
Themenheft 65: Tagungsband des 1. AK Mediendidaktik der DGfE-Sektion Medienpädagogik (MEDIDA24). Herausgegeben von Maria Klar, Josef Buchner, Barbara Getto, Marco Kalz und Michael Kerres

Unterrichtet vom Avatar?

Studierenden-Präferenzen bei Erklärvideos

Julia Jochim¹ , Cecilia Post¹ und Ricarda Bolten-Bühler¹

¹ Europäische Fernhochschule Hamburg

Zusammenfassung

Zur Produktion von Erklärvideos steht Hochschullehrenden im 21. Jahrhundert eine Vielzahl von Optionen zur Verfügung, vom Studiodreh über Animationssoftware bis hin zu KI-gestützten Lösungen für Stimme und Darsteller:innen. Der Beitrag untersucht mit einem Mixed-Methods-Ansatz die Auswirkungen unterschiedlicher Videoformate auf die wahrgenommene Lernwirksamkeit, das Engagement und das Interesse von Hochschulstudierenden und versucht, Formatpräferenzen zu identifizieren. Die Ergebnisse der quantitativen Erhebung zeigen mit einer Ausnahme keine signifikanten Unterschiede in Lernwirksamkeit, Engagement und Interesse sowie in der Bewertung der Darreichungsform. Die qualitative Erhebung hingegen zeigt Präferenzen für reale Darsteller:innen sowie Cartoon-Animationen; KI-generierte Stimmen und Avatare werden jedoch überwiegend nicht als negativ oder störend wahrgenommen. Zudem geht aus den Ergebnissen hervor, dass die Formatpräferenz vom Lehrinhalt abhängt: Während menschliche Akteur:innen für komplexe oder zwischenmenschliche Themen bevorzugt werden, eignen sich animierte Formate für allgemeine Inhalte. Die Ergebnisse bieten wichtige Impulse für die Videogestaltung im Hochschulkontext.

Taught by an Avatar? Student Preferences for Instructional Videos

Abstract

In the 21st century, educators in higher education may choose from a diverse array of tools for creating instructional videos, ranging from professionally produced studio recordings to animation software and AI-driven solutions featuring synthetic voices and avatars. This study employs a mixed-methods approach to explore how different video formats influence students' perceived learning effectiveness, engagement, and interest, with the goal of identifying format preferences. The quantitative analysis reveals no significant differences in these factors across most formats, with only one exception. However, qualitative findings indicate a preference for videos featuring real-life actors

and cartoon-style animations, although AI-generated voices and avatars are generally not perceived to be disruptive or distracting. Moreover, format preferences appear to be content-dependent: real actors are favoured for complex or interpersonal topics, whereas animated formats are preferred for more general content. These findings provide valuable insights for educational media creators, helping them to design more effective instructional videos for higher education.

1. Einleitung

Erklärvideos gehören zu den beliebtesten digitalen Lehrmedien im Hochschulumfeld. Sie werden von Studierenden vielfach als nützlich empfunden und stark nachgefragt (Busse und Bargel 2017; Findeisen, Horn, und Seifried 2019; Schaarschmidt, Albrecht, und Börner 2016; Zawacki-Richter, Hohlfeld, und Müskens 2014). Empirische Ergebnisse weisen darauf hin, dass Erklärvideos positive Effekte auf die Lernleistung haben (van der Meij und van der Meij 2014; Lloyd und Robertson 2012; Ali 2019) sowie Aufmerksamkeit und Engagement der Lernenden (Hartsell, Yuen, und Yuen 2006) unterstützen.

Im Jahr 2025 steht Lehrenden zur Videoproduktion eine Vielzahl von Optionen zur Verfügung; diese reichen vom herkömmlichen «Profi-Dreh» im Studio oder in Eigenregie mittels Computer- oder Handykamera über Animationssoftware bis hin zur KI-gestützten Produktion mit künstlichen Stimmen und Avataren, sodass menschliche Darsteller:innen nicht mehr zwingend erforderlich sind. Vielfach sind die Programme anwenderfreundlich und ermöglichen auch Nicht-Medienexpert:innen eine qualitativ hochwertige Umsetzung.

Offen ist, welche Umsetzungsformen bzw. Ästhetiken die Präferenzen von Studierenden am ehesten treffen sowie geeignet sind, Interesse zu wecken, die Aufmerksamkeit zu halten und so Inhalte effizient zu vermitteln. Insbesondere der Nutzung KI-generierter Stimmen und Figuren begegnen Lehrende mitunter mit Zurückhaltung; befürchtet wird eine negative Reaktion Lernender bedingt durch das Gefühl, mit Material «aus der Konserve» abgespeist zu werden, oder auch aufgrund des «Uncanny-Valley-Effekts», der Misstrauen bzw. Unbehagen gegenüber künstlichen Figuren beschreibt, die dem Menschen sehr ähnlich sind (Mori 2012).

Die vorliegende Studie untersucht die Formatpräferenzen Studierender sowie die Wirkung verschiedener Videoformate auf das Lernerlebnis. Dabei werden Realfilme, animierte Videos (Cartoon-Optik) und KI-Avatar-Videos miteinander verglichen. Durch die Analyse sollen geeignete Formate identifiziert und somit ein Beitrag zur Gestaltung effektiver Erklärvideos geleistet werden.

2. Theoretische Grundlagen

Die Studie stützt sich auf die Definition von Erklärvideos von Findeisen, Horn und Seifried, die das Genre als «eigenproduzierte, kurze Filme [beschreiben], in denen Inhalte, Konzepte und Zusammenhänge erklärt werden (Erklärvideos im engeren Sinne) oder Tätigkeiten und Prozesse demonstriert und kommentiert werden (Tutorial), jeweils mit der Intention, beim Betrachter ein Verständnis zu erreichen bzw. einen Lernprozess auszulösen» (Findeisen, Horn, und Seifried 2019, 18). Zentral für diese Definition ist eine Didaktisierung des Inhalts; charakterisiert sind Videoerklärungen durch Multimedialität, d. h. den «gleichzeitige[n] Einsatz von Audio und visuellen Gestaltungselementen» (Wolf 2015, 129).

Zu berücksichtigen sind bei der Gestaltung von Erklärvideos die Informationsverarbeitung der Lernenden sowie Faktoren, die das Lernen erleichtern oder behindern können. Einen geeigneten Bezugsrahmen stellt die Cognitive Load Theory (Sweller 1988; Sweller, Ayres, und Kalyuga 2011) dar, die zwischen drei Arten kognitiver Belastung beim Lernen unterscheidet: intrinsischer, extrinsischer und lernbezogener kognitiver Belastung. Die intrinsische kognitive Belastung bezieht sich auf die Komplexität des Lernmaterials selbst, die extrinsische ist bedingt durch die Darstellung und Gestaltung des Lernmaterials. Die lernbezogene kognitive Belastung wiederum bezieht sich auf die Belastung, die durch die geistigen Prozesse entsteht, die für das Verständnis der Inhalte erforderlich sind. Sweller et al. (1998) betonen, dass insbesondere die Begrenzungen des menschlichen Arbeitsgedächtnisses Implikationen für das Instruktionsdesign haben und zur Optimierung von Lernprozessen die extrinsische kognitive Belastung weitestmöglich zu reduzieren ist. Eine gezielte Erhöhung der lernbezogenen kognitiven Belastung wiederum kann den Lernprozess befördern. Daraus folgt für die Gestaltung von Erklärvideos, dass zur Minimierung der extrinsischen kognitiven Belastung die Inhalte klar, konsistent und ohne ablenkende oder störende Elemente präsentiert werden; die lernbezogene kognitive Belastung wiederum sollte gefördert werden (ebd.).

Die Cognitive Theory of Multimedia Learning (Mayer 2014) baut auf der Cognitive Load Theory auf. Im Zentrum der Theorie stehen drei Grundannahmen: visuelle und auditive Informationen werden in zwei verschiedenen Informationskanälen verarbeitet (dual-channel assumption); jeder von diesen hat nur eine begrenzte Aufnahmekapazität (limited-capacity assumption); aktives Lernen erfordert koordinierte kognitive Prozesse beim Lernen und die Strukturierung von Informationen (active processing assumption). Die Theorie definiert drei Instruktionsziele und mehrere Grundsätze zur Erleichterung der kognitiven Last, von denen einige für die vorliegende Untersuchung relevant sind: die Kombination von Text und Bild, der Verzicht auf überflüssige Informationen, die Nutzung von entweder Text oder grafischer

Darstellung im Bild, die gleichzeitige Präsentation von Erzählung und dazugehörigen Bildern, die Nutzung menschlicher Stimmen sowie menschlich anmutender Gestik für Figuren.

Da das Projekt eine hohe Praxisorientierung aufweist und die beiden Theorien hier anschlussfähig sind, fiel die Wahl auf diese Zugänge. Bei einer weiteren Betrachtung der Ergebnisse sollen darüber hinaus ergänzende theoretische Rahmungen einbezogen werden.

Im Kontext des Forschungsansatzes, in dessen Rahmen KI-generierte Stimmen und Avatare zum Einsatz kommen, ist zudem der Uncanny Valley-Effekt (Mori 2012) zu berücksichtigen. Diese Theorie besagt, dass die Affinität zu Objekten, die menschenähnlich wirken, in negative Gefühle umschlägt, sobald Objekte fast, aber nicht vollständig menschlich wirken (z. B. humanoide Roboter). Sie ist relevant für die Bewertung oder Prognose der Effekte von KI-generierten Figuren in Erklärfilmen. Ein mögliches Unbehagen bei Lernenden könnte ablenken und sich lernhemmend auswirken. Zu beachten ist, dass Folgeforschungen die Theorie nicht komplett bestätigen konnten und nicht jede nahezu menschliche Darstellung Unbehagen auslöst, sondern der Effekt nur unter bestimmten Bedingungen auftritt. Nachgewiesen werden konnte er u. a. bei inkonsistent gestalteten Figuren, bei denen bestimmte Merkmale stärker menschlich als andere erschienen (Kätsyri et al. 2015).

Diese theoretischen Ansätze sind zentral für die Interpretation der folgenden Untersuchungsergebnisse, da sie helfen können, Reaktionen und Einschätzungen Lernender zu erklären und einzuordnen.

3. Gestaltungselemente bei Erklärvideos: Forschungsansätze

Die Wirkung verschiedener Gestaltungselemente in Erklärvideos ist Gegenstand zahlreicher wissenschaftlicher Studien. Findeisen, Horn und Seifried (2019) werteten 24 empirische Studien aus. Als bedeutsam für eine positive Wirkung erwiesen sich der Einsatz interaktiver Elemente, die Videoperspektive (bei Demonstration einer Handlung), das Alter der erklärenden Person, die Videodauer sowie das Design (wahrgenommene Ästhetik). Plass et al. (2014) stellten fest, dass bestimmte Designelemente Emotionen und Lernleistung beeinflussen. Heidig, Müller und Reichelt (2015) ermittelten in einer Folgestudie keine Einflüsse objektiver Unterschiede in der Ästhetik, jedoch einen Einfluss der subjektiv wahrgenommenen Ästhetik sowie der Nutzerfreundlichkeit auf das emotionale Befinden und damit auf die intrinsische Motivation der Lernenden.

Das Forschungsfeld zur Wirkung von KI-generierten Stimmen und Figuren im Bildungskontext ist noch vergleichsweise jung; dennoch liegen bereits Studien vor, auf die aufgebaut werden kann. Kit et al. (2023) untersuchten die Wirkung KI-generierter Stimmen in Erklärvideos auf Studierende und stellten fest, dass die Teilnehmenden

zwar menschliche Stimmen bevorzugten, aber KI-generierte Stimmen nicht massgeblich negativ beurteilten, was die Autor:innen schlussfolgern liess, dass KI-Stimmen stärker im Bildungsumfeld zum Einsatz kommen könnten. Kim et al. (2022) stellten fest, dass Studierende einer KI-Stimme, die stark menschlich klang, eine höhere Glaubwürdigkeit attestierten als einer maschinenartigen Stimme.

Vallis et al. (2024) untersuchten die Wirkung von KI-Avataren auf Studierende im Wirtschaftsbereich und stellten weder stark positive noch stark negative Einstellungen fest; einige Proband:innen nahmen den Avatar überhaupt nicht als solchen wahr. Leiker et al. (2023) stellten in einer Studie, in der eine Gruppe erwachsener Lernender ein Video mit einem menschlichen Darsteller und eine Kontrollgruppe ein Video mit KI-Figur präsentiert bekam, keine qualitativen Unterschiede beim gefühlten Lernerlebnis fest. Dem steht die Studie von Netland et al. (2024) entgegen, in der die Teilnehmenden zwar gleiche Lernergebnisse für Videos mit menschlichem Darsteller und solche mit Avatar erzielten, bezüglich des Lernerlebnisses jedoch Videos mit menschlichen Darstellern bevorzugten. Eine weitere Studie stellte bei Videos mit menschlichem Darsteller ein höheres Engagement seitens der Lernenden fest als bei einem KI-generierten Darsteller, letzterer wurde als ablenkend, unangenehm bzw. distanziert empfunden (Arkün-Kocadere und Çağlar-Özhan 2024). Tan (2024) untersuchte die Rezeption eines KI-generierten Avatars in einem Lernmanagementsystem und stellte eine im Gesamten positive Wahrnehmung durch Studierende fest, wenn auch Bedenken hinsichtlich der Natürlichkeit des Avatars geäussert wurden.

Animierte Erklärvideos im Cartoon-Stil in der Hochschulbildung sind bisher vergleichsweise wenig erforscht. Einen wichtigen Beitrag leisteten Liu und Elms (2019) in einer Studie mit australischen Hochschulstudierenden, in der unter anderem eine Steigerung von Interesse und Engagement am Lernprozess sowie verbessertes Verständnis der Inhalte durch Cartoon-Videos festgestellt wurde. Das Design der Charaktere, Drehbuch und Stimmen erwiesen sich als wichtigste Determinanten für Interesse, Spass und Engagement im Lernprozess. Für animierte Charaktere konnte Jintapitak (2018) eine Verbesserung des Lernerfolgs nachweisen sowie erhöhte Motivation.

Studien, die die Perzeption von Realfilmen, animierten Cartoon-Videos und solchen mit KI-Avataren vergleichen, fehlen zum gegenwärtigen Zeitpunkt. Die vorliegende Studie zielt darauf, einen Beitrag zur Schliessung dieser Forschungslücke zu leisten und widmet sich der folgenden Fragestellung:

Welche Auswirkungen haben unterschiedliche Videoformate bei Erklärvideos (Realfilm, animiertes Cartoon-Video mit menschlicher bzw. künstlicher Stimme, KI-Avatar mit künstlicher Stimme) auf die selbstberichtete Lernwirksamkeit, das Engagement und das Interesse von Studierenden, und welche Formatpräferenzen lassen sich bei den Studierenden feststellen?

4. Methodik

Es wurde ein Mixed-Methods-Ansatz in Form eines sequenziellen Vertiefungsdesigns (Kuckartz 2014, 77–78) gewählt: Eine quantitative Fragebogenerhebung wurde mit einer anschließenden qualitativen Vertiefungsstudie in Form von leitfadengestützten Interviews kombiniert. Die Erhebungen wurden im Frühjahr und Sommer 2024 an einer deutschen Fernhochschule durchgeführt.

Ziel dieser Methodentriangulation war, unterschiedliche methodische Zugänge zu eröffnen, um den Forschungsgegenstand aus mehreren Perspektiven zu beleuchten (Denzin 1970; Flick 2011). Mit der quantitativen Erhebung sollte eine empirische Grundlage geschaffen werden, um Muster und Zusammenhänge zu identifizieren und Präferenzen, wahrgenommene Lernwirksamkeit, Engagement und Interesse für verschiedene Videovarianten messbar zu machen. Die Erhebung beruht auf Einzelitems, die keiner gesonderten Testung unterzogen wurden, und diente als Vorstudie zur Vorbereitung der qualitativen Erhebung. Die qualitative Erhebung wurde auf dieser Grundlage aufgebaut und zielte darauf ab, einzelne Aspekte vertiefend zu explorieren und dabei individuelle Wahrnehmungen und Meinungen sowie weiterführende Perspektiven einzubeziehen. Diese Kombination trägt zur Erhöhung der Validität und Erweiterung des Erkenntnisgewinns bei, da sowohl statistische Trends als auch subjektive Sichtweisen erfasst werden (Jick 1979).

4.1 Erhebungsmaterial

Für die Untersuchung wurde ein etwa zehnminütiges Erklärvideo mit dem Thema «Notizen machen» entwickelt, welches mit identischem Drehbuch in vier verschiedenen Versionen produziert wurde (siehe auch Abb. 1):

1. Realfilm: mit menschlichem Akteur im Studio (unterstützt durch Einblendung von PowerPoint-Folien und Fotos)
2. KI-Avatar: KI-generierter Film mit KI-Avatar und künstlicher Stimme (unterstützt durch Einblendung von PowerPoint-Folien und Fotos)
3. Cartoon mit Realstimme
4. Cartoon mit KI-Stimme

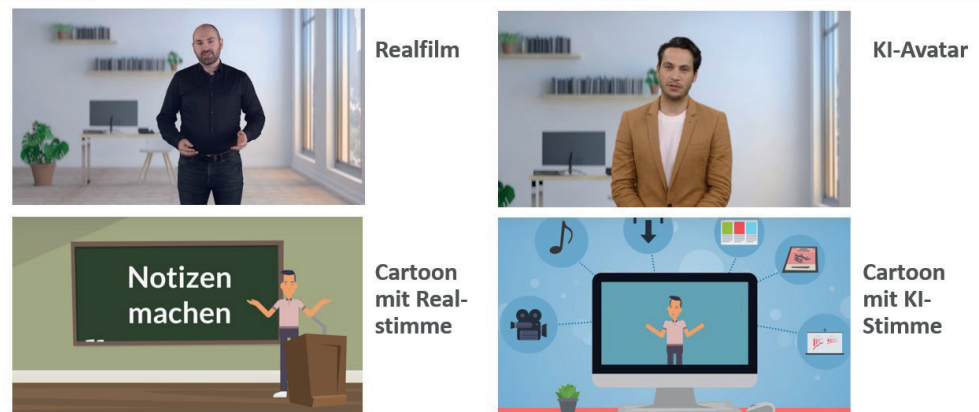


Abb. 1: Eingesetzte Videotypen, eigene Darstellung.

Für Version 2 kam die Software Synthesia¹ zum Einsatz, für die Versionen 3 und 4 die Animationssoftware Vyond.² Version 3 und 4 unterschieden sich ausschliesslich bezüglich der eingesetzten Stimme; in Version 3 kam die Stimme des Darstellers aus dem Realfilm zum Einsatz, in Version 4 eine KI-Stimme des Programms Vyond. Diese war nicht mit der Stimme aus Version 2 identisch.

Alle Videos enthielten einen kurzen animierten Abschnitt im Whiteboard-Stil, in dem die Anwendung der vorgestellten Notizmethoden praktisch gezeigt wurde (s. Abb. 2).

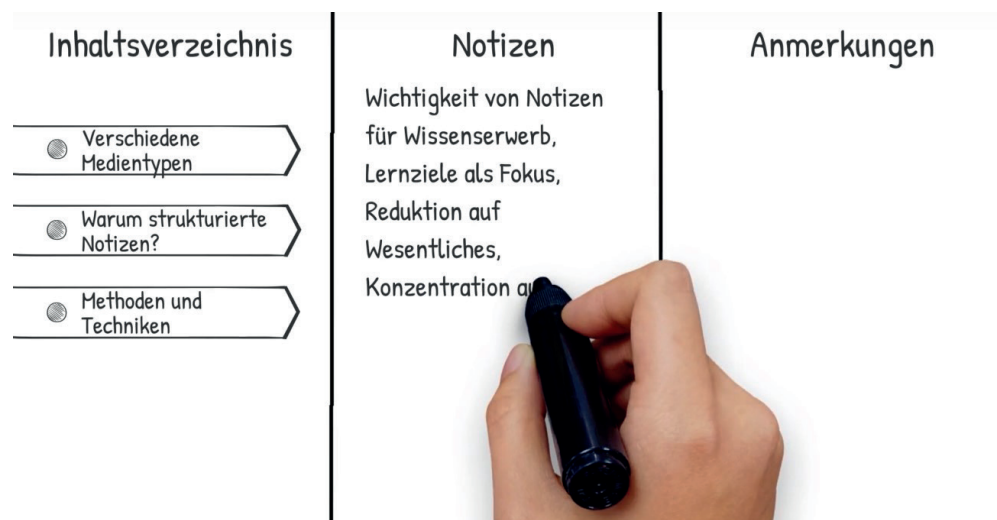


Abb. 2: Whiteboard-Darstellung im Erklärvideo, eigene Darstellung.

- 1 Synthesia ist eine KI-gestützte Software, die ermöglicht, aus Texten automatisiert Videos mit realistisch menschlich aussehenden, computergenerierten Avataren zu erstellen.
- 2 Die Software Vyond ermöglicht die Erstellung von animierten Videos in verschiedenen Comic- und Zeichenstilen.

4.2 Quantitative Erhebung

Der Erhebungsfragebogen wurde unter 695 Studierenden verteilt. Sämtliche Teilnehmenden absolvierten ein Fernstudienprogramm mit zwei wöchentlichen Onlinevorlesungen, durch die ein geeigneter Kontaktpunkt gegeben war. Die Befragten bekamen an zwei Tagen im Mai 2024 in ihren jeweiligen Onlinevorlesungen jeweils eine Version des Videos vorgeführt und wurden anschliessend gebeten, ihre Eindrücke zu dem Video anonymisiert in einem Fragebogen zu schildern.

Insgesamt nahmen 114 Personen an der Befragung teil. 17 von ihnen wurden nachträglich aus dem Datensatz entfernt, da sie entweder nicht alle relevanten Fragen der Erhebung beantwortet hatten (12) oder die Zuordnung des gesehenen Videotyps auf Basis der Angaben nachträglich nicht mehr möglich war (5). 97 Personen wurden demnach in die Datenauswertung eingeschlossen.

Das Fragebogeninstrument erfasste neben einigen soziodemografischen Variablen verschiedene Aspekte zum selbstberichteten Interesse am gezeigten Video, zur eingeschätzten Lernwirksamkeit, der Videopräsentation sowie zum Vorwissen und der Erfahrungen der Teilnehmenden in Bezug auf Lernvideos und das vorgestellte Thema.

Zur Abfrage der genannten Konstrukte wurden jeweils quantitative Items (geschlossenes Antwortformat) mit qualitativen Items (offenes Antwortformat) kombiniert, um eine möglichst ganzheitliche Einschätzung der Studierenden zu den Videos zu erhalten. Items mit geschlossenem Antwortformat wurden überwiegend anhand einer 6-stufigen Likertskala beantwortet (1 = überhaupt nicht [interessant/motiviert/ansprechend etc.] bis 6 = sehr [interessant/motiviert/ansprechend etc.]). Es handelt sich bei den erhobenen Variablen jeweils um Einzelitems, die nicht die Voraussetzungen für eine Zusammenführung zu psychometrischen Skalen aufweisen. Die erhobenen Daten wurden mittels der Software SPSS deskriptiv und inferenzstatistisch ausgewertet.

Die qualitativen Items wurden als Grundlage für die nachfolgend beschriebene qualitative Erhebung verwendet; ihre Auswertung wird hier nicht im Detail dargestellt.

4.3 Qualitative Erhebung

Nach Auswertung der quantitativen Erhebung wurde im Juli 2024 eine vertiefende leitfadengestützte Interviewstudie mit acht Studierenden durchgeführt, um die Erkenntnisse der Fragebogenstudie zu validieren und zu erweitern. Die Teilnehmenden wurden per Aufruf im Online-Campus der Fernhochschule rekrutiert.

In einem Online-Meeting wurde den Studierenden zunächst eine Version des Videos komplett vorgeführt. Die Zuteilung erfolgte per Los; jede Version wurde von zwei Personen gesehen.

Im Anschluss an die Videopräsentation wurden im persönlichen Gespräch Einschätzungen und Präferenzen des Studenten bzw. der Studentin exploriert. Anhand eines halbstandardisierten Leitfadens wurde hierbei auf allgemeine Eindrücke des Videos, die Verständlichkeit der gezeigten Informationen, die visuelle und akustische Umsetzung, die moderierende Person bzw. Figur, die Passung des Formats für verschiedene Lehrinhalte, den Einfluss des Formats auf den eigenen Lernerfolg sowie auf störende und hilfreiche Aspekte des Videos eingegangen.

Im weiteren Verlauf der Interviews wurden den Studierenden Ausschnitte der drei verbleibenden Videoversionen präsentiert, zu denen sie erneut um eine Einschätzung sowie zusätzlich um die Angabe ihrer Präferenz gebeten wurden.

Die Interviewlänge betrug ca. 30–40 Minuten. Die Auswertung des transkribierten Datenmaterials erfolgte mittels qualitativer Inhaltsanalyse nach Kuckartz und Rädiker (2022) unter Nutzung der Software MAXQDA 24. Gearbeitet wurde mit einem deduktiven Codesystem (siehe Tabelle 1).

Code	Definition
Allgemeiner Eindruck	Allgemeine Äusserungen und Beobachtungen zum Video
Verständlichkeit	Angaben zur Verständlichkeit der im Video vorgetragenen Inhalte
Moderator	Angaben zur vortragenden Figur
Visuelle Gestaltung	Angaben zur visuellen Gestaltung des Videos
Stimme/Akustik*	Angaben zur vortragenden Stimme, Begleitmusik etc.
Struktur	Äusserungen zur Struktur des Videos
Verbesserungsvorschläge	Vorschläge zur Verbesserung des Videos
Eignung Lernformate	Aussagen zur Eignung verschiedener Videoformate zum Lernen und für bestimmte Themen
Präferenz	Angaben, welches Videoformat bevorzugt wird
* Der Code Stimme/Akustik setzt sich aus den beiden Subcodes «Stimme/Akustik KI-Stimme» sowie «Stimme/Akustik menschliche Stimme» zusammen.	

Tab. 1: Codesystem.

5. Ergebnisse der quantitativen Erhebung

5.1 Stichprobe

Von den 97 Personen, die in die Datenauswertung eingeschlossen wurden, ordneten sich 71 Personen dem weiblichen und 25 Personen dem männlichen Geschlecht zu. Eine Person machte keine Angabe zu ihrer Geschlechtszugehörigkeit. Das Alter der Proband/innen lag zwischen 20 und 59 Jahren, wobei sich die deutliche Mehrheit

der Befragten (75%) in die Altersspanne von 20–40 Jahren einordnete. Die Mehrheit der Befragten studierte zum Zeitpunkt der Erhebung Soziale Arbeit (58.8%) (s. Tab. 2).

14 der befragten Personen hatten den Realfilm gesehen, 20 die animierte Version mit menschlicher Stimme, 18 die animierte Version mit künstlicher Stimme sowie 45 die KI-Avatar-Version.

Variable	Ausprägung	n	Anteil
Alter in Jahren	20–29	39	40.2%
	30–39	34	35.1%
	40–49	20	20.6%
	50–59	4	4.1%
Geschlecht	weiblich	71	73.2%
	männlich	25	25.8%
	divers	1	1.0%
Studiengang	Psychologie	3	3.1%
	BWL & Wirtschaftspsychologie	29	29.9%
	BWL	6	6.2%
	Soziale Arbeit	57	58.8%
	Kindheits- und Jugendpädagogik	2	2.1%
Präsentierter Videotyp	Realfilm	14	14.4%
	KI-Avatar	45	46.4%
	Cartoon mit Realstimme	20	20.6%
	Cartoon mit KI-Stimme	18	18.6%

Tab. 2: Ausprägung soziodemografischer Variablen der Stichprobe der quantitativen Befragung.

5.2 Deskriptive Beschreibung der erhobenen Konstrukte

Etwa die Hälfte der Befragten (49.4%) gab an, dass sie Lernvideos gern oder sehr gern zum Lernen verwendeten. Die Frage, ob sie dem Video ein anderes Lernmedium vorgezogen hätten, verneinten 81.4% der Teilnehmenden. Hinsichtlich der vorherigen Erfahrungen mit Lehrvideos zum Thema Notizenmachen gab die grosse Mehrheit der Befragten (83.5%) an, noch keine entsprechenden Videos gesehen zu haben. Etwas mehr als die Hälfte der Befragten (60.8%) berichtete, vor dem Ansehen des Videos bereits eine mittlere Vertrautheit mit den Methoden des Notizenmachens gehabt zu haben (s. Tab. 3).

		Videotyp											
		Realfilm			KI-Avatar			Cartoon mit Realstimme			Cartoon mit KI-Stimme		
	Item	M	MD	SD	M	MD	SD	M	MD	SD	M	MD	SD
1	Interesse	3.71	4.0	1.44	3.82	4.0	1.54	3.65	4.0	1.46	4.06	4.0	1.21
2	Aufmerksamkeit	3.86	4.0	1.51	3.87	4.0	1.46	3.90	4.0	1.45	3.89	4.0	1.32
3	Motivation	4.07	4.0	1.27	3.80	4.0	1.55	4.40	5.0	1.60	3.78	4.0	1.11
4	Bild & Ton	5.29	5.5	.83	4.93	5.0	1.16	5.50	6.0	.83	5.06	5.0	1.06
5	Visuelle Gestaltung	5.00	5.0	1.18	4.58	5.0	1.14	5.15	5.0	.88	4.61	5.0	1.29
6	Stimme	5.21	5.5	1.12	4.69	5.0	1.38	5.15	5.0	.75	4.44	5.0	.72
7	Figur/ Person	4.57	5.0	1.40	4.53	5.0	1.41	4.95	5.0	.83	4.11	4.0	1.13
8	Verständlichkeit	5.14	5.0	.86	5.33	6.0	.93	5.70	6.0	.47	5.06	5.0	.80
9	Struktur	4.86	5.0	.86	5.09	5.0	.95	5.25	5.0	.55	4.89	5.0	.96
10	Methoden einsetzbar	4.36	4.0	1.01	4.38	4.0	1.15	4.75	5.0	.79	4.17	4.0	1.25
11	Präferenz Lernvideos	4.21	4.5	1.63	4.07	4.0	1.32	4.80	5.0	1.40	4.44	5.0	1.34
12	Vorab Vertrautheit	4.07	4.0	1.39	3.56	4.0	1.27	4.00	4.0	1.52	3.39	3.0	1.04

Anmerkung. 1=Wie interessant fanden Sie das Video? 2=Hat das Video Ihre Aufmerksamkeit gehalten? 3=Wie motiviert fühlen Sie sich, die vorgestellten Methoden auszuprobieren? 4=Wie bewerten Sie die Qualität des Videos in Bezug auf Bild u. Ton? 5=Wie ansprechend fanden Sie die visuelle Gestaltung des Videos? 6=Wie ansprechend fanden Sie die vortragende Stimme? 7=Wie ansprechend fanden Sie die vortragende Figur/Person? 8=Wie klar und verständlich waren die Informationen? 9=Wie bewerten Sie die Strukturierung der Inhalte im Video? 10=Wie gut glauben Sie, die vorgestellten Methoden nach dem Ansehen des Videos anwenden zu können? 11=Wie gerne nutzen Sie Videos zum Lernen? 12=Wie vertraut waren Sie vor dem Ansehen des Videos bereits mit den Methoden des Notizenmachens? Antwortformat: 6-stufige Likertskala.

Tab. 3: Deskriptive Statistiken der erhobenen Items.

5.3 Inferenzstatistische Datenanalyse

Um festzustellen, inwieweit sich signifikante Gruppenunterschiede zwischen den Videotypen in Bezug auf die Lernwirksamkeit, das Interesse, Engagement oder die bewertete Präsentation zeigen, wurden mehrere Kruskal-Wallis-Tests durchgeführt. Der KW-Test diente dabei als non-parametrisches Pendant zur klassischen einfaktoriellen Varianzanalyse. Zwar lagen die quantitativen Daten auf Intervallskalenniveau

vor, was u. a. die Berechnung von Mittelwerten ermöglichte, jedoch zeigte die Prüfung der Voraussetzungen für eine ANOVA Verletzungen der Normalverteilungsannahme.

5.3.1 Lernwirksamkeit

Der KS-Test bezüglich der Verständlichkeit der präsentierten Informationen zeigte signifikante Gruppenunterschiede für die vier Videotypen ($\chi^2(3)=8.06$, $p=.045$). Anschliessend durchgeführte Post-hoc-Tests (Dunn-Bonferroni-Tests) deckten auf, dass sich die Gruppen «Cartoon KI-Stimme» ($M=5.06$) und «Cartoon Realstimme» ($M=5.70$) signifikant voneinander unterschieden ($z=3.927$, $p=.028$). Es handelt sich jedoch um einen schwachen Effekt ($\epsilon^2=.08$). Für die bewertete Strukturierung der Inhalte im Video ($\chi^2(3)=2.09$, $p=.055$) sowie die Einschätzung, die Methoden des Notizenmachens nach dem Ansehen des Videos anwenden zu können ($\chi^2(3)=2.80$, $p=.423$), ergaben sich keine signifikanten Gruppenunterschiede.

5.3.2 Interesse und Engagement

Für das bewertete Interesse am Video ($\chi^2(3)=.09$, $p=.823$), die Lenkung der Aufmerksamkeit ($\chi^2(3)=.05$, $p=.997$) sowie die Motivation, die vorgestellten Methoden selbst auszuprobieren ($\chi^2(3)=3.39$, $p=.335$) konnten anhand der Kruskal-Wallis-Tests keine signifikanten Gruppenunterschiede zwischen den Videoversionen festgestellt werden.

5.3.3 Präsentation

Ähnlich zeigt sich das Bild in Bezug auf die zu bewertenden Aspekte der Video-Präsentation und -Qualität. Sowohl für die bewertete Qualität des Videos in Bezug auf Bild und Ton ($\chi^2(3)=4.36$, $p=.225$), die visuelle Gestaltung des Videos ($\chi^2(3)=4.57$, $p=.206$), die vortragende Stimme ($\chi^2(3)=2.84$, $p=.418$) als auch die vortragende Figur/Person ($\chi^2(3)=4.53$, $p=.210$) unterschieden sich die vier Video-Typen nicht signifikant voneinander.

6. Ergebnisse der qualitativen Erhebung

6.1 Stichprobe

Insgesamt nahmen acht Personen an der Interviewstudie teil. Von diesen ordneten sich sechs dem weiblichen und zwei dem männlichen Geschlecht zu. Das Alter der Teilnehmenden lag zwischen 25 und 52 Jahren ($M=37.6$, $SD=8.99$, $MD=36$). Vier

Personen studierten zum Zeitpunkt der Erhebung Soziale Arbeit (50.0%), jeweils eine Person studierte Psychologie, Kindheits- und Jugendpädagogik, Personalmanagement & Corporate Learning sowie Psychologische Beratung.

6.2 Verständlichkeit

Die Teilnehmenden gaben unabhängig vom Format grösstenteils an, der Videopräsentation gut folgen zu können. Eine Teilnehmende fasste zusammen: «klar und verständlich, ich wusste, um was es geht, was der Sinn des Videos ist» (P2, Pos. 10); lediglich eine Person äusserte, es sei «ein bisschen schnell» gegangen (P8, Pos. 3). Die praktischen Beispiele in Form der Animation im Whiteboard-Stil, in der die Anwendung der vorgestellten Methoden gezeigt wurde, wurden von mehreren Teilnehmenden als besonders positiv für das Verständnis empfunden. So befand eine Teilnehmende:

«Am verständlichsten und einprägsamsten waren [...] zum Schluss diese Beispiele.» (P3, Pos. 14).

Die Aussagen deuten darauf hin, dass die Verständlichkeit unabhängig vom verwendeten Format ist, was die Ergebnisse der quantitativen Befragung bestätigt.

6.3 Visuelle Gestaltung

Beim Realfilm, der, abgesehen von der Animation der Beispiele, in einem klassischen Vortrag zur Kamera mit gelegentlichem Einblenden von PowerPoint-Folien und Fotos bestand, wurde von einigen Teilnehmenden das Fehlen von visuellen Darstellungen zur Verbesserung von Verständnis und Merkfähigkeit bemängelt. So äusserte eine Teilnehmende:

«Manche Sachen wären für mich leichter gewesen, wenn die noch mal visuell dargestellt gewesen wären.» (P2, Pos. 3)

Von der gleichen Person wurde zudem der Wunsch nach mehr Farbe und Lebendigkeit geäussert, es hätte «ein bisschen fröhlicher bunt sein» können (Pos. 22–23). Kritisiert wurde zudem, dass einzelne Inhalte zu lange gezeigt wurden, «ohne dass sich was [...] getan hat» (P2, Pos. 3), andere dafür zu schnell ein- und wieder ausgeblendet wurden: Das «ständige schnelle Wechseln von den Bildern» (P5, Pos. 48) wurde als anstrengend empfunden, aber auch, wenn «Schrift zu kurz eingeblendet war» (P3, Pos. 10). Bei Einstellungen, die den Sprecher zeigen, wurde ein ruhiger und ablenkungsfreier Hintergrund als positiv bewertet, «die Aufmerksamkeit war auf ihm» (P8, Pos. 27).

Die KI-Avatar-Version war im Grundsatz ähnlich gestaltet wie der Realfilm. Auch hier wurde ein «Fokus [...] auf den Herrn [...] und nicht so viel Raum nebenbei» positiv empfunden (P8, Pos. 36), schnelle Wechsel von Bildern, Folien und visuellen Elementen als anstrengend oder störend: «Dass während des Gesprochenen die Sachen sich visuell so aufgemacht haben, [da] ist bei mir nichts hängengeblieben» (P6, Pos. 12).

Die Cartoon-Versionen wurden gemischt aufgenommen. Einige Teilnehmende empfanden die animierten visuellen Elemente, «dass das noch mal alles untermalt wird» (P1, Pos. 41), als hilfreich, andere fühlten sich durch «zu viel Unnötiges» (P8, Pos. 35), zu viele Animationen oder bewegte Elemente, abgelenkt, insbesondere wenn diese als zu «abstrakt» (P3, Pos. 4, 17) empfunden wurden. So störte sich eine Teilnehmende u. a. an dem sich bewegenden Schaum auf einer Kaffeetasse:

«[A]ls sich dieser Schaum auf der Kaffeetasse bewegte, konnte ich nicht woanders mehr ins Bild gucken.» (P4, Pos. 4)

Dies korrespondiert mit den Ergebnissen von Sung und Mayer (2012), die feststellten, dass inhaltlich irrelevante grafische Elemente sich negativ auf den Lernerfolg auswirken können.

Unabhängig vom vorgeführten Videotyp äusserten sich Teilnehmende positiv zu dem über alle Versionen gleichen Beispielteil, in dem im Whiteboard-Stil die Anwendung der vorgestellten Methoden vorgeführt wurde: «das war doch sehr, sehr einprägsam» (P3, Pos. 4), urteilte eine Teilnehmende. Einige Teilnehmende, die eine der animierten Versionen komplett gesehen hatten, bewerteten jedoch den «Stilbruch» (P3, Pos. 5) zwischen dem Whiteboard-Stil und dem sonst verwendeten Animationsstil als störend.

6.4 Moderator

Die vortragende Person im Realfilm, ein kameraerfahrener Mitarbeiter des eMedienzentrums der Hochschule, hinterliess überwiegend einen «sympathischen [...], aber auch [...] kompetenten Eindruck» (P8, Pos. 23). Er wirkte «ruhig und gelassen» (P2, Pos. 25), die Teilnehmenden fühlten sich von ihm angesprochen. Dies wurde verstärkt durch die Tatsache, dass einige Teilnehmende den Mitarbeiter bereits in anderen Videos gesehen hatten und durch den «Wiedererkennungseffekt» (P5, Pos. 47) ein Vertrautheitsgefühl eintrat.

Der KI-Avatar wurde von einigen Teilnehmenden als solcher erkannt, andere nahmen ihn nicht als Avatar wahr. Einige Teilnehmende äusserten sich überrascht über die Qualität:

«Wär mir echt nicht aufgefallen. [...] Also, erschreckend, erschreckend gut.» (P8, Pos. 40)

«Das hat man ein bisschen gemerkt, aber es hat mich jetzt nicht gestört. Also, der wirkte trotzdem professionell und trotzdem auch, wie soll ich sagen, trotzdem wie ein Mensch.» (P1, Pos. 16–17)

Andere Teilnehmende bemängelten Manierismen der KI-Figur, er habe «den Kopf zu viel bewegt [...] und zu viel gezwinkert» (P2, Pos. 48), was als ablenkend empfunden wurde. Ebenso wurden nicht synchrone Lippenbewegungen moniert:

«[I]ch konnte dann überhaupt nicht mehr so richtig gut zuhören, [...] weil mich das so irritiert hat.» (P6, Pos. 6)

Zudem empfanden einzelne Teilnehmende den Avatar als nicht menschlich:

«Ich fands sehr unlebendig. [...] Ich fand den tot.» (P6, Pos. 20)

Die Cartoon-Moderatorenfigur wiederum wurde von einigen Teilnehmenden als ablenkend bzw. irritierend empfunden: «Der hat doch ziemlich stark den Fokus auf sich gezogen.» (P3, Pos. 2) Kritisiert wurden unter anderem unnatürliche Bewegungen sowie Probleme mit der Lippensynchronität: «immer so ein Gefühl, irgendwas stimmt da nicht.» (P3, Pos. 3, Pos. 24). Visuelle Details wie fehlende Finger – «ich hätte gerne einen Daumen» –, «Haarfarbton» oder Hautton (P5, Pos. 9, 19) wurden ebenfalls kritisch erwähnt.

Eine Teilnehmende sah die Möglichkeit, dass die animierte Figur die Einprägsamkeit befördern und als eine Art Ankerpunkt, ggf. auch für verschiedene Themenkreise, dienen könne: «Vielleicht erinnert man sich gerade deswegen dann vielleicht an diese Figur, wenn man daran denkt, was hat der überhaupt geredet?» (P7, Pos. 7) Die Teilnehmende schlug vor, bestimmte Avatare bestimmten Themenbereichen zuzuordnen, um für Wiedererkennung zu sorgen.

Die Interviewteilnehmenden äusserten sich am positivsten über den Darsteller des Realfilms aufgrund seiner natürlichen und authentischen Präsentation. Während einige Teilnehmende keine Probleme mit dem KI-Avatar hatten, führten Unnatürlichkeiten in Bewegung und Sprache bei anderen zu Ablenkung. Hierzu muss angemerkt werden, dass mit hoher Wahrscheinlichkeit zukünftige technologische Verbesserungen diese Probleme reduzieren werden. Die animierte Moderatorenfigur wurde teils als ablenkend, teils als potenziell einprägsam empfunden.

6.5 Stimme/Akustik

Die Stimme des Darstellers aus dem Realfilm wurde von den meisten Teilnehmenden positiv bewertet. Kritisiert wurde lediglich in einem Fall ein etwas zu hohes Sprechtempo.

Die Stimme des KI-Avatars wurde gemischt beurteilt. Mehrere Teilnehmende äusserten, dass ihnen die Stimme entweder nicht als KI-generiert aufgefallen war oder sie diese als nicht negativ oder störend empfunden hatten:

«[G]estört hat es mich jetzt [...] ehrlich gesagt nicht. Die KIs sind ja doch nicht ganz so verkehrt mittlerweile, was sie können.» (P7, Pos. 43)

«[D]ie Stimme fand ich tatsächlich nicht unangenehm. [...] insgesamt hätte ich gedacht, ja, [das ist] eine echte Person, die da gerade spricht.» (P8, Pos. 42)

Eine andere Teilnehmende empfand die Stimme als «am aufgeregtesten» bzw. «unter Druck gesetzt» (P2, Pos. 55), wieder eine andere Person monierte sie als «ein bisschen monoton» (P4, Pos 32). Auch die Betonung wurde kritisiert:

«[W]enn ich jetzt wahrscheinlich noch ein bisschen länger zugehört hätte, [hätte ich] Ihnen auch sagen können, genau was an der Stimme komisch ist. Aber irgendwas war da.» (P5, Pos. 61)

Bei den Aussagen der Teilnehmenden zu den beiden Cartoon-Versionen fiel auf, dass mehrere Personen keinen grossen Unterschied zwischen KI- und Menschenstimme wahrnahmen. Dies legt den Schluss nahe, dass KI-erzeugte Stimmen bei einer animierten Figur, die per se nicht als menschlich empfunden wird, weniger auffällig für den Zuschauenden sind. Eine gewisse Künstlichkeit der Stimme wird hier offenbar weniger wahrgenommen als in Formaten mit menschlich anmutenden Figuren oder sogar als passender empfunden. So äusserte eine Teilnehmende zur menschlichen Stimme in der Cartoonversion, diese habe «nicht zu diesem Männchen» gepasst (P6, Pos. 39). Die Person bestätigte auf Nachfrage, dass ihr bei der Cartoon-Version die KI-Stimme besser gefallen hatte. Die KI-Stimme als solche wurde durchweg positiv kommentiert:

«[D]ie Stimme, von der Stimmfarbe, von der Geschwindigkeit, fand ich so angenehm.» (P3, Pos. 22)

«Das war wie so eine Erzählerstimme.» (P6, Pos. 35)

Die Real-Stimme wurde wie bereits in der Realversion mehrheitlich als angenehm empfunden, ein Teilnehmender kritisierte jedoch eine gewisse Monotonie:

«[Ich habe] diese Sprecherstimme teilweise [als] monoton empfunden [...]. Also, da kam irgendwann so der Punkt, wo ich dachte, [ein] bisschen ändern dürfte die sich gerne zwischendurch.» (P4, Pos. 17)

Die Akzeptanz für KI-Stimmen, so lässt sich aus den Aussagen der Teilnehmenden schliessen, nimmt zu, und Unterschiede zwischen KI-generierten und menschlichen Stimmen wurden teilweise nicht bemerkt. Auffällig war die Tatsache, dass die KI-Stimme in der Cartoon-Version gleichwertig oder sogar teilweise positiver als die menschliche bewertet wurde. Möglicherweise wird bei einer als künstlich wahrgenommenen Comicfigur keine perfekte menschliche Authentizität der Stimme erwartet; eine leicht künstlich klingende Stimme könnte daher als zum Gesamteindruck der Animation passend und als weniger störend empfunden werden. Die Diskrepanz zwischen einer nicht ganz natürlichen Stimme und einer animierten Figur ist hier möglicherweise geringer als die zwischen einer künstlichen Stimme und einer menschlich wirkenden Figur wie dem KI-Avatar.

6.6 Eignung verschiedener Videoformate für verschiedene Einsatzzwecke

Befragt, ob unterschiedliche Umsetzungsformate sich für unterschiedliche Zwecke oder Themen eignen könnten, gaben die Teilnehmenden mehrheitlich an, dass die Wahl des Videoformats von dem vermittelten Inhalt abhängen sollte.

6.6.1 Menschliche Vortragende für anspruchsvolle Inhalte

Mehrere Personen betonten die Bedeutung menschlicher Vortragender für Themen, die ein tiefes fachliches Verständnis erfordern, menschliche Interaktionen zeigen, komplexe Erklärungen beinhalten oder bei denen die individuelle Lehrperson entscheidend ist. So erwähnte eine Teilnehmende, dass z. B. bei Themen, bei denen auch «Mimik und Gestik» (P1, Pos. 51). eine Rolle spielen, echte Menschen vor der Kamera besser geeignet seien:

«[Z]um Beispiel in dem [...] Modul [...] Kommunikation und Konfliktmanagement gab es [...] Online-Lernvideos. Und da fand ich sehr hilfreich, dass die Dozenten selber auch Szenen in den Videos nachgestellt haben. Und da weiss ich nicht, wie weit man das jetzt mit der KI halt auch so gut hinbekommen würde.» (P1, Pos. 23)

Eine andere Teilnehmende betonte die Wichtigkeit einer vortragenden Person, der die Fachkompetenz für das Thema zuzutrauen ist:

«[B]ei sehr wirtschaftlichen Themen würde ich eher auf die Mensch-Variante gehen. [B]ei allgemeineren [...], übergreifenden Themen, da fände ich jetzt ne Mischung nicht schlimm. [...] Wirtschaftsrecht, Unternehmensführung usw., da könnte ich zum Beispiel wenig mit Comics anfangen, [...] da brauche ich einfach auch den Input von jemanden, wo ich das Gefühl habe, der kennt sich aus.» (P2, Pos. 61)

Eine gewisse Glaubwürdigkeit der/des Vortragenden spielt offensichtlich eine Rolle, insbesondere in fachlich anspruchsvollen Bereichen, in denen das Vertrauen in die Kompetenz und Expertise der Lehrperson entscheidend ist. Bei Inhalten, die sich mit zwischenmenschlicher Interaktion befassen, wurde ebenfalls für eine reale Person vor der Kamera plädiert.

6.6.2 Animierte Videos für allgemeine und einführende Inhalte

Animierte bzw. Cartoon-Videos konnten sich Teilnehmende zum Beispiel für allgemeine oder einführende Videos zum Studium, allgemeine Konzepte oder Definitionen vorstellen:

«[W]enn jetzt [...] Definitionen oder einfach nur der sachliche Inhalt vermittelt werden soll, [...] vielleicht gerade wenn es sehr nüchterner Stoff ist, finde ich es dann nicht schlecht, wenn es [ein] Comicformat ist, damit das nicht so dröge ist und trocken vielleicht.» (P1, Pos. 54)

«Also um grundsätzliche Modulinhalte zu vermitteln, finde ich das mit der Realperson am besten. [S]o einzelne Begrifflichkeiten, Begriffe erklären oder auch Formeln, [...] Merksätze oder sowas kann ich mir sehr gut als Cartoonartiges vorstellen.» (P3, Pos. 47)

Daraus lässt sich ableiten, dass animierte Formate insbesondere für die Vermittlung grundlegender und einführender Inhalte als geeignet angesehen werden. Sie können dazu beitragen, trockenen oder nüchternen Stoff lebendig zu gestalten, zudem können Animationen das Verständnis befördern. Die Teilnehmenden sehen in solchen Formaten eine Möglichkeit, den Lernprozess aufzulockern und Inhalte leichter erfassbar zu machen.

6.6.3 Animationen zur Visualisierung komplexer Inhalte

Auch für Visualisierungen wurde das animierte Format als passend empfunden, insbesondere bei komplexen Theorien oder Strukturen. Dabei wurde auch die Vermutung geäußert, dass die animierten Visualisierungen beim Merken der Inhalte hilfreich sein könnten. So überlegte eine Teilnehmerin:

«[F]ür manche Sachen ist dieses Comicformat tatsächlich aus dem Grunde einfach gar nicht schlecht, weil man super schön irgendwie Bilder dazu gestalten kann und super gut Zusammenhänge darstellen kann. Also [...] falls [es] jetzt um irgendwelche Theorien geht, die halt einfach sehr verzweigt sind oder sowas, würde sich das ja tatsächlich anbieten.» (P4, Pos. 38–39)

Betont wurde allerdings auch, dass Effekte und Farben ablenken und das Lernerlebnis beeinträchtigen können:

«[Da] wäre ich mehr von den Effekten abgelenkt.» (P8, Pos. 25).

Eine weitere Teilnehmerin brachte die Idee auf, dass animierte Figuren als Anker für bestimmte Themenbereiche dienen könnten:

«[I]ch glaub so diese Comicvariante gerade für diese allgemeinen Sachen ist vielleicht auch nicht so die verkehrteste Idee, um das einfach im Kopf zu behalten. Dass man einfach das [...] mit einer Figur verbinden kann.» (P7, Pos. 3)

Aus den Kommentaren der Teilnehmenden lässt sich schliessen, dass die Eignung des Videoformats stark vom Thema abhängig ist und dies bei der Formatwahl berücksichtigt werden sollte. Realfilme werden für komplexe, technische oder wissenschaftliche Inhalte bevorzugt; animierte bzw. Comic-Formate eignen sich nach Ansicht der Teilnehmenden für allgemeine, einführende oder weniger komplexe Themen. Ebenso werden bei solchen Formaten Vorteile in punkto Visualisierung erkannt, die das Verständnis und die Merkfähigkeit unterstützen. Von einigen Teilnehmenden wurde auch eine Kombination aus Real- und Animationsfilm ins Spiel gebracht, um die Vorteile beider Welten zu vereinen und die Konzentration aufrechtzuerhalten:

«Was ich halt auch immer gut finde, wenn so Abwechslung im Video ist. Also man muss sich ja dann immer zwingen, sich zu konzentrieren. Und wenn dann immer mal wieder was anderes ist, ein anderer Input kommt, dann ist man wieder wach und wieder bei dem Video sozusagen.» (P1, Pos. 51)

6.7 Präferenz für ein Format

Befragt nach einer Entscheidung, welches der vier angebotenen Formate sie bevorzugen würden, entschieden sich vier Teilnehmende für den Realfilm sowie vier für die Cartoon-Version, wobei hier lediglich nach der visuellen Umsetzung gefragt wurde, es wurde keine Entscheidung für die KI- oder menschliche Stimme eingefordert. Eine Präferenz der Teilnehmenden für die zu Anfang des Interviews komplett gesehene Videoversion war nicht erkennbar.

6.7.1 Begründung für die Wahl des Realfilms

Die Teilnehmenden, die dem Realfilm zuneigten, begründeten dies mehrheitlich mit der darstellenden Person, die als sympathisch und angenehm empfunden wurde – «dem konnte ich am besten zuhören» (P2, Pos. 50). Der Vortrag wurde als unaufgeregt und leicht aufzunehmen empfunden, zudem wurde die menschliche Ansprache als positiv hervorgehoben: «Das transportiert was.» (P6, Pos. 43) Daraus lässt sich schliessen, dass die Teilnehmenden sich von dem Darsteller angesprochen und auf menschlicher Ebene abgeholt fühlten.

6.7.2 Begründung für die Wahl der Animationsvariante

Eine Teilnehmende bezeichnete das Animationsformat als «fetzig» und betonte, die Abwechslung und visuelle Dynamik helfe, die Konzentration aufrechtzuerhalten: «dann ist man wieder wach» (P1, Pos. 49). Eine weitere nutzt Videos häufig nebenbei oder in erhöhter Geschwindigkeit und ging davon aus, dass die visuellen Elemente der Comic-Version dabei helfen, neue Abschnitte zu erkennen und Informationen schneller zu verarbeiten:

«[I]ch glaube [...] fast, dass es mir mit der Comicvariante einfach schneller als Bild im Kopf bleibt.» (P7, Pos. 45)

Eine einheitliche Präferenz liess sich somit aus den Interviews nicht ableiten. Allerdings fällt auf, dass die KI-Avatar-Version von keiner Person als erste Wahl benannt wurde. Die Präferenzen scheinen von individuellen Faktoren abhängig, z. B. von Lerngewohnheiten, dem Wunsch nach Abwechslung, Bedarf an Visualisierung oder auch dem persönlichen Bedürfnis nach menschlicher Interaktion und Ansprache.

7. Diskussion

Weder die quantitative noch die qualitative Erhebung zeigten eine eindeutige Präferenz der Studierenden für eines der vier untersuchten Videoformate. In der quantitativen Analyse konnten bis auf eine Ausnahme keine signifikanten Unterschiede in der Bewertung der Formate festgestellt werden.

Die Auswertung der Interviews ergab erste Hinweise zu möglichen Ursachen dieses Ergebnisses und legte nahe, dass die Formatpräferenz der Studierenden stark von individuellen Faktoren und Vorlieben abhängt. Während der Realfilm mit menschlichem Darsteller auf starke Sympathie stiess und für seine Authentizität gelobt wurde, wurden den Cartoon-Versionen Vorteile in der Visualisierung attestiert, die vielfach als lernunterstützend gesehen wurde. Dies entspricht der eingangs erwähnten Cognitive Theory of Multimedia Learning, die davon ausgeht, dass Visualisierung die kognitive Last mindern kann (Mayer 2014). Gleichzeitig wurde jedoch von Studierenden thematisiert, dass Visualisierungen ablenkend wirken können, insbesondere wenn diese eher dekorativ (z. B. Schaum auf einer Kaffeetasse) sind. Dieses Phänomen wurde bereits von Sung und Mayer (2012) untersucht, die feststellten, dass Grafiken zwar zur Zufriedenheit von Lernenden beitragen, jedoch der Lerneffekt von der inhaltlichen Relevanz der visuellen Elemente abhängig ist. Inhaltlich nicht relevante grafische Elemente erhöhen somit die extrinsische kognitive Belastung.

Mittels KI erzeugte Figuren und Stimmen wurden unterschiedlich wahrgenommen; teilweise wurden sie gar nicht als nicht-menschlich erkannt, wie auch in der Studie von Vallis et al. (2024). Andere Teilnehmende waren von Unnatürlichkeiten irritiert. Möglicherweise kommt hier der Uncanny Valley-Effekt (Mori 2012) zum Tragen. Eine klare Ablehnung von KI-erzeugten Inhalten im Vergleich zu realen Akteuren und Stimmen war jedoch nicht zu erkennen, was sich mit den Erkenntnissen von Kit et al. (2023) deckt. Im Fall der animierten Version wurde die KI-Stimme sogar vereinzelt als passender bewertet.

Zudem ergab die qualitative Erhebung, dass für viele Teilnehmende die Eignung des Videoformats vom individuellen Thema abhängt. Übereinstimmend gaben sie an, reale Personen für komplexe Inhalte und Animationen für allgemeine oder einführnde Themen zu bevorzugen.

8. Limitationen und Ausblick für weitere Forschung

Auch wenn die Studie wichtige erste Erkenntnisse zur genannten Forschungslücke beiträgt, muss die Generalisierbarkeit der Ergebnisse aufgrund der kleinen Stichproben in beiden Teilerhebungen als eingeschränkt betrachtet werden. Geringe Gruppengrößen und der Stichprobengröße geschuldete Abweichungen von der Normalverteilung können ggf. zu einer Verzerrung der Ergebnisse führen und möglicherweise vorhandene kleine Effekte verdecken. Darüber hinaus könnte ein Selektionseffekt in der Stichprobe aufgetreten sein, indem insbesondere Studierende mit hohem Interesse an Lernvideos bevorzugt an den Erhebungen teilgenommen haben.

Hinzu kommt, dass die Teilnehmenden Studierende einer Fernhochschule sind. Diese sind im Durchschnitt älter als Studierende an Präsenzhochschulen, häufig berufstätig und bringen daher andere Erfahrungen und Bedürfnisse in ihr Studium ein, was Präferenzen und Wahrnehmungen in Bezug auf Erklärvideos beeinflussen könnte.

Der methodische Ansatz konzentrierte sich auf selbstberichtete Daten; objektive Lernleistungsmessungen wurden nicht einbezogen. Es kamen lediglich Einzelitems zum Einsatz, was die Validität der Ergebnisse ggf. beeinträchtigt haben könnte.

Zur Erhöhung der Generalisierbarkeit der Erkenntnisse sollten weitere Studien mit grösseren und ausgewogeneren Stichproben durchgeführt werden, die sowohl Fern- als auch Präsenzstudierende umfassen und validierte psychometrische Skalen verwenden. Weiterhin sollten Längsschnittstudien zur Untersuchung langfristiger Effekte der verschiedenen Videoformate auf den Lernerfolg durchgeführt werden. Hierbei sollten neben selbsteingeschätzten Kennwerten auch objektivere

Indikatoren wie Prüfungsleistungen einbezogen werden. Kontrollierte Experimente könnten den Einfluss spezifischer Elemente, z. B. Stimme oder vortragende Figur, isoliert betrachten.

9. Praktische Implikationen

Hochschullehrende und Medienspezialist:innen an Hochschulen können die Ergebnisse der hier dargestellten Erhebungen in ihre Arbeit einbeziehen und darauf basierend ihre Formatwahl treffen. Die Formatwahl kann entsprechend den Aussagen der Interviewteilnehmenden vom Inhalt abhängig gemacht werden, sodass für komplexe und zwischenmenschliche Inhalte reale Personen zum Einsatz kommen, für allgemeine, einführende oder weniger komplexe Themen animierte Videos. Mischformate aus Realfilm und Animation können die Vorteile beider Formate vereinen.

Gezielte Visualisierungen haben sich in der Erhebung als nützlich erwiesen, wobei auf eine massvolle und bewusste Verwendung visueller Elemente geachtet werden sollte, um die Lernenden nicht zu überfrachten und vom Inhalt abzulenken. Anknüpfend an die Ergebnisse von Sung und Mayer (2012) ist ein massvoller Einsatz von inhaltlich relevanten Grafiken zu empfehlen, während auf rein dekorative visuelle Elemente, die die Aufmerksamkeit vom Lehrinhalt ablenken, verzichtet werden sollte.

Videoproduktionsprozesse können dank KI-Technologien beschleunigt und verbilligt werden, sodass mehr Videomaterial für Studierende zur Verfügung gestellt werden kann. Zudem setzen solche Produktionsmethoden Kapazitäten bei Lehrenden frei, die ermöglichen, den Fokus stärker auf Inhalte und zwischenmenschliche Kommunikation mit den Lernenden zu setzen. Die Erkenntnis aus der Erhebung, dass KI-Figuren sowie KI-Stimmen von Lernenden weitgehend akzeptiert und nur von wenigen Personen als störend empfunden werden, kann helfen, Vorbehalte unter Lehrenden gegenüber neuen Video- und Voice-Technologien abzubauen. So können diese animiert werden, KI-Produktionen zunehmend in die Lehrpraxis einzubeziehen.

Literatur

- Ali, Shoukot. 2019. «IMPACTS OF WATCHING VIDEOS ON ACADEMIC PERFORMANCE AT UNIVERSITY LEVEL», Juni. <https://doi.org/10.5281/zenodo.3244393>.
- Arkün-Kocadere, Selay, und Şeyma Çağlar-Özhan. 2024. «Video Lectures With AI-Generated Instructors: Low Video Engagement, Same Performance as Human Instructors». *International Review of Research in Open and Distributed Learning* 25 (3): 350–69. <https://doi.org/10.19173/irrodl.v25i3.7815>.

- Busse, Birgitta, und Tino Bargel. 2017. «Befragungen zu E-Learning an Hochschulen – Erfahrungen und Sicht der Studierenden». WORKINGPAPER 96. Hefte zur Bildungs- und Hochschulforschung / AG Hochschulforschung. Konstanz: Universität Konstanz. <https://kops.uni-konstanz.de/handle/123456789/47470>.
- Denzin, Norman K. 1970. *The Research Act: A Theoretical Introduction to Sociological Methods*. London: Aldine Publishing Company.
- Findeisen, Stefanie, Sebastian Horn, und Jürgen Seifried. 2019. «Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos». *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, Oktober, 16–36. <https://doi.org/10.21240/mpaed/00/2019.10.01.X>.
- Flick, Uwe. 2011. *Triangulation*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92864-7>.
- Hartsell, Taralynn, Steve Yuen, und Yin Yuen. 2006. «Video streaming in online learning». *AACE Journal* 14 (Januar):31–43.
- Heidig, Steffi, Julia Müller, und Maria Reichelt. 2015. «Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning». *Computers in Human Behavior* 44 (März):81–95. <https://doi.org/10.1016/j.chb.2014.11.009>.
- Jick, Todd D. 1979. «Mixing Qualitative and Quantitative Methods: Triangulation in Action». *Administrative Science Quarterly* 24 (4): 602. <https://doi.org/10.2307/2392366>.
- Jintapitak, Manissaward. 2018. «Use of animation characters to motivate students in a higher education class». In *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)*: 280–84. <https://doi.org/10.1109/ICDAMT.2018.8376540>.
- Kätsyri, Jari, Klaus Förger, Meeri Mäkräinen, und Tapio Takala. 2015. «A Review of Empirical Evidence on Different Uncanny Valley Hypotheses: Support for Perceptual Mismatch as One Road to the Valley of Eeriness». *Frontiers in Psychology* 6 (April): 1-16. <https://doi.org/10.3389/fpsyg.2015.00390>.
- Kim, Jihyun, Kelly Merrill Jr., Kun Xu, und Stephanie Kelly. 2022. «Perceived credibility of an AI instructor in online education: The role of social presence and voice features». *Computers in Human Behavior* 136 (November):107383. <https://doi.org/10.1016/j.chb.2022.107383>.
- Kit, Leong Wai, Chew Yui-Y, Balqis Zulkifli, und Who Suet Nie. 2023. «Perception of University Students towards the Use of Artificial Intelligence-Generated Voice in Explainer Videos – Taylor’s FSLM Journals». *SEARCH Journal of Media and Communication Research*, 77–89.
- Kuckartz, Udo. 2014. *Mixed Methods. Methodologie, Forschungsdesign und Analyseverfahren*. Wiesbaden: Springer VS.
- Kuckartz, Udo, und Stefan Rädiker. 2022. *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz Juventa. https://www.beltz.de/produkt_detailansicht/47304-qualitative-inhaltsanalyse-methoden-praxis-computerunterstuetzung.html.

- Leiker, Daniel, Ashley Ricker Gyllen, Ismail Eldesouky, und Mutlu Cukurova. 2023. «Generative AI for Learning: Investigating the Potential of Learning Videos with Synthetic Virtual Instructors». In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, herausgegeben von Ning Wang, Genaro Rebolledo-Mendez, Vania Dimitrova, Noboru Matsuda, und Olga C. Santos, 523–29. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_81.
- Liu, Chelsea, und Philip Elms. 2019. «Animating Student Engagement: The Impacts of Cartoon Instructional Videos on Learning Experience». *Research in Learning Technology* 27 (April). <https://doi.org/10.25304/rlt.v27.2124>.
- Lloyd, Steven A., und Chuck L. Robertson. 2012. «Screencast Tutorials Enhance Student Learning of Statistics». *Teaching of Psychology* 39 (1): 67–71. <https://doi.org/10.1177/0098628311430640>.
- Mayer, Richard E., Hrsg. 2014. *The Cambridge Handbook of Multimedia Learning*. 2. Aufl. Cambridge Handbooks in Psychology. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>.
- Meij, Hans van der, und Jan van der Meij. 2014. «A comparison of paper-based and video tutorials for software learning». *Computers & Education* 78 (September):150–59. <https://doi.org/10.1016/j.compedu.2014.06.003>.
- Mori, Masahiro. 2012. «The Uncanny Valley». Übersetzt von Karl MacDorman und Nori Kageki. *IEEE Robotics & Automation Magazine* 19 (2): 98–100. <https://doi.org/10.1109/MRA.2012.2192811>.
- Netland, Torbjørn, Oliver von Dzengelevski, Katalin Tesch, und Daniel Kwasnitschka. 2024. «Comparing Human-made and AI-generated Teaching Videos: An Experimental Study on Learning Effects». *Computers & Education*, September, 105164. <https://doi.org/10.1016/j.compedu.2024.105164>.
- Plass, Jan L., Steffi Heidig, Elizabeth O. Hayward, Bruce D. Homer, und Enjoon Um. 2014. «Emotional design in multimedia learning: Effects of shape and color on affect and learning». *Learning and Instruction* 29 (Februar):128–40. <https://doi.org/10.1016/j.learninstruc.2013.02.006>.
- Schaarschmidt, Nadine, Claudia Albrecht, und Claudia Börner. 2016. «Videoeinsatz in der Lehre. Nutzung und Verbreitung in der Hochschule». In *Teaching Trends 2016. Digitalisierung in der Hochschule: Mehr Vielfalt in der Lehre.*, herausgegeben von Wolfgang Pfau, Caroline Baetge, Svenja Mareike Bedenlier, Carina Kramer, und Joachim Stöter, 5:39–48. Digitale Medien in der Hochschullehre. Waxmann. <http://nbn-resolving.de/urn:nbn:de:0111-pe-docs-188846>.
- Sung, Eunmo, und Richard E. Mayer. 2012. «When graphics improve liking but not learning from online lessons». *Computers in Human Behavior* 28 (5): 1618–25. <https://doi.org/10.1016/j.chb.2012.03.026>.
- Sweller, John. 1988. «Cognitive Load During Problem Solving: Effects on Learning». *Cognitive Science* 12 (2): 257–85. https://doi.org/10.1207/s15516709cog1202_4.

- Sweller, John, Paul Ayres, und Slava Kalyuga. 2011. *Cognitive Load Theory*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4419-8126-4>.
- Sweller, John, Jeroen J. G. van Merriënboer, und Fred G. W. C. Paas. 1998. «Cognitive Architecture and Instructional Design». *Educational Psychology Review* 10 (3): 251–96. <https://doi.org/10.1023/A:1022193728205>.
- Tan, Saw Fen. 2024. «Perceptions of students on artificial intelligence-generated content avatar utilization in learning management system». *Asian Association of Open Universities Journal* 19 (2): 170–85. <https://doi.org/10.1108/AAOUJ-12-2023-0142>.
- Vallis, Carmen, Stephanie Wilson, Daniel Gozman, und John Buchanan. 2024. «Student Perceptions of AI-Generated Avatars in Teaching Business Ethics: We Might Not Be Impressed». *Postdigital Science and Education* 6 (2): 537–55. <https://doi.org/10.1007/s42438-023-00407-7>.
- Wolf, Karsten. 2015. «Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung». In *Filmbildung im Wandel*, herausgegeben von Anja Hartung, Thomas Ballhausen, Christine Trültzsch-Wijnen, Alessandro Barberi, und Katharina Kaiser-Müller, 121–31. Wien: New Academic Press.
- Zawacki-Richter, Olaf, Günter Hohlfeld, und Wolfgang Müskens. 2014. «Mediennutzung im Studium». *Schriftenreihe zum Bildungs- und Wissenschaftsmanagement*, Januar.