörigkeitsgefühl trifft.

3 Forschungsfragen

Bisherige Studien zeigen, dass digitale Lernumgebungen durch audiovisuelle Medien wie Erklärvideos und interaktive Aufgaben Lernprozesse unterstützen und die Motivation der Lernenden fördern können (vgl. Cheryan et al., 2009; Conradt, 2009; Sterzig, 2022; Henukh et al., 2022; Arslan et al., 2020).

Zugleich belegen Forschungsarbeiten zu Diversität im Bildungskontext, dass vielfältige Repräsentationen von Geschlechtern, kulturellen Hintergründen oder Lebensrealitäten die Identifikation und das Zugehörigkeitsgefühl erleben von Lernenden fördern können (vgl. Tajmel, 2024; Tajfel & Turner, 1997; McKinley et al., 2014). Diese Befunde legen nahe, dass das Bedürfnis nach sozialer Eingebundenheit im Sinne der Selbstbestimmungstheorie nach Deci und Ryan (1993) nicht nur durch reale Interaktionen, sondern auch durch symbolische Repräsentationen in medialen Lernformaten angesprochen werden kann. Aus dieser theoretischen Annahme lässt sich ableiten, dass eine diversitätssensible Gestaltung digitaler Lernumgebungen wesentliche Faktoren der Motivation positiv beeinflussen kann.

Die dargestellten theoretischen Überlegungen verdeutlichen auch, dass Diversität ein zentrales Element einer zukunftsorientierten Physikdidaktik ist (vgl. Tajmel, 2024; Tolksdorf, 2017; Schnirch, 2006; Krebs, 2023). Unterschiede in der Wahrnehmung und Repräsentation sozialer Gruppen beeinflussen, wer sich mit Physik identifiziert und welche Lernenden sich im Fach wiederfinden. In realen Lernkontexten treten Diversitätsdimensionen jedoch selten isoliert auf, sondern wirken häufig intersektional, also in ihrem Zusammenspiel (vgl. Krebs, 2023). Trotz dieser theoretischen Grundlage und einzelner empirischer Hinweise existieren bislang keine

systematischen Untersuchungen, die den Einfluss von Darstellungen von Diversitätsdimensionen innerhalb digitaler Lernumgebungen auf die intrinsische Motivation von Schüler:innen analysieren. Besonders im Fach Physik, das nachweislich durch eingeschränkte Repräsentationsvielfalt und stereotype Rollenbilder geprägt ist, fehlt es an empirischen Befunden dazu, inwiefern visuelle oder narrative Diversität in Erklärvideos motivationale Prozesse beeinflusst.

Zusätzlich zeigt die Studie von Mulvey et al. (2022), dass Jugendliche mit zunehmendem Alter sensibler auf Aspekte von Inklusion, Zugehörigkeit und erlebter Ungleichbehandlung reagieren. Da diese Faktoren zentrale Prädiktoren für das Engagement in naturwissenschaftlichen Lernumgebungen darstellen, erscheint es plausibel, dass Diversitätsdarstellungen in einer digitalen Lernumgebung je nach Jahrgangsstufe unterschiedlich auf die Motivation wirken können.

Die vorliegende Arbeit greift diese Forschungslücke auf, indem sie den Einfluss von Diversitätsdarstellungen in einer digitalen Physik-Lernumgebung empirisch untersucht und damit einen Beitrag zur Verbindung von Motivationsforschung, Mediendidaktik und Diversity Education im naturwissenschaftlichen Lernen leistet. Daraus ergeben sich die folgenden Forschungsfragen:

- (F1) Wie wirken diversitätsberücksichtigende digitale Lernumgebungen auf die intrinsische Motivation von Schüler:innen?
- (F2) Inwiefern unterscheiden sich die Wirkungen diversitätsberücksichtigender digitaler Lernumgebungen auf die Motivation von Schüler:innen je nach Jahrgangsstufe?

4 Methode

Dieses Kapitel beschreibt das methodische Vorgehen der vorliegenden Arbeit. Ziel ist es, das Forschungsdesign, die Rahmenbedingungen der Datenerhebung sowie die eingesetzten Instrumente und Auswertungsverfahren darzustellen.

Zunächst wird das Forschungsdesign im Abschnitt 4.1 erläutert, das den Aufbau und die Struktur der Untersuchung beschreibt. Anschließend werden im Abschnitt 4.2 das Setting und unter 4.4 die Zielgruppe vorgestellt. Darauf folgt die Beschreibung der Intervention (Abschnitt 4.4), in der die Konzeption und Darstellung der digitalen Lernumgebung erläutert wird. Im Anschluss wird im Abschnitt 4.5 das verwendete Messinstrument zur Erfassung der Motivation der Lernenden vorgestellt, erläutert und das Vorgehen zur Überprüfung seiner Gütekriterien dargelegt. Abschnitt 4.7 erläutert den Ablaufplan der Durchführung und welche Besonderheiten und Probleme sich bei der Durchführung gezeigt haben. Im letzten Abschnitt 4.7 werden die Datenaufbereitung und -auswertung beschrieben, um das methodische Vorgehen der Untersuchung nachvollziehbar darzustellen.

4.1 Forschungsdesign

Beim Forschungsdesign handelt es sich um eine quantitativ-experimentelle Untersuchung mit qualitativen Ergänzungen, bei der eine Einzelmessung der intrinsischen Motivation jeder/jedes

Teilnehmenden mittels Fragebogen nach Abschluss der digitalen Lernumgebung durchgeführt wurde. Das Forschungsdesign basiert auf den Vergleich zweier Gruppen.

Zur besseren Veranschaulichung zeigt Abbildung 3 den zeitlichen Ablauf der Intervention. Die Untersuchung umfasste eine 45-minütige Unterrichtseinheit. Zu Beginn erhielten die Lernenden eine etwa fünfminütige Einführung, in der der Untersuchungsablauf erläutert und die Gruppeneinteilung vorgenommen wurde. Anschließend bearbeiteten die Lernenden rund 25 Minuten lang die digitale Lernumgebung zum Thema Feinstaub und beantworteten abschließend den Fragebogen zur Erfassung ihrer intrinsischen Motivation.

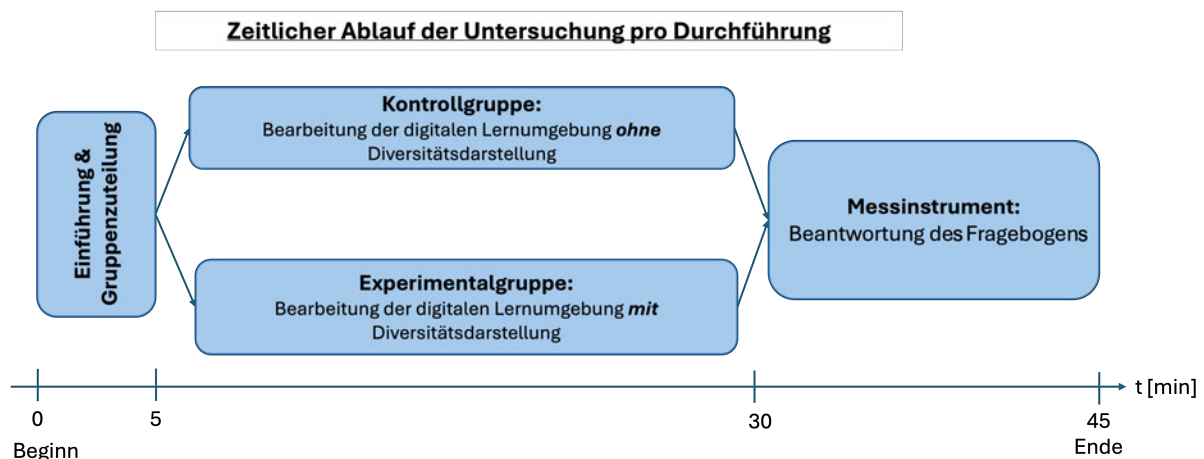


Abbildung 3: Zeitlicher Ablauf der Untersuchung pro Durchführung

Die unabhängige Variable stellt die Gestaltung der digitalen Lernumgebung dar. Beide Gruppen erhielten inhaltlich identische Materialien, die sich lediglich in der Darstellung von Diversität unterscheiden. Die Zuteilung zu den Gruppen erfolgte randomisiert, um systematische Verzerrungen zu vermeiden.

- **Experimentalgruppe:** Diese Gruppe nutzte eine Version mit sichtbaren Diversitätsdarstellungen, die verschiedene Dimensionen umfassten: Geschlecht, ethnische Zugehörigkeit, physische Fähigkeiten, Religion und sexuelle Orientierung. Die Diversität zeigte sich in den dargestellten Personen innerhalb der Lernumgebung.
- **Kontrollgruppe:** Diese Gruppe arbeitete mit einer inhaltlich identischen, jedoch anders gestaltete Version, in der keine expliziten Diversitätsmerkmale vorkamen. Die abgebildeten Personen entsprachen dabei überwiegend dem traditionellen Bild weißer, cis-männlicher Akteure.

Im Folgenden werden zur besseren Lesbarkeit vereinfachte Bezeichnungen verwendet: Die Experimentalgruppe, die eine digitale Lernumgebung mit Diversitätsdarstellung erhielt, wird als *Diversitäts-Gruppe* bezeichnet, während die Kontrollgruppe ohne Diversitätsdarstellung im Weiteren als *Ohne-Diversität-Gruppe* bezeichnet wird.

4.2 Settings: Humboldt Explorers

In diesen Abschnitt werden die Rahmenbedingungen dieser Untersuchung näher erläutert. Die durchzuführende digitale Lernumgebung zum Thema Feinstaub wurde im Rahmen des Citizen-Science-Schul-Projekts *Humboldt Explorers* entwickelt, das von der Bayer Foundation getragen wird. Das Projekt richtet sich an Berliner Lerngruppen der 5. bis 10. Jahrgangsstufe und verfolgt das Ziel, das Interesse und die Motivation bei den Lernenden für die Naturwissenschaften zu fördern. Dabei liegt ein Fokus auf der Förderung von 21st Century Skills (vgl. Hosang, 2021), also Lernkompetenz, digitale Kompetenz und Lebenskompetenz, sowie der Förderung der Fähigkeit der Data Literacy, Daten kritisch zu erfassen, zu analysieren und reflektiert zu nutzen (vgl. Statistisches Bundesamt, o.D.). Die thematischen Schwerpunkte der Lerneinheiten orientieren sich an der Lebenswelt der Schüler:innen und greifen gesellschaftlich relevante Fragen des urbanen Klima- und Umweltschutzes auf.

Bereits bestehende Lerneinheiten zu Themen wie Geräuschpegel oder pflanzliche Artenvielfalt ermöglichen es Lernenden, mithilfe digitaler Werkzeuge wie Tablets und *senseBoxen* eigene Messdaten zu erheben, auszuwerten und zu interpretieren. Ergänzt werden sie durch Erklärvideos und interaktive Quizaufgaben, die Wissenserwerb und Motivation fördern. In der nächsten Projektphase sollen das Humboldt-Explorers-Projekt internationalisiert und neue Lernumgebungen entwickelt werden. Dazu entsteht derzeit eine Einheit zum Thema Feinstaub, die sowohl Kenntnisse zu Ursachen, Auswirkungen und Maßnahmen zur Reduktion vermittelt als auch das Projekt auf die Sekundarstufe II ausweitet. Die neue Lernumgebung umfasst ein Erklärvideo sowie dazugehörige interaktive Quizaufgaben, die sowohl dem im Humboldt-Explorers-Projekt etablierten Lernkonzept folgen als auch die in der Theorie im Abschnitt 2.2.1 beschriebenen motivationalen Vorteile digitaler Lernformate aufgreifen. Dieser Teil der Lernumgebung ist bewusst zu Beginn platziert, da er kein spezifisches Vorwissen der Schüler:innen voraussetzt. Dadurch kann er auch unabhängig von den weiteren Abschnitten der Lerneinheit durchgeführt werden und dient als thematische Einführung in das Gesamtthema. Somit wurde dieser Teil der Lernumgebung, bestehend aus einem Erklärvideo und Quizaufgaben, sowohl im Rahmen des nächsten Projektziels des *Humboldt-Explorers-Projekts* als auch im Kontext dieser Arbeit entwickelt und als Pilotversuch umgesetzt.

4.3 Zielgruppe

Neben der inhaltlichen Bedeutung des Themas bietet Feinstaub auch methodisch ein großes Potenzial für den Physikunterricht. Die Beschäftigung mit realen Messwerten und deren Auswertung fördert bei Schüler:innen zentrale Kompetenzen der Erkenntnisgewinnung, insbesondere den reflektierten Umgang mit Daten, also das Überprüfen, Strukturieren und Interpretieren von Messwerten. In der digitalen Lernumgebung geschieht dies unter anderem durch interaktive Quizaufgaben, in denen reale Messdaten analysiert und in neue Zusammenhänge eingeordnet werden. Zugleich werden Elemente der Kommunikationskompetenz angesprochen, da die Lernenden physikalische Phänomene aus ihrer eigenen Lebenswelt kritisch analysieren und die gewonnenen

Erkenntnisse sachgerecht darstellen und begründen. Die Auseinandersetzung mit authentischen Datensätzen unterstützt damit nicht nur den Erwerb fachlicher Fähigkeiten, sondern stärkt auch grundlegende Aspekte des wissenschaftlichen Denkens und Arbeitens (vgl. SenBJF, 2015).

Die Jahrgangsstufen ab der 9. Klassen sind besonders geeignet, da die Lernenden gemäß dem Rahmenlehrplan über die grundlegenden Kompetenzen im Umgang mit physikalischen Messdaten verfügen (vgl. SenBJF, 2021). Sie sind in der Lage, Messwerte zu erfassen, auszuwerten und in verschiedenen Darstellungsformen (z. B. Tabellen, Diagrammen oder Texten) zu interpretieren (vgl. SenBJF, 2021). Der Bildungsplan betont in den prozessbezogenen Kompetenzen des Bereichs Erkenntnisgewinnung und Kommunikation, dass Schüler:innen Experimente durchführen, Messdaten erfassen und auswerten sowie Informationen aus unterschiedlichen Darstellungsformen entnehmen und überführen können (vgl. SenBJF, 2021). Somit stellt die Arbeit mit realen Messdaten in dieser Untersuchung keine neue Kompetenz dar, sondern knüpft an bereits erworbene Fähigkeiten der Schüler:innen an.

Die deutsche Auslandsschule in [REDACTED] verfügte zum Zeitpunkt der Untersuchung im Vergleich zu Berliner Schulen über eine deutlich geringere Schüler:innenzahl (221), sodass pro Jahrgangsstufe jeweils nur eine Klasse existierte. Daher werden die Begriffe Jahrgangsstufe und Klasse in dieser Arbeit synonym verwendet.

Die gesamte Anzahl der Schüler:innen ab der 9. Klassenstufe an der Schule in [REDACTED] betrug insgesamt 52, wobei 50 von ihnen die digitale Lernumgebung bearbeitet hatten. Die Zuteilung zu den jeweiligen Experimental- bzw. Kontrollgruppen erfolgte über individuell generierte QR-Codes, die in Papierform ausgeteilt wurden. Durch dieses Verfahren wurde eine zufällige (randomisierte) Zuordnung der Teilnehmenden zu den beiden Bedingungen sichergestellt. Zu Beginn jeder Durchführung wurde darauf geachtet, dass die Anzahl der ausgeteilten QR-Codes in beiden Gruppen ausgeglichen war, um eine gleichmäßige Verteilung zwischen der Diversitäts- und der Ohne-Diversitäts-Gruppe zu gewährleisten.

In Tabelle 1 ist die Verteilung der Teilnehmenden nach Klassenstufen, Klassengröße sowie Zuordnung zu den beiden teilgenommenen Untersuchungsgruppen dargestellt.

Klassenstufe	Klassenstärke	Ohne- Diversität-Gruppe	Diversitäts-Gruppe
9	13	7	6
10	13	6	7
11	12	6	6
12	14	6	6
Gesamt Anzahl	52	25	25

Tabelle 1: Übersicht der Teilnehmenden nach Klassenstufe, Klassenstärke und Gruppenzugehörigkeit

4.4 Intervention

In diesen Abschnitt werden die Maßnahmen beschrieben, die im Rahmen der experimentellen Untersuchung eingesetzt wurden. Sowohl für die Experimentalgruppe (Diversitäts-Gruppe) als auch die Kontrollgruppe (Ohne-Diversitäts-Gruppe) wurde das entsprechende Erklärvideo in einem eigenen Abschnitt der digitalen Lernumgebung (erstellt mit *WordPress*) eingebettet. Unterhalb des Videos befanden sich eigens entwickelte Quizaufgaben zum Thema Feinstaub, die sich inhaltlich auf das Video bezogen.

Die digitale Lernumgebung der Ohne-Diversität-Gruppe (Kontrollgruppe) ist unter <https://www.humboldt-explorers.de/lessons/feinstaub-2/> abrufbar, und die der Diversitäts-Gruppe (Experimentalgruppe) unter <https://www.humboldt-explorers.de/lessons/feinstaub/>.

4.4.1 Erklärvideo

Die inhaltliche und didaktische Gestaltung orientierte sich an den von Kulgemeyer (2020) formulierten sieben Kernideen zur Qualität von Erklärvideos (siehe Tabelle 14, S. 75). Diese Kriterien bildeten die Grundlage für die Konzeption des Storyboard (siehe Abbildung 18, S. 85) und des Voice-over-Skripts (siehe Abbildung 17, S. 78), die beide im Anhang dokumentiert sind. Das Storyboard wurde auf Basis des Voice-over-Skripts entwickelt und diente als visuelle Planungshilfe für die spätere Umsetzung des Videos. Diese soll zunächst neutral dargestellt werden, so dass es sowohl für die Diversitäts-Gruppe als auch Ohne-Diversität-Gruppe eingesetzt werden kann. Erwähnenswert ist, dass das Voice-over in beiden Videovarianten von derselben Person aus der Arbeitsgruppe der Humboldt-Explorers gesprochen wurde, sodass die Sprecherstimme keinen Einfluss auf die Untersuchungseffekte hinsichtlich der Diversitätsdarstellung haben.

Inhaltlich: Inhaltlich soll das Video zentrale umweltrelevante Aspekte von Feinstaub, darunter die Feinstaubpartikel, Entstehung, gesundheitliche und klimatische Auswirkungen, Messverfahren sowie die WHO Grenzwerte vermitteln. Dabei wurde eine Sprachebene gewählt, die für Lernende ohne vertiefte Vorkenntnisse verständlich ist und mögliche Fehlvorstellungen gezielt adressiert. So wird beispielsweise verdeutlicht, dass Feinstaub keine Gase, sondern sehr kleine Partikel sind, die für das menschliche Auge unsichtbar sind, und dass er weiträumig transportiert werden kann und dass selbst sehr kleine Mengen langfristig gesundheitsschädlich wirken können. Fachbegriffe wie PM10 oder PM2,5 werden schrittweise eingeführt und mit anschaulichen Beispielen erläutert, etwa durch den Vergleich der Feinstaubmasse mit dem Gewicht einer 1-Cent-Münze. Diese Veranschaulichungen dienen der kognitiven Entlastung und der besseren Vorstellbarkeit abstrakter Größenordnungen.

Das Erklärvideo folgt einer klaren inhaltlichen Struktur: Es beginnt mit einer Frageliste, wie zum Beispiel „Was ist Feinstaub?“, die im Laufe des Erklärvideos beantwortet werden soll. Diese Fragen führen die Lernenden über die Entstehung und Wirkung von Feinstaub bis hin zu seiner Messung und den Grenzwerten der WHO. Im Anschluss an das Video bearbeiten die Schüler:innen Aufgaben, in denen reale Feinstaubdaten aus Berlin ausgewertet werden sollen. Um

den Übergang zwischen Video und anschließender Aufgabenbearbeitung zu erleichtern, wird im Video die Webseite *Luftdaten.berlin.de* vorgestellt. Dabei werden exemplarisch Messstationen, Diagramme sowie die verschiedenen Einstellungsmöglichkeiten der Darstellungssequenz gezeigt. Durch diese Einbindung wird sichergestellt, dass die Lernenden mit der Struktur der Webseite und den dargestellten Daten bereits vertraut sind und die nachfolgenden Analyseaufgaben selbstständig und zielgerichtet bearbeiten können. In der abschließenden Sequenz des Videos werden die eingangs formulierten Fragen systematisch beantwortet und die zentralen Inhalte noch einmal zusammenfassend hervorgehoben. Die inhaltliche Information über das Thema Feinstaub, sowie das Voice-Over (sowohl inhaltlich als auch stimmlich) ob inhaltlich oder auch Stimme sind für beide Videoversionen identisch. Die Unterschiede liegen in der Darstellung der Personen.

Animation und Live-Action: Es stellt sich die Frage, ob die Figuren und die inhaltlichen Elemente zum Thema Feinstaub im Erklärvideo animiert oder als Live Action (mit realen Personen) dargestellt werden sollen.

Die Studie von Champoux (2005) betont in seiner vergleichenden Analyse, dass Live-Action-Filme einen Einblick in die Realität geben und durch reale Mimik, Körpersprache und authentische Interaktionen stärker zur emotionalen Bindung und sozialen Identifikation beitragen (vgl. Champoux, 2005). Animationsfilme hingegen eröffnen mehr gestalterische Freiheit und können die symbolische Bedeutungen von Theorien und Konzepten besser darstellen und eignen sich somit besonders, um abstrakte Ideen zu vermitteln. Während Live-Action-Darstellungen stärker an die soziale Wirklichkeit anschließen, regen animierte Szenen vor allem die Vorstellungskraft und kognitive Verarbeitung an. Aus diesem Grund wurde eine multimodale Gestaltung gewählt, bei der Animation und Live-Action-Sequenzen kombiniert werden, um komplexe Inhalte zugleich anschaulich und ansprechend zu vermitteln. So werden die alltägliche Aktivitäten, wie das Kochen, oder auf den Straßen zu sein in realitätsnahen Sequenzen dargestellt, während die Entstehung sekundärer Feinstaubpartikel (siehe linkes Bild in Abbildung 4) und der Einfluss auf die Gletscherschmelze durch die Verringerung der Albedo⁷ (siehe rechtes Bild in Abbildung 4) in rein animierten Szenen visualisiert wurden die mithilfe von PowerPoint erstellt wurden. Besonders das Eindringen der Feinstaubpartikel (PM_{2,5}) in die Atemwege bis hin zu Bronchien und Bronchiolen wird als Animation verdeutlicht. Darüber hinaus wurden gezielte Szenenwechsel und Sprechpausen eingeplant, um den Lernenden ausreichend Verarbeitungszeit zu geben und die Aufmerksamkeit auf neue Inhalte zu lenken.

⁷Die Albedo bezeichnet das Rückstrahlvermögen einer Oberfläche und gibt an, welcher Anteil der einfallenden elektromagnetischen Strahlung von ihr reflektiert wird.



Abbildung 4: Animationsszenen

Die audiovisuelle Gestaltung wurde mit der Software *Da Vinci Resolve 17* erstellt. Die Gesamtvideolänge beträgt 6 Minuten und 52 Sekunden.

Diversitätsdarstellung bei den beiden Gruppen: Das Erklärvideo der ohne-Diversität-Gruppe sollte in den Szenen mit Personen überwiegend weiße männliche Akteure zeigen. Eine bewusste Ausnahme ist die Küchenszene⁸, in der eine Frau kocht, während der Mann entspannt am Esstisch sitzt und auf das Essen wartet. Diese Darstellung greift gezielt ein traditionelles Hausfrauenbild auf, um gängige Rollenklischees in der Kontrollgruppe beizubehalten. Ebenso zeigt die Kaminszene ein heterosexuelles Paar, das gemeinsam Zeit verbringt. Die Wohnzimmer-Szene dient der Darstellung einer religiösen Zugehörigkeit. Für die Gruppe Ohne-Diversität wurde dabei das Christentum gewählt, da es in der deutschen Bevölkerung die Mehrheitsreligion darstellt (vgl. fowid, 2024). Diese Zugehörigkeit kann in der Szene beispielsweise durch eine auf dem Tisch liegende Bibel, ein Kreuz an der Wand oder ein Bild von Jesus symbolisiert werden.

Das Erklärvideo der Diversitäts-Gruppe wurde so konzipiert, dass fünf von sechs Diversitätsdimensionen (Geschlecht, ethnische Zugehörigkeit, physische Fähigkeit, sexuelle Orientierung, Religion) in unterschiedlichen Szenen gezielt aufgegriffen werden. Die Auswahl der in der Diversitätsversion des Erklärvideos dargestellten Dimensionen orientierte sich an etablierten Differenzlinien, die in der intersektionalen und bildungswissenschaftlichen Forschung als besonders relevant für schulische Teilhabe und Repräsentationsprozesse gelten (vgl. Crenshaw, 2020; Winker & Degele, 2020; Tajmel, 2024). Da Lernende innerhalb einer Klassenstufe in der Regel eine relativ homogene Altersstruktur aufweisen, wurde die Dimension „Alter“ in der Gestaltung der Videoversionen nicht als eigenständige Repräsentationskategorie hervorgehoben. Stattdessen lag der Schwerpunkt auf den anderen fünf Diversitätsdimensionen. Da Diversitätsdimensionen in der Realität häufig intersektional überlappen und nicht eindeutig voneinander zu trennen sind, wie etwa die Darstellung einer Frau mit Kopftuch, die sowohl eine ethnische Zugehörigkeit als auch eine religiöse Identität repräsentiert, wurde in der Gestaltung darauf geachtet, jede Dimension in einer eigenen Szene besonders hervorzuheben und visuell klar umzusetzen.

Ursprünglich war vorgesehen, beide Videovarianten mit realen Personen aus dem Freundes- und Bekanntenkreis bzw. aus dem Projekt *Humboldt Explorers* zu besetzen. Da sich jedoch nur wenige Personen für eine Mitwirkung, insbesondere für die Version ohne Diversitätsdarstellung,

⁸siehe Szene 5 auf S. 3 des Storyboards, das im Anhang auf S. 85 befindet

faden, wäre eine eindeutige Unterscheidung zwischen den beiden Varianten nur eingeschränkt möglich gewesen. Aus diesem Grund wurde entschieden, für einige Szenen auf das KI-basierte Video-Tool *Sora* zurückzugreifen.

Der Einsatz dieses Tools stellte sich als zeitaufwendig und technisch anspruchsvoll heraus, da die generierten Videosequenzen teilweise nicht den im Storyboard vorgesehenen Szenen entsprachen. Einzelne Bewegungsabläufe, Mimik, Interaktionen oder auch die Darstellung des menschlichen Körpers wirkten zum Teil unnatürlich, was auf die technischen Limitierungen des Programms *Sora* in seiner frühen Entwicklungsphase (Ende 2024) zurückzuführen ist. Aus diesem Grund mussten Sequenzen mehrfach angepasst oder vollständig neu erstellt werden. Infolgedessen weichen bei einigen Szenen die Darstellung, Handlung als auch bestimmte Diversitätsdimensionen von ursprünglichen Ideen und dem Storyboard ab. Diese Anpassungen erfolgten mit dem Ziel, die beabsichtigten Darstellungen im finalen Erklärvideo so authentisch wie möglich darzustellen. Allerdings zeigen sich bei einigen Szenen noch kleine Unstimmigkeiten in der Körperdarstellung von Figuren, wie zum Beispiel, dass in der Kaminszene der Ohne-Diversität-Gruppe der Mann mit Brüsten dargestellt wurde (siehe Tabelle 3, Szene *Kamin* der Ohne-Diversität-Gruppe). In einer weiteren Szene der Diversitäts-Gruppe, in der mehrere Personen im Park Karten spielen, ist zu erkennen, dass alle Spielenden die Rückseite der Karten zu sich selbst halten und eine Person mit einem Finger eine Handkarte einer anderen greift (siehe Tabelle 2, Szene *Park* der Diversitäts-Gruppe). Solche Abweichungen stellen jedoch keine inhaltliche Einschränkung dar, da die intendierten Darstellungsmerkmale der Szenen erhalten blieben.

In Tabelle 4 werden die jeweiligen Videoausschnitte beider Erklärvideoversionen präsentiert, die denselben Szenenabschnitt zur gleichen Spielzeit zeigen. Ergänzend werden die dargestellte Szene und die zugehörigen Diversitätsdimensionen beschrieben sowie eine Erläuterung der jeweiligen Szene im Hinblick auf Diversität gegeben.

Szene	Ohne-Diversität-Gruppe	Diversitäts-Gruppe
Straße I (0:12:20)		
	<p>Szenenbeschreibung: Ein Geschäftsperson steht an einer belebten Straße.</p> <p>Erläuterung: Die Diversitätsversion hebt die Sichtbarkeit unterschiedlicher ethnischer Gruppen im beruflichen Alltag hervor. Sie vermittelt das Bild einer offenen, vielfältigen Gesellschaft, in der Menschen verschiedener Herkunft selbstverständlich Teil des öffentlichen Lebens sind.</p>	
Park (0:54:17)		
	<p>Szenenbeschreibung: In einem Park spielen drei Personen ein Kartenspiel.</p> <p>Erläuterung: Die Diversitätsversion symbolisiert ein friedliches und gleichberechtigtes Zusammenleben verschiedener ethnischer Gruppen. Vielfalt wird hier als selbstverständlich und positiv erlebt.</p>	

Tabelle 2: Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 1/3

Szene	Ohne-Diversität-Gruppe	Diversitäts-Gruppe
Kamin (1:15:12)		
<p>Szenenbeschreibung: Ein Paar, das im Kaminzimmer die gemeinsam Zeit genießt.</p> <p>Erläuterung: In dieser Diversitätsversion wird die Dimension sexuelle Orientierung aufgegriffen. Die Szene soll die Selbstverständlichkeit gleichgeschlechtlicher Beziehungen im Alltag verdeutlichen. Durch die ruhige und harmonische Darstellung soll ein positives und normales Bild unterschiedlicher sexueller Orientierungen vermittelt werden.</p>		
Küche (1:20:00)		
<p>Szenenbeschreibung: Eine Person ist am Kochen, während der Partner bzw. die Partnerin auf dem Smartphone spielt.</p> <p>Erläuterung: In dieser Szene wird in der Diversitätsversion die Dimension Geschlechter aufgegriffen. Durch die bewusste Rollenverteilung, der Mann übernimmt die Kochrolle, werden traditionelle Geschlechterstereotype aufgebrochen und alternative Rollenbilder sichtbar gemacht. In der Version ohne Diversitätsdarstellung wird dagegen ein konventionelles Hausfrauenbild aufgegriffen, um klassische Rollenklischees beizubehalten.</p>		

Tabelle 3: Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 2/3

Szene	Ohne-Diversität-Gruppe	Diversitäts-Gruppe
Straße II (1:31:08)		
	<p>Szenenbeschreibung: An einer belebten Straße beobachtet eine Person den Verkehr.</p> <p>Erläuterung: In dieser Szene wird in der Diversitätsversion die Dimension Physische Fähigkeit aufgegriffen. Die Szene soll den selbstverständlichen und gleichberechtigten Umgang von Menschen mit und ohne körperliche Einschränkungen im öffentlichen Raum verdeutlichen.</p>	
Wohnzimmer (2:13:24)		
	<p>Szenenbeschreibung: Eine gläubige Person, die erkältet ist, sitzt auf einer Couch im Wohnzimmer.</p> <p>Erläuterung: In dieser Szene wird in der Diversitätsversion die Dimension Religion aufgegriffen. Die Wahl einer islamischen Darstellung erfolgte bewusst vor dem Hintergrund des Untersuchungsortes: Die Datenerhebung fand an einer deutschen, christlich geprägten Schule in Thailand, einem überwiegend buddhistischen Land, statt. Durch die Wahl einer Religion, die im jeweiligen kulturellen Umfeld weniger sichtbar ist, sollte religiöse Vielfalt hervorgehoben und wertschätzend repräsentiert werden.</p>	

Tabelle 4: Übersicht der Filmszenen der beiden Videovarianten mit Szenenbeschreibung und Erläuterungen - Teil 3/3

Erwähnenswert ist, dass das Erklärvideo ohne Diversitätsdarstellung vollständig KI-generiert wurde. Im Gegensatz dazu enthält die Version mit Diversitätsdarstellung zwei von sechs Filmszenen (Küchen- und Wohnzimmerszene) mit realen Personen, während die übrigen vier Sequenzen mithilfe von KI generiert wurden.

In einigen weiteren, insbesondere animierten Szenen wirkten die Darstellungen zunächst relativ leer. Deshalb wurden KI-generierte Personen integriert, um ein realistischeres und lebendigeres Gesamtbild zu schaffen. Diese ergänzenden Szenen sind in der folgenden Tabelle 5 und 6 dargestellt.



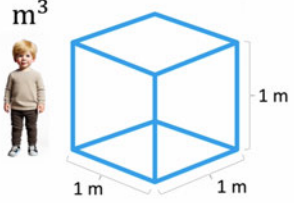
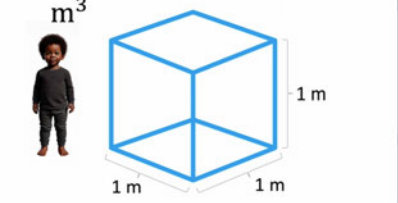


Szene	Ohne-Diversität-Gruppe	Diversitäts-Gruppe
Fragend (04:03:00)		
<p>Szenebeschreibung: Eine Person, die fragend blickt.</p> <p>Erläuterung: In dieser Szene wird die Diversitätsdimension ethnische Zugehörigkeit aufgegriffen. Die Szene verdeutlicht, dass Fragen und Neugier im Lernprozess universell sind und unabhängig von Herkunft oder Aussehen eine zentrale Rolle im Verstehen naturwissenschaftlicher Inhalte spielen.</p>		
Einheit (04:10:10)		
<p>Szenebeschreibung: Ein Kind, das neben einem Ein-Meter-Würfel steht.</p> <p>Erläuterung: Diese Szene greift die Diversitätsdimension ethnische Zugehörigkeit auf. Durch die Darstellung eines Kindes mit anderer Hautfarbe entsteht ein vielfältigeres Bild der Lernenden. Zugleich wird verdeutlicht, dass naturwissenschaftliche Themen wie Maßeinheiten und Volumen universelle Lerninhalte sind, die für alle Kinder gleichermaßen bedeutsam sind.</p>		
Münzen (04:31:07)		
<p>Szenebeschreibung: Eine Person, die das Gewicht einer Ein-Cent-Münze auf einer Waage präsentiert.</p> <p>Erläuterung: Hier wird die Diversitätsdimension religiöse bzw. kulturelle Zugehörigkeit thematisiert. Die Darstellung einer Frau mit Kopftuch macht ein alltägliches Bild religiöser Vielfalt sichtbar und betont Normalität und Akzeptanz unterschiedlicher kultureller Ausdrucksformen.</p>		

Tabelle 5: Szenen mit KI-generierten Bilder von Personen - Teil 1/2

Szene	Ohne-Diversität-Gruppe	Diversitäts-Gruppe
Körperliche Schäden (05:40:00)		
	<p>Szenebeschreibung: Eine hustende Person, während gleichzeitig Krankheitsbezeichnungen eingeblendet werden, die durch Feinstaubbelastung entstehen oder verstärkt werden können.</p> <p>Erläuterung: In dieser Szene wird die Diversitätsdimension ethnische Zugehörigkeit aufgegriffen. Durch die bewusste Variation in der Darstellung der Personen – etwa durch unterschiedliche Hautfarben oder kulturelle Hintergründe – wird verdeutlicht, dass Umweltbelastungen wie Feinstaub alle Menschen gleichermaßen betreffen. Auf diese Weise fördert die Szene eine inklusive Sichtweise auf Umwelt- und Gesundheitsfragen und unterstreicht die globale Bedeutung des Themas.</p>	
WHO (05:51:00)		
	<p>Szenenbeschreibung: Repräsentiert die Arbeitsgruppe der WHO.</p> <p>Erläuterung: Diese Szene thematisiert die Diversitätsdimensionen Geschlecht, ethnische Zugehörigkeit und Alter. Durch die Darstellung einer vielfältigen Personengruppe wird das internationale und inklusive Selbstverständnis der WHO verdeutlicht. Die Szene zeigt, dass wissenschaftliche und gesundheitspolitische Organisationen von globaler Zusammenarbeit und Vielfalt geprägt sind und vermittelt so ein Bild, das über traditionelle, einseitige Darstellungen hinausgeht.</p>	

Tabelle 6: Szenen mit KI-generierten Bilder von Personen- Teil 2/2

4.4.2 Interaktive digitale Quizaufgaben

Im Anschluss an das Erklärvideo bearbeiteten die Schüler:innen ein digitales Quiz, das insgesamt 13 Aufgaben umfasste. Die Aufgaben dienten dazu, das im Video vermittelte Wissen zu überprüfen und zugleich die Anwendung auf reale Datenkontexte zu fördern. Der Inhalt der Quizaufgaben war in beiden Gruppen identisch. Unterschiede bestanden ausschließlich in der bildlichen Darstellungen der Figuren mit oder ohne Diversität, wie sie bereits im Video vorka-

men.

Die ersten acht Aufgaben waren als Single-Choice- bzw. Multiple-Choice-Items gestaltet und bezogen sich direkt auf die Inhalte des Erklärvideos. Thematisch behandelten sie grundlegende Aspekte zu Feinstaub, wie die Definition und Einheit von Feinstaubpartikeln, deren Entstehungsquellen sowie Auswirkungen auf Gesundheit und Klima. Abbildung 5 zeigt exemplarisch die Darstellung einer dieser Quizaufgaben für beide Gruppen, links die ohne Diversitätsdarstellung und rechts die der Diversitäts-Gruppe.

Abbildung 5: Quizaufgabe - Einfluss auf die Gesundheit

Die darauffolgenden drei Aufgaben bezogen sich auf zwei Berliner Messstandorte: *Wedding* und *Grunewald*. Zu den Aufgaben mit dem jeweiligen Standort wurde eine kurze Beschreibung der Umgebung in Text- und Bildform sowie einen Link zu den entsprechenden Messdaten dargestellt (siehe Abbildung 6). Dadurch sollten die Schüler:innen einen authentischen Einblick in die unterschiedlichen Umweltbedingungen innerhalb einer Großstadt erhalten.

Abbildung 6: Beschreibung der Bezirke: Wedding (links) und Grunewald (rechts)

Die Aufgabe für den Messstandort *Wedding* ist als Lückentext gestaltet und zielte darauf ab, die Schüler:innen mit der Webseite Luftdaten.berlin vertraut zu machen. Für einen vorgegebenen Tag sollten sie die statistische Kennzahlen der beiden Feinstaubpartikel (PM10 und PM2,5) des Berliner Standorts *Wedding* ablesen und in die entsprechenden Felder eintragen (siehe Abbildung 7). Als Hintergrundbild wurde die im Video dargestellte Straßenkreuzungsszene gewählt. Zwar befindet sich die tatsächliche Messstation dieses Standortes nicht direkt an der Straße, sondern in einem Wohngebiet und zählt zu den sogenannten *Hintergrundmessstationen* (vgl. Deutsche Umwelthilfe, 2021). Diese gestalterische Entscheidung sollte den thematischen Zusammenhang zwischen städtischem Verkehr und Feinstaubbelastung visuell verdeutlichen und damit den Lernenden eine anschauliche Kontextualisierung der Messdaten ermöglichen. Zudem zählt

die Messstation Wedding laut Deutsche Umwelthilfe (2021) zu den Orten mit der höchsten Feinstaubbelastung (PM_{2,5}) in Deutschland und belegte im entsprechenden Ranking den achten Platz.

Abbildung 7: Quizaufgabe - Statische Kennzahlen des Messstation Wedding

In den Aufgaben zum Standort Grunewald sollten die Lernenden statische Kennzahlen der beiden Berliner Standorte (Wedding und Grunewald) miteinander vergleichen und mögliche Erklärungen für die unterschiedlichen Feinstaubemissionen in Multiple-Choice-Aufgaben auswählen. Als Hintergrundbild wurde die im Video gezeigte Parkszone verwendet (siehe Abbildung 8). Zwar stellt diese Szene nicht den tatsächlichen Ort Grunewald dar, die Wahl dieses Motivs erfolgte jedoch bewusst, um den Kontrast zwischen urban geprägten und naturhaften Gebieten visuell hervorzuheben.

Abbildung 8: Quizaufgabe: Wedding und Grunewald vergleichen

Die letzten beiden Quizaufgaben bezogen sich auf Messdaten der Schule. Anhand von Diagrammen sollten die Schüler:innen den dargestellten Peaks den Messstandort der Schule zuordnen und eine Vermutung zur Ursache der erhöhten Werte in einem Textfeld formulieren. Ziel dieser Aufgaben war es, lokale Messdaten zu interpretieren und den Schüler:innen durch die Einbindung der an ihrer Schule installierten Messstationen die Relevanz von Feinstaubbelastung im eigenen Alltag bewusst zu machen. Gleichzeitig sollten sie verschiedene Darstellungsformen von Messdaten kennenlernen, um ihren kompetenten Umgang mit Diagrammen, digitalen Medien und Umweltinformationen zu fördern. Auf die Einbettung von Figuren wurde bewusst verzichtet, um die Übersichtlichkeit der Aufgaben zu behalten.

4.5 Messinstrument

Der Fragebogen wurde digital am Ende der Lernumgebung eingebettet, sodass die Teilnehmenden ihn direkt nach Abschluss der Lernaktivität ausfüllen konnten. Die Bearbeitungszeit betrug etwa fünfzehn Minuten.

Zu Beginn des Fragebogens wurden die Schüler:innen nach ihrer Klassenstufe sowie nach ihrem allgemeinen Eindruck von der aktuellen Lerneinheit gefragt. Diese einleitenden Fragen dienten dazu, Hintergrundinformationen zu erheben und die späteren Antworten im Kontext der jeweiligen Lerngruppe und des individuellen Erlebens der Lerneinheit einordnen zu können.

Ziel der eingesetzten Items im Fragebogen war es, die motivationalen Aspekte innerhalb der digitalen Lernumgebung zu erfassen. Dafür wurde das Intrinsic Motivation Inventory (IMI) eingesetzt, ein bekanntes und theoretisch fundiertes Messinstrument zur Erfassung der intrinsischen Motivation. Ergänzend wurden zwei offene Fragen („*Was hat euch besonders gut gefallen?*“ und „*Was hat euch nicht gefallen?*“) zur Erfassung individueller Rückmeldungen integriert.

Im folgenden Abschnitt (4.5.1) werden die Auswahl der Subskalen des IMI für diese Untersuchung erläutert. Daran anschließend werden in Abschnitt 4.5.2 die für diese Untersuchung verwendeten Übersetzungen sowie die ausgewählten Items erläutert. Anschließend erfolgt in 4.5.3 eine Beschreibung der methodische Überprüfung der Gütekriterien der eingesetzten Items. Im letzten Abschnitt 4.5.4 werden die offenen Fragen und ihre Funktion beschrieben.

4.5.1 Auswahl der Subskala

Auf Grundlage der theoretischen Ausführungen in 2.1.4 wurde für die vorliegende Untersuchung das *Intrinsic Motivation Inventory* (IMI) als Messinstrument der intrinsischen Motivation gewählt.

Für die empirische Untersuchung der vorliegende Arbeit wurden drei Subskalen des IMI ausgewählt – *Interesse*, *Anstrengung* und *Nützlichkeit*. Diese Auswahl orientiert sich an den inhaltlichen Zielsetzungen der Untersuchung und am Kontext des Untersuchungssettings.

Die Formulierung der englischen Items der Subskala *relatedness* (*soziale Eingebundenheit*) bezieht sich auf die Interaktionen zwischen echten Personen, wie „*I'd like a chance to interact with this person more often.*“ (vgl. IMI-english, 2022). Aus der Schüler:innenperspektive beschreibt diese Subskala das Gefühl der Zusammenarbeit mit bestimmten Personen, entweder mit den Gruppenmitglieder:innen oder auch der Lehrkraft. Eine passende Adaption dieser Subskala auf das Setting dieser digitalen Lernumgebung liegt nicht vor. Da für diese Untersuchung die digitale Lernumgebung individuell und ohne sozialen Austausch bearbeitet wurde, wurde die Subskala *soziale Eingebundenheit* nicht untersucht. Da den Lernenden weder die Wahl der Lernumgebung noch der Bearbeitungsschritte offenstand, wurde die Subskala *wahrgenommene Wahlfreiheit* ebenfalls nicht einbezogen. Zudem wurde die Teilnahme als freiwillig und nicht bewertungsrelevant kommuniziert, sodass kein Leistungsdruck zu erwarten war. Vor diesem Hintergrund wurde auf die Erhebung der Subskala *Druck/Anspannung* verzichtet. Diese Reduktion

folgt dem Vorgehen vergleichbarer Studien von Wilde et al. (2009) und ermöglicht eine theoriegeleitete, zugleich aber ökonomische Erfassung der motivationalen Dimensionen. Wie bereits dargestellt, liegen bislang keine validierten deutschsprachigen Versionen der Subskalen *Anstrengung* und *Nützlichkeit* vor. Die englische Originalversion des IMI zeigte hingegen in mehreren empirischen Studien, unter anderem in der Untersuchung von Bosch (2024), eine hohe Validität und Reliabilität.

In einer deutschen empirischen Untersuchung von Schnirch (2006) zur Wirkung einer multimedial unterstützten Lernumgebung auf das Interesse und die Lernmotivation von Schüler:innen wurde eine übersetzte und für den schulischen Kontext angepasste Version des englischen Originalfragebogens des IMI verwendet (vgl. Schnirch, 2006). Auch in der Studie von Mézes (2016), die den Einfluss unterschiedlicher Formen der Anleitung beim Experimentieren im Physikunterricht auf die Motivation untersuchte, wurde die IMI-Originalversion ins Deutsche übertragen und kontextbezogen angepasst, indem jeweils das Wort „*activity*“ durch „*Experiment*“ ersetzt wurde (vgl. Mézes, 2016, S. 116). Beide Studien verwendeten zudem eine fünfstufige Likert-Skala anstelle der siebenteiligen Skala des englischen Originals, um die Verständlichkeit und Differenzierungsfähigkeit für den schulischen Einsatz zu erhöhen.

Vor diesem Hintergrund ist es erforderlich die vorgenommenen Übersetzungen, kontextuellen Anpassungen und die reduzierte Auswahl an Subskalen zu überprüfen, sodass die verwendeten Items die intendierten Konstrukte zuverlässig und valide abbilden. Daher werden zunächst die übersetzten Items der drei Subskalen dargestellt und anschließend das methodische Vorgehen zur Prüfung der Gütekriterien des eingesetzten IMI-Instruments beschrieben.

4.5.2 Eingesetzte Items

Für die Umsetzung in dieser Untersuchung wurde auf die Struktur und Formulierungen des englischen Originalinstruments zurückgegriffen. Um eine sprachlich und kontextuell angemessene Version für den deutschsprachigen Schulkontext zu erhalten, erfolgte eine Übersetzung der Items der drei ausgewählten Subskalen - *Interesse* (7 Items), *Anstrengung* (5 Items) und *Nützlichkeit* (7 Items) - in Zusammenarbeit mit einer Englischlehrkraft. Bei der Übersetzung der Items wurden auch bestehenden Übersetzungen aus der deutschsprachigen Fassung des IMI vom Homepage der Center for Self-Determination Theory (o.D.) sowie aus der Kurzskala intrinsischer Motivation (KIM) nach Wilde et al. (2009) integriert. Auf diese Weise entstand ein konsistentes Instrument, das sowohl theoretisch fundiert als auch sprachlich und inhaltlich an die Zielgruppe angepasst ist.

Interesse: Bei der Übersetzung der Items der Subskala *Interesse* zeigte sich, dass die Items Nr. 1 und 2 aus Tabelle 7 denselben inhaltlichen Aspekt der Freude an der Tätigkeit abdecken, da beide nahezu synonym formuliert sind und positive Emotionen gegenüber der Bearbeitung der Lerneinheit ausdrücken („Freude daran haben“ bzw. „Spaß gemacht haben“). Aus diesem Grund wurde auf eine Doppelung verzichtet und Item 2 beibehalten. Darüber hinaus erwies sich die deutsche Übersetzung von Item 7 als sprachlich schwierig, da sie für Schüler:innen

potenziell missverständlich gewesen wäre. Eine sinnngemäße Übersetzung hätte eine sprachlich stark vereinfachte oder kontextverändernde Formulierung erfordert. Um Missverständnisse oder kognitive Überforderung durch verschachtelte Satzstrukturen zu vermeiden, wurde das Item ausgeschlossen. Somit verbleiben für die Untersuchung fünf Items der Subskala *Interesse*, die in

Subskala: <i>Interesse</i>		
Nr.	Englisch	deutsche Übersetzung
1	I enjoyed doing this activity very much.	Ich hatte viel Freude daran, die Lerneinheit „Feinstaub“ zu bearbeiten.
2	This activity was fun to do.	Die Lerneinheit „Feinstaub“ hat mir Spaß gemacht.
3	I thought this was a boring activity. (R)	Ich finde die Lerneinheit „Feinstaub“ langweilig. (R)
4	This activity did not hold my attention at all. (R)	Die Lerneinheit „Feinstaub“ konnte meine Aufmerksamkeit überhaupt nicht aufrecht halten. (R)
5	I would describe this activity as very interesting.	Ich fand die Lerneinheit sehr interessant.
6	I thought this activity was quite enjoyable.	Die Lerneinheit war unterhaltsam.
7	While I was doing this activity, I was thinking about how much I enjoyed it.	Während ich die Lerneinheit „Feinstaub“ bearbeitete, dachte ich darüber nach, wie sehr sie mir gefiel.
Die Items mit der Kennzeichnung (R) weisen eine negativ Polung auf.		

Tabelle 7: Alle Items der Subskala *Interesse*

Anstrengung: Die fünf originalen englischen Items sowie die selbst übersetzten deutschen Items der Subskala *Anstrengung* sind in Tabelle 8 dargestellt. Bei der Übersetzung zeigte sich, dass die Items 1 und 3 inhaltlich und sprachlich nahezu identische sind. Daher wurde eines der beiden ausgeschlossen, um den Umfang des Fragebogens zu reduzieren. Die Entscheidung fiel zugunsten von Item 3, da dessen Formulierung im Deutschen natürlicher und altersangemessener wirkt als Item 1. Somit verbleiben für die Untersuchung vier Items der Subskala *Anstrengung*, die in Tabelle 7 fett dargestellt wurde.

Subskala: <i>Anstrengung</i>		
Nr.	Englisch	deutsche Übersetzung
1	I put a lot of effort into this.	Ich habe viel Anstrengung in die Lerneinheit „Feinstaub“ investiert.
2	I didn't try very hard to do well at this activity. (R)	Ich habe mich nicht sehr angestrengt, um bei der Lerneinheit gut abzuschneiden. (R)
3	I tried very hard on this activity.	Ich habe mich bei der Lerneinheit sehr angestrengt.
4	It was important to me to do well at this task.	Es war mir wichtig, bei den Quizaufgaben gut abzuschneiden.
5	I didn't put much energy into this. (R)	Ich habe nicht sehr viel Energie in die Durchführung der Lerneinheit „Feinstaub“ aufgewendet. (R)
Die Items mit der Kennzeichnung (R) weisen eine negativ Polung auf.		

Tabelle 8: Alle Items der Subskala *Anstrengung*

Nützlichkeit: Tabelle 9 präsentiert die originalen englischen Items und die deutsche Übersetzung der Subskala *Nützlichkeit*. Die Items 6 und 7 wurden nicht übernommen, da sie sprachlich sehr nahe bei Item 1 liegt. Die Entscheidung für Item 1 ist, dass die Wichtigkeit für einen selbst mehr betont als die andere Items. Somit verbleiben für die Untersuchung die ersten fünf Items der Subskala *Nützlichkeit*, die in Tabelle 9 dargestellt ist.

Subskala: Nützlichkeit		
Nr.	Englisch	deutsche Übersetzung
1	I believe this activity could be of some value to me.	Ich bin überzeugt, dass die Lerneinheit „Feinstaub“ sehr wertvoll für mich ist.
2	I think that doing this activity is useful for ...	Ich denke, dass die Lerneinheit „Feinstaub“ wichtig ist, weil sie alltagsrelevant ist.
3	I think this is important to do because it can ...	Ich denke, dass die Lerneinheit „Feinstaub“ mir helfen kann, verantwortungsbewusster mit meiner Gesundheit umzugehen.
4	I would be willing to do this again because it has some value to me.	Ich wäre bereit, die Lerneinheit noch einmal zu durchlaufen, weil sie für mich wertvoll ist.
5	I think doing this activity could help me to ...	Ich denke, dass die Lerneinheit „Feinstaub“ hilfreich ist, um ein besseres Bewusstsein für Umweltschutz zu entwickeln.
6	I believe doing this activity could be beneficial to me.	Ich bin überzeugt, dass die Lerneinheit „Feinstaub“ mir nützlich sein kann.
7	I think this is an important activity.	Ich denke, dass die Lerneinheit „Feinstaub“ wichtig ist.

Tabelle 9: Alle Items der Subskala *Nützlichkeit*

Insgesamt umfasste der Fragebogen 14 Items (5 Items zur Subskala *Interesse*, 4 zur Subskala *Anstrengung* und 5 zur Subskala *Nützlichkeit*). Die Items wurden in einer nicht nach Subskalen geordneten Reihenfolge präsentiert (siehe Abbildung 19 auf Seite 88). Für die Beantwortung wurde eine 6-stufige Likert-Skala verwendet (1 = stimmt gar nicht, 2 = stimmt wenig, 3 = stimmt ein bisschen, 4 = stimmt teils-teils, 5 = stimmt ziemlich, 6 = stimmt völlig). Im Gegensatz zur originalen 7-stufigen Skala des IMI und der 5-stufigen Skala des KIM wurde in dieser Untersuchung bewusst eine 6-stufige Variante verwendet, um die Tendenz zur neutralen Mittelposition zu vermeiden (vgl. Menold & Bogner, 2016). Eine vollständige Darstellung aller Items sowie der digitalen Umsetzung des Fragebogens ist im Anhang auf Seite 86 zu finden.

4.5.3 Gütekriterien der IMI-Items

Zur Einschätzung der Qualität des eingesetzten Messinstruments wird im Folgenden dargestellt, wie die Gütekriterien Objektivität, Reliabilität und Validität geprüft wurden. An dieser Stelle erfolgt ausschließlich die Beschreibung des methodischen Vorgehens; die Ergebnisse der Prüfungen werden erst im Ergebnisteil (Abschnitt 5.1) dargestellt.

Objektivität: Objektivität bezeichnet die Unabhängigkeit der Ergebnisse von der Person, die ein Messverfahren durchführt, auswertet oder interpretiert. Sie gliedert sich in Durchführungs-, Auswertungs- und Interpretationsobjektivität.

Die Durchführung lag zwar vollständig bei der forschenden Person, durch die standardisierte digitale Lernumgebung und den IMI-Fragebogen erhielten jedoch alle Lernenden identische Instruktionen, Abläufe und Bearbeitungszeiten, ohne dass inhaltliche Unterstützung gegeben wurde. Auch die Auswertung erfolgte nach einem klar definierten und dokumentierten Schema, das eine einheitliche Kodierung der Likert-Antworten, den Umgang mit fehlenden Werten sowie die Mittelwertbildung der Subskalen festlegt, sodass eine andere Person bei gleicher Vorgehensweise zu denselben Ergebnissen gelangen würde. Die Interpretation der Befunde orientierte sich schließlich theoriegeleitet an der Selbstbestimmungstheorie von Deci und Ryan (1993) sowie an festgelegten statistischen Kriterien wie Signifikanzniveaus und Effektstärken, wodurch auch die Interpretationsobjektivität gewährleistet ist.

Insgesamt weist diese Untersuchung durch die standardisierte digitale Durchführung als auch durch regelgeleitete Auswertung und transparente Interpretationskriterien eine hohe Objektivität auf.

Reabilität: Zur Bestimmung der internen Konsistenz der drei IMI-Subskalen (*Interesse, Anstrengung, Nützlichkeit*) wurden für die gesamte Stichprobe ($n=46$) mithilfe des Programms **R** zwei statistische Maße zur Zuverlässigkeit des Instruments berechnet: McDonald's ω und das ordinale Cronbach- α . Die Ergebnisse werden im Abschnitt 5.1 (S. 49) präsentiert.

Es wurde McDonald's ω gewählt, da dieser als theoretisch fundierter und robuster Indikator der internen Konsistenz gilt (vgl. Dunn et al., 2014). Das ordinale Cronbach- α wurde ergänzend ausgewiesen, um eine Vergleichbarkeit mit bestehenden Studien herzustellen und zugleich das ordinale Messniveau der Likert-Skalen angemessen zu berücksichtigen (vgl. Gadermann et al., 2012).

Validität: Zur Überprüfung der zugrunde liegenden Faktorstruktur wurde eine explorative Faktorenanalyse (EFA) mit der gesamten Stichprobe ($n=46$) durchgeführt, da die übersetzten Items bislang keiner empirischen Prüfung unterzogen wurden (vgl. Gäde et al., 2020). Eine konfirmatorische Faktorenanalyse (CFA) wäre aufgrund der geringen Stichprobengröße und der Sensitivität konfirmatorischer Modelle gegenüber kleinen Stichproben nicht angemessen (vgl. Gäde et al., 2020).

Da die theoretische Modellstruktur des IMI drei Subskalen vorsieht, wurde die Validität primär hinsichtlich dieser Drei-Faktoren-Struktur geprüft. Diese Vorgehensweise ist insbesondere bei kleineren Stichproben sinnvoll, da laut Winter et al. (2009) auch bei $N < 50$ verlässliche Ergebnisse erzielt werden können, sofern die Faktorstruktur einfach und die Faktorladungen ausreichend hoch sind ($\lambda > 0,8$) (vgl. Winter et al., 2009). Nach Winter et al. (2009) verbessert ein höheres Verhältnis von Items zu Faktoren(Subskalen) (p/f) die Qualität der Faktorenanalyse, während ein niedrigeres Verhältnis zu instabileren Ergebnissen führt (vgl. Winter et al., 2009).

In ihren Simulationen zeigte sich, dass bereits bei p/f-Verhältnissen von etwa sechs eine stabile Fakturstruktur auch bei kleinen Stichproben möglich war. Mit einem Verhältnis von $p/f \approx 4,7$ (14 Items, 3 Faktoren) liegt die vorliegende Untersuchung in einem vergleichbaren Bereich und die Faktorladungen haben überwiegend hohen Ladungen zwischen 0,7 und 0,9 und erfüllen damit die von Winter et al. (2009) beschriebenen günstigen Bedingungen für eine zuverlässige Faktoranalyse.

Zudem zeigte Winter et al. (2009) in Simulationsstudien, dass sich die Qualität einer Faktoranalyse verbessert, wenn pro Faktor mehrere Items vorliegen (ein höheres Verhältnis von Items zu Faktoren). Bereits bei einem Verhältnis von etwa sechs Items pro Faktor konnten zuverlässige Ergebnisse erreicht werden. Mit einem Verhältnis von rund 4,7 (14 Items auf 3 Faktoren) liegt die vorliegende Untersuchung in einem ähnlichen Bereich.

Neben der Ermittlung der Faktorladungen wurden zusätzlich Eigenwerte, Varianz und Varianzanteile innerhalb der Faktoren berechnet, um die Bedeutung der einzelnen Faktoren zu beurteilen und die Angemessenheit der faktoriellen Lösung empirisch abzusichern. Da die unrotierten Faktorladungen jedoch häufig Überschneidungen zwischen den Faktoren aufweisen und somit nur eingeschränkt interpretierbar sind, wurde ergänzend eine rotierte Faktorenstruktur herangezogen (vgl. Universität Zürich, o.D.). Durch die Rotation werden hohe Ladungen verstärkt und niedrige Ladungen abgeschwächt, sodass eine klarere Einfachstruktur entsteht und die inhaltliche Zuordnung der Items zu den Faktoren erleichtert wird.

4.5.4 Offene Fragen

Ergänzend zu den Item-Skala wurden zwei offene Fragen „*Was hat euch besonders gut gefallen?*“ und „*Was hat euch nicht gefallen?*“ am Ende des Fragebogens aufgenommen, um qualitative Rückmeldungen der Schüler:innen zu erfassen. Ziel war es, selbst formulierte Rückmeldungen der Schüler:innen zu erfassen, um individuelle Eindrücke zur digitalen Lernumgebung zu gewinnen und besser zu verstehen, welche Aspekte als besonders positiv hervorgehoben oder als weniger ansprechend wahrgenommen wurden. Die Anzahl der offenen Fragen wurde bewusst auf zwei reduziert. Der Schwerpunkt der Untersuchung lag auf der Erfassung der intrinsischen Motivation mithilfe der standardisierten Item-Skala des IMI; die offenen Fragen dienten lediglich der ergänzenden qualitativen Einordnung der Ergebnisse. Eine größere Anzahl offener Fragen hätte den Rahmen des Fragebogens erweitert und das Risiko einer kognitiven Überforderung der Schüler:innen erhöht. Durch die Beschränkung auf zwei kurze und leicht verständliche Fragen konnte gewährleistet werden, dass die Befragung für die Teilnehmenden überschaubar blieb.

4.6 Durchführung

In diesem Abschnitt wird zunächst der Ablauf der Durchführung chronologisch beschrieben (Abschnitt 4.6.1). Im anschließenden Abschnitt 4.6.2 werden die auftretenden technischen Probleme beschrieben, die während der Durchführung der Intervention auftraten.

4.6.1 Ablaufplan der Durchführung

Die Durchführungen sollte innerhalb einer Woche, an unterschiedlichen Tagen stattfinden. Sowohl Klassenlehrkräfte als auch Fachlehrkräfte, die in den jeweiligen Klassen unterrichten, wurden im Vorfeld gebeten, eine 45-minütige Unterrichtsstunde für die Untersuchung zur Verfügung zu stellen. Dadurch sollte gewährleistet werden, dass keine wesentlichen Unterrichtsinhalte versäumt werden.

Da die Untersuchung im Rahmen einer digitalen Lernumgebung stattfand, wurden die Schüler:innen bereits eine Woche vor dem Untersuchungstag mehrfach darauf hingewiesen, sowohl mündlich als auch schriftlich, ihr eigenes Endgerät und Kopfhörer mitzubringen. Für Schüler:innen, die kein eigenes Gerät zur Verfügung hatten, standen insgesamt fünf schulische Notebooks als Ersatz bereit. Da für die Durchführung der Lernumgebung ein Internetzugang erforderlich war, wurden im Vorfeld ausreichende Zugangscodes für das Schul-WLAN vorbereitet.

Zusätzlich wurde vor der Durchführung die Zustimmung der Schulleitung eingeholt. Die Schulleitung wurde darüber informiert, dass die Untersuchung anonym durchgeführt wurde; sämtliche personenbezogenen Daten wurden vertraulich behandelt und ausschließlich für den Zweck dieser Arbeit verwendet, im Einklang mit den Datenschutzrichtlinien der Schule und den ethischen Standards empirischer Bildungsforschung (vgl. Bildungsserver Berlin-Brandenburg, o.D.).

Die Durchführungen fanden an zwei aufeinanderfolgenden Tagen statt (16.-17. Januar 2025). Am ersten Tag nahmen die Klassenstufen 10, 11 und 12 in der jeweils ersten, zweiten und achten Unterrichtsstunde teil. Am zweiten Tag folgte die Durchführung mit der 9. Klasse in der siebten Unterrichtsstunde. Die Klassenstärken der einzelnen Jahrgangsstufen sind in Tabelle 1 dargestellt. Insgesamt umfassten die vier Klassen 52 Schüler:innen, von denen zwei Lernende der 12. Klasse am Durchführungstag abwesend waren. Damit nahmen insgesamt 50 Schüler:innen an der Untersuchung teil. Alle Durchführungen fanden in den Klassenräumen der jeweiligen Klassen statt. Jede Einheit umfasste eine Unterrichtsstunde von 45 Minuten.

Die Datenerhebung wurde nicht im Rahmen des regulären Physikunterrichts durchgeführt, sondern in verschiedenen Unterrichtsstunden anderer Fächer (u. a. Musik, Geografie und Englisch) umgesetzt.

Zu Beginn wurden alle Teilnehmenden über den Ablauf und die Dauer der Durchführung informiert. Es wurde erläutert, dass sie im Rahmen dieser Masterarbeit einen Abschnitt der neu entwickelten digitalen Lerneinheit zum Thema Feinstaub erproben und anschließend ein Feedback dazu geben können. Zudem wurden sie darauf hingewiesen, dass im Anschluss ihre Motivation im Zusammenhang mit der Nutzung der Lernumgebung erhoben wird. Die Teilnahme war freiwillig; eine Nichtteilnahme hatte keinerlei Konsequenzen.

Anschließend erfolgte eine etwa fünfminütige organisatorische Einführung. In diesem Zeitraum wurde der Ablauf erläutert, die Verbindung mit dem schulischen WLAN hergestellt und bei Bedarf Schülerinnen und Schülern ohne eigenes Endgerät ein Schulnotebook bereitgestellt. Danach riefen die Teilnehmenden über den bereitgestellten QR-Code beziehungsweise den angegebenen

Hyperlink die Webseite der digitalen Lernumgebung auf. Im Anschluss bearbeiteten sie die digitale Lerneinheit und füllten abschließend den am Ende eingebetteten digitalen Fragebogen aus.

4.6.2 Technische Probleme bei der Durchführung

Während der ersten beiden Durchführungen mit den Jahrgangsstufen 10 und 11 traten technische Schwierigkeiten bei der Darstellung von drei Quizaufgaben auf, insbesondere bei jenen, die der Wiederholung des Videoinhalts dienten. Zum Zeitpunkt der ersten beiden Durchführungen war die Ursache dieses Problems noch nicht bekannt, sodass die Aufgaben auch während der zweiten Durchführung nicht angezeigt werden konnten. Einige Schüler:innen bemerkten die fehlenden Aufgaben, meldeten dies der begleitenden Person und erwähnten es zusätzlich in den offenen Rückmeldungen. Für die Analyse stellt dieses Problem jedoch keine Beeinträchtigung dar, da ausschließlich Übungsaufgaben betroffen waren, die keine Darstellungen von Personen enthielten. Im Anschluss an die zweite Durchführung konnte die Ursache identifiziert und behoben werden. Weitere technische oder internetbezogene Probleme traten während der Durchführung nicht auf.

4.7 Datenaufbereitung und -auswertung

In diesem Abschnitt wird das Vorgehen bei der Datenaufbereitung und -auswertung erläutert. Dabei werden sowohl die quantitativen als auch die qualitativen Items hinsichtlich ihrer jeweiligen Verarbeitungsschritte und Analysemethoden dargestellt.

4.7.1 Quantitativ

Nach Abschluss der Datenerhebung wurden alle eingegangenen Fragebögen zunächst in Excel auf Vollständigkeit überprüft. Dabei zeigte sich, dass einige Teilnehmende den Fragebogen entweder nicht abgesendet oder unvollständig ausgefüllt hatten. Datensätze ohne jegliche Angaben wurden von der weiteren Analyse ausgeschlossen. Aus diesem Grund unterscheidet sich die Gesamtanzahl der Teilnehmenden von der tatsächlich ausgewertete Stichprobe. Die folgende Tabelle zeigt die Anzahl der ausgewertete Stichproben pro Klassenstufe sowie deren Zuordnung zur Diversitäts- bzw. Ohne-Diversitäts-Gruppe.

Klassenstufe	Ohne- Diversität-Gruppe	Diversitäts-Gruppe
9	6	6
10	5	7
11	5	6
12	6	5
Gesamt Anzahl	22	24

Tabelle 10: Übersicht der ausgewertete Stichprobe nach Klassenstufe und Gruppenzugehörigkeit
Einzelne fehlende Antworten auf einzelnen Item führten hingegen nicht zum Ausschluss der jeweiligen Person. Fehlte beispielsweise nur die Antwort auf ein einzelnes Items, wurde dieses Item

bei der Berechnung des Skalenmittelwerts ausgeschlossen; die übrigen Antworten der betreffenden Person gingen weiterhin in die Auswertung ein. Auf diese Weise konnten alle verfügbaren Daten in die Analyse einbezogen und Datenverluste minimiert werden.

Im nächsten Schritt wurde alle negativ gepolten Items rekodiert, sodass höhere Werte durchgängig eine höhere Zustimmung bzw. Ausprägung der jeweiligen Dimension widerspiegeln. Anschließend wurde für jede Versuchsperson die Mittelwertaggregation pro Subskala durchgeführt. Dabei wurden die Antworten aller Items einer Subskala aufsummiert und durch die Anzahl der beantworteten Items geteilt. Dieses Vorgehen ist insbesondere dann sinnvoll, wenn Subskalen unterschiedlich viele Items umfassen, da es die resultierenden Werte auf der ursprünglichen Likert-Skala (1–6) belässt und die Interpretierbarkeit erhöht.

Anschließend erfolgte eine Übertragung der Daten in das Statistikprogramm **R** zur weiteren Auswertung.

Zur statistischen Überprüfung der Forschungsfragen wurde ein vergleichendes Untersuchungsdesign gewählt, da zwei unabhängige Gruppen – die Experimentalgruppe (Diversitäts-Gruppe) und die Kontrollgruppe (ohne Diversitätsdarstellung) – miteinander verglichen wurden. Ziel war es, Unterschiede in den Mittelwerten der drei IMI-Subskalen (*Interesse*, *Nützlichkeit* und *Anstrengung*) zwischen beiden Gruppen zu identifizieren. Zur Überprüfung wurden t-Tests für unabhängige Stichproben durchgeführt.

Darüber hinaus wurde geprüft, inwiefern sich die Ausprägungen dieser drei Dimensionen zwischen den verschiedenen Jahrgangsstufen unterscheiden. Aufgrund der begrenzten Stichprobengröße in den einzelnen Klassenstufen wurde auf eine zweifaktorielle Varianzanalyse verzichtet. Stattdessen wurden separate t-Tests für unabhängige Stichproben innerhalb jeder Klassenstufe durchgeführt, um Unterschiede zwischen der Diversitäts- und der Kontrollgruppe auf den drei Subskalen zu prüfen.

Eine der Voraussetzungen für die Anwendung des t-Tests ist ein metrisches Skalenniveau der analysierten Variablen. Obwohl die Items des IMI formal lediglich ordinalskaliert sind, werden aggregierte Skalen-Mittelwerte in der empirischen Motivations- und Bildungsforschung üblicherweise als intervallskaliert interpretiert, sodass der Einsatz parametrischer Verfahren (z. B. t-Tests) als zulässig gilt.

Eine weitere Voraussetzungen betrifft die Varianzhomogenität zwischen den Vergleichsgruppen. Zur Überprüfung dieser Annahme wurde im Programm **R** ein Levene-Test durchgeführt. Dieser berechnet eine F-Statistik aus dem Verhältnis der Varianz zwischen den Gruppen und innerhalb der Gruppen und ermöglicht damit eine Aussage über die Homogenität der Varianzen. Die Ergebnisse zeigen, dass für keine der drei Subskalen signifikante Unterschiede zwischen den Gruppen vorliegen ($F(1,44) = 0,50 - 1,08$ und $p = 0,31 - 0,49$)⁹. Damit kann von homogenen Varianzen ausgegangen werden, womit die Voraussetzung für die Anwendung des t-Tests erfüllt ist.

Zur ergänzenden Bewertung der Ergebnisse wurden neben den Signifikanzniveau (p-Werten)

⁹Die Ergebnisse des Levene-Tests sind in Tabelle 21 auf Seite 91 zu dargestellt. Ein F-Wert nahe 1 weist auf ähnliche Varianzen zwischen den Gruppen hin, während höhere Werte auf größere Unterschiede hindeuten.

auch Effektstärken nach *Hedges' g* berechnet, um die praktische Relevanz der Unterschiede zwischen den Untersuchungsgruppen präziser einschätzen zu können. Im Gegensatz zu Cohen's *d*, das insbesondere für größere Stichproben entwickelt wurde, berücksichtigt Hedges' *g* eine Korrektur für kleine Stichprobenumfänge und reduziert somit die Überschätzung des Effekts, die bei Cohen's *d* auftreten kann. Daher gilt Hedges' *g* als die geeignetere Kennzahl bei Stichproben mit weniger als etwa 20 Teilnehmenden pro Gruppe (vgl. Lakens, 2013).

Darüber hinaus erfolgte ein deskriptiver Vergleich der Mittelwerte und Effektstärken zwischen den Jahrgangsgruppen, um mögliche alters- oder stufenspezifische Tendenzen aufzuzeigen, ohne daraus inferenzstatistische Schlüsse zu ziehen.

Um Aussagen zur motivationalen Wirkung innerhalb der einzelnen Jahrgangsstufen treffen zu können, wurden Mediane der beiden Gruppen pro Jahrgang miteinander verglichen. Aufgrund der sehr kleinen Stichprobengrößen pro Gruppe (zwischen fünf und sieben Personen) erwies sich der Median als robuster gegenüber Ausreißern, weshalb auf die Betrachtung von Mittelwerten verzichtet wurde.

4.7.2 Qualitativ

Neben den quantitativen Analysen wurden auch die beiden offenen Fragen („*Was hat euch besonders gut gefallen?*“ und „*Was hat euch nicht gefallen?*“) am Ende des Fragebogens qualitativ ausgewertet.

Die Auswertung erfolgte auf Grundlage der inhaltlich-strukturierenden qualitativen Inhaltsanalyse nach Kuckquartz (2018) und orientierte sich zugleich an der Verbindung qualitativer und quantitativer Inhaltsanalyse nach Scheijderberg et al. (2022). Alle Antworten der Schüler:innen wurden in einer Excel-Tabelle getrennt nach den beiden offenen Fragen gesammelt und anschließend sorgfältig gelesen. Die Auswertung der beiden offenen Fragen erfolgte getrennt voneinander, jedoch nach dem gleichen mehrstufigen Verfahren.

Es erfolgte eine induktive Kategorisierung, das heißt, die Kategorien wurden direkt aus dem Antwortmaterial entwickelt. Ähnliche Aussagen wurden zu inhaltlich übergeordneten Kategorien zusammengefasst und eindeutig benannt.

Jede Aussage wurde anschließend einer oder mehreren passenden Kategorien zugeordnet (Kodierung). Anschließend wurde die Häufigkeit der Nennungen pro Kategorie gezählt.

Zur besseren Vergleichbarkeit der beiden Untersuchungsgruppen wurden die entwickelten Kategorien exemplarischen Aussagen von Schüler:innen gegenübergestellt. Dadurch konnte eine direkte Gegenüberstellung und inhaltliche Vergleichbarkeit der Gruppenergebnisse gewährleistet werden.

Anschließend wurde der prozentuale Anteil berechnet, indem die Anzahl der Nennungen einer Kategorie durch die Gesamtzahl der abgegebenen Antworten auf die jeweilige Frage dividiert wurde. Die Häufigkeiten der einzelnen Kategorien wurden ergänzend quantifiziert, um die relative Bedeutung wiederkehrender Themen aufzuzeigen. Dabei gibt der berechnete Anteil an, wie viele

Teilnehmende einen bestimmten Aspekt im Verhältnis zur Gesamtzahl der Antworten auf die jeweilige Frage nannten.

5 Darstellung der Ergebnisse

In diesem Kapitel werden die Ergebnisse der Datenauswertung dargestellt. Zunächst werden die Reliabilität und Konstruktvalidität der eingesetzten IMI-Subskalen im Abschnitt 5.1 überprüft, um die Qualität des Messinstruments abzusichern. Anschließend werden in 5.2 die quantitativen Ergebnisse der Untersuchung präsentiert, bevor im letzten Abschnitt 5.3 die qualitativen Rückmeldungen der Lernenden dargestellt werden.

5.1 Reliabilität und Validität der IMI-Subskalen

Im Abschnitt 2.1.4 wurde bereits erläutert, dass für die deutsche Version des Intrinsic Motivation Inventory (IMI) bislang keine gesicherten Angaben zur Messgenauigkeit vorliegen. Auch die Kurzskala Intrinsischer Motivation (KIM) von Wilde et al. (2009) umfasst lediglich drei der insgesamt sieben Items der Subskala *Interesse*, während die Dimensionen *Anstrengung* und *Nützlichkeit* nicht berücksichtigt werden. Aus diesem Grund wurde in der vorliegenden Arbeit die Messqualität der insgesamt 14 eingesetzten Items (5 der Subskala *Interesse*, 4 der Subskala *Anstrengung* und 5 der Subskala *Nützlichkeit*) eigenständig überprüft, mit dem Ziel, die theoretisch angenommene Faktorstruktur des IMI empirisch zu prüfen und ihre Eignung für die nachfolgenden Gruppenvergleiche abzusichern.

Objektivität Die Objektivität der Untersuchung ist durch die standardisierte digitale Durchführung, den einheitlichen IMI-Fragebogen sowie ein klar definiertes und dokumentiertes Auswertungsverfahren gewährleistet, sodass andere Personen bei identischer Vorgehensweise zu denselben Ergebnissen gelangen würden. Auch die theoriegeleitete und an festen statistischen Kriterien orientierte Interpretation der Befunde stellt sicher, dass die Ergebnisse unabhängig von der auswertenden Person sind.

Reliabilität: Die Ergebnisse der beiden Maße zur Reliabilität (Cronbach's α und McDonald's ω) sind in Tabelle 11 dargestellt. Wie im Methodenteil Abschnitt 4.5.3 (S. 42) beschrieben, wurde die interne Konsistenz der drei Subskalen anhand von Cronbach's α und McDonald's ω mit der gesamten Stichprobe ($n=46$) bestimmt. McDonald's ω wurde aufgrund seiner höheren Robustheit gewählt, während das ordinale Cronbach- α ergänzend zur Wahrung der Vergleichbarkeit und zur Berücksichtigung des ordinalen Likert-Niveaus berechnet wurde.

Insgesamt weisen die Subskalen *Interesse* und *Nützlichkeit* hohe interne Konsistenzen auf (α , $\omega \approx 0,83$ – $0,85$).¹⁰ Dagegen weist die Subskala *Anstrengung* niedrige Reliabilität auf ($\alpha = 0,22$; $\omega = 0,61$).

¹⁰Ordinales Cronbach- α gilt ab 0,70 als akzeptabel und ab 0,80 als gut (vgl. Flandorfer, 2019). Für McDonald's ω wird ebenfalls ein Wert ab 0,70 als ausreichend angesehen; zudem reagiert ω weniger empfindlich auf ungleiche Ladungen (vgl. Stensen & Lydersen, 2022).

Subskala	Ordinales Cronbach- α	McDonald's ω
Interesse	0.83	0.84
Anstrengung	0.22	0.61
Nützlichkeit	0.84	0.85

Tabelle 11: Cronbach's α und McDonald's ω der drei IMI-Subskalen

Validität: Zur Überprüfung der zugrunde liegenden Struktur wurde eine explorative Faktorenanalyse (EFA) durchgeführt, die sich auf die erwartete dreifaktorielle Faktorstruktur konzentrier-

Zusätzlich wurde ein Scree-Test^a durchgeführt, um Hinweise zu erhalten wie viele Faktoren sinnvoll extrahiert werden sollten (vgl. Ledesma et al., 2015). Der Scree-Plot des Scree-Tests in Abbildung 9 zeigt einen möglichen Knick nach dem vierten Faktor. Daher wurden die Eigenwerte sowohl für die dreifaktorielle als auch für die vierfaktorielle Faktorstruktur berechnet. Die Eigenwerte geben an, wie viel Varianz jeder Faktor erklärt, und stellen damit ein zentrales Kriterium für die Bedeutung eines Faktors innerhalb der Faktoranalyse dar.

^aDer Scree-Test basiert auf der grafischen Darstellung der Eigenwerte (Scree-Plot) in absteigender Reihenfolge. Der Punkt, an dem der Kurvenverlauf deutlich abflacht („Knickpunkt“), liefert Hinweise darauf, wie viele Faktoren sinnvoll extrahiert werden sollten.

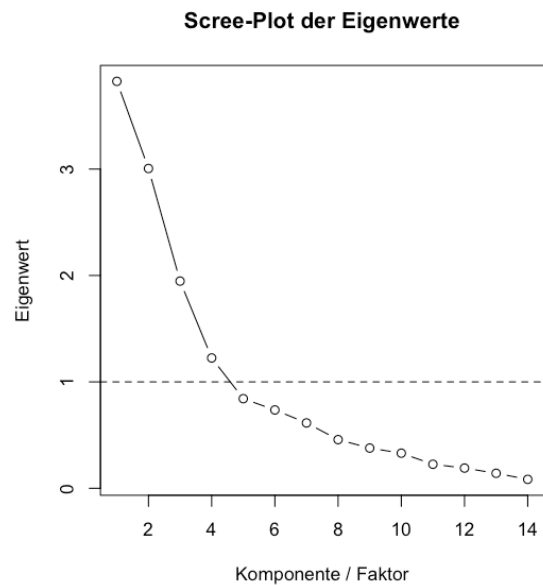


Abbildung 9: Scree-Plot der Eigenwerte-Faktoren-Diagramm

Die Eigenwerte der dreifaktoriellen Lösung liegen alle über 1 und sprechen damit nach dem Kaiser-Kriterium¹¹ für die Beibehaltung aller drei Faktoren (siehe Tabelle 17, S. 89).

Bei der vierfaktoriellen Lösung überschreiten hingegen nur drei der vier Faktoren den Schwellenwert (vgl. Tabelle 20, S. 90). Der vierte Faktor weist mit einem Eigenwert von 0,82 eine geringere Varianzaufklärung auf und erfüllt damit nicht die Anforderungen des Kaiser-Kriteriums. Daher wird im Folgenden ausschließlich die dreifaktorielle Lösung weiterverfolgt, da sie sowohl theoretisch als auch empirisch die überzeugendere und statistisch tragfähigere Struktur darstellt.

Für die dreifaktorielle Analyse wurden zunächst die Faktorladungen der einzelnen Items ermittelt. Diese sind in Abbildung 10 als Balkendiagramm dargestellt und basieren auf der dreifaktoriellen Lösung.¹²

¹¹Kaiser-Kriterium besagt, dass nur Faktoren extrahiert werden sollen, deren Eigenwert größer als 1,0 ist (vgl. Universität Zürich, o.D.).

¹²Diese Lösung ist auch tabellarisch im Anhang ist dargestellt (siehe 15 auf Seite 89).

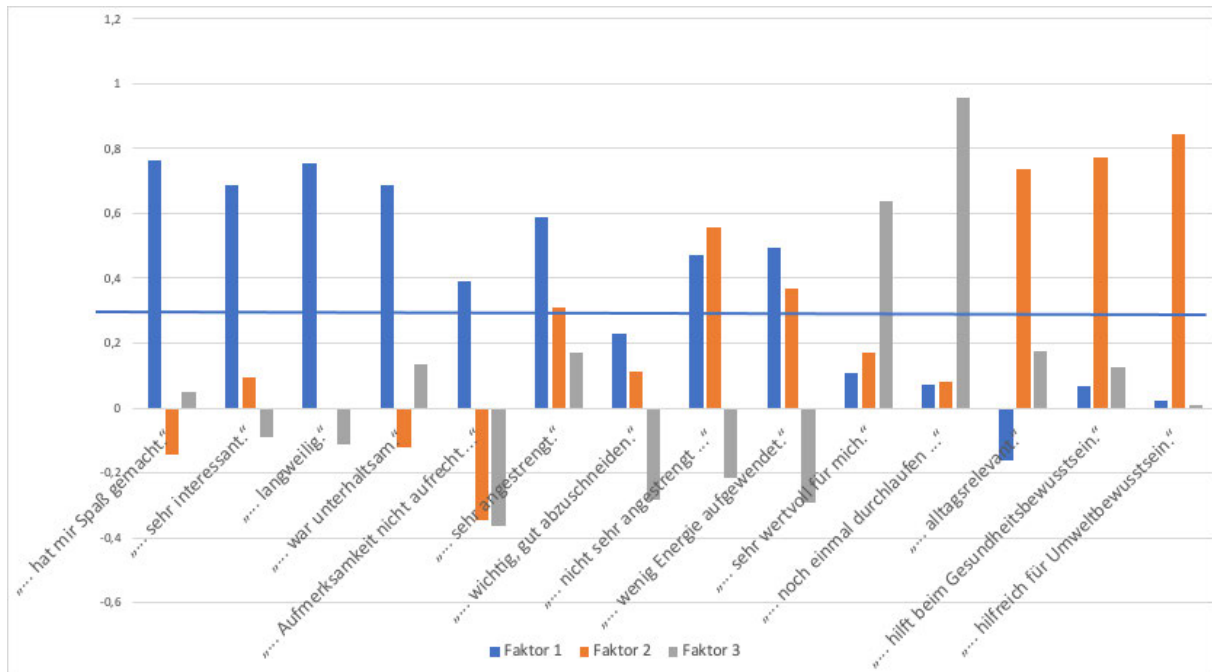


Abbildung 10: Faktorenloadungen der Items für eine dreifaktorielle Faktorstruktur

Die x-Achse zeigt die gekürzten Itembezeichnungen. Die ersten fünf Items (von links nach rechts) gehören theoretisch zur Subskala *Interesse*, die folgenden vier zu den Subskala *Anstrengung* und die letzten fünf zu den Subskala *Nützlichkeit*. Die drei farblich unterschiedlichen Balken pro Item repräsentieren die drei Faktoren, die in der explorativen Faktoranalyse identifiziert wurden. Die y-Achse stellt die Höhe der Faktorenladung zwischen -0,6 bis +1,0 dar. Eine positive Ladung zeigt an, dass ein Item inhaltlich mit dem jeweiligen Faktor übereinstimmt und positiv mit diesem korreliert, während eine negative Ladung darauf hinweist, dass das Item dem Faktor inhaltlich invers mit ihm zusammenhängt. Die horizontale Linie bei 0,3 dient als Orientierung für relevante Faktorladungen (vgl. Universität Zürich, o.D.). Nach Universität Zürich zufolge ist ein Faktor interpretierbar, wenn mindestens vier Variablen eine Faktorladung von $\geq 0,60$ aufweisen.¹³

Die ersten vier Items der Subskala *Interesse* laden deutlich und homogen auf Faktor 1 (Ladungen zwischen 0,688 und 0,764), während das Item „... Aufmerksamkeit nicht aufrecht ...“ (0,398) eine moderatere, aber interpretierbare Hauptladung aufweist. Somit sind die genannten Kriterien nach Universität Zürich erfüllt.

Die nächsten vier Items der Subskala *Anstrengung* zeigen dagegen ein heterogeneres Muster: Während einzelne Items moderat auf Faktor 1 (blaue Balken) oder Faktor 2 (orange Balken) laden, weist kein Item eine dominante Hauptladung auf dem dritten Faktor auf. Dieses gemischte Ladungsmuster deutet auf eine eingeschränkte, aber empirisch nachvollziehbare konsistente Faktorenstruktur dieser Subskala hin.

Die drei letzten Items, die zur *Nützlichkeit* im Sinne alltags-, gesundheits- oder umweltbezogener Relevanz gehören, weisen klare Hauptladungen auf Faktor 2 (orange Balken) auf (zwischen

¹³Zusätzlich gilt nach Universität Zürich Ladungen $\leq 0,20$ als irrelevant; Items ohne klare Hauptladung sollten entfernt werden. Ladungen ab $\pm 0,30$ bis $\pm 0,40$ sind minimal akzeptabel. Ein Faktor ist interpretierbar bei ≥ 4 Items $\geq \pm 0,60$ oder ≥ 10 Items $\geq \pm 0,40$. Bei $n < 300$ sollten Faktoren mit überwiegend niedrigen Ladungen nicht interpretiert werden.

0,735 und 0,843). Die beiden Items zur persönlichen Wertigkeit („wertvoll“, „noch einmal durchführen“) laden hingegen auf Faktor 3 (graue Balken) (0,637 bzw. 0,957) und bilden damit eine thematische Unterfacette innerhalb der Subskala *Nützlichkeit*.

Um die Faktorstruktur interpretierbarer zu machen, wurde anschließend eine Rotation durchgeführt. Diese erleichtert die Zuordnung der Items zu den Faktoren, indem sie hohe Ladungen hervorhebt und Kreuzladungen reduziert. Die resultierende dreifaktorielle Lösung erklärt insgesamt 54 % der Gesamtvarianz. Faktor 1 (24 %, $\lambda = 3,35$) erweist sich als varianzstärkster Faktor, gefolgt von Faktor 2 (19 %, $\lambda = 2,61$) und Faktor 3 (11 %, $\lambda = 1,60$).¹⁴

Für die vorliegende Arbeit werden Gruppenvergleiche auf Basis der theoretischen Subskalenstruktur des IMI durchgeführt (*Interesse*: 5 Items; *Anstrengung*: 4 Items; *Nützlichkeit*: 5 Items). Obwohl die explorative Faktoranalyse vereinzelt abweichende Ladungsmuster zeigte, wurden keine empiriebasierten Umgruppierungen der Items vorgenommen. Ziel der EFA war die prüfende Betrachtung der Faktorstruktur und nicht die Neukonzeption des Instruments. Die Faktorenbildung erfolgte daher theoriegeleitet, um die inhaltliche Konsistenz und Vergleichbarkeit mit bisherigen Studien sicherzustellen.

5.2 Quantitative Ergebnisse

In die quantitative Analyse gehen die Daten von insgesamt $n = 46$ Schüler:innen aus den Jahrgangsstufen 9 bis 12 ein. Die drei IMI-Subskalen setzen sich aus jeweils fünf Items für *Interesse*, vier Items für *Anstrengung* und fünf Items für *Nützlichkeit* zusammen, das entspricht insgesamt 14 Items. Für jede Versuchsperson wurden die Antworten pro Subskala mittels Mittelwertaggregation zusammengefasst, sodass ein jeweils durchschnittlicher Subskalenwert pro Person vorlag.

Im Folgenden Abschnitt 5.2.1 werden zunächst die Ergebnisse der beiden Untersuchungsgruppen über alle Jahrgangsstufen hinweg dargestellt. Anschließend werden in 5.2.2 für jede Subskala die Ausprägungen aller Jahrgangsstufen getrennt nach den beiden Untersuchungsgruppen dargestellt, sodass sowohl innerhalb der einzelnen Jahrgangsstufen als auch über alle Jahrgänge hinweg ein direkter Gruppenvergleich möglich ist. Abschließend werden in 5.2.3 die Mediane der beiden Gruppen nach Jahrgangsstufe präsentiert.

In allen Abschnitten werden hierfür Boxplots eingesetzt, die direkte visuelle Vergleiche sowohl zwischen den Untersuchungsgruppen als auch zwischen den Klassenstufen ermöglichen. Alle Boxplots zeigen die zentralen Tendenzen und die Verteilung der Ausprägungen auf der sechsstufigen Likert-Skala (1 = „stimmt gar nicht“ bis 6 = „stimmt völlig“) und beinhalten zusätzlich die jeweiligen Mittelwerte sowie das Signifikanzniveau ($*=p < 0,05$), die durch einen t-Tests ermittelt wurde.

¹⁴Die Rotationsmatrix der Varimax-Rotation ist in Tabelle 16 (S. 89) dargestellt.

5.2.1 Über alle Jahrgangsstufen hinweg:

Die Betrachtung über alle Jahrgangsstufen hinweg ist wichtig, da sie die statistische Aussagekraft erhöht und jahrgangsübergreifende Muster sichtbar macht. So lässt sich überprüfen, ob Unterschiede zwischen Diversitäts- und Kontrollgruppe konsistent über verschiedene Altersgruppen hinweg auftreten.

Von der Gesamtstichprobe entfielen 24 Teilnehmende auf die Diversitätsgruppe, während die Vergleichsgruppe aus 22 Schüler:innen bestand. Die Ausprägungen der drei IMI-Subskalen *Interesse*, *Anstrengung* und *Nützlichkeit* zwischen den beiden Untersuchungsgruppen über alle Jahrgangsstufen hinweg ist in Abbildung 11 als Boxplot dargestellt. Ergänzend zeigen die Rauten jeweils die Mittelwerte der Gruppen. Die oberhalb der Gruppenpaare angegebenen Signifikanzniveaus (ns, **) zeigen an, ob sich die Mittelwerte der beiden Gruppen statistisch signifikant unterscheiden.

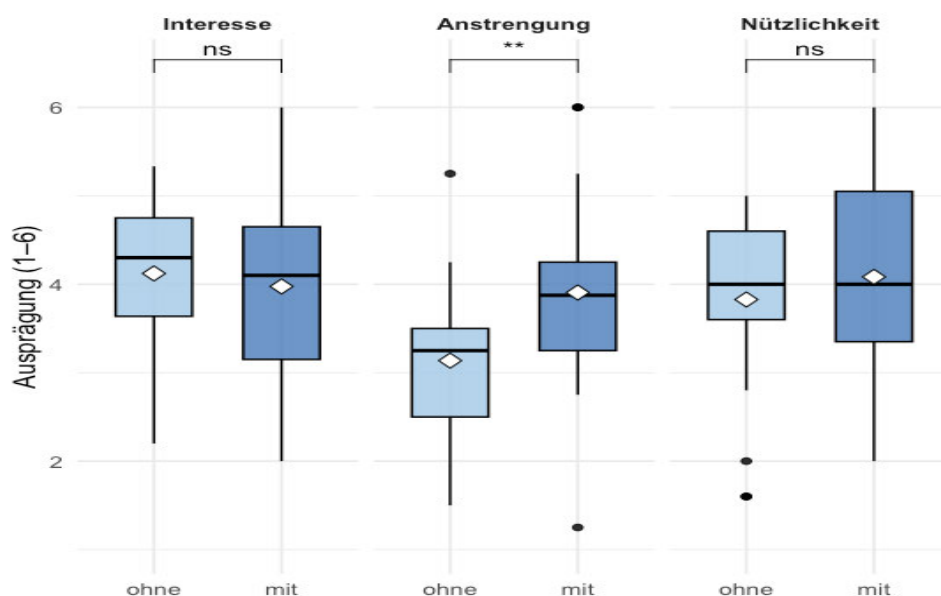


Abbildung 11: Vergleich der Ausprägungen der drei IMI-Subskalen zwischen Diversitäts- und Kontrollgruppe (inkl. Signifikanzniveaus)

Die Medianwerte der beiden Untersuchungsgruppen unterscheiden sich bei den Subskalen *Interesse* ($\tilde{x}_{mit} = 4,10$; $\tilde{x}_{ohne} = 3,90$) und *Nützlichkeit* ($\tilde{x}_{mit} = 4,00$; $\tilde{x}_{ohne} = 3,90$) kaum. Für die Subskala *Anstrengung* zeigt sich hingegen ein deutlicherer Abstand: In der Diversitätsgruppe lag der Median bei $\tilde{x}_{mit} = 3,90$, während er in der Gruppe ohne Diversitätsdarstellung bei $\tilde{x}_{ohne} = 3,25$ lag. Über alle Klassen hinweg liegen die Mittelwerte für die Subskala *Interesse* der beiden Gruppen eng beieinander und erzielen jeweils mittlere bis höhere Ausprägungen ($M_{ohne} = 4,09$; $M_{mit} = 3,98$). Sowohl die t-Tests ($p = 0,0705$) als auch die Effektstärken (Hedges' g : $g = -0,11$)¹⁵ weisen darauf hin, dass weder statistisch signifikante noch praktisch bedeutsame Unterschiede vorliegen. Bei der Subskala *Nützlichkeit* lagen die Mittelwerte der beiden Gruppen nah beieinander und zeigten mittlere bis höhere Ausprägungen. Auch hier zeigen sich weder statistisch signifikante ($p = 0,429$) noch praktisch bedeutsame Unterschiede

¹⁵Konvention der Interpretation von g : $|g| < 0,2$ - kein Effekt; $0,2 \leq |g| \leq 0,5$ - kleinen Effekt; $0,5 \leq |g| \leq 0,8$ - moderater Effekt; $|g| \geq 0,8$ - großer Effekt (vgl. Lenhard & Lenhard, 2022).

($g = 0,23$). Für die Subskala *Anstrengung* zeigt sich ein deutlich höherer Mittelwert in der Diversitäts-Gruppe. Der t-Test ergab einen signifikanten Unterschied ($p = 0,009$), begleitet von einer mittleren Effektstärke ($g = 0,79$), was auf einen moderaten Unterschied zwischen den Gruppen hinweist.

Insgesamt zeigen beide Gruppen in allen Subskalen positive Ausprägungen. Das bedeutet, dass die Lernenden die Lernumgebung in beiden Gruppen überwiegend zustimmend bewerteten, da ihre Mittelwerte auf der sechsstufigen Skala jeweils über dem Skalenmittelpunkt lagen. Besonders bei der Subskala *Anstrengung* zeigt sich in der Gruppe mit Diversitätsdarstellung ein deutlich höherer Median und auch die Mittelwerte ($M_{mit} = 3,91$, $M_{ohne} = 3,14$). Bei der Subskala *Nützlichkeit* fällt der Unterschied hingegen geringer aus, was sich in den nah beieinanderliegenden Mittelwerten widerspiegelt ($M_{mit} = 4,08$, $M_{ohne} = 3,83$). Für *Interesse* liegen die Mittelwerte der beiden Gruppen hingegen näher beieinander, mit nur leichten Vorteilen für die Kontrollgruppe ($M_{mit} = 3,98$, $M_{ohne} = 4,09$).

5.2.2 Gruppenvergleich über alle Klassenstufen je Subskala:

Diese jahrgangsbezogenen Vergleiche sind relevant, weil sie Hinweise darauf geben, ob der Einfluss der Diversitätsdarstellung möglicherweise alters- oder jahrgangsstufenabhängig ausfällt. Durch die getrennte Betrachtung wird sichtbar, ob bestimmte Altersgruppen stärker oder schwächer auf die Gestaltung der digitalen Lernumgebung reagieren. Darüber hinaus ermöglichen die jahrgangsbezogenen Mittelwerte eine differenziertere Interpretation der Gesamtergebnisse: Muster, die in der Gesamtstichprobe nicht deutlich hervortreten, können in einzelnen Jahrgängen klarer erkennbar werden und so wichtige Hinweise für zielgruppenspezifische Gestaltungsansätze digitaler Lernmaterialien liefern.

In der Tabelle 12 zeigt die Gruppeneinteilung der Stichprobe nach Klassenstufe.

Gruppe	9. Klasse	10. Klasse	11. Klasse	12. Klasse
N_{ohne}	6	5	5	5
N_{mit}	6	7	6	5

Tabelle 12: Stichproben nach Gruppen und Klassenstufe

Im Folgenden werden die drei Subskalen getrennt voneinander ausgewertet. Für jede Subskala wird ein eigenes Diagramm in Form von Boxplots für die Jahrgangsstufen 9 bis 12 getrennt nach Diversitäts- und Kontrollgruppe dargestellt, das die jeweiligen Verteilungen einschließlich Median, Quartilen, Mittelwert (Raute) und Signifikanzniveau zeigt. Auf diese Weise lässt sich für jede Dimension nachvollziehen, wie sich die Ausprägungen der Diversitäts- und der Kontrollgruppe innerhalb der einzelnen Jahrgangsstufen sowie über alle Klassen hinweg unterscheiden. Zunächst werden die Ergebnisse der Subskala *Interesse* dargestellt, gefolgt von *Nützlichkeit* und abschließend *Anstrengung*.

Interesse: Die Abbildung 12 stellt die Ausprägungen der Subskala *Interesse* dar. Insgesamt zeigen die Mittelwerte beider Gruppen in allen Jahrgangsstufe positive Ausprägungen. In den Jahrgangsstufen 9 und 10 lagen die Mittelwerte ($M_{9_{mit}} = 3,47$, $M_{9_{ohne}} = 3,87$; $M_{10_{mit}} = 4,34$,

$M_{10_{ohne}} = 4,35$) und der Median beider Gruppen ($\tilde{x}_{9_{mit}} = 3,30$, $\tilde{x}_{9_{ohne}} = 3,70$; $\tilde{x}_{10_{mit}} = 4,60$, $\tilde{x}_{10_{ohne}} = 4,40$) nahezu gleichauf. In der Jahrgangsstufe 11 erreichte die Diversitätsgruppe höhere Mittelwerte ($M_{mit} = 4,37$, $M_{ohne} = 3,64$), während sich in Klasse 12 das umgekehrte Muster zeigte ($M_{mit} = 3,60$, $M_{ohne} = 4,47$) (siehe Abbildung 12, S. 55). In der Gruppe mit diversitätssensibler Darstellung lag der Median der Jahrgangsstufe 12 im Skalenmittelpunkt ($\tilde{x} = 3,0$), während er in der Kontrollgruppe deutlich höher ausfiel ($\tilde{x} = 4,5$).

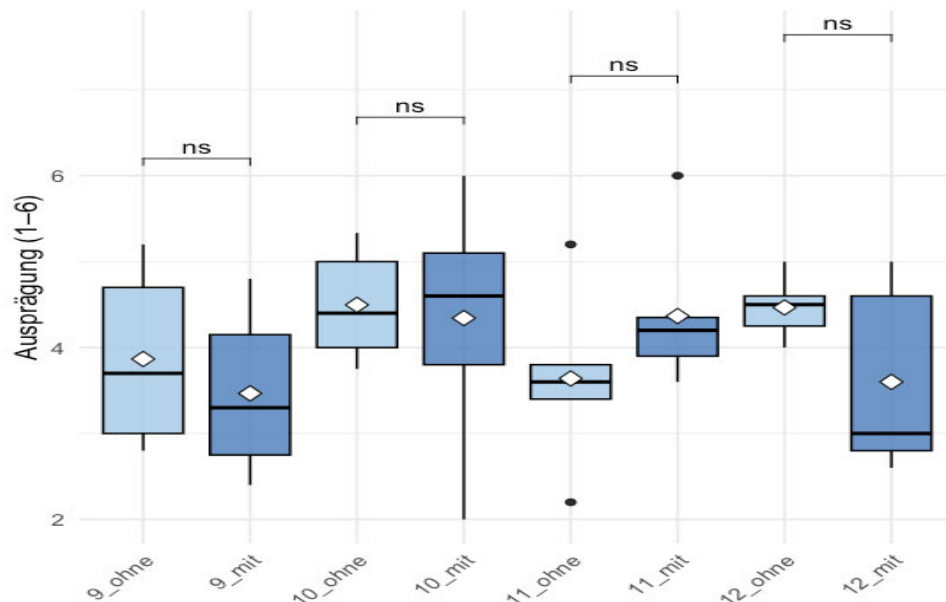


Abbildung 12: Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala *Interesse*

Die t-Tests zeigen durchgängig keine signifikanten Unterschiede ($p_9 = 0,51$; $p_{10} = 0,996$; $p_{11} = 0,241$; $p_{12} = 0,102$). Die Effektstärken in den Klassenstufen 9 und 10 fielen klein bis gering aus ($g_9 = -0,37$; $g_{10} = 0,00301$) und weisen daher auf kaum relevante Unterschiede hin. In Klasse 11 zeigt sich eine mittlere Stärke ($g_{11} = -0,695$), was auf einen moderaten Unterschied hindeutet, während in Klasse 12 ein großer Effekt ($g_{12} = 1,01$) beobachtet wurde.

Nützlichkeit:

Bei der Subskala *Nützlichkeit* lagen die Mittelwerte beider Gruppen der Jahrgangsstufen 9, 10 und 12 nahezu gleichauf (siehe Abbildung 13). Auch bei dieser Subskala zeigt die Jahrgangsstufe 11 bei der Diversitätsgruppe eine deutlich höhere Mittelwert ($M_{mit} = 4,27$, $M_{ohne} = 3,48$) und leicht höheren Median ($\tilde{x}_{mit} = 3,40$; $\tilde{x}_{ohne} = 3,70$). In der Jahrgangsstufe 10 weist die Gruppe ohne Diversitätsdarstellung einen deutlich höheren Median auf, während in der Jahrgangsstufe 12 ebenfalls ein höherer Median zu beobachten ist, jedoch mit einem geringeren Abstand zwischen den Gruppen.

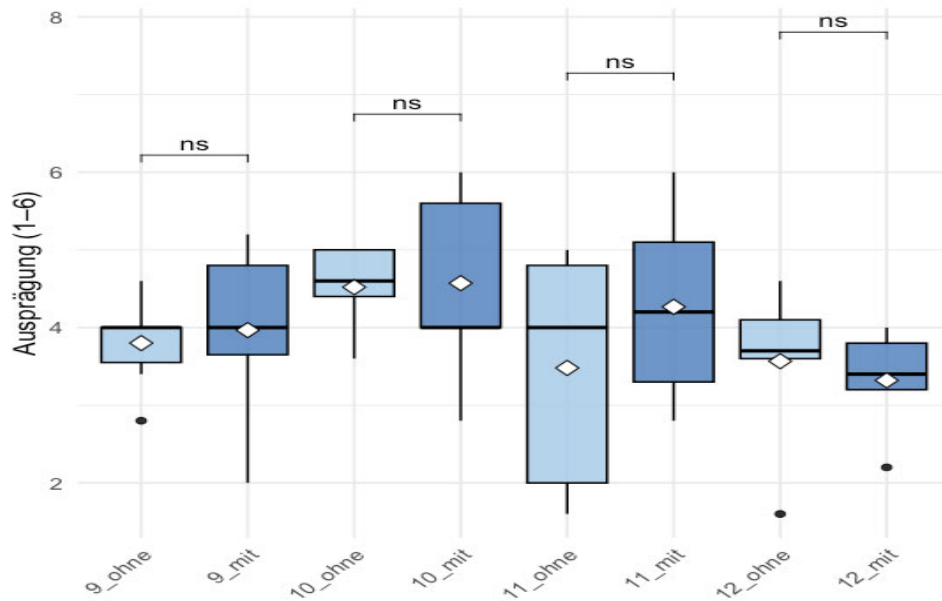


Abbildung 13: Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala *Nützlichkeit*

Die t-Tests zeigen in allen Jahrgangsstufen durchgängig keine signifikanten Unterschiede ($p_9 = 0,762$; $p_{10} = 0,931$; $p_{11} = 0,382$; $p_{12} = 0,663$). Die Effektstärken fielen in allen Klassenstufen mit Ausnahme der 11. Jahrgangsstufe gering aus ($g_9 = 0,17$; $g_{10} = -0,048$; $g_{12} = 0,249$). Für die 11. Klasse zeigte sich hingegen eine mittlere Effektstärke ($g = -0,509$).

Anstrengung:

In der Abbildung 14 zu Subskala *Anstrengung* ist zu erkennen, dass Gruppen mit diversitätssensiblen Darstellungen in allen Jahrgangsstufen außer der 12. einen höheren Mittelwert und Median aufweisen. In der 12. Klasse liegen diese Kennwerte nahe beieinander und der t-Test zeigt keinen signifikanten Unterschied zwischen den Gruppen ($p = 0,820$). Allerdings weist die Effektstärke mit $g = 0,130$ lediglich auf einen vernachlässigbaren Effekt hin.

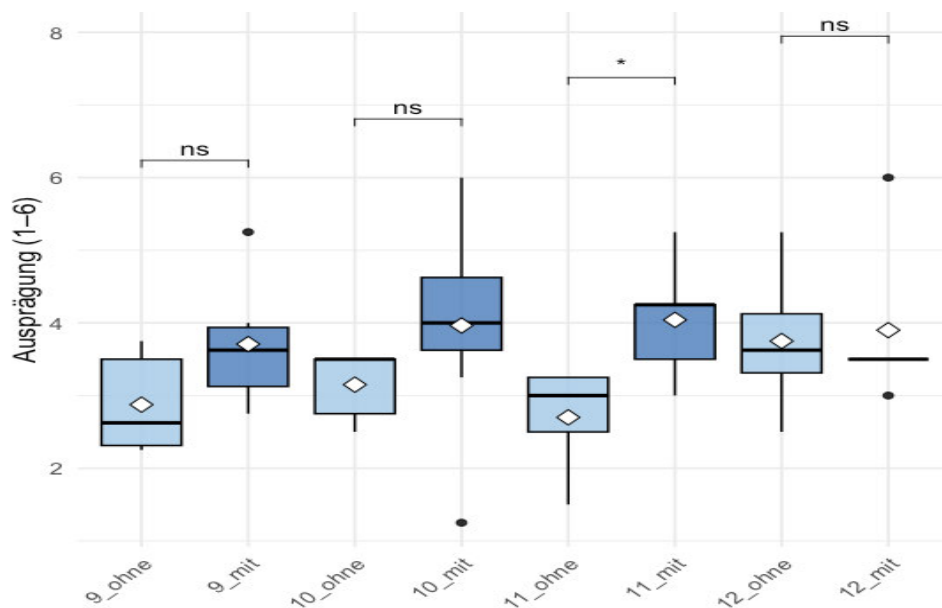


Abbildung 14: Vergleich der Gruppen der jeweiligen Jahrgangsstufe für den Subskala *Anstrengung*.

In den 9. und 10. Jahrgangsstufen erreichte die Diversitätsgruppe höhere Mittelwerte ($M_{9_{mit}} =$

3,71, $M_{9_{ohne}} = 2,88$; $M_{10_{mit}} = 3,96$, $M_{10_{ohne}} = 3,15$), allerdings ohne statistische Signifikanz ($p_9 = 0,101$; $p_{10} = 0,269$). Die Effektstärke deutet in Klasse 9 auf einen großen Effekt hin ($g = 0,96$), während in Klasse 10 ein mittlerer Effekt erkennbar ist ($g = -0,632$). In Klasse 11 zeigt sich ein statistisch signifikanter Unterschied zugunsten der Diversitätsgruppe ($p = 0,019$) mit einem sehr großen Effekt ($g = -1,57$).

5.2.3 Median der Gruppenunterschiede nach Jahrgangsstufe

Um Aussagen zur motivationalen Wirkung innerhalb der einzelnen Jahrgangsstufen treffen zu können, wurden die Mediane der beiden Gruppen pro Jahrgang miteinander verglichen. Aufgrund der sehr kleinen Stichprobengrößen pro Gruppe (zwischen fünf und sieben Personen) erwies sich der Median als robuster gegenüber Ausreißern, weshalb auf die Betrachtung von Mittelwerten verzichtet wurde. Die jeweils höhere Median-Ausprägung sowie die zugehörigen Werte für jede Jahrgangsstufe sind in Tabelle 13 dargestellt.

Jahrgangsstufe	Interesse		Nützlichkeit		Anstrengung	
	Gruppe	Median	Gruppe	Median	Gruppe	Median
9	ohne	3,70	gleich	4,00	mit	3,63
10	mit	4,60	ohne	4,60	mit	4,00
11	mit	4,20	mit	4,20	mit	4,25
12	ohne	4,50	ohne	3,70	ohne	3,63

Tabelle 13: Gruppe mit dem höheren Median in der jeweiligen Jahrgangsstufe, sowie die Medianwerte

In Jahrgang 11 weist die Diversitätsgruppe in allen drei Subskalen die höheren Medianwerte auf, während sich in Jahrgang 12 das entgegengesetzte Muster zeigt. In Jahrgang 10 liegen die Medianwerte der Diversitätsgruppe bei den Subskalen *Interesse* und *Anstrengung* höher, wohingegen mehr als 50 % der Lernenden der Ohne-Diversität-Gruppe höhere Werte in der Subskala *Nützlichkeit* abgeben. In Jahrgang 9 liegt der Median der Subskala *Interesse* der Ohne-Diversität-Gruppe höher, während die Diversitätsgruppe bei *Anstrengung* den höheren Median aufweist. In der Subskala *Nützlichkeit* entsprechen sich die Medianwerte beider Gruppen, da sie denselben Wert aufweisen.

5.3 Qualitative Ergebnisse

Neben den quantitativen Ergebnissen wurden auch die qualitativen Rückmeldungen der Lernenden ausgewertet, um ein umfassenderes Bild ihrer Erfahrungen mit der digitalen Lernumgebung zu gewinnen. Die offenen Fragen ermöglichen es, Aspekte zu erfassen, die durch die standardisierten IMI-Items nicht oder nur teilweise abgebildet werden können. Damit dient die qualitative Analyse als Ergänzung der quantitativen Befunde.

Die qualitativen Antworten der Schüler:innen auf die offenen Fragen („Was hat euch besonders gut gefallen?“ und „Was hat euch nicht gefallen?“) wurden inhaltlich analysiert und thematisch kategorisiert. Im Folgenden werden die Ergebnisse der beiden Gruppen dargestellt. Für

beide Fragen wird die Bildung der Kategorien erläutert und eine Auswertung vorgenommen. Abbildungen 15 und 16 zeigen die prozentuale Verteilung der Kategorien für beide Fragen.

„Was hat euch besonders gut gefallen?“

Die Antworten der Schüler:innen auf diese offene Frage wurden thematisch ausgewertet und zu inhaltlich ähnlichen Aussagen zusammengefasst. Insgesamt gaben 16 von 24 Schüler:innen aus der Diversitätsgruppe und 15 von 22 Schüler:innen aus der Gruppe ohne Diversitätsdarstellung zu dieser Frage eine Antwort.

Bei der Kategorisierung der Antworten, konnte eine Antwort mehr als einer Kategorie zugeordnet werden, wenn sie inhaltlich mehrere Aspekte enthielt (z. B. „*Es hat mir Spaß bereitet und ich habe mehr gelernt.*“ \Rightarrow Kategorien *Interesse* sowie *Lernzuwachs*). Die Kategorien sind somit nicht disjunkt, sondern teilweise überlappend. Es wurden keine Punktwerte vergeben, stattdessen wurde die Häufigkeit der Nennungen innerhalb jeder Kategorie erfasst.¹⁶ Insgesamt ergaben sich sechs Hauptkategorien:

- **Interesse:** In dieser Kategorie wurden Antworten erfasst, die auf ein gesteigertes Interesse am Thema oder an der Bearbeitung der Aufgaben hinweisen. Mehrere Schüler:innen äußerten, dass sie das Thema spannend oder die Bearbeitung der Quizaufgaben besonders motivierend fanden ($n_{ohne} = 0$; $n_{mit} = 2$).
- **Allgemeine Bewertung:** Hierunter wurden allgemeine, weniger inhaltlich spezifizierte Urteile zur Lernumgebung gefasst. Diese reichten von kurzen positiven Bewertungen („cool“, „nett“) bis zu neutralen Einschätzungen oder Hinweisen auf organisatorische Aspekte („kein Unterricht“) ($n_{ohne} = 1$; $n_{mit} = 2$).
- **Nützlichkeit:** Hierzu zählen Aussagen, die die Relevanz des Themas für den Alltag oder die Gesellschaft betonen. Die Lernenden beschrieben das Thema als wichtig und sahen in der Einheit einen praktischen Nutzen, etwa durch ein besseres Bewusstsein für Umwelt- und Gesundheitsaspekte ($n_{ohne} = 3$; $n_{mit} = 1$).
- **Lernzuwachs:** Diese Kategorie umfasst Äußerungen, in denen die Schüler:innen angaben, Neues gelernt oder ihr Wissen erweitert zu haben. Viele beschrieben, dass sie in kurzer Zeit viel gelernt und ihr Verständnis für das Thema Feinstaub vertieft hätten ($n_{ohne} = 4$; $n_{mit} = 4$).
- **Interaktive Elemente:** Unter dieser Kategorie wurden Rückmeldungen zusammengefasst, die sich auf die interaktiven Bestandteile der Lernumgebung beziehen, wie das Quiz oder die Arbeit mit realen Messdaten. Besonders hervorgehoben wurden dabei die Abwechslung, die unmittelbare Rückmeldung zu den Aufgaben und die Möglichkeit, aktiv am Lernprozess teilzunehmen ($n_{ohne} = 4$; $n_{mit} = 9$).
- **Darstellungsform der Lernumgebung:** Diese Kategorie umfasst Aussagen, die sich auf die Gestaltung und Verständlichkeit des Erklärvideos sowie der Aufgaben beziehen. Die

¹⁶Die Kodierungen der einzelnen Antworten der Diversitätsgruppe bzw. der Gruppe ohne Diversitätsdarstellung sind in den Tabellen 26 und 25 (S. 93&92) dargestellt.

Schüler:innen betonten insbesondere die einfache, anschauliche und informative Aufbereitung der Inhalte, wodurch das Verständnis erleichtert wurde ($n_{ohne} = 6$; $n_{mit} = 2$).

Die prozentualen Anteile der Nennungen beider Gruppen sind in der Abbildung 15 dargestellt. Die Darstellung in Prozentwerten ermöglicht es, die relativen Häufigkeiten der genannten Kategorien trotz unterschiedlicher Antwortzahlen klar und unmittelbar zwischen den Gruppen zu vergleichen.

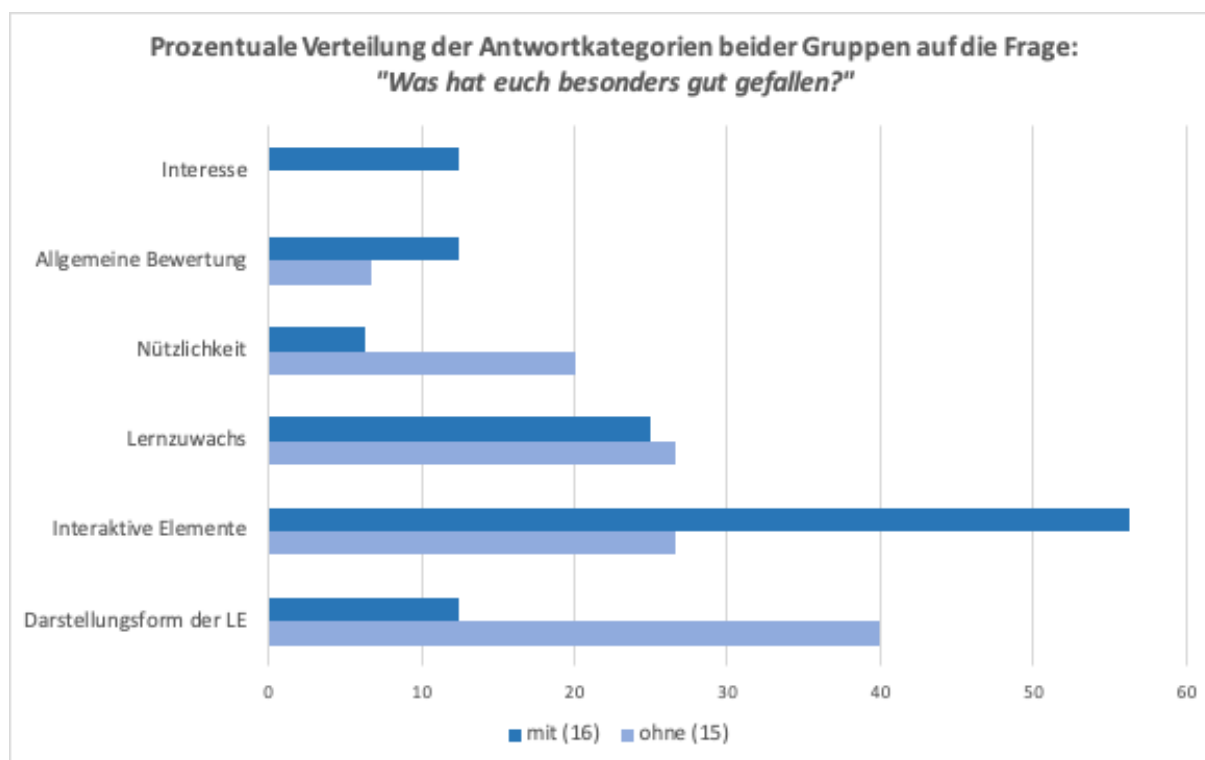


Abbildung 15: Prozentuale Verteilung der Antwortkategorien beider Gruppen auf die Frage: „Was hat euch besonders gut gefallen?“

In der Diversitätsgruppe wurden vor allem die interaktiven Elemente der Lernumgebung positiv hervorgehoben (56 %), beispielsweise „die Vielfalt an aufgaben Arten, nicht nur Auswahl Fragen...“. In der Kontrollgruppe spielte dieser Aspekt mit 27 % eine geringere Rolle. Beide Gruppen betonten gleichermaßen den Lernzuwachs (ca. 25 %), etwa durch Aussagen wie „Man hat in kürzester Zeit viel gelernt“, was auf einen insgesamt positiven wahrgenommenen Lernerfolg hinweist.

Die Darstellungsform der Lernumgebung wurde dagegen von der Kontrollgruppe geschätzt (40%). Hier lobten Teilnehmende insbesondere „Die kreative Gestaltung, die Verständlichkeit des Videos (das Video generell“ (siehe Tabelle 26 Seite 93), während dieser Aspekt in der Diversitätsgruppe nur vereinzelt genannt wurde (13 %). Einzelne Teilnehmende der Diversitätsgruppe nannten zudem *Interesse* oder eine allgemein positive Bewertung als Gründe für ihre Zufriedenheit. Die Kategorie *Nützlichkeit* wurde in beiden Gruppen genannt; jedoch wurde sie in der Gruppe ohne Diversitätsdarstellung häufiger erwähnt (20% zu 6,25%).

Insgesamt deuten die Ergebnisse darauf hin, dass beide Lernumgebungen von den Teilnehmenden positiv wahrgenommen wurden. Während die Diversitätsgruppe stärker die interaktive Kompo-

nente der Lerneinheit hervorhob, betonte die Kontrollgruppe vor allem die gestalterische Umsetzung der digitalen Lernumgebung.

„Was hat euch nicht gefallen?“

Analog zur vorangegangenen Frage wurden auch die Antworten auf diese offene Frage inhaltlich analysiert und zu thematisch ähnlichen Aussagen zusammengefasst. Insgesamt gaben 11 von 24 Schüler:innen aus der Diversitätsgruppe und 15 von 22 Schüler:innen aus der Gruppe ohne Diversitätsdarstellung zu dieser Frage eine Antwort. Es wurde beispielsweise die Aussage „*Es waren zu viele Zahlen und ein bisschen zu sachlich*“ sowohl der Kategorie *Schwierigkeiten bzw. Überforderung* als auch der Kategorie *Inhalt bzw. Thema* zugeordnet. Dieses Beispiel verdeutlicht, dass die Schüler:innen in ihren Rückmeldungen häufig sowohl auf den Umfang oder Schwierigkeitsgrad der Aufgaben als auch auf inhaltliche Aspekte und die Art der Themenaufbereitung Bezug nahmen. Die Kategorien sind somit auch in dieser Auswertung nicht disjunkt. Die Kodierung der einzelnen Antworten der Diversitätsgruppe bzw. Gruppe ohne Diversitätsdarstellung ist in den Tabellen 29 und 28 auf Seite 96 und 95 dargestellt. Tabelle 30 auf Seite 97 zeigt die Häufigkeiten und prozentualen Anteile der Nennungen beider Gruppen sowie exemplarische Schüleraussagen, die den jeweiligen Kategorien zugeordnet wurden. Dabei ergaben sich insgesamt **acht Hauptkategorien**:

- **Kein Lernzuwachs:** In dieser Kategorie wurden Äußerungen zusammengefasst, in denen Schüler:innen angaben, nichts Neues gelernt zu haben oder das Lernmaterial als zu bekannt empfanden ($n_{ohne} = 0$; $n_{mit} = 1$).
- **Desinteresse:** Diese Kategorie umfasst Rückmeldungen mit Desinteresse an der Lerneinheit, wie die Äußerung „ich will nach Hause“ ($n_{ohne} = 1$; $n_{mit} = 0$).
- **Keine Kritik:** Diese Kategorie beinhaltet Antworten, in denen die Schüler:innen ausdrücklich angaben, keine Kritikpunkte zu haben oder mit der Lernumgebung zufrieden zu sein ($n_{ohne} = 3$; $n_{mit} = 1$).
- **Gestaltung der KI-Bilder:** Diese Kategorie umfasst kritische Rückmeldungen zur Darstellung der im Video verwendeten KI-generierten Personen und Szenen. Mehrere Schüler:innen empfanden die Abbildungen als künstlich oder wenig authentisch und gaben an, dass diese die Konzentration eher erschwerten als unterstützten ($n_{ohne} = 3$; $n_{mit} = 5$).
- **Technische Probleme:** Diese Kategorie umfasst Hinweise auf technische Schwierigkeiten bei der Nutzung der Webseite oder des Quiz. Mehrere Schüler:innen berichteten, dass einzelne Aufgaben oder Antwortfelder nicht funktionierten oder die Seite teilweise fehlerhaft war ($n_{ohne} = 2$; $n_{mit} = 2$).
- **Didaktische Umsetzung:** Hierzu zählen Rückmeldungen zur Sozialform und unklare Aufgabenstellungen. Manche Schüler:innen wünschten sich mehr Interaktion oder Gruppenarbeit und empfanden die Einzelarbeit als weniger motivierend ($n_{ohne} = 2$; $n_{mit} = 2$).

- **Inhalt bzw. Thema:** Diese Kategorie enthält kritische Anmerkungen zur thematischen Ausrichtung oder zur inhaltlichen Relevanz der Lerneinheit. Einzelne Schüler:innen äußerten Skepsis gegenüber bestimmten inhaltlichen Aussagen, wie etwa dem Zusammenhang zwischen Feinstaub und Klimawandel ($n_{ohne} = 3$; $n_{mit} = 0$).
- **Schwierigkeiten bzw. Überforderung:** Hierunter wurden Aussagen zusammengefasst, die auf eine kognitive oder inhaltliche Überforderung hinweisen. Einige Schüler:innen berichteten von zu vielen Fragen, zu vielen Zahlen. Auch die selbstständige Datensuche oder Rechenaufgaben wurden teils als anstrengend beschrieben ($n_{ohne} = 4$; $n_{mit} = 3$).

Die prozentualen Anteile der Nennungen beider Gruppen sind in der Abbildung 16 dargestellt.

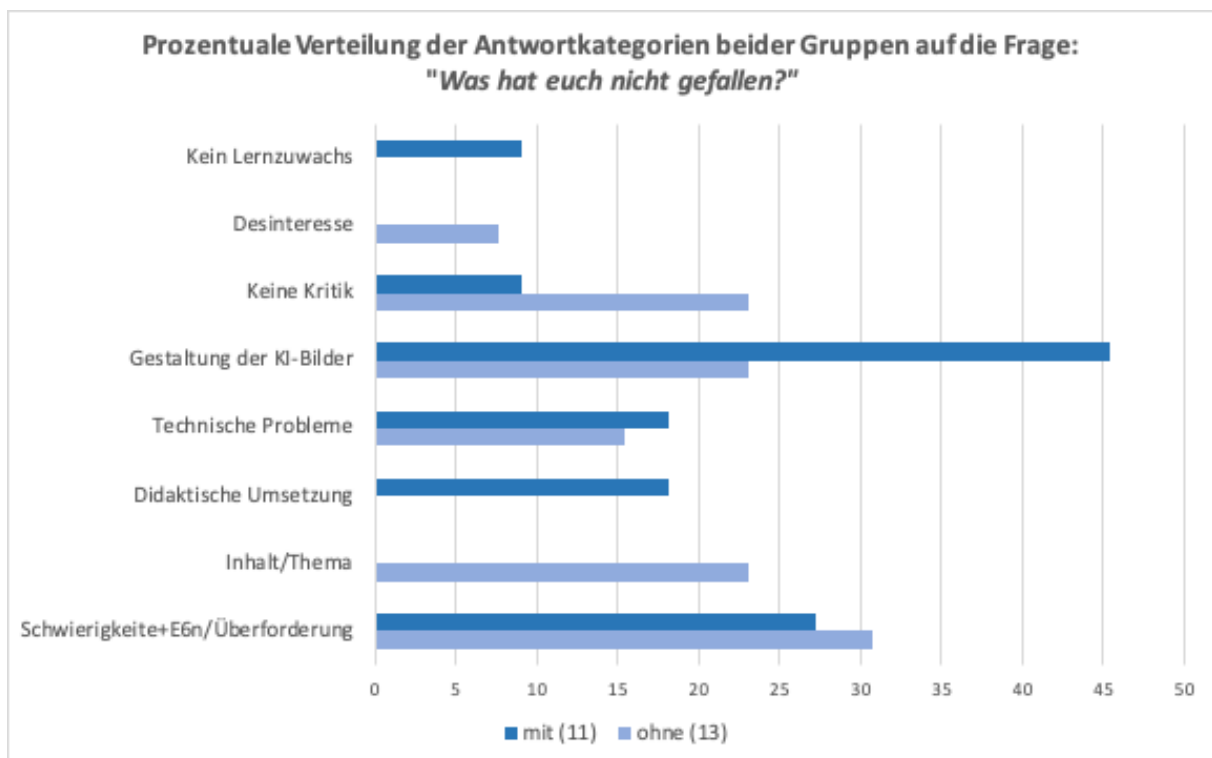


Abbildung 16: Prozentuale Verteilung der Antwortkategorien beider Gruppen auf die Frage: „Was hat euch nicht gefallen?“

In beiden Gruppen wurde am häufigsten die Kategorie Schwierigkeiten bzw. Überforderung genannt (Kontrollgruppe: 31 %, Diversitätsgruppe: 27,27 %), etwa durch Aussagen wie „Das ich viel lesen musste und fast alles falsch hatte“. In der Diversitätsgruppe wurde besonders häufig die Gestaltung der KI-Bilder kritisiert (45 %), beispielsweise mit Kommentaren wie „KI im Video, es fühlt sich nicht authentisch an.“. Dieser Aspekt wurde auch von Teilnehmenden der Kontrollgruppe angesprochen (23 %), jedoch deutlich seltener. Darüber hinaus äußerten Lernende der Kontrollgruppe Kritik am Inhalt bzw. Thema der Lerneinheit („Ich glaube nicht, dass Feinstaub zum Klimawandel beiträgt“), was in der Diversitätsgruppe nicht vorkam.

Beide Gruppen nannten vereinzelt Punkte wie die didaktische Umsetzung („Es war sehr leise und es hätte mehr Spaß gemacht, hätte man in der Gruppe gelernt“) oder technische Probleme („Die Webseite ist zum Teil kaputt“). Einzelne Teilnehmende gaben an, keine Kritik zu haben („Alles war in Ordnung“). Nur vereinzelt wurden Desinteresse oder fehlender Lernzuwachs thematisiert.

Insgesamt zeigen die Ergebnisse, dass die Lernenden der Diversitätsgruppe vor allem die künstlichen Bildelemente als störend empfanden, während die Kontrollgruppe häufiger inhaltliche oder technische Aspekte kritisierten. Beide Gruppen berichteten ähnliche Schwierigkeiten hinsichtlich der Aufgabenbewältigung, was auf einen vergleichbaren kognitiven Anspruch der Lerneinheit schließen lässt.

6 Diskussion der Ergebnisse

Die vorliegende Untersuchung verfolgte das Ziel zu prüfen, ob die Integration von Diversitätsdarstellungen in einem Erklärvideo und den zugehörigen digitalen Quizaufgaben einen Einfluss auf die intrinsische Motivation von Schüler:innen in einer digitalen Lernumgebung zum Thema Feinstaub hat (F1) und inwiefern sich diese Wirkungen zwischen verschiedenen Jahrgangsstufen unterscheiden (F2).

Generell ist zu berücksichtigen, dass diese Untersuchung auf einer vergleichsweise kleinen Stichprobe ($N = 46$) basiert, die zudem in zwei Gruppen unterteilt wurde (Diversitätsgruppe: $N = 24$; Ohne-Diversitätsgruppe: $N = 22$). Besonders gering fallen die Fallzahlen in den gruppenspezifischen Jahrgangsstufen aus, sodass die erzielten Befunde lediglich als erste Hinweise bzw. Tendenzen zu verstehen sind und nicht als belastbare generalisierbare Aussagen interpretiert werden sollten.

Im Folgenden werden die Ergebnisse zu den beiden Forschungsfragen diskutiert und beantwortet (Abschnitt 6.1 und 6.2). Abschließend werden in 6.3 die zentralen Limitationen dieser Untersuchung dargestellt und kritisch eingeordnet.

6.1 (F1) Wie wirken diversitätsberücksichtigende digitale Lernumgebungen auf die intrinsische Motivation von Schüler:innen?

Digitale Lernumgebungen durch audiovisuelle Medien, wie Erklärvideos und interaktive Aufgaben können die Motivation der Schüler:innen fördern (vgl. Bogedan, 2017; Conradt, 2009; Findeisen et al., 2019; Arslan et al., 2020). Studien zur Diversität im Bildungskontext zeigen zugleich, dass vielfältige Repräsentationen von Geschlechtern, kulturellen Hintergründen oder Lebensrealitäten Identifikation und Zugehörigkeit fördern (vgl. Tajmel, 2024; Tajfel & Turner, 1997; McKinley et al., 2014). Vor diesem Hintergrund lässt sich annehmen, dass soziale Eingebundenheit im Sinne der Selbstbestimmungstheorie nach Deci und Ryan (1993) auch durch symbolische Repräsentationen in digitalen Medien adressiert werden kann.

Daraus ergibt sich die theoretische Erwartung, dass eine digitale diversitätssensible Gestaltung tendenziell zu höheren Ausprägungen motivationaler Erlebensdimensionen führen könnte, etwa zu stärkerem Interesse, höherer Anstrengungsbereitschaft oder größerer wahrgenommener Nützlichkeit.

Die quantitativen Ergebnisse dieser Untersuchung in Abschnitt 5.2.1 zeigen jedoch, dass die diversitätssensible Gestaltung keine durchgehend stärkende motivationale Wirkung entfaltet hat,

sondern lediglich einzelne Facetten beeinflusste. Für die Subskalen *Interesse* und *Nützlichkeit* ergaben sich keine signifikanten Unterschiede zwischen der Diversitätsgruppe und der Kontrollgruppe. Lediglich für die Subskala *Anstrengung* zeigte sich ein statistisch signifikanter Unterschied zugunsten der Diversitätsgruppe, begleitet von einer mittleren Effektstärke.

Die Ergebnisse zur Subskala *Nützlichkeit* und *Interesse* lassen sich gut durch das thematische Setting erklären. Die fehlenden Gruppenunterschiede bei der *Nützlichkeit* deuten darauf hin, dass die Bewertung dieser Subskala primär inhaltlich geprägt war. Feinstaub besitzt aufgrund der jährlichen Burning Season, der sichtbaren Echtzeitdaten und der direkten Alltagsrelevanz eine hohe persönliche Bedeutung für die Lernenden, sodass das Thema von beiden Gruppen gleichermaßen als nützlich eingeschätzt wurde. Auch die qualitativen Rückmeldungen stützen diese Interpretation. Insgesamt drei Schüler:innen beider Gruppen erwähnten die *Nützlichkeit* des Themas explizit positiv. Gleichzeitig äußerten einige Lernende auch Kritik an der inhaltlichen Darstellung des Feinstaubthemas. So bemängelte eine Person etwa, dass „die Quelle begrenzt und nicht differenziert“ (siehe Tabelle 28) gewesen sei, während eine andere Rückmeldung lautete, das Video sei „ein bisschen zu sachlich“ gestaltet. Diese Äußerungen lassen erkennen, dass sich einzelne Lernende bereits zuvor mit dem Thema Feinstaub auseinandergesetzt hatten und daher eine differenziertere oder abwechslungsreichere Darstellung erwarteten.

Diese Befunde weisen darauf hin, dass situativ stark eingebettete und lebensweltlich bedeutsame Themen das *Interesse* der Lernenden bereits so intensiv aktivieren können, dass gestalterische Unterschiede im Erklärvideo, wie Diversitätsrepräsentationen, kaum einen zusätzlichen Einfluss entfalten. Diversität scheint im Fall relevanter Umweltthemen für diese Schüler:innen somit keinen unmittelbaren Mehrwert für das situative Interesse zu bieten. Vielmehr rückt der inhaltliche Bezug zum eigenen Alltag der Lernenden in den Vordergrund und wird zum dominierenden Einflussfaktor für ihre motivationale Bewertung. Ergänzend dazu zeigen die qualitativen Daten, dass zwei Lernende der Diversitätsgruppe ihr Interesse ausdrücklich und ohne entsprechende Aufforderung äußerten, etwa durch Aussagen wie „Es hat mir Spaß bereitet“ (siehe Tabelle 26). Diese Rückmeldungen beziehen sich auf die gesamte digitale Lernumgebung und nicht ausschließlich auf die Gestaltung des Erklärvideos. Trotz dieser expliziten Interessensbekundungen einzelner Lernender spiegelt sich dieser Eindruck nicht in den quantitativen Ergebnissen wider. Dies deutet darauf hin, dass die positiven Einzeläußerungen nicht ausreichten, um das gruppenweite Antwortmuster messbar zu beeinflussen, und somit eher individuelle Wahrnehmungen als kollektive Effekte der Diversitätsdarstellung widerspiegeln.

Vor dem Hintergrund der Selbstbestimmungstheorie nach Deci und Ryan (1993), die betont, dass Bedeutsamkeit und lebensweltliche Verankerung kognitive Aktivierung und Interesse fördern, erscheint es plausibel, dass die starke thematische Relevanz potenzielle Diversitätseffekte überlagert hat.

Zusätzlich ist zu erwähnen, dass die explorative Faktoranalyse ergab, dass die Subskala *Nützlichkeit* nicht vollständig homogen ist. Zwei Items bilden offenbar eine eigenständige Unterfacette ab. Dies könnte erklären, warum Unterschiede zwischen den Gruppen in dieser Subskala weniger klar ausfallen, da die Skala mehrere inhaltliche Bedeutungsbereiche bündelt.

Nur bei der Subskala *Anstrengung* zeigte sich ein statistisch signifikanter Unterschied zugunsten der Diversitätsgruppe, begleitet von einer mittleren Effektstärke. Dies lässt darauf schließen, dass die Lernenden dieser Gruppe ihre Mitarbeit als engagierter und bedeutsamer wahrnahmen und bereit waren, mehr Aufwand in die Aufgaben der Lernumgebung zu investieren (vgl. Deci & Ryan, 1993). Ergänzend dazu verdeutlichen die qualitativen Rückmeldungen, dass mehr Schüler:innen der Diversitätsgruppe (56,25 %) ihr Interesse und ihre aktive Auseinandersetzung explizit hervorhoben als in der Kontrollgruppe (26,67 %). Genannt wurden dabei Aspekte wie die „Vielfalt an aufgaben Arten, nicht nur Auswahl Fragen“ oder „Das man auch selber Daten raussuchen sollte“. Diese Aussagen beziehen sich vor allem auf die interaktiven und handlungsorientierten Bestandteile der Lernumgebung und deuten darauf hin, dass die diversitätssensible Version diese Elemente für einige Lernende besonders ansprechend oder aktivierend erscheinen ließ.

Die Theorie besagt, dass Erklärvideos allein häufig nicht ausreichen, um Lernmotivation oder Lernwirksamkeit nachhaltig zu fördern, sondern durch aktivierende Elemente wie interaktive Aufgaben ergänzt werden sollten (vgl. Kulgemeyer, 2020; Findeisen et al., 2019). Da beide Gruppen dieselben fachlichen Inhalte, Quizformate und Arbeitsaufträge erhielten und sich ausschließlich die Darstellung der Personen unterschied, deutet der Unterschied darauf hin, dass die diversitätssensible Darstellung einen zusätzlichen motivationalen Impuls gegeben haben könnte. Dies bietet eine mögliche Erklärung dafür, warum diese Gruppe insgesamt höhere Werte in der Subskala *Anstrengung* berichtete.

Zu berücksichtigen ist zudem, dass die interaktiven Quizaufgaben erst nach dem Ansehen des Erklärvideos bearbeitet wurden. Es ist daher möglich, dass sich Lernende der Diversitätsgruppe durch die diversitätssensible Darstellung im Video stärker angesprochen oder einbezogen fühlten und mit einer höheren Grundmotivation in die anschließenden Aufgaben gingen. Dies bietet eine weitere plausible Erklärung dafür, warum diese Gruppe insgesamt höhere Werte in der Subskala *Anstrengung* berichtete. Eine solche Wirkung würde darauf hindeuten, dass die Diversitätsdarstellung das Gefühl von Zugehörigkeit stärken konnte, ein Aspekt, der nach der Selbstbestimmungstheorie wesentlich zur motivationalen Aktivierung beiträgt und Lernende dazu anregen kann, sich intensiver mit den Aufgaben auseinanderzusetzen.

Die Interpretation der Ergebnisse zur Subskala *Anstrengung* ist jedoch durch die geringe Reliabilität dieser Messung eingeschränkt. Cronbachs Alpha ($\alpha = 0,22$) weist auf eine sehr niedrige interne Konsistenz hin, und auch McDonald's Omega ($\omega = 0,61$) signalisiert lediglich eine grenzwertige Zuverlässigkeit. Das heterogene Muster der explorativen Faktorenanalyse zeigt zudem, dass die Items dieser Subskala nicht eindeutig ein gemeinsames Konstrukt abbilden. Damit ist der gefundene signifikante Unterschied zwar theoretisch plausibel und wird durch qualitative Rückmeldungen gestützt, seine Stabilität ist jedoch begrenzt.

Insgesamt deuten die Befunde darauf hin, dass die eingesetzte Diversitätsdarstellung keine eindeutig motivational relevante Wirkung entfaltet hat. Verschiedene inhaltliche und methodische Faktoren können zu diesem Ergebnis beigetragen haben.

Für die Beantwortung der Forschungsfrage bedeutet dies, dass diversitätssensible Darstellungen

keine konsistenten Effekte auf *Interesse* oder *Nützlichkeit* zeigten, jedoch Hinweise auf eine positive Wirkung im Bereich der *Anstrengung* und ihre Bereitschaft erkennbar sind. Aufgrund der eingeschränkten Messpräzision dieser Subskala sollte dieser Befund jedoch vorsichtig interpretiert werden und vielmehr als erster Hinweis denn als eindeutiger Effekt verstanden werden.

6.2 (F2) Inwiefern unterscheiden sich die Wirkungen diversitätsberücksichtigender digitale Lernumgebungen auf die Motivation von Schüler:innen je nach Jahrgangsstufe?

Die theoretische Betrachtung ergab, dass Jugendliche Diversitäts- und Zugehörigkeitsaspekte nicht unabhängig vom Alter wahrnehmen, sondern diese im Verlauf der Adoleszenz zunehmend differenzierter wahrnehmen. Empirische Befunde zeigen, dass insbesondere ältere Schüler:innen sensibler auf Inklusion, Zugehörigkeit und erlebte Ungleichbehandlung reagieren (Mulvey et al., 2022), während internationale Analysen belegen, dass soziale Unterschiede im Zugehörigkeitsgefühl vor allem in der Sekundarstufe deutlich hervortreten und mit zunehmendem Alter an Bedeutung gewinnen (Kleemola et al., 2025). Diese altersbezogenen Entwicklungen lassen erwarten, dass ältere Schüler:innen Diversitätsdarstellungen bewusster wahrnehmen und stärker positiv in ihre Bewertung der Lernumgebung einbeziehen könnten.

Der Vergleich der einzelnen Subskalen zwischen den beiden Gruppen innerhalb der einzelnen Jahrgangsstufen zeigt, dass weder für die Subskala *Interesse* noch für *Nützlichkeit* in einem der Jahrgänge signifikante Unterschiede festgestellt werden konnten. Für die Subskala *Anstrengung* ist lediglich in der Jahrgangsstufe 11 ein signifikanter Unterschied mit einer großen Effektstärke zu finden, während bei den anderen Jahrgangsstufen kein Unterschied erkennbar ist.

Der Vergleich der Kenngrößen der Subskala *Interesse* zeigte zwischen den Jahrgangsstufen leichte Unterschiede, jedoch ohne erkennbares Muster. Beim Vergleich der Kenngrößen der zwei jüngeren Klassenstufen (9. und 10. Jahrgang) mit den Oberstufen ist kein Muster zu erkennen, ob die Diversitätsgruppe eine höhere *Interesse* aufweist, da die Verteilung sehr heterogen ist. Daraus lässt sich ableiten, dass diversitätssensible Darstellungen das *Interesse* bei Schüler:innen unabhängig vom Alter nicht beeinflusst.

Auch die wahrgenommene *Nützlichkeit* zeigte zwischen den Jahrgangsstufen leichte Schwankungen. Die Kennwerte der Subskala *Nützlichkeit* liegen in allen Gruppen und Jahrgangsstufen über dem Skalenmittelpunkt, was auf einen insgesamt positiven persönlichen Bezug zur Relevanz des Feinstaubthemas hinweist. Auffällig ist jedoch, dass der Median in Jahrgang 12, im Vergleich zu den übrigen Jahrgangsstufen, deutlich niedriger ausfällt. Eine mögliche Erklärung könnte sein, dass sich die Lernenden dieser Klassenstufe bereits intensiver mit dem Thema Feinstaub beschäftigt haben, wodurch das Thema für sie weniger neu oder weniger bedeutsam erschien als für jüngere Jahrgänge. Insgesamt bewegen sich die Unterschiede jedoch im moderaten Bereich, ein Indiz dafür, dass die Einschätzung des thematischen Nutzens nicht stark altersabhängig ist.

Bei der Subskala *Anstrengung* zeigt sich lediglich in der Jahrgangsstufe 11 ein signifikanter Unterschied mit einer sehr großen Effektstärke ($g = -1,57$) zugunsten der Diversitätsgruppe.

Während in den übrigen Jahrgängen keine vergleichbaren Unterschiede auftraten.

Auffällig ist jedoch, dass die Kennwerte der Diversitätsgruppe, mit Ausnahme der 12. Jahrgangsstufe, in allen Jahrgangsstufen höhere Ausprägungen in *Anstrengung* aufweisen. Im 12. Jahrgang liegen die Kennwerte beider Gruppen sehr nah beieinander ($M_{ohne} = 3,75$, $M_{mit} = 3,90$; $\tilde{x}_{ohne} = 3,63$, $\tilde{x}_{mit} = 3,50$), sodass die Diversitätsdarstellung hier keinen erkennbaren Einfluss auf die *Anstrengung* hatte. Eine mögliche Erklärung dafür ist, dass Schüler:innen im Abschlussjahr stärker durch andere Faktoren geprägt sind, etwa durch Prüfungsdruck, die Priorisierung anderer Fächer oder ihre Zukunftsplanung nach der Schulzeit. Solche übergeordneten Anforderungen könnten die Bedeutung gestalterischer Merkmale im Lernvideo überlagern und damit das Engagement unabhängig von der präsentierten Darstellungsform beeinflussen.

Wie bereits zuvor diskutiert, ist die Interpretation der Ergebnisse dieser Subskala jedoch aufgrund der geringen Reliabilität und der eingeschränkten Validität mit Vorsicht vorzunehmen. Die Messung der *Anstrengung* weist deutliche Unsicherheiten auf, sodass die beobachteten Unterschiede zwar theoretisch plausibel, empirisch jedoch nur eingeschränkt belastbar sind.

Der Vergleich von Median zwischen den beiden Gruppen für die einzelnen Subskalen in den jeweiligen Jahrgangsstufen zeigt ein heterogenes Bild. Während in einigen Jahrgängen die Diversitätsgruppe in allen drei Subskalen höhere Medianwerte aufweist, finden sich in anderen Jahrgängen keine Unterschiede oder sogar höhere Werte in der Ohne-Diversitätsgruppe. Dieses uneinheitliche Muster legt nahe, dass die Wirkung der Diversitätsdarstellung nicht über alle Altersgruppen hinweg konsistent ausfällt, sondern möglicherweise von jahrgangsbezogenen Faktoren beeinflusst wird, wie etwa dem Vorwissen, der thematischen Vorerfahrung oder der individuellen Lernorientierung.

Die ausgeprägte Heterogenität der Ergebnisse über die Jahrgänge hinweg verdeutlicht, dass der Effekt kontextabhängig ist und nicht als allgemeiner Befund verallgemeinert werden sollte. Insgesamt lässt sich festhalten, dass diversitätssensible digitale Lernumgebungen nicht altersunabhängig wirksam sind. Methodisch ist zudem zu berücksichtigen, dass die Subskala *Anstrengung* nur eine eingeschränkte Reliabilität aufweist, was die Belastbarkeit der Aussagen zu den Effekten weiter begrenzt. Trotz dieser Einschränkungen liefern die Ergebnisse wertvolle Hinweise darauf, dass diversitätssensible Darstellungen das engagierte Lernverhalten bestimmter Lerngruppen fördern können und deshalb ein relevantes Gestaltungselement in digitalen Lernumgebungen darstellen.

6.3 Limitierung

Um die Befunde der vorliegenden Untersuchung angemessen einzuordnen, ist es notwendig, mehrere methodische und inhaltliche Einschränkungen zu berücksichtigen. Diese Limitierungen betreffen sowohl die Stichprobenstruktur und Messinstrumente als auch die Gestaltung der Intervention sowie den thematischen Kontext der Lerneinheit. Sie begrenzen die Aussagekraft der Ergebnisse und sollten bei der Interpretation der Wirkungen diversitätssensibler digitaler Lernumgebungen bedacht werden.

Begrenzte Stichprobengröße und Gruppenaufteilung

Die Gesamtstichprobe war mit $N = 46$ relativ klein und verteilte sich zudem auf zwei experimentelle Gruppen sowie vier Jahrgangsstufen. Dadurch verringern sich statistische Power und Robustheit der Befunde erheblich. Insbesondere die jahrgangsbezogenen Analysen basieren auf sehr kleinen Teilgruppen, sodass beobachtete Unterschiede eher als Tendenzen und nicht als belastbare Effekte zu interpretieren sind.

Eingeschränkte psychometrische Eigenschaften der IMI-Subskala

Die Interpretation der Ergebnisse wird durch die psychometrischen Eigenschaften einzelner IMI-Subskalen begrenzt. Während die Subskalen *Interesse* und *Nützlichkeit* insgesamt zufriedenstellende interne Konsistenzen zeigten, wies die Subskala *Anstrengung* nur eine eingeschränkte Reliabilität auf. Dadurch sind Interpretationen signifikanter Gruppenunterschiede, insbesondere für diese Subskala, nur mit Vorsicht möglich.

Zudem liegen für die Subskala *Nützlichkeit* und *Anstrengung* die deutsche IMI-Version bislang keine abgesicherte Validitäts- und Strukturinformationen vor. Daher bleibt offen, in welchem Ausmaß die Skalen die intendierten Konstrukte im Kontext digitaler Lernumgebungen adäquat erfassen. Zusätzlich zeigten in der explorativen Faktorenanalyse die Subskalen *Nützlichkeit* und *Anstrengung* eine inhomogene Struktur. Zwei der fünf Items der Subskala *Nützlichkeit* bilden eine eigenständige Unterfacette ab, was auf eine inhaltliche Heterogenität innerhalb dieser Skala hinweist. Diese Strukturmerkmale könnten die Sensitivität der Skala reduzieren und zu weniger klaren Gruppenunterschieden beigetragen haben.

Die Subskala *Anstrengung* erwies sich ebenfalls als strukturell instabil. Drei der vier Items luden in der Faktorenanalyse stärker auf andere Subskalen als auf den intendierten Faktor, was auf eine unzureichende Zuordnung der Konstrukte hinweist. Besonders auffällig war das Item „Es war mir wichtig, bei den Quizaufgaben gut abzuschneiden“, das in keinem der extrahierten Faktoren eine Ladung von über 0,30 erreichte. Nach den Empfehlungen der Universität Zürich (o.D.) sollte ein solches Item entfernt werden, da es weder konzeptionell noch empirisch einen klaren Beitrag zum Zielkonstrukt leistet.

Diese Befunde verdeutlichen, dass die psychometrischen Schwächen der beiden Subskalen die Messgenauigkeit und Interpretierbarkeit der Ergebnisse eingeschränkt haben könnten. Insbesondere die strukturelle Heterogenität könnte dazu beigetragen haben, dass vorhandene Effekte abgeschwächt oder verzerrt abgebildet wurden.

Ausgewählte Subskalen des IMI

Schließlich wurde nur ein Teil der IMI-Skalen eingesetzt (*Interesse*, *Nützlichkeit*, *Anstrengung*). Andere Facetten der intrinsischen Motivation, wie *wahrgenommene Kompetenz*, *wahrgenommene Wahlfreiheit*, *Druck* oder *soziale Eingebundenheit*, wurden nicht erhoben. Damit bleibt unklar, ob Diversitätsdarstellungen auf diese weiteren motivationalen Dimensionen gewirkt hätten.

Qualität der visuellen Umsetzung (KI-Bilder)

In beiden Gruppen wurde Kritik an unnatürlichen Bewegungen oder untypischen Körperhal-

tungen in KI-generierten Szenen geäußert. Dies deutet darauf hin, dass die wahrgenommene Glaubwürdigkeit der Darstellungen möglicherweise einen stärkeren Einfluss auf das emotionale Erleben hatte als die Frage, ob sich Lernende den dargestellten Personengruppen zugehörig fühlen.

Auffällig ist, dass der Anteil entsprechender Nennungen in der Diversitätsgruppe bei etwa 45 % lag und damit nahezu doppelt so hoch wie in der Gruppe ohne Diversitätsdarstellung. Eine mögliche Erklärung hierfür ist, dass die Experimentalgruppe sowohl KI-generierte Szenen als auch Sequenzen mit echten Menschen sah und somit einen direkten Vergleich ziehen konnte. In der Kontrollgruppe hingegen bestanden alle Szenen ausschließlich aus KI-generierten Darstellungen, wodurch potenzielle Auffälligkeiten weniger stark hervortraten.

Dieser Befund legt nahe, dass künstlich wirkende oder ästhetisch inkonsistente KI-Bilder die wahrgenommene Glaubwürdigkeit der Darstellung beeinträchtigen und damit das emotionale Erleben der Lernenden beeinflussen können. Inwieweit diese Wahrnehmungen im Vergleich zu den inhaltlichen Darstellungsmerkmalen tatsächlich motivationale Prozesse beeinflusst haben, lässt sich auf Basis der vorliegenden Daten jedoch nicht bestimmen.

Kontextabhängigkeit des Themas Feinstaub

Die Lerneinheit thematisiert Feinstaub in [REDACTED], wo Luftverschmutzung ein stark lebensweltlich verankertes Problem darstellt. Die hohe Relevanz des Themas könnte andere Einflussfaktoren, darunter die Diversitätsdarstellungen, überlagert haben. Die Ergebnisse lassen daher keine Rückschlüsse darauf zu, wie Diversität in weniger alltagsnahen Themenbereichen wirken würde.

Kurzfristiges Messdesign

Die Erhebung der intrinsischen Motivation erfolgte einmalig und unmittelbar nach Abschluss der Lerneinheit. Dadurch lassen sich weder Veränderungen im motivational-affektiven Erleben vor und nach der Intervention noch längerfristige Entwicklungen erfassen. Unklar bleibt daher, ob diversitätssensible Darstellungen in längerfristigen Lernsettings, über mehrere Sitzungen hinweg oder bei wiederholter Exposition stärkere motivationale Effekte erzielen würden.

Mögliche Sensitivität der Teilnehmenden

Einige Lernende schienen bereits während der Durchführung die unterschiedlichen Videoverversionen bewusst zu erkennen. Diese Reaktivität könnte sich auf Bewertungen oder Antworten ausgewirkt haben und stellt eine mögliche Ursache für Verzerrungen dar.

7 Ausblick

Die Ergebnisse dieser Arbeit geben erste Hinweise darauf, wie Diversität in digitalen Lernumgebungen wahrgenommen wird und vor allem, wann sie eine Wirkung entfalten kann. Gleichzeitig zeigen sie auch deutlich, wo zukünftige Arbeiten ansetzen sollten.

Ein wichtiger Punkt betrifft die Größe der Stichprobe. Mit mehr Schüler:innen pro Jahrgang ließen sich Effekte zuverlässiger nachweisen. Größere Gruppen oder zusammengefasste Alters-

bereiche könnten helfen, Muster klarer herauszuarbeiten. Darüber hinaus deutet vieles darauf hin, dass eine einmalige, eher subtile Darstellung von Diversität nicht ausreicht, um Motivation spürbar zu beeinflussen. Zukünftige Lernumgebungen könnten Diversität sichtbarer und über längere Zeit hinweg einbinden, zum Beispiel durch wiederkehrende Figuren, kleine Geschichten oder Beispiele aus verschiedenen Lebenswelten.

Auch die Subskala *Anstrengung* sollte in zukünftigen Studien überarbeitet oder durch alternative Items ergänzt werden, um die interne Konsistenz und Validität zu erhöhen. Eine größere Stichprobe würde darüber hinaus robustere psychometrische Analysen ermöglichen.

Während der Durchführung in der 12. Klasse fiel auf, dass ein Schüler beim Austausch des Videostands mit seinem Sitznachbarn den Unterschied zwischen den beiden Videoversionen bemerkte und sofort die Verbindung zur Untersuchung erkannte. Dieser Moment macht deutlich, dass einige Lernende Diversitätsaspekte bewusst wahrnehmen, auch wenn dies nicht immer in den quantitativen oder offenen Antworten sichtbar wird. Gerade deshalb wäre es sinnvoll, künftig genauer zu erforschen, wie Schüler:innen Diversität in Lernmaterialien wahrnehmen, welche Elemente ihnen besonders auffallen und unter welchen Bedingungen Repräsentation tatsächlich Motivation beeinflussen kann.

Aus den Ergebnissen lässt sich ebenfalls ableiten, dass die Qualität der Darstellung eine große Rolle spielt: Wenn KI-generierte Szenen künstlich wirken, lenkt das eher ab, als dass es zum Lernen beiträgt. Hier besteht noch viel Potenzial, Lernmaterialien so zu gestalten, dass sie glaubwürdiger und ansprechender wirken. Schließlich könnte es spannend sein, die Frage der Diversität auch in anderen Themenbereichen zu untersuchen. Das Thema Feinstaub ist für die Schüler:innen in [REDACTED] besonders relevant, vielleicht so relevant, dass es andere Einflüsse überlagert. In weniger lebensweltbezogenen Themen könnten Diversitätsdarstellungen eventuell stärker wirken.

Auch für die Weiterentwicklung der im Rahmen des *Humboldt-Explorers*-Projekts konzipierten digitalen Lerneinheit zum Thema Feinstaub bieten die Befunde dieser Untersuchung wichtige Ansatzpunkte. Die Rückmeldungen der Lernenden deuten darauf hin, dass insbesondere KI-generierte Bildsequenzen aufgrund ihrer teils uneinheitlichen oder wenig authentischen Darstellung als irritierend wahrgenommen wurden. Für zukünftige Revisionen erscheint es daher sinnvoll zu prüfen, ob entweder ein konsistent umgesetzter Animationsstil oder der verstärkte Einsatz realer filmischer Aufnahmen mit authentischen Handlungssituationen zu einer höheren visuellen Kohärenz und stärkeren Identifikationsmöglichkeiten beitragen kann. Auf diese Weise können die gewonnenen Erkenntnisse unmittelbar in die Weiterentwicklung der Lerneinheit einfließen und dazu beitragen digitale Bildungsformate im Rahmen des *Humboldt-Explorers*-Projekts weiter zu verbessern.

Ein weiterer bemerkenswerter qualitativer Befund war, dass nur Lernende der Diversitätsgruppe den Wunsch nach kooperativen Sozialformen äußerten. Das weist auf Aspekte hin, die über die drei erhobenen IMI-Subskalen hinausgehen. Da dieser Wunsch theoretisch der Selbstbestimmungstheorie der *sozialen* Eingebundenheit zuzuordnen ist, könnte zukünftige Forschung prüfen,

inwiefern Diversitätsdarstellungen nicht nur *Interesse*, *Anstrengung* und *Nützlichkeit* beeinflussen, sondern auch sozialbezogene Bedürfnisse und Präferenzen für gemeinsames Lernen. Eine Erweiterung zukünftiger Studien um eine validierte, deutschsprachige IMI-Subskala *relatedness* (*soziale Eingebundenheit*), die auf den Kontext solcher digitaler Lernumgebungen angepasst ist, könnte hier wertvolle Einblicke liefern.

Insgesamt zeigt diese Arbeit, dass Diversität in digitalen Lernumgebungen ein wichtiges Thema ist, das jedoch nicht automatisch zu mehr Motivation führt. Vielmehr hängt die Wirkung davon ab, wie sichtbar, wie authentisch und wie passend Diversität eingebunden wird. Die Ergebnisse unterstreichen damit, was auch die Forschung zu Repräsentation zeigt: Entscheidend sind Wiederholung, Kontext, Relevanz und die stimmige Einbettung in das Lernmaterial. Genau hier bieten sich viele spannende Ansatzpunkte für zukünftige Entwicklungen, sowohl in der Forschung als auch in der praktischen Gestaltung moderner Lernangebote, wie etwa durch sichtbarer gestaltete und wiederkehrende Diversitätsfiguren, den Einsatz narrativer Elemente, qualitative Vertiefungen wie Interviews zur Wahrnehmung von Diversität oder vergleichende Studien zu unterschiedlichen Gestaltungsformen von Lernvideos.

Literatur

- Arslan, B., Y. Jiang, M. Keehner, T. Gong, I. R. Katz und F. Yan (2020). The Effect of Drag-and-Drop Item Features on Test-Taker Performance and Response Strategies. In: *Educational Measurement* 39 (2), S. 96–106. DOI: 10.1111/emip.12326.
- Attali, Y., C. Laitusis und E. Stone (2015). Differences in Reaction to Immediate Feedback and Opportunity to Revise Answers for Multiple-Choice and Open-Ended Questions. In: *Educational and Psychological Measurement* 76 (5), S. 787–802. DOI: 10.1177/0013164415612548.
- Bernstein, F. und T. Wilhelm (2023). Digitale Messwerterfassung mit Mikrocontrollern: Konnektivität, Datenauswertung, Nutzeroberflächen. In: *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrstagung* 1 (1), S. 281–286.
- Bildungsserver Berlin-Brandenburg (o.D.) *Datenschutz*. URL: <https://bildungsserver.berlin-brandenburg.de/daten-schutz> letzter Zugriff: 20. Oktober 2025.
- Bogedan, C. (2017). *Bildung in der digitalen Welt - Strategie der Kulturministerkonferenz*. Techn. Ber. Kultusministerkonferenz.
- Bosch, C. (2024). Assessing the psychometric properties of the intrinsic motivation inventory in blended learning environments. In: *Journal of Education and e-Learning Research* 11 (2), S. 263–271. DOI: 10.20448/jeelr.v11i2.5468.
- Champoux, J. E. (2005). Comparative analyses of live-action and animated film remake scenes: finding alternative film-based teaching resources. In: *Educational Media International* 42 (1), S. 49–69. DOI: 10.1080/09523980500116662.
- Cheryan, S., V. C. Plaut, P. G. Davies und C. M. Steele (2009). Ambient Belonging: How Stereotypical Cues Impact Gender Participation in Computer Science. In: *Journal of Personality and Social Psychology* (6), S. 1045–1060. DOI: 10.1037/a0016239.
- Cohen, D. und I. Sasson (2016). Online quizzes in a virtual learning environment as a tool for formative assessment. In: *Journal of Technology and Science Education* 6 (3), S. 96–106. DOI: 10.3926/jotse.217.
- Conradty, C. (2009). *Multimedial unterstütztes Lenen: Intrinsische Motivation & kognitiver Lernerfolg*. Dissertation. University Bayreuth.
- Crenshaw, K. W. (2020). Die Intersektion von »Rasse« und Geschlecht demarginalisieren. In: *Feministische Theorie und kritische Medienkulturanalyse*. Hrsg. von T. Thomas und U. Wischermann. (1. Auflage). Bielefeld: transcript Verlag, S. 179–188. ISBN: 9783839440841 (ebook).
- CSDT (2022). *Intrinsic Motivation Inventory (IMI) - Scale Description*. Jahr aus dem Upload-Datum der Datei abgeleitet; im PDF nicht angegeben.
- Deci, E. L. und R. M. Ryan (1985). *Intinsic Motivation and Self-Determination in Human Behavior*. New York: Springer Science+Business Media. DOI: 10.1007/978-1-4899-2271-7.
- Deci, E. L. und R. M. Ryan (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. In: *Zeitschrift für Pädagogik* 39 (2), S. 223–238. DOI: 10.25656/01:11173.
- Deutsche Umwelthilfe (2021): *Orte mit der höchsten Feinstaubbelastung (PM_{2,5}) in Deutschland*. URL: https://www.duh.de/fileadmin/user_upload/download/Projektinformation/Verkehr/Luftreinhaltnung/Orte_h%C3%B6chster_Belastung_Gesundheitsklage.pdf letzter Zugriff: 20. Oktober 2025.
- Dunn, T. J., T. Baguley und V. Brunnsden (2014). From alpha to omega: A practical solution to pervasive problem of internal consistency. In: *The British Psychological Society* (3), S. 399–412. DOI: 10.1111/bjop.12046.
- Eckert, C. (2014): *Stereotype Threat*. URL: <https://www.gender-glossar.de/post/stereotype-threat> at Zugriff am 30. Oktober 2025.
- Findeisen, S., S. Horn und J. Seifried (2019). Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos. In: *Zeitschrift für Theorie und Praxis der Medienbildung*, S. 16–36. DOI: 10.21240/mpaed/00/2019.10.01.X.
- Flandorfer, P. (2019): *Cronbachs Alpha berechnen und interpretieren - mit Beispiel*. URL: <https://www.scribbr.ch/statistik-ch/cronbachs-alpha/> letzter Zugriff: 08. November 2025].
- fowid (2024): *Data Literacy - Kompetenzen für ein digitales Zeitalter Religionszugehörigkeiten 2023*. URL: <https://fowid.de/meldung/religionszugehoerigkeiten-2023> letzter Zugriff: 09. Dezember 2024.
- Frazer, L. (2017). Diversity in Physics Education Research: A Decade of Inclusion and Future Challenges. In: *arXiv: Physics Education*, S. 1–9. DOI: 10.48550/arXiv.1710.07863.
- Frühwirth, G. (2020). *Selbstbestimmt unterrichten dürfen – Kontrolle unterlassen können*. Wiesbaden: Springer Verlag.

- Gäde, J. C., K. Schermelleh-Engel und H. Brandt (2020). Konfirmatorische Faktorenanalyse (CFA). In: *Testtheorie und Fragebogenkonstruktion*, hrsg. von H. Moosbrugger und A. Kelava. Springer Verlag, S. 615–659. DOI: 10.1007/978-3-662-61532-4_24.
- Gadermann, A., M. Guhn und B. Zumbo (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. In: *Practical Assessment, Research & Evaluation* 17, S. 1–13.
- González-Pérez, S., R. M. de Cabo und M. Sáinz (2020). Girls in STEM: Is It a Female Role-Model Thing? In: *Frontiers in Psychology* 11, S. 1–21. DOI: 10.3389/fpsyg.2020.02204.
- Hannover, B. (2011). Geschlecht und soziale Ungleichheit. In: *Empirische Bildungsforschung*. Hrsg. von H. Reinders, H. Ditton, C. Gräsel und B. Gniewosz. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 169–180. DOI: 10.1007/978-3-531-93021-3.
- Hennecke, M. und V. Brandstätter (2016). Intrinsische Motivation. In: *Soziale Motive und soziale Einstellungen (Enzyklopädie der Psychologie, Serie Sozialpsychologie, Bd. 2.)* hrsg. von H.-W. Bierhoff und D. Frey. Göttingen: Hogrefe Verlag, S. 1–30. ISBN: 9783840905643.
- Henukh, A., I. M. Astra, Supriyadi, A. Reski und M. M. S. Hidayatullah3 (2022). The Effectiveness of Using Quizizz in Fundamental Physics Learning in the Era of the Covid-19 Pandemic. In: *Journal of Physics: Conference Series* 2309 (012054), S. 1–7. DOI: 10.1088/1742-6596/2309/1/012054.
- Homepage der Center for Self-Determination Theory (o.D.). *Center for Self-Determination Theory (CSDT)*. URL: <https://selfdeterminationtheory.org/intrinsic-motivation-inventory/> letzter Zugriff: 13. Oktober 2025.
- Honkomp-Wilkens, V. und K. D. Wolf (2024). Diversität in Erklärvideos auf YouTube. Dekonstruktion oder Fortführung einer genderspezifischen Ordnung in informellen audio-visuellen Bildungsräumen? In: *UNGEN in der Medienpädagogik* 53, S. 198–225. DOI: 10.21240/mpaed/53/2024.02.25.X.
- Hosang, K. (2021): *[21st Century Skills] Kompetenzen für die Zukunft*. URL: <https://karlhosang.de/21st-century-skills/> letzter Zugriff: 20. Oktober 2025.
- IMI-english (2022): *Intrinsic Motivation Inventory IMI*. URL: https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_Complete.pdf Jahr aus dem Upload-Datum der Datei abgeleitet; im PDF nicht angegeben, letzter Zugriff: 10.11.2024.
- Kleemola, K., A. Toom, H. Hyttinen, T. Tuononen, I. Karacay, T. Kroezen, I. Kogan und A. Alieva (2025). Exploring intersectional inequalities in students’ sense of belonging in education across educational pathways and across educational contexts. In: *frontiers* 10 (1502115), S. 1447–1464. DOI: 10.3389/feduc.2025.1502115.
- KMK (2025). *Diversitätsbewusste und interkulturelle Bildung und Erziehung in der Schule*. Amtliche Fassung.
- Krebs, A.-K. (2023). *Vielfalt im Physikunterricht Zur Wirkung von Lehrkräftefortbildungen unter Diversitätsaspekten*. Studien zum Physik- und Chemielernen, Band 361. Hrsg. von M. Hopf und L. V. Mathias Ropohl. Berlin: Lagos Verlag. ISBN: 978-3-8325-5672-3.
- Kuckquartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. (4. Auflage). Bochum. ISBN: 978-3-7799-4683-0 (e-book).
- Kulgemeyer, C. (2020). Erklären im Physikunterricht. In: *Physikdidaktik - Grundlagen*. Hrsg. von E. Kircher, R. Girwidz und H. E. Fischer. 4. Auflage. Würzburg, S. 403–426. DOI: 10.1007/978-3-662-59490-2.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. In: *frontiers in Psychology*. DOI: 10.3389/fpsyg.2013.00863.
- Ledesma, R. D., P. Valero-Mora und G. Macbeth (2015). The Scree Test and the Number of Factors: a Dynamic Graphics Approach. In: *Spanish Journal of Psychology* 8 (e11), S. 1–10. DOI: 10.1017/sjp.2015.13.
- Lenhard, W. und A. Lenhard (2022). *Berechnung von Effektstärken*. DOI: 10.13140/RG.2.2.17823.92329.
- Mayer, R. E. (2002). Multimedia Learning. In: *The psychology of learning and motivation* 41, S. 85–139.
- McAuley, E., T. Dungan und V. V. Tammen (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. In: *Research Quarterly for Exercise and Sport* 60 (1), S. 48–58.
- McKinley, C. J., D. Mastro und K. M. Warber (2014). Social Identity Theory as a Framework for Understanding the Effects of Exposure to Positive Media Images of Self and Other on Intergroup Outcomes. In: *International Journal of Communication* 8, S. 1–20. DOI: 10.1037/0003-066X.55.1.68.
- Meller, S. (2024). *Erklärvideos im Sachunterricht. Eine explorative Studie zum Umgang von Lehrkräften mit dem audiovisuellen Medium*. Wiesbaden: Springer VS. DOI: 10.1007/978-3-658-43856-2.

- Menold, N. und K. Bogner (2016). *Design of Rating Scales in Questionnaires*. Techn. Ber. Mannheim: GESIS – Leibniz-Institute for the Social Sciences. DOI: 10.15465/gesis-sg_en_015.
- Mézes, C. (2016). Zur Motivation beim Experimentieren im Physikunterricht. Dissertation. Pädagogischen Hochschule Schwäbisch Gmünd.
- Mulvey, K. L., C. J. Mathews, J. Knox, A. Joy und J. Cerda-Smith (2022). The role of inclusion, discrimination, and belonging for adolescent science, technology, engineering and math engagement in and out of school. In: *Journal of research in science* 59 (8), S. 1447–1464. DOI: 10.1002/tea.21762.
- Obermoser, S. (2018). Einsatz moderner Medien im Unterricht. Unterstützung von Lernprozessen durch Lehr- und Lernvideos? In: *Haushalt in Bildung & Forschung*. 4. Leverkusen: Verlag Barbara Budrich, S. 59–74. DOI: 10.3224/hibifo.v7i4.05.
- OECD (2025). *Reducing Inequalities by Investing in Early Childhood Education and Care*. Starting Stron. Techn. Ber. Paris. DOI: 10.1787/b78f8b25-en.
- Ostrow, K. S. und N. T. Heffernan (2018). Testing the Validity and Reliability of Intrinsic Motivation Inventory Subscales Within ASSISTments. In: C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren und B. du Boulay (Hrsg.): *Artificial Intelligence in Education*. Cham: Springer International Publishing, S. 381–394. ISBN: 978-3-319-93843-1.
- Özdemir, O. (2024). Kahoot! Game-based digital learning platform: A comprehensive meta-analysis. In: *Journal of Computer Assisted Learning* 41 (1), S. 1–24. DOI: 10.1111/jcal.13084.
- Pope, E. E. (2022). *Foreword*. letzter Zugriff: 25. November 2025. New York: UNDP.
- Rheinberg, F. und R. Vollmeyer (2019). *Motivation*. (9. überarbeitete und erweiterte Auflage). Stuttgart. DOI: 10.17433/978-3-17-032955-3.
- Rose, J. (2015). The Gamification of Physics Education: A Controlled Study of the Effect on Motivation of First Year Life Science Students. Masterarbeit. Guelph, Ontario, Canada: The University of Guelph.
- Ryan, R. M. und E. L. Deci (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. In: *American Psychologist* 55 (1), S. 68–78. DOI: 10.1037/0003-066X.55.1.68.
- Sato, K. (2022). Designing effective vocabulary videos: comparing animated and live-action movies. Dissertation. University of Nebraska – Lincoln. DOI: 10.13140/RG.2.2.19254.27204.
- Scheijderberg, C., O. Wiczorek und I. Steinhardt (2022). *Qualitative und quantitative Inhaltsanalyse. Digital und automatisiert. Eine anwendungsorientierte Einführung mit empirischen Beispielen und Softwareanwendungen*. Weinheim. DOI: 10.25656/01:29209.
- Schiefele, U. und E. Schaffner (2015). Motivation. In: *Pädagogische Psychologie*. Hrsg. von E. Wild und J. Möller. (3. überarbeitete und erweiterte Auflage). Berlin/Heidelberg: Springer Verlag, S. 153–175. DOI: 10.1007/978-3-662-61403-7.
- Schnirch, A. (2006). *Gendergerechte Interessen- und Motivationsförderung im Kontext naturwissenschaftlicher Grundbildung*. Studien zum Physik- und Chemielernen, Band 54. Hrsg. von H. Niedderer, H. Fischler und E. Sumfleth. Berlin: Logo-Verlag. ISBN: 3-8325-1334-5.
- Self-Determination Theory (2022): *Intrinsic Motivation Inventory (IMI) - German*. URL: https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_German.pdf Jahr aus dem Upload-Datum der Datei abgeleitet; im PDF nicht angegeben, letzter Zugriff: 10.11.2024.
- SenBJF (2015). *Rahmenlehrplan Teil C Physik. Jahrgangsstufen 7-10*. Amtliche Fassung. Bildungsserver Berlin-Brandenburg.
- SenBJF (2021). *Rahmenlehrplan für die gymnasiale Oberstufe*. Amtliche Fassung. Bildungsserver Berlin-Brandenburg.
- Statistisches Bundesamt (o.D.) *Data Literacy - Kompetenzen für ein digitales Zeitalter*. URL: https://www.destatis.de/DE/Service/DataLiteracy/_inhalt.html letzter Zugriff: 20. Oktober 2025.
- Stensen, K. und S. Lydersen (2022). Internal consistency: from alpha to omega? In: *Tidsskr Nor Lægeforen*, S. 1045–1060. DOI: 10.4045/tidsskr.22.0112.
- Sterzig, F. (2022). Zur Lernwirksamkeit von Erklärvideos in der Physik. Studien zum Physik- und Chemielernen, Band 346. Dissertation. Berlin: Universität Paderborn; Logos Verlag. DOI: 10.30819/5576.
- Tajfel, H. und J. C. Turner (1997). An integrative theory of intergroup conflict. In: S. Worchel und W. G. Austin (Hrsg.): *The social psychology of intergroup relations*. Brooks/Cole, S. 33–47.
- Tajmel, T. (2009). Ein Beispiel: Physikunterricht. In: *Migration und schulischer Wandel: Unterricht*. Hrsg. von S. Fürstenau und M. Gomolla. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 139–156. DOI: 10.1007/978-3-531-91724-5.

- Tajmel, T. (2024). Diversity/Diversität und Migration in der Physikdidaktik. In: *Gender & Diversity Studies in MINT meets Naturwissenschaftsdidaktik*. Hrsg. von S. Huch und M. Erlemann. Wiesbaden: Springer Verlag, S. 173–192. DOI: 10.1007/978-3-658-43616-2.
- Tolksdorf, F. (2017). Diversity in Physik-Schulbüchern. Masterarbeit. Humboldt-Universität zu Berlin.
- UNESCO (2009). *Investing in cultural diversity and intercultural dialogue: UNESCO world report*. ISBN: 978-92-3-104077-1.
- Universität Zürich (o.D.) *Faktorenanalyse*. URL: https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/interdependenz/reduktion/faktor.html#2.5._Schritt_3:_Bestimmen_der_Anzahl_Faktoren letzter Zugriff: 23. November 2025.
- Wedl, J. (2018). Diversity – Intersektionalität: Ein kurzer Vergleich anhand von zwei Modellen. In: *Schule lehrt / lernt Vielfalt*. Hrsg. von A. Spahn und J. Wedl. (1. Auflage). 18. Göttingen: Waldschlösschen Verlag, S. 56–61. ISBN: 978-3-937977-10-4.
- WHO (2023): *Disability*. URL: <https://www.who.int/news-room/fact-sheets/detail/disability-and-health> letzter Zugriff: 25. November 2025.
- Wilde, M., K. Bätz, A. Kovaleva und D. Urhahne (2009). Überprüfung einer Kurzskala intrinsischer Motivation (KIM). In: *Zeitschrift für Didaktik der Naturwissenschaften*, S. 31–45. DOI: 10.25656/01:31663.
- Winker, G. und N. Degele (2020). Intersektionalität. Zur Analyse sozialer Ungleichheiten. In: *Feministische Theorie und kritische Medienkulturanalyse*. Hrsg. von T. Thomas und U. Wischermann. (1. Auflage). Bielefeld: transcript Verlag, S. 179–188. ISBN: 9783839440841 (ebook).
- Winter, J. de, D. Dodou und P. Wieringa (2009). Exploratory Factor Analysis With Small Sample Sizes. In: *Multivariate Behavioral Research* 44 (2), S. 147–181. DOI: 10.1080/00273170902794206.
- Wlodkowski, R. J. (2002). Motivation and Diversity: A Framework for Teaching. In: *New directions for teaching & learning* 1999 (78), S. 5–16. DOI: <https://doi.org/10.1002/tl.7801>.
- Wolf, K. D. (2015a). Produzieren Jugendliche und junge Erwachsene ihr eigenes Bildungsfernsehen? Erklärvideos auf Youtube. In: *Television* (1), S. 35–39.
- Wolf, K. D. (2015b). Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung. Band 2 der Reihe "Mediale Impulse". In: *Filmbildung im Wandel*. Hrsg. von A. Hartung, T. Ballhausen, C. Trültzsch-Wijnen, A. Barberi und K. Kaiser-Müller. Wien: New Academic Press, S. 121–131. ISBN: 978-3-7003-1928-3.
- Wolf, K. D. und C. Kulgemeyer (2022). Lehren und Lernen mit Erklärvideos im Fachunterricht. In: *Handbuch - Lernen mit digitalen Medien*. Hrsg. von G. Brägger und H.-G. Rolff. (2. korrigierte Auflage). Weinheim: Beltz Verlag. Kap. 19, S. 474–487. ISBN: 978-3-407-83232-0 (e-Book).

A Intervention

A.1 Qualitätskriterien für ein gutes Erklärvideo nach Kulgemeyer

Kernidee	Kriterium	Beschreibung
Adaption	Adaption an Vorwissen, Fehlvorstellungen und Interesse	Das Video bezieht sich auf gut beschriebene Eigenschaften einer Adressatengruppe (wahrscheinliches Vorwissen, Interessen, Schülervorstellungen)
Veranschaulichungswerkzeuge nutzen	Beispiele	Das Video nutzt Beispiele, um das Erklärte zu veranschaulichen
	Analogien und Modelle	Das Video nutzt Analogien und Modelle, um die neue Information mit bekannten Wissensbereichen zu verbinden
	Darstellungsformen und Experimente Sprachebene	Das Video nutzt Darstellungsformen und Experimente zur Veranschaulichung
	Mathematisierungsgrad	Das Video wählt eine Sprachebene passend zur beschriebenen Adressatengruppe Das Video wählt einen Mathematisierungsgrad passend zur beschriebenen Adressatengruppe
Relevanz verdeutlichen	Prompts zu relevanten Inhalten geben Direkte Ansprache des Adressaten	Das Video betont, a) warum das Erklärte wichtig für die Adressatengruppe ist und b) gibt Prompts zu besonders wichtigen Teilen Das Video involviert die Adressaten durch Handlungsaufforderungen und direkte Ansprache (statt unpersönlichem Passiv)
Struktur geben	Regel-Beispiel oder Beispiel-Regel Zusammenfassungen geben	Wenn Fachwissen das Lernziel ist, wird eine Regel-Beispiel-Struktur bevorzugt, bei Routinen eine Beispiel-Regel-Struktur Das Video fasst die wesentlichen Aspekte zusammen
Präzise und kohärent erklären	Exkurse vermeiden Hohe Kohärenz des Gesagten	Das Video fokussiert auf die Kernidee, vermeidet Exkurse und hält den cognitive load gering. Insbesondere verzichtet es auf zu viele Beispiele, Analogien, Modelle oder Zusammenfassungen Das Video verbindet Sätze durch Konnektoren, insbesondere „weil“
Konzepte und Prinzipien erklären	Neues und komplexes Prinzip als Thema	Das Video bezieht sich auf ein neues Prinzip, das zu komplex zur Selbsterklärung ist
In Unterrichtsgang einbetten	Anschließende Lernaufgaben	Das Video beschreibt eine Lernaufgabe, mit der das Erklärte selbst vertieft werden kann

Tabelle 14: Kriterien für gute Erklärvideos nach Kulgemeyer (2020, S. 423)

A.2 Voice-over-Skript

Treffen 14. 11. 24

Legende:

- fette und graue **Wörter** sind die Beschreibung der Abschnitte, diese bitte **NICHT** einsprechen
- grün- sprechpausen, hier wird einen Szenenwechsel geben

Storyboard:

Was ist Feinstaub?

Was ist Feinstaub? Wie entsteht er, und welche Auswirkungen hat er auf den menschlichen Körper? Diese Fragen werden wir in diesem Video beantworten.

“Feinstaub” sind winzige und feine Partikel, die in der Luft schweben. Sie sind so winzig, dass sie eingeatmet werden können. Die Feinstaubpartikel werden nach Größen kategorisiert:

- PM 10, sind Partikel mit einem Durchmesser von weniger als 10 Mikrometer.
- PM_{2,5} sind Partikel, die kleiner als 2,5µm sind.

Die kleineren Partikel, also PM_{2,5}, bleiben länger in der Luft und können weitere Entfernungen zurücklegen als PM₁₀.

Übrigens steht PM für *particulate matter*, was auf Englisch Feinstaub bedeutet.

Wo kommt Feinstaub vor?

An einer stark befahrenen Straße ist die Feinstaubbelastung höher (Pause) als in einem Park am Stadtrand.

Aber warum ist das so?

Um das zu verstehen, schauen wir uns an, wie Feinstaub entsteht.

Wie entsteht Feinstaub?

Die Entstehung von Feinstaub lässt sich in zwei Weisen unterteilen.

Primärer Feinstaub: sind Staubpartikel, die von einer Quelle direkt in die Luft abgegeben werden, wie zum Beispiel durch Verbrennungsprozesse, wie Autos, (Pause) Holzöfen (Pause) oder beim Kochen.

Sekundärer Feinstaub entsteht aus gasförmigen Stoffen, wie Schwefel- und Stickoxiden oder Ammoniak, die durch chemische Reaktionen in der Atmosphäre Feinstaubpartikel bilden.

Warum ist Feinstaub gefährlich?

Feinstaubpartikel sind so klein, dass wir sie mit bloßem Auge nicht sehen können. Beim Einatmen gelangen sie in unseren Körper.

- Partikel, die größer als 5µm sind, werden von gesunden Menschen meist beim Ausatmen wieder ausgeschieden.
- Kleinere Partikel können jedoch in die Bronchien und Bronchiolen vordringen, (Pause)
- und die Partikel, die kleiner als 2,5 µm sind, können in die Lungenbläschen (Alveolen) gelangen (Pause) und dort viele Jahre verbleiben.

Gesundheitliche Auswirkungen:

Hier können die Feinstaubpartikel über die Schleimhautgewebe in die Lymphknoten oder sogar ins Blut gelangen.

Das kann vorübergehend die Lungenfunktion negativ beeinträchtigen, Lungenentzündungen oder Herzrhythmusstörungen auslösen. Besonders gefährdet sind Menschen mit bestehenden Lungen- oder Herzerkrankungen oder auch allergischen Reaktionen, da Feinstaub ihre Beschwerden verschlimmern kann.

Klimatische Auswirkungen:

Feinstaub schadet nicht nur der Gesundheit, sondern auch dem Klima. Die Feinstaubpartikel können zum Beispiel durch Wind über weite Strecken transportiert und verbreitet werden und sich sogar an den Polen oder in Gletschern ablagern. Dort reduzieren sie die sogenannte Albedo, (Pause) das Reflexionsvermögen der Eisflächen. Dadurch wird weniger Sonnenlicht vom Gletscher reflektiert, stattdessen absorbiert er mehr Wärme. Das beschleunigt das Abschmelzen der Gletscher und trägt erheblich zum Klimawandel bei.

Wie wird Feinstaub gemessen?

In Berlin wird die Feinstaubbelastung an verschiedenen Orten stündlich gemessen. Die Messdaten können zum Beispiel auf der Webseite "Luftdaten.berlin.de." gelesen werden. Auf dieser Seite sind die Messstationen auf einer Stadtkarte angezeigt.

- Stationen "V" messen den Luftqualitätsindex (LQI) an verkehrsreichen Straßen
- Stationen "H", messen in Hinterhöfen
- und Stationen "S" am Stadtrand.

Unter der Karte sind Messdaten der einzelnen Stationen aufrufbar. Schauen wir zum Beispiel die Station an der Karl-Marx-Straße an.

Die x-Achse im Diagramm zeigt den jeweiligen Tag, und die y-Achse gibt die gemessene Feinstaubkonzentration in $\mu\text{g}/\text{m}^3$ an. Zum Beispiel betrug am 6.11.2024 die maximale Feinstaubbelastung der Partikel PM10 35 Mikrogramm pro Kubikmeter und PM2,5 2,7 Mikrogramm pro Kubikmeter.

Die Einheit Mikrogramm pro Kubikmeter gibt an, wie viele Feinstaubpartikel in einem Kubikmeter Luft enthalten sind.

Aber was bedeutet das genau?

Die Feinstaubkonzentration wird in Mikrogramm pro Kubikmeter angegeben. Ein Kubikmeter ist ein Würfel, der 1 Meter hoch, breit und tief ist. Angenommen, die Konzentration beträgt 50 Mikrogramm pro Kubikmeter, dann befinden sich in diesem gesamten Würfel Luft gerade einmal 50 Mikrogramm Feinstaub.

Jetzt stellen wir uns eine 1-Cent-Münze vor. Sie wiegt im Durchschnitt 2,3 Gramm – das entspricht 2.300.000 Mikrogramm. Um das Gewicht einer einzigen 1-Cent-Münze zu erreichen, bräuchte man die gesamte Feinstaubmenge aus 46.000 Kubikmetern Luft! Das ist

ungefähr das Volumen von 18 olympischen Schwimmbecken. Obwohl diese Mengen so winzig erscheinen, können sie erhebliche Auswirkungen auf unsere Gesundheit haben.

Definition Mittelwerte:

Die Feinstaubkonzentration schwankt im Laufe des Tages. Zum Beispiel gibt es in der Nacht weniger Verkehr als am Tag. Daher kann die Feinstaubkonzentration am Morgen kleiner sein als am Nachmittag, wo Kraftfahrzeuge den abgelagerten Feinstaub wieder aufwirbeln und durch den Verkehr neuer entsteht.

Aus diesem Grund wird die Feinstaubkonzentration stündlich gemessen und am Ende des Tages ein Durchschnittswert berechnet. Dieser Wert wird *Tagesmittelwert* genannt. Er zeigt uns, wie stark die Luft an diesem Tag im Durchschnitt verschmutzt war.

Über ein Jahr hinweg werden die Tagesmittelwerte gesammelt. Deren Durchschnitt ergibt den *Jahresmittelwert*, (*Pause*) er beschreibt die durchschnittliche Luftqualität in dem Jahr.

Grenzwertempfehlung:

Feinstaub kann den menschlichen Körper schädigen – wie sehr, hängt von der Konzentrationsmenge und der Dauer der Belastung ab. Um die gesundheitlichen Risiken zu minimieren, empfiehlt die Weltgesundheitsorganisation, abgekürzt WHO, folgende Grenzwerte:

- Für PM10:
 - soll der Tagesmittelwert 45 Mikrogramm pro Kubikmeter nicht überschreiten. Wobei an maximal 18 Tagen im Jahr dieser Grenzwert überschritten werden darf. Allerdings soll insgesamt der Jahresmittelwert: 20 Mikrogramm pro Kubikmeter nicht überschritten werden.
- Für PM2,5:
 - beträgt der Tagesgrenzwert: 25 Mikrogramm pro Kubikmeter
 - und der Jahresgrenzwert: 10 Mikrogramm pro Kubikmeter

Zusammenfassung:

Wir halten fest:

Feinstaub entsteht durch Verbrennungsprozesse und chemische Reaktionen in der Atmosphäre.

Die kleineren Partikel, PM 2,5, sind besonders gefährlich, da sie tief in die Lunge eindringen können und langfristig die Gesundheit schädigen.

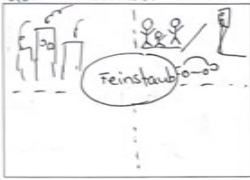
Neben den gesundheitlichen Risiken beeinflusst Feinstaub auch das Klima, etwa durch die Beschleunigung des Gletscherschmelzens.

Um diese gesundheitlichen und klimatischen Folgen zu reduzieren und um die Luftqualität zu verbessern, empfiehlt die WHO Grenzwerte.

A.3 Storyboard

PERSONA:


Bilder nacheinander einblenden



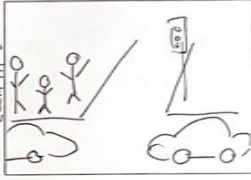
Was ist Feinstaub?
Wie entsteht er und ...

USER STORY/SCENARIO:

Karl-Mark Strafe




... welche Auswirkungen hat er auf den menschlichen Körper?



Die Fragen werden wir in diesem Video beantworten.

Storyboard Panel 4:

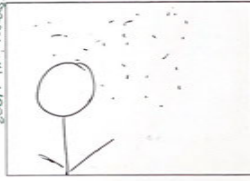
Zoom in: Monolog



Wie bereits im Name enthalten ist, ist Feinstaub sehr feine und kleine Luftpartikel.

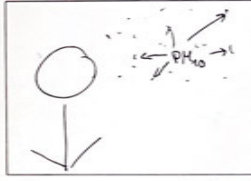
Storyboard Panel 5:

Zoom in: Wirkung auf Person



Die Feinstaubarten werden in verschiedene Größenkategorien eingeteilt.

Storyboard Panel 6:

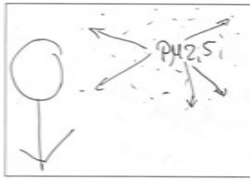


PM₁₀ Das sind Feinstaubpartikel, die einen Durchmesser kleiner als 10 µm haben.

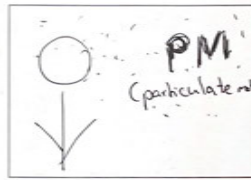
PAGE # 1 PROJECT/TEAM: DATE: STORYBOARD NINGROUP.COM

PERSONA:

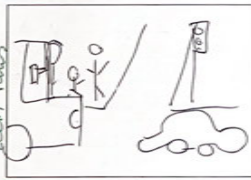
USER STORY/SCENARIO:



PM_{2,5} sind Partikel, die kleiner als 2,5 µm, wobei der feineren Feinstaubpartikel, also PM_{2,5} länger in der Luft verbleiben als PM₁₀.



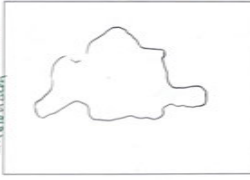
PM ist die englische Abkürzung für "particulate matter", was auf Deutsch "Feinstaub" bedeutet.



An einer stark befahrenen Straße ist die Feinstaubbelastung hoch.

Storyboard Panel 4:

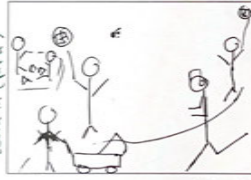
Einfluss von Verkehr auf Feinstaubbelastung



Im Vergleich zu einem grünen Teil am Stadtrand...

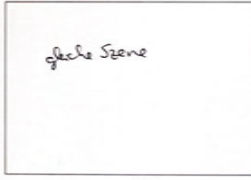
Storyboard Panel 5:

Zoom in (Rück)



... ist die Feinstaubbelastung geringer

Storyboard Panel 6:



Was ist das so? Um diese Frage beantworten zu können, müssen wir wissen wie Feinstaub entsteht.

PAGE # 2 PROJECT/TEAM: DATE: STORYBOARD NINGROUP.COM

PERSONA:

USER STORY/SCENARIO:

Entstehung von Feinstaub

primär sekundär

Die Entstehung von Feinstaub kann in primäre und sekundäre Entstehungsweise unterteilt werden.

primäre

Feinstaub

Der primäre Feinstaub sind Staubpartikel, die von einer Quelle direkt in die Luft abgegeben wird.

Einblenden

wie zum Beispiel Verbrennungsprozessen von Kraftfahrzeuge,...

Letzte Szene oben rechts platzen und neue Szene einblenden

Stoßsenne Holzofen Szene

... Holzofen ...

... oder auch beim kochen.

Stoßsenne Holzofen Szene

Leichen Szene.

PAGE # 3 PROJECT/TEAM: DATE: STORYBOARD NINGROUP.COM

PERSONA:

USER STORY/SCENARIO:

sekundär

Der sekundäre Feinstaub sind gasförmige Vorläufersubstanzen, die bedeutet...

Pflanzenschutzmittel

... sie entsteht durch komplexe chemische Reaktionen.

gleiche Szene

Diese sind zum Beispiel Schwefeloxid, Stickoxid und Ammoniak.

Szene "Vor- und Nach-Straße"

PM10 und PM25 Feinstaubpartikel sind so klein...

Zoom in Personen, die joggen

... dass wir sie mit bloßem Auge nicht sehen können.

Somit gelangen sie beim Einatmen in unseren Körper.

PAGE # 4 PROJECT/TEAM: DATE: STORYBOARD NINGROUP.COM

PERSONA:

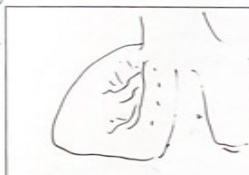
USER STORY/SCENARIO:



Zoom in: Verlauf weg von Feinstaubpartikel zeigen (Anatomie)



Zoom in: 1. Lungenflügel



Partikel größer als 5 μm scheiden gesunde Menschen meist beim Ausatmen wieder aus.

Kleinere Partikel können jedoch in die Bronchien...

... und die Bronchiolen vordringen.

Zoom in Bronchien



gleiche Szene

Zoom out



Die Partikel, die kleiner 2,5 μm sind, können in die Lungenbläschen gelangen...

...und dort viele Jahre verbleiben.

Hier können sie über die Schleimhautgäste in den Lymphknoten oder sogar ins Blut gelangen

PAGE # 5

PROJECT/TEAM:

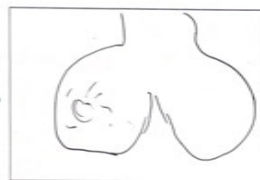
DATE:

STORYBOARD NINGROUP.COM

PERSONA:

USER STORY/SCENARIO:

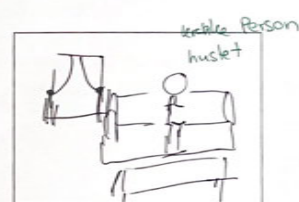
Darstellung entzündete Lunge



Das kann vorübergehend die Lungenfunktion negativ beeinträchtigen...

gleiche Szene

Lungenentzündungen oder Herzrhythmusstörungen auslösen.

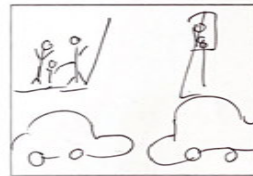


Besonders gefährdet sind Menschen mit bestehenden Lungen- oder Herzkrankungen...

Allergie durch Pollen

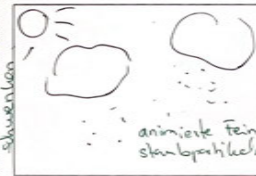


... oder auch Allergien, da Feinstaub ihre Beschwerden verschlimmern kann.



Feinstaub wickelt sich nicht nur negativ auf unsere Gesundheit aus, sondern auch auf das Klima.

Kommt nach oben (Himmel) streuen



Die feinen Feinstaubpartikel werden durch Wind und andere Faktoren über weite Strecken transportiert...

PAGE # 6

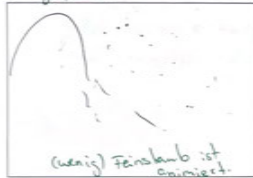
PROJECT/TEAM:

DATE:

STORYBOARD NINGROUP.COM

PERSONA:

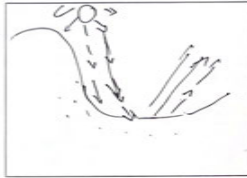
Gletscher



(wenig) Feinstaub ist
angeweht.

... und können sich sogar
an den Felsen und in
Gletschergebieten ablagern.

USER STORY/SCENARIO:

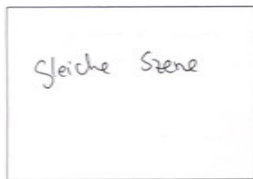


Dort verringern sie die
sogenannte Albedo, also
das Reflexionsvermögen der
Eisflächen.

mehr Feinstaub



Das bedeutet, dass weniger
Sonnenlicht vom Gletscher zurück
in die Atmosphäre reflektiert wird.

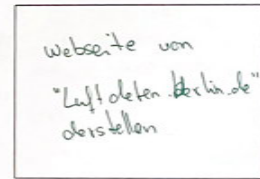


Gleiche Szene

Stattdessen wird mehr
Wärme absorbiert, was
das Abschmelzen...



... der Gletscher zusätzlich
beschleunigt und somit den
Klimawandel verstärkt.



Webseite von
"Luftdaten.berlin.de"
darstellen

In Berlin wird an verschiedenen
Orten die Feinstaubbelastung stündlich
gemessen. Zum Beispiel auf der
Webseite luftdaten.berlin.de.

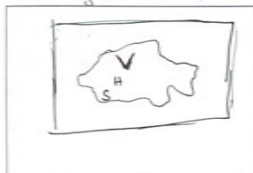
PAGE # 7 PROJECT/TEAM:

DATE:

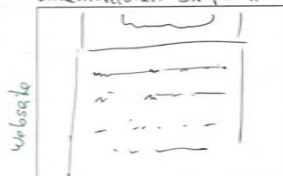
STORYBOARD NINGROUP.COM

PERSONA:

Abbildung der Seite Luftdaten.berlin.de



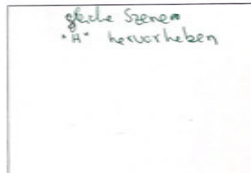
Auf einer Stadtkarte sind die
Messstationen verzeichnet.
Stationen verlassen die
Luftqualitätsindex an verschiedenen
verkehrsreichen Straßen...



Unter der Karte können die
Daten der einzelnen Messstationen
eingesehen werden. Zum Bsp. an
der Karl-Marx-Straße.

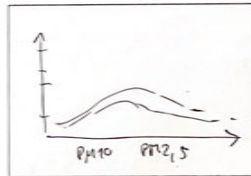
PAGE # 8 PROJECT/TEAM:

USER STORY/SCENARIO:

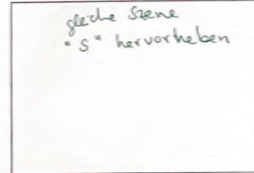


gleiche Szene
"H" hervorheben

... Stationen "H"
messen ~~so~~ in Hinterhöfen...

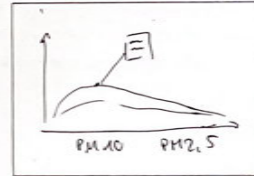


Die x-Achse stellt den Tag
dar und die y-Achse gibt die gemessene
Feinstaubkonzentration in $\mu\text{g}/\text{m}^3$



gleiche Szene
"S" hervorheben

... und Stationen "S" an
Stadttrand.

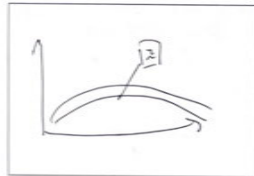


Zum Bsp. betrug am
6.11.2024 die maximale
Feinstaubbelastung der Parkhal
PM10 $35\mu\text{g}/\text{m}^3$...

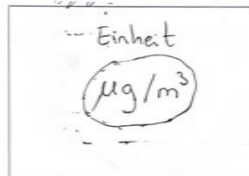
STORYBOARD NINGROUP.COM

PERSONA:

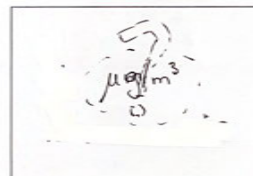
USER STORY/SCENARIO:



und PM_{2.5} 2,7 µg/m³ dargestellt.



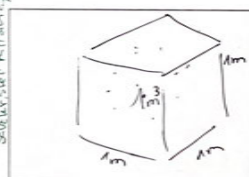
Die Einheit µg/m³ zeigt, wie viele Feinstaubpartikel in einem Kubikmeter Luft enthalten ist.



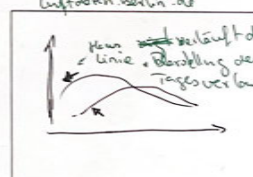
Aber was bedeutet das genau?



Stellen wir uns einen kleinen Salzstapel vor, der 100 µg wiegt.



Wenn wir dieses Salz gleichmäßig in einem Würfel mit einer Kantenlänge von 1 Meter, also 1 m³ verteilen, dann hätten wir eine Konzentration von 100 µg/m³.



Über den Tag verteilt kann die Feinstaubbelastung schwanken.

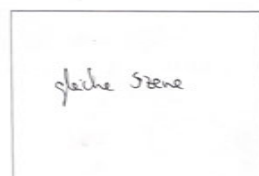
PAGE # 9

PROJECT/TEAM:

STORYBOARD NINGROUP.COM

PERSONA:

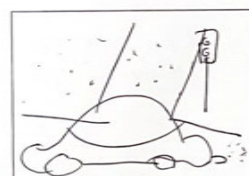
USER STORY/SCENARIO:



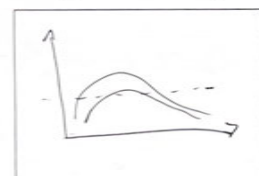
zum Beispiel kann sie am Nachmittag höher sein als am Morgen, da in der Nacht kaum bis gar kein Verkehr fahr...



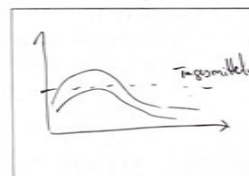
und dadurch der Feinstaub sich auf dem Boden ablagert.



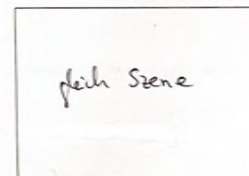
Über den Tag kann dieser Feinstaub aufgewirbelt werden und zusätzlich durch den Verkehr neu entstehen.



Aus diesem Grund wird der Feinstaubkonzentration stündlich gemessen und anschließend der Durchschnitt aller 24 Std. berechnet.



Am Ende des Tages berechnet man den Durchschnitt aller 24 Std. das nennt man den Tagesmittelwert.



Es zeigt uns, wie stark die Luft an diesem Tag im Durchschnitt verschmutzt war.

PAGE # 10

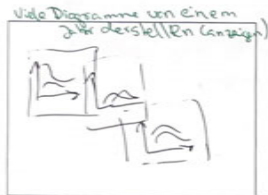
PROJECT/TEAM:

DATE:

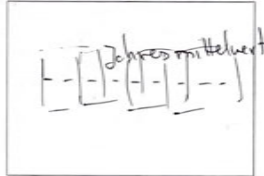
STORYBOARD NINGROUP.COM

PERSONA:

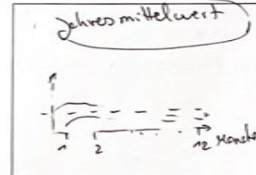
USER STORY/SCENARIO:



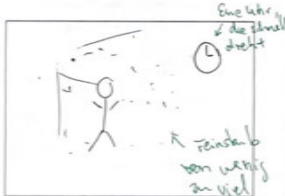
Über ein ganzes Jahr hinweg werden viele Tagesmittelwerte gemittelt.



Der Durchschnitt aller Tagesmittelwerte eines Jahres wird Jahresmittelwert genannt.



Es hilft uns zu verstehen, wie gut oder schlecht die Luftqualität im gesamten Jahr war.



Feinstaub kann den menschlichen Körper schädigen, wie sehr, hängt von der Konzentration und der Dauer der Belastung ab.



Deshalb gibt es Grenzwerte, die nicht überschritten werden sollen.



Die Weltgesundheitsorganisation, abgekürzt als WHO empfiehlt folgende Grenzwerte

PAGE # 11 PROJECT/TEAM:

DATE:

STORYBOARD NINGROUP.COM

PERSONA:

USER STORY/SCENARIO:

WHO		
Zeit	Mittelwert	Grenzwert
PM10	Tages	45 $\mu\text{g}/\text{m}^3$

Für PM10:

Tagesgrenzwert: 45 $\mu\text{g}/\text{m}^3$, an höchstens 18 Tagen im Jahr darf es überschritten werden

gleiche Szene		
---------------	--	--

Die WHO möchte diese Grenzwerte ab 2030 weiter senken, und die Luftqualität zu verbessern: ...

PAGE # 12 PROJECT/TEAM:

WHO		
Zeit	Mittelwert	Grenzwert
PM10	Tages	45 $\mu\text{g}/\text{m}^3$
	Jahres	20 $\mu\text{g}/\text{m}^3$

Jahresgrenzwert: 20 $\mu\text{g}/\text{m}^3$

WHO		
Zeit	Mittelwert	Grenzwert
PM10	Tages	45 $\mu\text{g}/\text{m}^3$
PM2.5	Jahres	15 $\mu\text{g}/\text{m}^3$

PM10: Der Tagesgrenzwert bleibt 45 $\mu\text{g}/\text{m}^3$, aber es soll nur noch an 3 Tagen im Jahr überschritten gehen dürfen.

DATE:

STORYBOARD NINGROUP.COM

WHO		
Zeit	Mittelwert	Grenzwert
PM10	Tages	45 $\mu\text{g}/\text{m}^3$
PM2.5	Jahres	10 $\mu\text{g}/\text{m}^3$

Für PM2.5: Tagesgrenzwert 25 $\mu\text{g}/\text{m}^3$
Jahresgrenzwert 10 $\mu\text{g}/\text{m}^3$

WHO		
Zeit	Mittelwert	Grenzwert
PM10	Tages	45 $\mu\text{g}/\text{m}^3$
PM2.5	Jahres	15 $\mu\text{g}/\text{m}^3$

Der Jahresgrenzwert soll auf 15 $\mu\text{g}/\text{m}^3$ gesenkt werden.

PERSONA:

USER STORY/SCENARIO:

Ab Jahr 2030 (WHO)

PM ₁₀	Tagesgrenze	5 µg/m³
PM _{2.5}	Jahresgrenze	15 µg/m³

PM_{2.5}: Ein neuer Tagesgrenzwert von 5 µg/m³ und ein Jahresgrenzwert von 15 µg/m³ sind geplant.

Szenen aus der Seite 5

Die Kleinen Asthmatiker sind besonders gefährdet, da sie tief in die Lunge eindringen...

Szene 6 von Seite 3

Wir helfen bei:
Feinstaub entsteht durch Verbrennungsprozesse oder...

Szene 2 von Seite 4

...chemische Reaktionen in der Atmosphäre.

Szene 3 Seite 6

... und die Gesundheit schädigen können.

Szene 3 Seite 6

Kurzfristig führen hohe Feinstaubkonzentrationen zu Beschwerden wie Husten, Atemnot oder Augenreizungen, während...

DATE:

PAGE # 13 PROJECT/TEAM:

STORYBOARD NINGROUP.COM

PERSONA:

USER STORY/SCENARIO:

... eine langfristige Belastung des
Risiko für chronische Erkrankungen
erhöht

Darüber hinaus beeinflusst
Feinstaub das Klima, etwa
durch Abschmelzung von
Gletschern.

Um diese gesundheitlichen
und klimatischen Folgen
zu reduzieren, empfiehlt die
WHO strengere Grenzwerte.

PAGE # 14

PROJECT/TEAM:

DATE:

STORYBOARD HINGROUP.COM

Abbildung 185 Storyboard

B Messinstrument: Fragebogen

Feedback

Liebe Humboldt-Explorers,

im Rahmen meiner **Masterarbeit** möchte ich **eure intrinsische Motivation** im Zusammenhang mit der **Lerneinheit** zum Thema **Feinstaub** besser verstehen. Mit dem Ausfüllen des folgenden Fragebogens unterstützt ihr nicht nur meine Masterarbeit, sondern helft auch dabei, das Projekt *Humboldt-Explorers* weiterzuentwickeln und die Lerneinheit für zukünftige Schülerinnen und Schüler zu verbessern.

Eure Teilnahme ist selbstverständlich **freiwillig** und eure Angaben werden **anonym** erfasst. Es ist nicht möglich, eure Antworten einer bestimmten Person zuzuordnen. Deshalb möchte ich euch bitten, die Fragen so ehrlich wie möglich zu beantworten.

Vielen Dank, dass ihr mich bei meiner Arbeit unterstützt! Ich schätze eure Hilfe sehr.

Ich besuche die Klasse:

[Wähle deine Klassenstufe aus.](#)

Wie war euer Gesamteindruck heute?

- ☐ 😞
- ☐ 😐
- ☐ 😊
- ☐ 😄
- ☐ 😁

☐ keine Angabe

Die Lerneinheit "Feinstaub" hat mir Spaß gemacht.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich fand die Lerneinheit sehr interessant.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich habe mich bei der Lerneinheit sehr angestrengt.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Es war mir wichtig, bei den Quizaufgaben gut abzuschneiden.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich finde die Lerneinheit "Feinstaub" langweilig.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich habe mich nicht sehr angestrengt, um bei der Lerneinheit gut abzuschneiden.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich bin überzeugt, dass die Lerneinheit "Feinstaub" sehr wertvoll für mich ist.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich wäre bereit, die Lerneinheit noch einmal zu durchlaufen, weil sie für mich wertvoll ist.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich denke, dass die Lerneinheit "Feinstaub" wichtig ist, weil sie alltagsrelevant ist.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

Ich denke, dass die Lerneinheit "Feinstaub" mir helfen kann, verantwortungsbewusster mit meiner Gesundheit umzugehen.

- ☐ 1 – stimmt gar nicht
- ☐ 2 – stimmt wenig
- ☐ 3 – stimmt ein bisschen
- ☐ 4 – stimmt teils-teils
- ☐ 5 – stimmt ziemlich
- ☐ 6 – stimmt völlig

<p>Ich habe nicht sehr viel Energie in die Durchführung der Lerneinheit "Feinstaub" aufgewendet.</p> <p><input type="radio"/> 1 – stimmt gar nicht</p> <p><input type="radio"/> 2 – stimmt wenig</p> <p><input type="radio"/> 3 – stimmt ein bisschen</p> <p><input type="radio"/> 4 – stimmt teils-teils</p> <p><input type="radio"/> 5 – stimmt ziemlich</p> <p><input type="radio"/> 6 – stimmt völlig</p>	<p>Ich denke, dass die Lerneinheit "Feinstaub" hilfreich ist, um ein besseres Bewusstsein für Umweltschutz zu entwickeln.</p> <p><input type="radio"/> 1 – stimmt gar nicht</p> <p><input type="radio"/> 2 – stimmt wenig</p> <p><input type="radio"/> 3 – stimmt ein bisschen</p> <p><input type="radio"/> 4 – stimmt teils-teils</p> <p><input type="radio"/> 5 – stimmt ziemlich</p> <p><input type="radio"/> 6 – stimmt völlig</p>
<p>Die Lerneinheit war unterhaltsam.</p> <p><input type="radio"/> 1 – stimmt gar nicht</p> <p><input type="radio"/> 2 – stimmt wenig</p> <p><input type="radio"/> 3 – stimmt ein bisschen</p> <p><input type="radio"/> 4 – stimmt teils-teils</p> <p><input type="radio"/> 5 – stimmt ziemlich</p> <p><input type="radio"/> 6 – stimmt völlig</p>	<p>Die Lerneinheit "Feinstaub" konnte meine Aufmerksamkeit überhaupt nicht aufrecht halten.</p> <p><input type="radio"/> 1 – stimmt gar nicht</p> <p><input type="radio"/> 2 – stimmt wenig</p> <p><input type="radio"/> 3 – stimmt ein bisschen</p> <p><input type="radio"/> 4 – stimmt teils-teils</p> <p><input type="radio"/> 5 – stimmt ziemlich</p> <p><input type="radio"/> 6 – stimmt völlig</p>
<p>Was hat euch besonders gut gefallen?</p> <div></div>	<p>Was hat euch nicht gefallen?</p> <div></div>
<div>Absenden</div>	

Abbildung 19: Erhebungsmessinstrument: Fragebogen

C Ergebnisse der quantitativen Analyse

C.1 Validitäts- und Reliabilitätsprüfung

Items (gekürzt)	Faktorladungen		
	Faktor ₁	Faktor ₂	Faktor ₃
„... hat mir Spaß gemacht.“	0,764	-0,145	0,051
„... sehr interessant.“	0,688	0,094	-0,091
„... langweilig.“	0,755	-0,006	-0,114
„... war unterhaltsam.“	0,689	-0,122	0,135
„... Aufmerksamkeit nicht aufrecht ...“	0,392	-0,344	-0,365
„... sehr angestrengt.“	0,590	0,310	0,172
„... wichtig, gut abzuschneiden.“	0,231	0,114	-0,281
„... nicht sehr angestrengt ...“	0,472	0,559	-0,214
„... wenig Energie aufgewendet.“	0,492	0,369	-0,290
„... sehr wertvoll für mich.“	0,107	0,171	0,637
„... noch einmal durchlaufen ...“	0,074	0,083	0,957
„... alltagsrelevant.“	-0,163	0,735	0,174
„... hilft beim Gesundheitsbewusstsein.“	0,067	0,773	0,124
„... hilfreich für Umweltbewusstsein.“	0,024	0,843	0,010

Tabelle 15: Faktorladung einzelner Items der dreifaktoriellen Lösung

Analysewerte	Faktor ₁	Faktor ₂	Faktor ₃
Eigenwert	3,35	2,61	1,60
Anteil erklärte Varianz	0,24	0,19	0,11
Kumulative erklärte Varianz	0,24	0,43	0,54
Varianzanteil innerhalb der Faktoren	0,44	0,34	0,21
Kumulative Varianzanteile	0,44	0,79	1,00

Tabelle 16: Rotierte Ladungsmatrix der dreifaktoriellen Lösung

Analysewerte	Faktor ₁	Faktor ₂	Faktor ₃
Eigenwert	3.17	2.60	1.78
Anteil erklärte Varianz	0,23	0,19	0,13
Kumulative erklärte Varianz	0,23	0,41	0,54
Varianzanteil innerhalb der Faktoren	0,42	0,34	0,24
Kumulative Varianzanteile	0,42	0,76	1,00

Tabelle 17: Varianzaufklärung der dreifaktoriellen Lösung

Items (gekürzt)	Faktorladungen			
	Faktor ₁	Faktor ₂	Faktor ₃	Faktor ₄
„... hat mir Spaß gemacht.“	0,714	-0,190	0,107	0,180
„... sehr interessant.“	0,773	0,113	-0,126	0,056
„... langweilig.“	0,803	-0,013	-0,119	0,109
„... war unterhaltsam.“	0,667	-0,142	0,157	0,094
„... Aufmerksamkeit nicht aufrecht ...“	0,308	-0,427	-0,258	0,270
„... sehr angestrengt.“	0,571	0,274	0,208	0,172
„... wichtig, gut abzuschneiden.“	0,010	-0,082	-0,057	0,670
„... nicht sehr angestrengt ...“	0,317	0,406	-0,016	0,631
„... wenig Energie aufgewendet.“	0,313	0,197	-0,072	0,693
„... sehr wertvoll für mich.“	0,058	0,148	0,682	-0,007
„... noch einmal durchlaufen ...“	0,036	0,078	0,996	-0,145
„... alltagsrelevant.“	-0,102	0,792	0,111	-0,047
„... hilft beim Gesundheitsbewusstsein.“	0,048	0,732	0,156	0,199
„... hilfreich für Umweltbewusstsein.“	0,056	0,859	-0,009	0,128

Tabelle 18: Faktorladung einzelnen Items der vierfaktoriellen Lösung

Analysewerte	Faktor ₁	Faktor ₂	Faktor ₃	Faktor ₄
Eigenwert	2,84	2,47	1,68	1,56
Anteil erklärte Varianz	0,20	0,18	0,12	0,11
Kumulative erklärte Varianz	0,20	0,38	0,50	0,61
Varianzanteil innerhalb der Faktoren	0,33	0,29	0,20	0,18
Kumulative Varianzanteile	0,33	0,62	0,82	1,00

Tabelle 19: Rotierte Ladungsmatrix der vierfaktoriellen Lösung

Analysewerte	Faktor ₁	Faktor ₂	Faktor ₃	Faktor ₄
Eigenwert	3,4	2,64	1,66	0,82
Anteil erklärte Varianz	0,24	0,19	0,12	0,06
Kumulative erklärte Varianz	0,24	0,43	0,55	0,61
Varianzanteil innerhalb der Faktoren	0,40	0,31	0,19	0,10
Kumulative Varianzanteile	0,40	0,71	0,90	1,00

Tabelle 20: Varianzaufklärung der vierfaktoriellen Lösung

C.2 Levene-Test

Subskala	F(1,44)	p-Wert	Ergebnis
Interesse	0,95	0,34	n.s.
Anstrengung	1,08	0,31	n.s.
Nützlichkeit	0,50	0,49	n.s.

Tabelle 21: Ergebnisse des Levene-Tests zur Überprüfung der Varianzhomogenität

C.3 Mittelwertaggregation und deskriptiven Statistik

Interesse											
Klasse	N _{ohne}	M _{ohne}	SD _{ohne}	\tilde{x}_{ohne}	N _{mit}	M _{mit}	SD _{mit}	\tilde{x}_{mit}	t(44)	p - Wert	g(Hedges)
9	6	3,87	1,06	3,70	6	3,47	0,96	3,30	-0,69	0,508	-0,37
10	5	4,35	0,91	4,40	7	4,34	1,31	4,60	-0,006	0,996	0,00301
11	5	3,64	1,07	3,60	6	4,37	0,85	4,20	-1,26	0,241	-0,695
12	5	4,47	0,35	4,50	5	3,60	1,11	3,00	1,82	0,102	1,01
Gesamt	22	4,09	0,89	4,30	24	3,98	1,10	4,10	-0,38	0,705	-0,11

Tabelle 22: Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala *Interesse*

Anstrengung											
Klasse	N _{ohne}	M _{ohne}	SD _{ohne}	\tilde{x}_{ohne}	N _{mit}	M _{mit}	SD _{mit}	\tilde{x}_{mit}	t(44)	p - Wert	g(Hedges)
9	6	2,88	0,70	2,63	6	3,71	0,89	3,63	1,80	0,101	0,96
10	5	3,15	0,49	3,50	7	3,96	1,48	4,00	-1,17	0,269	-0,632
11	5	2,70	0,74	3,00	6	4,04	0,81	4,25	-2,84	0,019	-1,57
12	6	3,75	0,94	3,63	5	3,90	1,19	3,50	-0,234	0,820	0,130
Gesamt	22	3,14	0,80	3,25	24	3,91	1,07	3,88	2,73	0,009	0,79

Tabelle 23: Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala *Anstrengung*

Nützlichkeit											
Klasse	N _{ohne}	M _{ohne}	SD _{ohne}	\tilde{x}_{ohne}	N _{mit}	M _{mit}	SD _{mit}	\tilde{x}_{mit}	t(44)	p - Wert	g(Hedges)
9	6	3,80	0,62	4,00	6	3,97	1,16	4,00	0,31	0,762	0,17
10	5	4,52	1,58	4,60	7	4,57	1,18	4,00	-0,0893	0,931	-0,0482
11	5	3,48	0,58	4,00	6	4,27	1,26	4,20	-0,92	0,382	-0,509
12	6	3,57	1,04	3,70	5	3,32	0,70	3,40	0,450	0,663	0,249
Gesamt	22	3,83	1,03	4,00	24	4,08	1,146	4,00	0,80	0,429	0,23

Tabelle 24: Mittelwerte, ihre Standardabweichungen und Median der jeweiligen Gruppen und Klassenstufe der Subskala *Nützlichkeit*

D Kodierung der offenen Fragen

D.1 Frage: Was hat euch besonders gut gefallen?

Gruppe Ohne-Diversität: <i>Was hat euch besonders gut gefallen?</i>			
Nr.	Schüler:innenaussage	mögliche Unterkategorie	Zugeordnete Hauptkategorie
1	Wie simpel das Video ist und trotzdem informativ.	Verständlichkeit und informativ des Erklärvideos	Darstellungsform der LE
2	Das erklär Video war kreativ und hilfreich gestaltet. Man konnte alle Informationen für die Fragen gut heraus hören.	Hilfreiches und informatives Video	Darstellungsform der LE
3	Der Quiz	Quiz	Interaktive Elemente
4	Verständliche-s Video und Aufgabenstellungen.	Verständliches Video und Aufgaben	Darstellungsform der LE, Interaktive Elemente
5	Es ist für den Alltag relevant und ich finde es gut, dass man mehr Bewusstsein über Feinstaub dadurch haben kann	Alltagsrelevanz, Umweltbewusst, Lernzuwachs	Lernzuwachs, Nützlichkeit
6	Praktische Beispiele benutzt werden	Alltagsrelevanz	Darstellungsform der LE
7	Das Video ist nett.	Video	Darstellungsform der LE
8	Man hat in kürzester Zeit viel gelernt.	Wissenserwerb	Lernzuwachs
9	Das Thema ist ganz interessant und ziemlich wichtig	Interesse zum Thema und Wichtigkeit des Themas	Nützlichkeit
10	Die Chat gpt Generierten Bilder	KI Bilder	Darstellungsform der LE
11	Die Websites mit den messwerten	Arbeiten mit Daten	Interaktive Elemente
12	Das erklär Video und die Erklärung von den verschiedenen Erkrankungen	Erklärvideo	Darstellungsform der LE
13	Learn about dust	Wissenserwerb	Lernzuwachs
14	Ich finde gut das wir einfach unsere Aufgaben machen konnten	Selbstständiges Arbeiten	Interaktive Elemente
15	Es it cool hab viele Sachen gelernt	Bewertung und Wissenserwerb	Lernzuwachs, Allgemein Bewertung

Tabelle 25: Kodierung der Antworten der Ohne-Diversität-Gruppe auf die Frage „*Was hat euch besonders gut gefallen?*“

Gruppe Diversität: Was hat euch besonders gut gefallen?			
Nr.	Schüler:innenaussage	mögliche Unterkategorie	Zugeordnete Hauptkategorie
1	Gute fragen und interessante daten	Quizaufgaben	Interaktive Elemente
2	Das Video	Video	Darstellungsform der LE
3	Der verpasste Unterricht	kein Unterricht	Allgemeine Bewertung
4	Das Quiz war eine gute Überprüfung, ob man das Video verstanden hatte.	Verständnis-überprüfung	Interaktive Elemente
5	Ich fand es cool wie umfangreich das quiz war und es war sehr spannend zu bearbeiten! Ich habe noch nie ein Quiz wie das vorherige beantwortet und finde dass alle Quizze sich als Beispiel das vorgegebene Quiz bearbeiten um wirklich zu wissen wie man richtig gute Quizze herstellt!Meiner Meinung hat es mir auch geholfen zu verstehen wie schlecht die Verschmutzung der Luft in [REDACTED] eigentlich ist! Ich wusste garnicht dass es so schädlich sein würde...	Quizgestaltung, Alltagbezug, Gesundheit	Interaktive Elemente, Nützlichkeit, Lernzuwachs, Interesse
6	Vielfalt an aufgaben Arten, nicht nur Auswahl Fragen Gestaltung der Quizaufgaben Interaktive Elemente Dass man die Antwort zu den Fragen direkt sehen konnte.	Darstellung der Antworten von Quizaufgaben	Interaktive Elemente
7	Dass es nicht nur ankreuzen war, sonder verschiedene Fragestellungen und verschiedene Antwortmöglichkeiten. Abwechslungsreich 👍	Aufgabenvielfalt	Interaktive Elemente
8	Die interaktive Fragen	Quiz	Interaktive Elemente
9	die Form / Aufbau des Quizzes - Nutzung des Websites	Darstellung der Quizaufgaben und Nutzung der Webseite	Interaktive Elemente
10	Das man auch selber Daten raussuchen sollte	Arbeiten mit Daten	Interaktive Elemente
11	Die kreative Gestaltung, die Verständlichkeit des Videos (das Video generell)	Darstellung des Videos	Darstellungsform der LE
12	Ich konnte mehr über Feinstaub lernen und mein Wissen verbessern	Wissenserweitern	Lernzuwachs
13	Über Staubt zu lernen	Wissenserwerb	Lernzuwachs
14	Ich bin ziemlich Neutral gegen über alles	neutrale Einschätzung	Allgemeine Bewertung
15	Es hat mir Spaß bereitet und ich habe mehr gelernt.	Freude am Lernen und Wissenszuwachs	Lernzuwachs, Interesse

Tabelle 26: Kodierung der Antworten der Diversität-Gruppe auf die Frage „Was hat euch besonders gut gefallen?“

Beide Gruppen: Was hat euch besonders gut gefallen?					
Kategorien	Ohne (n=15)	mit (n=16)	Ohne (%)	mit (%)	Beispiel Aussagen
Interesse	-	2	-	12,5	Es hat mir Spaß bereitet und ich habe mehr gelernt.
Allgemeine Bewertung	1	2	6,67	12,5	Ich bin ziemlich Neutral gegen über alles
Nützlichkeit	3	1	20,00	6,25	... Meiner Meinung hat es mir auch geholfen zu verstehen wie schlecht die Verschmutzung der Luft in [REDACTED] eigentlich ist! Ich wusste garnicht dass es so schädlich sein würde...
Lernzuwachs	4	4	26,67	25	Man hat in kürzester Zeit viel gelernt.
Interaktive Elemente	4	9	26,67	56,25	Vielfalt an aufgaben Arten, nicht nur Auswahl Fragen
Darstellungsform der LE	6	2	40	12,5	Die kreative Gestaltung, die Verständlichkeit des Videos (das Video generell)

Tabelle 27: Kategorien der Antworten beider Gruppe auf die Frage „Was hat euch besonders gut gefallen?“

D.2 Frage: Was hat euch nicht gefallen?

Gruppe Ohne-Diversität: <i>Was hat euch nicht gefallen?</i>			
Nr.	Schüler:innenaussage	mögliche Unterkatgorie	Zugeordnete Hauptkategorie
1	Die Darstellungen wie der alten Mann ist manchmal eher irrelevant.	Personendarstellung ist irritieren	Gestaltung der KI-Bilder
2	Das Ausrechnen	Anstrengung	Schwierigkeiten/Überforderung
3	Ich glaube nicht das Feinstaub zum Klimawandel beiträgt.	Skepisis gegenüber dem Inhalt	Inhalt/ Thema
4	vielleicht Unterricht interaktiver gestalten?	Sozialform	Didaktisches Umsetzung
5	Die Webseite war nicht geeignet, um effizient zu arbeiten. Die Beschreibung der Aufgaben macht es schwer, genau die richtige Angehensweise zu nutzen. Die Quelle war begrenzt und nicht differenziert auch nicht sehr ueberzeugend mit den KI-Bilder	Webseite-Fehler, Arbeitsauftrag unklar, Skepsis gegenüber Inhalt, nicht authentische KI-Bilder	Technische Probleme, Didaktische Umsetzung, Inhalt/Thema, Gestaltung der KI-Bilder
6	Die Webseite ist zum Teil kaputt.	Webseite-Fehler	Technische Probleme
7	Alles war in Ordnung	keine Kritik	Keine Kritik
8	Ich mochte nicht wie die Bilder so AI Generated aussahen	KI Bilder	Gestaltung der KI-Bilder
9	Zu viele fragen	Zu viele Fragen	Schwierigkeiten/Überforderung
10	Es waren zu viele Zahlen und ein bisschen zu sachlich	Überforderung, Kritik zur Tiefe des Inhalts	Schwierigkeiten/Überforderung, Inhalt/Thema
11	Nothing	Keine Kritik	Keine Kritik
12	Das ich viel lesen musste und fast alles falsch hatte	Überföderung	Schwierigkeiten/Überforderung
13	Nichts ich will nach Hause	Keine Kritik, Desinteresse	Keine Kritik, Desinteresse

Tabelle 28: Kodierung der Antworten der Ohne-Diversität-Gruppe auf die Frage „*Was hat euch nicht gefallen?*“

Gruppe Diversität: Was hat euch nicht gefallen?			
Nr.	Schüler:innenaussage	mögliche Unterkategorie	Zugeordnete Hauptkategorie
1	Nach Daten suchen	Arbeiten mit Daten	Schwierigkeiten/ Überforderung
2	ich habe so gut wie nichts neues gelernt, ich habe bereits alles gewusst (mit Ausnahme von der Einheitskonversion von 46 000 m ³ zu Schwimmbaden. Ausserdem war die Auswahl von KI-Fotos eine ausserst schlechte	kein Lernzuwachs, Gestaltung der KI-Bilder	Kein Lernzuwachs, Gestaltung der KI-Bilder
3	Nichts.	Keine Kritik	keine Kritik
4	KI soll keine Zweck zum lernen dienen	Kritik zu KI-Bilder	Gestaltung der KI-Bilder
5	KI im video, es fühlt sich nicht authentisch an.	Unauthentische KI-Bilder	Gestaltung der KI-Bilder
6	Beim Video passierten manchmal zu viele Dinge auf einmal und man konnte sich nicht auf etwas konzentrieren. Zu viel AI.	Überladenes Video, zu viele KI-Bilder	Schwierigkeiten/ Überforderung, Gestaltung der KI-Bilder
7	AI Bilder/ Videos - Website hat ein paar Probleme, z.B. manche Antworten sind nicht sehbar	KI-Bilder, Webseite-Fehler	Gestaltung der KI-Bilder, Technische Probleme
8	Das die Fragen teilweise nicht funktioniert haben	Technische Fehler	Technische Umsetzung
9	Manche Quizfragen waren schwer	Schwierigkeit bei den Aufgaben	Schwierigkeit/ Überforderung
10	Einzel Arbeit	Sozialform	Didaktische Umsetzung
11	Es war sehr leise und es hätte mehr Spaß gemacht, hätte man in der Gruppe gelernt.	Sozialform	Didaktische Umsetzung

Tabelle 29: Kodierung der Antworten der Diversität-Gruppe auf die Frage „Was hat euch nicht gefallen?“

Beide Gruppen: Was hat euch nicht gefallen?					
Kategorien	Ohne (n=13)	mit (n=11)	Ohne (%)	mit (%)	Beispiel Aussagen
Kein Lernzuwachs	-	1	-	9,09	ich habe so gut wie nichts neues gelernt, ich habe bereits alles gewusst (mit Ausnahme von der Einheitskonversion von 46 000 m ³ zu Schwimmbaden...
Desinteresse	1	-	7,69	-	... ich will nach Hause
Keine Kritik	3	1	23,08	9,09	Alles war in Ordnung
Gestaltung der KI-Bilder	3	5	23,08	45,45	KI im video, es fühlt sich nicht authentisch an.
Technische Probleme	2	2	15,38	18,18	Die Webseite ist zum Teil kaputt.
Didaktische Umsetzung	2	2	15,38	18,18	Es war sehr leise und es hätte mehr Spaß gemacht, hätte man in der Gruppe gelernt.
Inhalt/Thema	3	-	23,08	-	Ich glaube nicht das Feinstaub zum Klimawandel beiträgt.
Schwierigkeiten/ Überforderung	4	3	30,77	27,27	Das ich viel lesen musste und fast alles falsch hatte

Tabelle 30: Kategorien der Antworten der beiden Gruppe auf die Frage „Was hat euch nicht gefallen?“

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und noch nicht für andere Prüfungen eingereicht habe. Für diese Arbeit wurde Künstliche Intelligenz ausschließlich zur sprachlichen Unterstützung verwendet. Sämtliche Quellen einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen, Bilder sowie die Nutzung von Künstlicher Intelligenz für die Erstellung von Abbildungen, sind als solche kenntlich gemacht. Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung eingeleitet wird.

Ort, Datum

Unterschrift
(Passlida Saila)