

Schwache KIs, starke Performanz? Form und Wirkung von KI-gestützten Erklärvideos im Englischgrammatikunterricht der Sekundarstufe I

Thomas Strasser¹

The primary goal of this contribution is to show to what extent certain grammatical skills of lower secondary English learners can be improved by using AI-powered visualizer tools compared to only analogue, teacher-centered grammar introduction processes.

The items to be investigated were syllabus-specific grammar topics (past simple vs. present perfect simple). The focus group was exposed to AI-powered visualizer tools and the control group received analogue teacher input. The quantitative measurement method found that the focus groups did not achieve a significantly better test score in the post-test than the control groups, who received analogue teaching input under almost identical conditions. However, the evaluated data suggest that especially learners with a lower language level (based on C-test measurements) achieved a significantly better result in the post-tests after the AI-video exposition than those learners with a similarly low language level who received the analog teacher input.

1. Einleitung: Ausgangslage und Forschungsbedarf

Das Zeitalter der Digitalisierung wurde in den letzten Jahren, fast schon Jahrzehnten, des Öfteren eingeläutet. Dabei gibt es unterschiedliche Ansätze, wie man Digitalisierung beschreiben kann. Angefangen von teilweise absurd anmutenden Superlativen, wie "digitale Optimierung", "digitale Höchstgeschwindigkeit", "digitale Superkräfte" (Selke & Klose 2016: 13) bis hin zu semantisch kulturpessimistisch formulierten Phrasen, wie "digitaler Demenz" (Spitzer 2016), oder der "digitale Untergang" (Glaser 2012). Generell kann konstatiert werden, dass auch nur ein oberflächlicher Blick in die Unmenge an Digi-Literatur eine bestimmte Überattribution des Themas Digitalisierung in unterschiedlichsten gesellschaftlichen Domänen (digitales Lernen, digitale Transformation) suggeriert.

Vor allem digitale Werkzeuge und Methoden sind für Stakeholder in der Bildung (Schulaufsichten, Verlage etc.) von großem Interesse, da digitale *gadgets* ob ihrer didaktischen Vielseitigkeit erfahrungsgemäß von Fremdsprachenlehrkräften sehr gut aufgenommen werden. Unterschiedliche Ansätze und Publikationen (Heil,

1 Korrespondenzadresse: Prof. Dr. Thomas Strasser, Pädagogische Hochschule Wien, Grenzackerstraße 18, 1100 Wien, E-Mail: thomas.strasser@phwien.ac.at

Wu, Lee & Schmidt 2016; Strasser 2018a, b; Cunningham, Rashid & Le 2019; Meurers & De Kuthy 2019) tragen dazu bei, das Thema der Digitalität im Fremdsprachenunterricht bewusst zu enttechnokratisieren.

Dass interaktive und kollaborative Apps, wie Padlet und Kahoot, den motivationalen *drive* bei Lehrkräften und Lernenden im Fremdsprachenunterricht steigern, kann zumindest aus einer erfahrungsbasierten, subjektiven Sichtweise festgehalten werden. Paradigmatische Ansätze, wie jene des SAMR- (Puentedura 2014) und ARCS-Modells (Keller 2016), versuchen, dies zumindest partiell wissenschaftlich zu untermauern. Und genau das dürfte eines der Probleme der digitalen Medien im Fremdsprachenunterricht sein: Bestimmte lernerfolgsspezifische Effekte und Auswirkungen diverser Tools im Kontext des kommunikativen Sprachunterrichts sind wenig bis kaum empirisch belegt. Die aktuelle Studienlage ist eher durch hermeneutisch-deskriptive Argumentationslinien charakterisiert, Publikationen, die oftmals einen sehr praxisorientierten Anwendungscharakter haben, werden vermehrt veröffentlicht, was für einen Multiplikatoreffekt (Stichwort: Einsatz von Tools in der Schule) sicherlich nicht von Nachteil ist.

Jedoch muss es auch das Desiderat fremdsprachendidaktischer Forschung sein, bestimmte Effekte der Toolanwendungen empirisch messen zu können (siehe Kapitel 5). Es wäre etwas vermessen anzunehmen, man könne nun in einem einzigen Beitrag empirisch belegen, ob ein bestimmtes Tool eine bestimmte Fähigkeit im Fremdsprachenunterricht fördert, da dies von so vielen Parametern, wie Zielgruppe, Unterrichtsmethode, Soziodemografie der Klasse, Vorbildung der Lernenden und Literalitäten bei den Lehrenden, abhängig ist. Was dieser Aufsatz aber versucht, ist ein sehr spezifisches digitales Tool (KI-gestützte Erklärvideoanwendung) in einem sehr spezifischen Anwendungskontext (laborähnliches Durchführungsdesign, siehe Kapitel 5) mit einer sehr konkreten grammatikalischen Performanz-Dichotomie (*past simple* vs. *present perfect simple*) in einem sehr spezifischen schulischen Kontext (gleichaltrige Schüler/innen aus demselben Klassenzug aus derselben Schule) zu untersuchen. Die Spezifität dieser Untersuchung resultiert daraus, eben keine generischen Interpretationen aufkommen zu lassen, sondern vielmehr in eine bestimmte empirische Spezifik zu gehen, die als basale Grundlage für mögliche, weitere explorative Studien fungieren kann.

Ziel dieses Aufsatzes ist es somit, in erster Linie den Begriff der Digitalisierung in einen konkreten empirischen fremdsprachendidaktischen Diskurs zu bringen, ohne dabei generische Konzepte digital-gestützten Lehrens und Lernens zu vergessen, um diese dann in Folge konziser in spezielle Domänen des *technology-enhanced language learning* (TELL), vor allem in jene der Künstlichen Intelligenzen einzuarbeiten.

2. Künstliche Intelligenzen im (Fremdsprachen-)Unterricht

2.1 KI kompakt. Ein kleiner, unvollständiger Überblick

Verfolgt man die Fachliteratur (Meurers u.a. 2019; Pandarova u.a. 2019: 343), wird deutlich, dass der Begriff der Künstlichen Intelligenzen (KIs) auch im Bereich des technologiegestützten bzw. digitalen Fremdsprachenunterrichts immer öfter erwähnt, analysiert und auch kritisch reflektiert wird. Eine differenzierte terminologische Klärung des Begriffes KI oder AI (engl. *artificial intelligence*) würde den Rahmen dieses Beitrages sprengen, jedoch erscheint es – auch im Kontext der definitorischen und anwendungsorientierten Verortung des Begriffes im Fremdspracheunterricht – wichtig, grundlegende Charakteristika zu skizzieren und redundante Fehlkonzeptionen bezüglich dieses umstrittenen Themas vor allem im Bereich des fremdsprachlichen Unterrichts aufzugreifen und argumentativ zu entkräften. KIs im Allgemeinen haben kaum etwas mit roboterähnlichen Maschinen zu tun, die die Weltherrschaft an sich reißen, auch wenn der sehr apokalyptische Diskurs in Film und Fernsehen das oftmals suggeriert. Zu viele unklare Begrifflichkeiten und Halbwahrheiten schwirren herum. Der Grund dafür ist, dass sehr oft Konzepte wie Künstliche Intelligenz, maschinelles Lernen, Deep Learning, Robotics, humanoide Computer etc. diskursiv in einen Topf geworfen werden (Baker & Smith 2019; Berent, Littlejohn & Blakemore 2020).

Vereinfacht und allgemein ausgedrückt, werden Künstliche Intelligenzen (u.a. im Alltag) folgendermaßen eingesetzt:

"Artificial Intelligence (AI) is now widely used for spam filtering, to speech recognition, to credit card fraud detection to face recognition" (Mitchell & Cohen 2018: 103). In dieser – zugegebenermaßen sehr generischen Definition – wird dennoch die Vielseitigkeit der Einsatzszenarien von KIs auch im gesellschaftlichen Leben dargelegt. Die prominentesten Einsatzszenarien von KIs im alltäglichen Leben sind die Gesichtserkennung bei Sicherheitskontrollen, Bild- und Bewegungsanalyse bei selbstfahrenden Autos oder die Sprachsteuerung bei Alexa, Siri und Co. Hier liegt eine Vielzahl an Daten (Bilder, Zahlen, Dokumente, etc.) zugrunde, die der sogenannte Algorithmus in den jeweiligen Kontext bringt, um daraus Schlüsse für das entsprechende Szenario zu ziehen. Viele KIs sind somit schon im Einsatz, ohne dass es den Nutzerinnen und Nutzern bewusst ist (Lotze 2018). Dabei wird die Anzahl an speziellen, KI-gestützten Applikationen und Programmen in den unterschiedlichsten bildungspolitischen und -ökonomischen Kontexten immer größer. Einige Anwendungsbereiche, vor allem im Fremdsprachenunterricht, seien hier nun erwähnt.

2.2 KI-gestützte Übersetzungsarbeit

KI-gestützte Übersetzungsarbeit hat in den letzten Jahren eine enorme qualitative Verbesserung erfahren (Merkert 2017). Google Translate und DeepL verwenden riesige Datensets, die mit korpuslinguistischen Informationen und Kontexten verknüpft werden, um somit die Übersetzungsqualität nicht auf ein bloßes Wort-für-Wort-Szenario zu reduzieren, sondern immer mehr in Richtung interkultureller Spezifität mit pragmatischer Kohärenz zu verbinden (Schmidt & Strasser 2022: 168). Dabei agieren Systeme wie DeepL konvolutional, das bedeutet, die Übersetzung erfolgt synchron, in Echtzeit, und kann sich den jeweiligen Inputs des Lernenden schnell anpassen. Eine der größten Herausforderungen hinsichtlich dieser Übersetzungs-KIs ist nicht die technologische Entwicklung, sondern vielmehr die Skepsis der Lehrkraft, wenn es darum geht, Sprache mithilfe solcher Anwendungen zu lernen (Pokrivcakova 2019: 140), obwohl zahlreiche Studien belegen, dass ein schieres Verbot solcher Technologien im Fremdsprachenunterricht kontraproduktiv sei (ebd.). Vielmehr ginge es darum, solche assistierenden Technologien im Sinne eines *scaffoldings* einzusetzen, da ohnehin noch interkulturell-pragmatische Kompetenzen des Lernenden notwendig sind, die eine Übersetzungssoftware noch nicht kohärent anwenden kann.

2.3 KI-gestützte Visualisierungstools

Ausgehend von den in der Literatur beschriebenen KI-spezifischen Technologien, unter anderem auch für den Fremdsprachenunterricht, widmet sich dieses Kapitel einer terminologischen und genrespezifischen Weiterentwicklung oben genannter Typologien. Ungeachtet dessen, ob es nun die Schreibassistenten mit den multimedialen Rückmelde-Pop-ups, die Chatbots bzw. Sprachlernapps mit integrierten Bilderwörterbüchern oder KI-gestützte Übersetzungsanwendungen mit Thesauri und Aussprache bzw. Transkriptions-Inserts sind, ein kleinster gemeinsamer konzeptioneller Nenner findet sich bei diesen KI-gestützten Algorithmen: der visuelle Impuls bzw. die visuelle Rückmeldung. In vielen anwendungsspezifischen Kontexten greift die KI auf eine Datenbank zu, um je nach Rückmelde- bzw. Outputprozessen ein Bild bzw. eine animierte Grafik zu zeigen (Bilder bei Vokabellernen, animierte Feedbackprozesse bei Sprachlernapps, animierte Chatbots, etc.). Deshalb ist im Folgenden von sogenannten KI-gestützten Visualisierungstools die Rede. Dies sind in der Regel Apps, Webseiten, Programme, die bestimmte textuelle Eingaben in Bilder, GIFs oder (animierte) Videos umwandeln. Dahinter steckt, wie bei all den anderen erwähnten KI-gestützten Werkzeugen, eine Datenbank, die konstant vom Algorithmus abgeglichen wird, um das sprachlich am

nächsten liegende Pattern in ein semantisch-kohärentes Bild umzuwandeln. Solche Performanzen sind nicht neu, da sie unter anderem in diesen Kontexten vorkommen:

- Fotosuche mittels Texteingabe beim Apple iPhone,
- Google Bildersuche,
- Collagengestaltung mit Canva durch automatisierte (bildindizierte) Vorlagengestaltung,
- Lernkartengestaltung mit automatisierter Text-Bild-Verknüpfung mit Quizlet,
- Automatisiertes Bild-Rendering basierend auf Spracheingabe mit Midjourney Bot oder Copymatic.

Eine weitere Domäne, bei der die KI-Visualisierungsimpulse für das Lehren und Lernen evident sind, ist jene der Erklärvideos. Gerade das Sprachenlernen mit visuellen Impulsen bietet enorme Potenziale (Wolf 2015: 33), was sich auch in der Tatsache bestätigt, dass über 86 % der rezipierten Wissensartefakte bei deutschen 12-19-jährigen Lernenden über Videos (vornehmlich YouTube) bezogen werden (Jörissen 2019). Das bewegte Bild ist somit als Lehr-/Lernmedium nicht mehr wegzudenken. Das Video (vor allem auch YouTube mit seiner hohen pädagogisch-wertvollen Schwankungsbreite) per se ist kein Garant für ein qualitativ zufriedenstellendes Lehr-/Lernmedium, aber in Kombination mit Prinzipien der Erklärdidaktik und KI-gestützten Visualisierungen durchaus ein Medium mit großem Potenzial, vor allem auch für den Fremdsprachenunterricht (Arnold & Zech 2019: 13; Strasser 2020: 29-30).

Eine Anwendung, die sich dieser visualisierten Erklärdidaktik verschrieben hat, ist Mysimpleshow, eine KI-gestützte Anwendung, die genau den oben beschriebenen Kriterien eines KI-gestützten Visualisierungstools entspricht. Hierbei können sogar ganze von den Lernenden bzw. vom Lehrenden verfasste Manuskripte oder Storyboards in das System eingespeist werden und die Anwendung produziert autonom ein animiertes Erklärvideo, basierend auf diesen Inputs. Wie ein solches Tool in Verbindung mit dem Erklären von grammatikalischen Strukturen im Englischunterricht – vor allem als lehrendenzentriertes Werkzeug – eingesetzt werden kann, wird in den folgenden Kapiteln genauer beschrieben. Zusammenfassend werden hier nochmals die Charakteristika von KI-gestützten Visualisierungstools dargelegt:

- Text-/Bild-Schnittstelle mit Suchbereichsfeld,
- Semantisch-generische Suchkompatibilität (d.h., wenn bestimmte Begriffe als Bild nicht gefunden werden, konzeptionell nächstmögliche Begriffe vorschlagen),
- Mediale Vielseitigkeit bei der Artefaktesuche (Bild, Video, GIFs, Animation),

- Extra-Feature: gesamte Manuskripte/Storyboards werden in Form eines Erklärvideos oder einer interaktiven Präsentation erstellt (Anwendungen: Mysimpleshow, Powtoon Cartoonsoftware, Design.ai Präsentationswerkzeug).

3. Erklärvideos

3.1 Relevanz/Bedeutung

KI-gestützte Visualisierungstools (darunter vor allem KI-Erklärvideos) an sich brauchen für nachhaltig curriculare Vermittlungsszenarien eine klare Struktur und eindeutig formulierte Didaktisierungsprozesse basierend auf den in der Wissenschaft formulierten Charakteristika – vor allem wenn es um Visualisierungen geht. Gerade im Kontext der Visualisierung bestimmter curricularer Artefakte im Sprachunterricht (Grammatik, Themenkomplexe, Wortschatz) kommen immer wieder Erklärvideos ins Spiel.

Folgende allgemeine Merkmale sind für Erklärvideos typisch (in Anlehnung an Brehmer & Becker 2017: 2-4):

- Inhaltlicher Facettenreichtum: Mit einem adäquaten, didaktisch-durchgedachten Storyboard lassen sich in der Regel fast zu jedem Thema Erklärvideos erstellen. Im Fremdsprachenunterricht können diese Videosequenzen über Grammatik, Wortschatz oder bestimmte lehrplanadäquate Themen produziert werden.
- Kreative Potenziale: Vor allem durch technologische Hilfsmittel (spezielle Stop-Motion-Apps, Software) lassen sich mittlerweile viele Inhalte auch visuell-dynamisch darstellen. Dies kann von der einfachen Legetricktechnik bis hin zur hochdynamischen KI-Visualisierung (siehe folgende Kapitel) reichen.
- Meist informeller Erzählstil: Die Präsentation bestimmter Inhalte erfolgt meistens sehr narrativ und "auf Augenhöhe", d.h. im jeweiligen Register der Zielgruppe. Meistens zeichnet sich das Gesprochene/Erklärte durch einen lockeren Erzählstil aus.

3.2 Wissenschaftliche Rahmenbedingungen

Ausgehend von diversen empirischen Studien (u.a. auch Krämer 2016: 29), die eine generelle Steigerung bestimmter kognitiver Wissensperformanzen nach Exposition mit Erklärvideos bei Lernenden attestieren, und die eher generisch-anmutende Annahme, dass dies unter anderem auf die bunte, dynamische Art vor allem der digitalen KI-gestützten Erklärvideoanwendungen zurückzuführen ist, braucht es vor allem Konzepte, die positiven Effekte der Videos auch in ein wissenschaftstheoretisches Narrativ einzubinden. Mayers (2014: 34) kognitive Theorie des multimedialen Lernens könnte sich möglicherweise als nützlich erweisen. Hierbei behauptet Mayer unter anderem, dass Menschen besser mit einer Kombination aus Grafiken und Narration als bloß mit Grafiken und Bildschirmtext lernen. Im Vordergrund stehe immer das "improved learning result", d.h. das bessere Lernergebnis (Bateman & Schmidt-Borcherding 2018: 2). Selbstverständlich ist hier zu kritisieren, dass ein Lernergebnis per se immer vom Kontext abhängt und allgemein schwer zu messen ist. Daher braucht die Applikation einer bestimmten Theorie immer einen konzisen Beobachtungsauftrag.

Beim "Spatial Contiguity"-Prinzip (Mayer 2009: 135) geht es vor allem darum zu zeigen, dass Menschen oftmals besser lernen, wenn bestimmte Inhaltswörter mit den entsprechenden Grafiken näher zueinander platziert werden. Das "Temporal Contiguity"-Prinzip (Mayer 2009: 153) suggeriert, dass Menschen besser lernen, wenn Inhaltswörter mit den jeweiligen Bildern gleichzeitig und eher nicht aufeinanderfolgend präsentiert werden, ein populäres Stilmerkmal eines Erklärvideos. Beide Prinzipien wurden in der Produktion der Erklärvideos für die empirischen Teile eingebaut. Ein weiterer Ansatz, den Mayer erwähnt und der auch eine anwendungsorientierte Kongruenz mit den allgemeinen Attributen eines Erklärvideos darstellt, ist zum Beispiel das Multimedia-Prinzip, das konstatiert, dass Menschen generell besser mit Wörtern und Bildern lernen als nur mit Wörtern alleine (vgl. Bateman & Schmidt-Borcherding 2018: 4).

Selbstverständlich schafft die explizite Granularität dieses Ansatzes auch einige Kontroversen (vgl. Bateman & Schmidt-Borcherding 2018), da multimodale Artefakte, und somit auch Erklärvideos, inhaltlich und strukturell zu divers und komplex sein können. Ferner bedarf es einer konziseren Definition bzw. Verortung, was denn letztendlich unter Bild und Text bzw. unter besserem Lernergebnis – auch im Kontext des Erklärvideos – zu verstehen sei:

Despite its relative complexity, the integrated model still simplifies things considerably and therefore needs further elaboration. For example, there might exist multiple levels of propositional representation within the verbal channel instead of only one level. The different levels might range from the micropropositions expressed in the text to various levels of macropropositions [...]. Similarly, there might exist multiple levels of mental models in the

pictorial channel ranging from coarse-grained overview models to detailed models with high granularity (Bateman & Schmidt-Borcherding 2018: 2).

Gerade bei digitalen bzw. multimedialen Erklärartefakten, wie Erklärvideos (*explainer videos*), fehlen oftmals die Evidenzen der Effektmessung des "besseren" Lernens. Um aber eine basale, generisch-antizipierende argumentative Grundlage zu Mehrwertspotenzialen von KI-gestützten Erklärvideos zu schaffen, kann Mayers *framework* als nützliche, wissenschaftlich-formulierte Orientierung dienen, um auch Parallelen bzw. Kontrastierungen zu kognitiven multimedialen Lehr-/Lernperformanzen (wie zum Beispiel bei englischen Grammatikvermittlungssperformanzen in der Sekundarstufe) herzustellen.

Was die fachspezifischen Qualitätsaspekte anbelangt, so sei auf das Folgekapitel verwiesen, in dem es vor allem darum geht, wie grammatikalische Phänomene im Unterricht bzw. in einer Lehr-/Lernsequenz curricular-adäquat und nachhaltig im Sinne eines messbaren Wissenstransfers eingeführt werden können.

3.3 Grammatik in Erklärvideos

Generell kann davon ausgegangen werden, dass eine multimediale Animation per se nicht unbedingt einen "Lernmehrwert" darstellt (Roche & Scheller 2004: 14). Ein überfrachtetes Erklärvideo kann in vielen Fällen auch kontraproduktiv sein,

[...] wenn die zu verarbeitende Information die Verarbeitungskapazität des Informationsverarbeitungssystems übersteigt, z.B. wenn zusammengehörige bildliche und verbale Information zeitlich und räumlich nicht aufeinander abgestimmt ist (ebd.).

Abgesehen davon, dass die Prinzipien des guten Erklärens auch bei animierten Erklärvideos unbedingt einzuhalten sind, ist bei der Gestaltung von grammatikalischen form- und strukturbasierenden Inhalten vor allem darauf zu achten, dass "[...] von visuellen Redundanzen, aussagelosen Begleitvisualisierungen und Reizüberflutungen, kurz einem visuellen Aktionismus [abzusehen ist]" (ebd.).

Es ist somit ersichtlich, dass man gerade bei bunten, höchst dynamischen Erklärvideo-Anwendungen mit niederschwelligem *user interface design* (UIX) vom *Bling-Bling*-Aktionismus mit potenziell drohender Visualisierungsredundanz abzurücken sollte, um sich vielmehr um die Prinzipien des guten Erklärens in Kombination mit einer kohärenten, zielgruppenadäquaten Grammatikeinführungsstrategie zu kümmern. Darüber hinaus sollte berücksichtigt werden, dass sich nicht jedes grammatikalische Thema für Erklärvideos eignet: "Dynamische Bilder (Animationen) eignen sich wesentlich besser zur Darstellung sequenzieller oder kausaler Sachverhalte, aber nur so lange auch dies nicht zu einer Reizüberflutung

oder Ablenkung führt" (Roche & Scheller 2004: 14). Somit wären eben sequenziell-kausal determinierte Zeitformen, wie jene der *past simple* und der *present perfect simple*, grundsätzlich für eine visuelle Immersion bzw. Didaktisierung im Englischunterricht geeignet.

Ferner sollte festgehalten werden, dass bestimmte grammatikalische Themen in Kombination mit Erklärvideos eher beim remedialen bzw. vorbereitenden Üben eingesetzt werden sollen, da ein gewisses Vorwissen nützlich ist:

Animationen sind tendenziell auch dann eher effizient, wenn die Lerner über das nötige Vorwissen zur Verankerung der neuen Information verfügen. Wo dies nicht gegeben ist, wird die präsentierte Information dagegen eher oberflächlich interpretiert und schafft zwar den subjektiven Eindruck, das Material gut verstanden zu haben, verhindert damit aber de facto ein tieferes Einsteigen in die Materie (Roche & Scheller 2004: 15).

Unter bestimmten Rahmenbedingungen eignen sich Animationen bzw. animierte Erklärvideos für die Vermittlung/Festigung von grammatikalischen Performanzen im Fremdsprachenunterricht. Welche grammatikalischen Bezugsrahmen und Vermittlungsmodelle nun in sinnvoller Kombination mit (KI-gestützten) Erklärvideos im Unterricht wirken können, wird im nächsten Kapitel erläutert.

4. Grammatikvermittlung im Fremdsprachenunterricht. Verortung bei KI-gestützten Erklärvideos

4.1 Unterrichtskonzeption zur Grammatikvermittlung

Um die in den vorigen Kapiteln erwähnten Potenziale KI-gestützter Erklärvideos in einen Kontext der Grammatikvermittlung im fremdsprachlichen Unterricht zu setzen, bedarf es einer strukturierten Herangehensweise. Es sollte herausgearbeitet werden, welche Bezugsrahmen zur Grammatikvermittlung die Wissenschaft anbietet, um diese dann in adaptierter Form für die Durchführung der KI-gestützten Erklärvideos anzuwenden.

Beim expliziten Grammatiklernen geht es um die "Bewusstmachung von Regeln zum Gebrauch bestimmter Formen" (Haß 2006: 128), so wie in unserem Beispiel mit einer 7. Klasse Gymnasium, wo das Festigen von *present perfect simple* vs. *past simple* im Vordergrund stand. Für die Vermittlung dieser beiden Zeitformen, die – wie bereits oben erwähnt – zu wiederholender curricularer Bestandteil in der 7. Klasse sind, müssen unter anderem die folgenden didaktischen Entscheidungen getroffen werden (vgl. Hinkel & Fotos 2002; Larsen-Freeman 2014: 252-54):

- Erklärungen zur Form, Bildung der *past simple* und *present perfect*

simple, einschließlich *irregular verbs*, Verneinungen sowie Fragestellungen;

- Informationen zum Funktionskomplex, d.h. Anbieten von Erklärsträngen, wann und wie die jeweiligen Zeiten verwendet werden.

Hier wird klar hervorgehoben, dass diese Funktionen dem Lernenden "sehr detailliert in den jeweils passenden Kontexten isoliert erläutert werden" (Haß 2006: 128). Dies ist in der folgenden empirischen Erhebung unter vergleichbar performativen Laborbedingungen (d.h. fast kongruente sequenzielle manuskriptbasierende Erklärperformanzen) beim Lehrendeninput und bei den KI-gestützten Erklärvideos der Fall. Somit gelten auch für das produzierte Erklärvideo folgende Vorgehensweisen:

- Informationen zum Frequenzwert einer Struktur, d.h., wie oft diese Zeiten verwendet werden (vor allem im Alltag). Diese Spezifika sind vor allem in den Erklärvideos evident.
- Verweise auf bewährte Lerntechniken, z.B. Eselsbrücken, oder im Falle der Instruktionsprozesse (Lehrendeninput und KI-Videos), sogenannte *Everyday-Life-Sätze*, die den Zeitgeist Jugendlicher unter Umständen treffen (*Yesterday I netfixed and chilled.*).

4.2 Design der instruktiven Prozesse im Rahmen der Erhebung

In Anlehnung an den wissenschaftlichen Diskurs hinsichtlich der Einführung sowie des Wiederholens von grammatikalischen Formen im Unterricht und vor allem durch die konkreten, in der wissenschaftlichen und praxisorientierten Literatur formulierten unterrichtlichen Performanzvorschläge im Bereich der Grammatikvermittlung (Scrivener 2012; Thaler 2012; Grimm, Meyer & Volkmann 2015: 89-102; Niemeier 2017; Berry 2021) wurde folgende Struktur bei den KI-gestützten Erklärvideos für die empirische Erhebung erarbeitet:

Kapitel 1: *Awareness (Bewusstsein)* (Thaler 2012: 238-40; Larsen-Freeman 2014: 252; Berry 2021)

Hier werden kontextualisierte Sätze aus dem alltäglichen Leben und angepasst an das Sprachregister des jeweiligen Lehrwerkes in Verbindung mit visuellen Anker (Bild-Text-Assoziation, vgl. Mayer 2014) und eingeblendeten Schlüsselwörtern präsentiert, die für die grammatikalische Struktur relevant sind, in Verbindung mit einem Erzähler visualisiert.

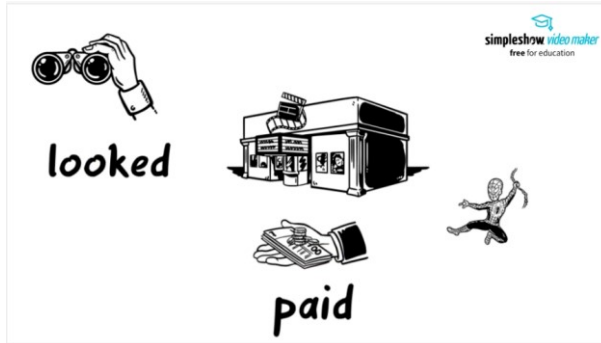


Abbildung 1: Kapitel 1 des Erklärvideos; Bild-Text-kombination mit *Everyday-Life*-Sätzen

Kapitel 2: *Explaining (Erklären der Strukturen)* (Haß 2006: 128; Thornbury 2011: 32; Scrivener 2012; Thaler 2012: 239; Larsen-Freeman 2014: 254; Berry 2021)

Hier wird das Regelwerk der jeweiligen Zeit näher erläutert. Dabei werden u.a. die Bildung, aber eben auch die Zeitpunkte, wann das Tempus verwendet wird, visuell dargestellt.



Abbildung 2: Kapitel 2 des Erklärvideos; Visualisierung der grammatikalischen Formen

Kapitel 3: *Real life situations* (Bezug zur Lebenswelt der Lernenden) (vgl. Keller 2016: 4)

Wichtig ist hierbei zu erwähnen, dass es sich oftmals um Beispielsätze handelt, die die Lebenswelten der jugendlichen Lernenden tangieren (vgl. Shifman 2014), um somit auch aus einer motivationalen Sichtweise heraus Relevanz zu generieren (vgl. Keller 2016: 5).



Abbildung 3: Kapitel 3 des Erklärvideos; "zeitgeistige" Beispielsätze

Kapitel 4: *Revising* (Wiederholen) (Haß 2006: 132; Scrivener 2012; Thaler 2012: 240; Grimm et al. 2015: 97; Schmidt & Strasser 2022: 175)

In dieser Sequenz werden die wichtigsten Regeln und Anwendungsszenarien der jeweiligen Zeit in textueller und visuell-konnotierter Form präsentiert und wiederholt.



Abbildung 4: Kapitel 4 des Erklärvideos; Wiederholen der wichtigsten Formen und Strukturen

Kapitel 5: *Testing and Mastering* (Überprüfen und Festigen) (Thornbury 2011: 141; Scrivener 2012; Thaler 2012: 240; Berry 2021)

Hierbei handelt es sich lediglich um eine Information innerhalb des Videos, dass die Lehrkraft nun Übungsblätter zur jeweiligen Zeitform austeilte und mit den Schülerinnen und Schülern durchgeht.

Generell kann festgehalten werden, dass das Design der Erklärvideos für die empirische Erhebung auf unterschiedlichen wissenschaftlichen Konzepten zur Grammatikvermittlung fußt. Welche Sequenzierung als *Instructional-Design*-Ansatz zum Einsatz im schulischen Unterricht kommt, hängt von vielen Parametern ab. Die Vermittlungsprinzipien innerhalb dieser Erklärvideos sind tendenziell deduktiv, ein Ansatz, der in der Literatur nicht immer unumstritten, jedoch hier im konkreten Kontext begründet ist. Das Ziel der Erhebung und der Erklärvideos ist es, bereits bekannte grammatikalische Strukturen zu festigen bzw. zu wiederholen, vor allem auch im Hinblick auf eine Schularbeitsvorbereitung; somit sind die Vorzüge für die deduktive Methode (zeitökonomisch, schnelle Regelbeschreibung) evident.

5. Empirischer Teil

5.1 Zielsetzung und Hypothesen

Im empirischen Teil wird nun detailliert darauf eingegangen, welche Form und Wirkung KI-gestützte Visualisierungstools (animierte Erklärvideos, wie *My-simpleshows*) zur Steigerung der Grammatikperformanz im Bereich *past simple* vs. *present perfect simple* bei Gymnasialschülerinnen und -schülern der Klasse 7 (3. Klasse Sekundarstufe 1 in Österreich) im Vergleich zu analogen, lehrkraft-zentrierten Grammatik-Instruktionsmethoden/-ansätzen haben können.

Die **Null-Hypothese (H0)** wird im Rahmen dieses Papers folgendermaßen formuliert: KI-gestützte Erklärvideos schaffen keine signifikant besseren Lerneffekte (*past simple/present perfect simple*) bei Klasse-7-Gymnasialschülerinnen und -schülern im Vergleich zu ähnlichen analogen Instruktionsmechanismen, die von der Lehrkraft durchgeführt werden.

Hypothese 1 (H1) versucht empirisch zu belegen, dass Lernendengruppen nach der Exposition mit KI-Erklärvideos ein signifikant besseres Lernergebnis im Post-Test erzielen als jene Gruppen, die einen analogen Lehrendeninput unter fast identischen Bedingungen (gleiche Sequenzierung, gleiches Instruktionsskript) erhielten.

Hypothese 2 (H2) intendiert einen empirischen Beleg dafür zu finden, dass vor allem Lernende mit niedrigerem Sprachniveau (ausgehend von C-Test-Erhebungen) nach der KI-Erklärvideo-Exposition ein signifikant besseres Ergebnis bei den Post-Tests erzielten als jene Lernenden mit einem ähnlich niedrigem Sprachniveau, die den analogen Lehrendeninput erhielten.

5.2 Durchführung der Studie

5.2.1 Stichprobe

Die Studie wurde an einem Wiener Innenstadtgymnasium im Zeitraum vom 15. April bis 30. Juni 2022 durchgeführt. Dabei wurden 76 Schülerinnen und Schüler der 3. Klasse Unterstufe (Klasse 7, zwischen 12-14 Jahre, Sekundarstufe 1) für unterschiedliche messmethodische Erhebungen herangezogen. Die generischen deskriptiven Statistiken zur Stichprobe können folgendermaßen artikuliert werden:

Der **KI-Gruppe (2 Fokusgruppen)** wurden drei KI-gestützte Erklärvideos mit Mysimpleshow in zeitlich ähnlich getakteten Abständen vorgespielt, während die **Kontrollgruppe (A-Gruppe)** drei zum Erklärvideo identische gesciptete Lehrendeninputs erhielt. Klassen 3A und Mischklasse 3ABC wurden jeweils mit ihrem Englisch-Stammlehrenden für die KI-Exposition und Klassen 3B und 3C jeweils mit ihren Englisch-Stammlehrenden für den analogen Unterricht ausgewählt. Somit blieben die Schüler/innen in ihren Klassen bei ihren Klassenkolleg/innen, in ihren Unterrichtsräumen und bei den gewohnten Lehrenden, mit dem einzigen Unterschied, dass innerhalb des gesamten partitativ aufgeteilten Jahrgangszuges die Expositionsarten variierten (2 Gruppen KI/2 Gruppen analog). Durch diese gruppendynamisch marginal interferierende Gruppenbildung wurde versucht, soziale Effekte weitestgehend zu minimieren, da eben das Setting der Schüler/innen nicht grob verändert wurde bzw. sie die Aufteilung vom regulären Englischunterricht gewohnt sind. Solche Aufteilungen innerhalb der Jahrgangszüge sind in Österreich innerhalb des fremdsprachlichen Unterrichts Usance.

Die Testphasierung bestand aus einem C-Test ($n = 76$), einem Pre-Test ($n = 76$) und einem Post-Test ($n = 72$). Insgesamt 35 (46,1%) SchülerInnen wurden der KI-Gruppe und 41 (53,9%) der Kontrollgruppe zugeordnet. Weiblichen Geschlechts waren 40 (52,6%), männlichen Geschlechts 35 (46,1%) und zu einem/einer Jugendlichen (1,3%) war das Geschlecht nicht bekannt. Die Verteilung auf die jeweiligen Klassenverbände war ähnlich: Der 3A waren 23 (30,3%) der befragten Kinder und Jugendlichen zuzuordnen, während 26 (34,2%) aus der 3B und 27 (35,5%) aus der 3C stammten. Und ebenso bei den Lehrenden wurde auf

eine Gleichverteilung geachtet: Lehrkraft "B" unterrichtete 20 (26,3%) der SchülerInnen, Lehrkraft "F" 21 (27,6%), Lehrkraft "H" 16 (21,1%) und Lehrkraft "S" 19 (25,0%) der SchülerInnen. Nachfolgende Tabelle zeigt die Mittelwerte und Standardabweichungen der erhobenen Leistungsmaße, nach KI- und Kontrollgruppe sowie für die gesamte Stichprobe.

	KI			KG			Gesamt		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
C-Test (184 Punkte)	35	141,29	17,65	41	137,71	31,11	76	139,36	25,70
Pre-Test (t1, 18 Punkte)	35	11,03	2,99	41	11,54	2,87	76	11,30	2,92
Post-Test (t2, 18 Punkte)	34	13,38	2,07	38	13,21	2,49	72	13,29	2,29
Differenz Post-Pre	34	1,97	3,44	38	0,71	4,30	72	1,29	3,95

Tabelle 1: Mittelwertsberechnung und Standardabweichungen der Test-Items

5.2.2 Erhebungsinstrumente

Folgende messmethodische Items wurden im Rahmen der Erhebung eingesetzt.

Online-Fragebogen

In einer Online-Umfrage mit der Software Limesurvey wurden 76 Schülerinnen und Schüler befragt. Das Ausfüllen der Online-Fragebögen erfolgte im Rahmen des Präsenzunterrichts unter operativer Aufsicht der jeweiligen Englischlehrerinnen im Zeitraum Anfang bis Mitte Mai 2022.

Der Fragebogen gliederte sich in drei Abschnitte. Der erste Abschnitt fragte vor allem die technischen Infrastrukturen innerhalb der schulischen und außerschulischen Lernumgebungen der Schülerinnen und Schüler ab. Im zweiten Teil ging es vor allem darum, wie für die Schülerinnen und Schüler ansprechender und motivierender Englischunterricht mit und ohne digitale Medien aussehen könnte und welche digitalen Anwendungen ihre Lehrkräfte bereits für den Grammatikunterricht einsetzten. Der dritte Teil beinhaltete soziodemografische Abfragen. Generelle Intention dieses Fragebogens war es, gegebenenfalls hypothesenstärkende bzw. -schwächende Einsichten aus Varianzanalysen mit Messwiederholungen und *t*-Test-Reihen in einem weiteren diskursiv-legitimierenden Kontext zu korrelieren.

C-Test

Der C-Test ist ein standardisiertes Verfahren (Raatz 2021), der als ein integrativer schriftlicher Test zu verstehen ist, der im Allgemeinen die Sprachbeherrschung der Probandinnen und Probanden misst. Durch vereinheitlichte Verfahren wird eine Art reduzierte Redundanz gewährleistet. Folgende Charakteristika sind beim C-Test von Bedeutung:

- Sie erfassen den allgemeinen Sprachstand in der Muttersprache oder in einer Fremdsprache,
- bestehen meist aus 5 Texten mit jeweils 20 oder 25 Lücken, die ergänzt werden müssen,
- dauern bei schriftlicher Bearbeitung maximal 30 Minuten
- messen objektiv und sehr genau,
- gehören inzwischen zu den meisterforschten Sprachtests
- werden weltweit angewendet,
- liegen inzwischen in vielen europäischen und außereuropäischen Sprachen vor (Raatz 2021).

Um bestehende Fähigkeitsunterschiede zwischen den Untersuchungsgruppen und auch um Lerneffekte in der methodischen Performanz der Pre- und Posttests erfassen zu können (u.a. auch um Hypothese 2 zu belegen), wurde der Sprachstandstest vorab durchgeführt (vor den Pre- und Post-Tests bzw. vor der Erklärvideoexposition). Hierbei wurden vier Texte mit durchschnittlich 45 Lücken pro Text mit einer C-Test-Software generiert. Die Anzahl der Lücken geht von einem standardisierten Minimalwert pro Text bei C-Tests von 20-25 aus, wurde aber erweitert, um curricular-spezifische Aufgabentexte inhaltlich in sich abschließen zu lassen. Um valide Vergleichswerte hinsichtlich des grammatikalischen Corpus zu erhalten, wurden die C-Tests mit sprachstandsadäquaten Textartefakten (Texte für 3. Klasse Gymnasium aus Lehrwerken) entwickelt. Die C-Tests wurden bei 76 Schülerinnen und Schülern im Zeitraum von Anfang bis Mitte Mai 2022 durchgeführt.

Pre-Test

Um bestimmte Messeffekte hinsichtlich einer möglichen Performanzsteigerung im Bereich des *present perfect simple* und des *past simple* vor und nach einer Exposition (sowohl KI als auch analoger Lehrendeninput) messen zu können, bieten sich ANOVA-Verfahren (*analysis of variance*; Varianzanalyse mit Messwiederholung) (Hain 2011) an. Die Schülerinnen und Schüler erhielten einen *cloze-test* mit antizipierbarem Schwierigkeitsgrad (Pandarova et al. 2019), der sich aus 18 Lücken mit einzusetzenden *past simple* oder *present perfect simple* Zeitformen zusammensetzte. Wichtig ist hierbei noch zu erwähnen, dass die Schülerinnen und Schüler bereits Vorwissen zu den beiden Zeitformen hatten.

Post-Test

Nach der Durchführung der C-Tests und des Pre-Tests wurden im Anschluss an die Exposition mit KI-gestützten Videos und Lehrendeninput die Post-Tests durchgeführt. Im Sinne einer empirischen Vergleichbarkeit wiesen die Pre- und Post-Tests die gleiche Länge (*cloze-test* mit 18 Items) und vor allem eine fast paritätische Frequenz an grammatikalischen Phänomenen (*form, usage, irregular forms* etc.) auf. Die Pre- und Post-Tests wurden – basierend auf curricular adäquaten Sprachregisterdesideraten – mit vergleichbaren Materialien aus Grammatikwerken für Klasse 7 erstellt. Die Verteilung der Gesamtpunkte ist nach grafischer Prüfung für Pre- und Post-Test ähnlich (d.h. zu beiden Zeitpunkten normalverteilt). Wäre einer der Tests zu einem Zeitpunkt zu leicht oder zu schwer, so wäre eine schiefe Verteilung zu diesem Zeitpunkt zu beobachten (z.B. mehr Schülerinnen und Schüler mit hoher Punktzahl oder vice versa). Tatsächlich unterscheiden sich die Mittelwerte von t1 zu t2. Dies liegt aber an der zwischenzeitigen Intervention.

5.2.3 Sequenzieller Ablauf der Erhebung

Nach dem C-Test und Pre-Test wurden im Zeitraum von acht Wochen (Anfang Mai bis Mitte Juni 2022) drei zentrale Inputphasen zur Grammatikvermittlung durchgeführt. Dabei wurden der Experimentalgruppe (Gruppe KI) drei KI-gestützte Erklärvideos vorgespielt. Video 1 behandelt das Tempus *past simple* (basierend auf den in Kapitel 4 erwähnten Grammatikvermittlungssequenzen; *forms, usage, exceptions*). Direkt im Anschluss an das Vorspielen des ersten Videos wurden Übungen zu den jeweiligen grammatikalischen Phänomenen durchgeführt. Diese erfolgten mit Übungshandouts im Lückentextformat (*cloze*). Zeitgleich instruierten die Lehrerinnen der A-Gruppe (analoge Kontrollgruppe) die Schülerinnen und Schüler mit einem "klassischen" Input, der aber mit der Struktur und dem Skript des Erklärvideos identisch war. Ziel dabei war es, dass sich die Lehrenden strikt an das eher instruktivistische Skript hielten, ohne dabei zu sehr in interaktionale Prozesse mit den Schülerinnen zu gehen, um auch eine gewisse Vergleichbarkeit der Expositionsdichotomien (KI vs. analog) zu gewährleisten. Diese Tatsache ist zugleich einer der evidentesten Limitationen der Studie, da davon ausgegangen werden kann, dass es durch die sehr strikten, linearen Vorgaben des Skriptes zu Verzerrungen in der Bewertung der Qualität der Lehrendeninstruktion kommen kann (Voss & Wittwer 2020). Die Empfehlung für eine weitere empirische Untersuchung von Instruktionseffekten bei KI-Videos und natürlichen, möglichst unbeeinflussten Unterrichtsinteraktionen von Lehrenden liegt somit nahe.

Nach durchschnittlich zwei bis drei Unterrichtsstunden des Übens des Tempus *past simple* und dem Erarbeiten weiterer curricularer Themen, wurde den Schülerinnen und Schülern das zweite Video vorgespielt bzw. der zweite Input dargelegt. Hierbei handelte es sich um das *present perfect simple* (*forms, usage, exceptions*). Im Anschluss wurde wieder mit analogen *Cloze*-Übungen gearbeitet. In der dritten Phase wurde ein Erklärvideo (KI-Gruppe) vorgespielt, das die Unterschiede zwischen *past simple* und *present perfect simple* darlegt bzw. fand ein Lehrendeninput zu diesem Phänomen mit der Kontrollgruppe (A-Gruppe) statt. Danach folgte das gewohnte Übungsprozedere, sodass sich ein Übungs- und Expositionszeitraum von durchschnittlich sieben bis acht Wochen ergab:

Zeitraum	Ereignis	Intention	KI-Gruppe	Gruppe A
Anfang-Mitte Mai 2022	Limesurvey-Erhebung	Erhebung digitale Kompetenzen SuS, Infrastruktur Schule/zuhause, Lerngewohnheiten	x	x
Mitte Mai 2022	C-Test	Sprachniveauerhebung für Effektmessungen	x	x
Mitte Mai 2022	Pre-Test (t1)	<i>past simple vs. present perfect simple</i> (Lückentext)	x	x
Mitte Mai – Ende Juni 2022	Exposition KI-Videos mit Folgeübungen	Festigen der grammatikalischen Strukturen mit KI-Tool	x	
Mitte Mai – Ende Juni 2022	Exposition mit analogem Lehrendeninput mit Folgeübungen	Festigen der grammatikalischen Strukturen mit Lehrer/inneninput		x
Ende Juni 2022	Post-Test (t2)	<i>past simple vs. present perfect simple</i> (Lückentext); ähnliche strukturelle, curriculare und phänomenologische Patterns	x	x

Tabelle 2: Phasierung des Forschungsdesigns – Durchführung der Erhebungen

5.3 Ergebnisse

Nachdem der Shapiro-Wilk-Test auf Normalverteilung für die Werte des C- sowie Pre-Tests signifikant ausfiel, $p < ,05$, wird zur statistischen Überprüfung auf signifikante Unterschiede der U-Test herangezogen (Bortz & Döring 2016).

Zunächst wurde betrachtet, ob sich zwischen der KI- sowie der Kontrollgruppe A ein genereller Unterschied im Fähigkeitsniveau beobachten lässt. Hierfür wurden die Ergebnisse des C-Tests zwischen beiden Gruppen verglichen: Es

zeigte sich kein signifikanter Unterschied, $U = 748,50$, $Z = 0,32$, $p = ,747$. Kontrollgruppe A ($M_{\text{Rang}} = 39,26$) sowie KI-Gruppe ($M_{\text{Rang}} = 37,61$) wiesen somit ein ähnliches Fähigkeitslevel, gemessen über den C-Test, vor der Pre- sowie Posttestung, auf.

Um zu betrachten, ob sich die Gruppen KI versus A über die zwei Messzeitpunkte (Zeitraum zwischen t1 und t2 durchschnittlich sechs Wochen) punktetechnisch unterschiedlich entwickelt (Verbesserung/Verschlechterung) haben, wurden für beide Messzeitpunkte ebenso U -Tests berechnet. Für den ersten Messzeitpunkt und somit den Pre-Test zeigte sich kein signifikanter Unterschied zwischen der Kontrollgruppe A ($M_{\text{Rang}} = 40,83$) sowie der Treatmentgruppe KI ($M_{\text{Rang}} = 35,77$), $U = 813,00$, $Z = 1,00$, $p = ,316$. Und auch beim Post-Test ließ sich kein auffälliger Unterschied zwischen Kontrollgruppe A ($M_{\text{Rang}} = 35,96$) und der KI-Gruppe ($M_{\text{Rang}} = 37,10$) beobachten, $U = 625,50$, $Z = -0,23$, $p = ,815$.

Zur Veranschaulichung der Entwicklung beider Gruppen zeigt die untenstehende Grafik das Verhältnis der Mittelwerte im Pre- und Posttest.

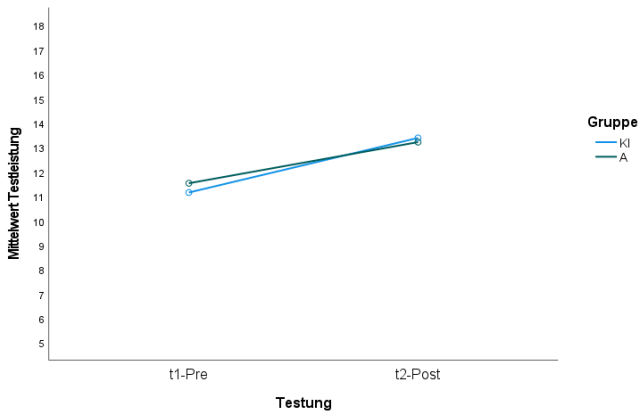


Abbildung 4: Mittelwertanalyse Testleistung KI-Gruppe vs. A-Gruppe (Lehrendeninput)

Wie sich erkennen lässt, haben sich beide Gruppen vom Pre-Test zum Post-Test punktetechnisch verbessert, die KI-Gruppe sogar marginal mehr als die A-Gruppe, da sie mit einem niedrigerem ("schlechteren") Mittelwert zum t1-Pre-Test startete, aber dann zum Zeitpunkt t2 (Post-Test) im Schnitt eine leicht bessere Leistung als die A-Gruppe zeigte. Die genauen Mittelwerte sind in der nachstehenden Tabelle aufgelistet: Es stellt sich somit die Frage, ob mit KI-gestützten Erklärvideos tatsächlich eine höhere Lernleistung (*superiorer Score*) im Vergleich zum analogen Lehrendeninput erzielt werden kann.

Deskriptive Statistiken der Testleistung Pre und Post

	Gruppe	<i>M</i>	<i>SD</i>	<i>N</i>
Pre-Test (t1)	KI	11,15	2,95	34
	A	11,53	2,94	38
	Gesamt	11,35	2,93	72
Post-Test (t2)	KI	13,38	2,08	34
	A	13,21	2,50	38
	Gesamt	13,29	2,29	72

Tabelle 3: Mittelwertsberechnung und Leistungssteigerungstendenzen (basierend auf t1 & t2)

Des Weiteren wurde ein *t*-Test der Differenz der Prüfungsleistung von Pre- zu Post-Test berechnet. Da die Differenzen der Punkte (Post-Test MINUS Pre-Test) laut Shapiro Wilk normalverteilt sind, $p = ,132$, wurde hier auf den teststärkeren *t*-Test zurückgegriffen (Bortz & Döring 2016). Die durchschnittliche Differenz von Pre- zu Post-Test beträgt in der Gruppe KI $M = 2,42$ ($SD = 2,96$) und in der Gruppe A $M = 1,68$ ($SD = 2,51$). Allerdings zeigte sich auch hier kein signifikanter Unterschied, $t(69) = 1,14$, $p = ,258$.

Die Differenzwerte wurden zusätzlich mit den C-Werten korreliert, um zu betrachten, ob es einen Zusammenhang zwischen dem Sprachstand und dem Ausmaß der Veränderung zwischen Pre- und Posttestung gibt. Der Korrelationskoeffizient wurde sowohl für die KI-Gruppe als auch die Kontrollgruppe A berechnet. Eine Spearman Korrelation zeigte hier, interpretiert nach Cohen (1988), einen mittelstarken und negativen Zusammenhang in der Gruppe KI, $\rho = -,27$, $p = ,127$, $n = 34$, und einen geringen negativen Zusammenhang in der Kontrollgruppe A, $\rho = -,15$, $p = ,375$, $n = 35$. Der negative Zusammenhang ist so zu deuten, dass mit schwächerem Sprachstand eine größere Verbesserung von Pre-Test auf Post-Test zu beobachten ist, wobei der Zusammenhang in der KI-Gruppe stärker ausfällt als in der Gruppe A, allerdings in beiden Fällen nicht signifikant ist.

Aufgrund der geringen Effekte und generischen Signifikanzen muss die Nullhypothese (H0) bestätigt werden:

KI-gestützte Erklärvideos schaffen keine signifikant besseren Lerneffekte (*past simple/present perfect simple*) bei Klasse-7-Gymnasialschülerinnen und -schülern im Vergleich zu ähnlichen analogen Instruktionsmechanismen der Lehrkraft.

Infolgedessen muss auch Hypothese 1 (H1) verworfen werden, die wie folgt lautet:

Lernendengruppen erzielen nach der Exposition mit KI-Erklärvideos ein signifikant besseres Lernergebnis im Post-Test als jene Gruppen, die einen analogen

Lehrendeninput unter fast identischen Bedingungen (gleiche Sequenzierung, gleiches Instruktionsskript) erhielten.

Im Rahmen der Analyse des Sprachstandstests C zeigte sich ein signifikanter Unterschied zwischen den untersuchten Klassenverbänden. Auch hier wurde wieder aufgrund der fehlenden Normalverteilung der C-Werte auf ein parameterfreies Verfahren, den Kruskal-Wallis-Test, zurückgegriffen (Bortz & Schuster 2016). Der Test zeigte einen signifikanten Wert, $H(2) = 7,86, p < ,05$. Bei den paarweisen Vergleichen zeigte sich, dass die Klasse 3C sich von der 3B signifikant unterscheidet, $p < ,01$, nicht aber von der 3A, $p = ,218$. Zwischen 3A und 3B zeigte sich ebenso kein signifikanter Unterschied, $p = ,142$. Abbildung 5 verdeutlicht das Verhältnis der Klassen im Sprachstand untereinander.

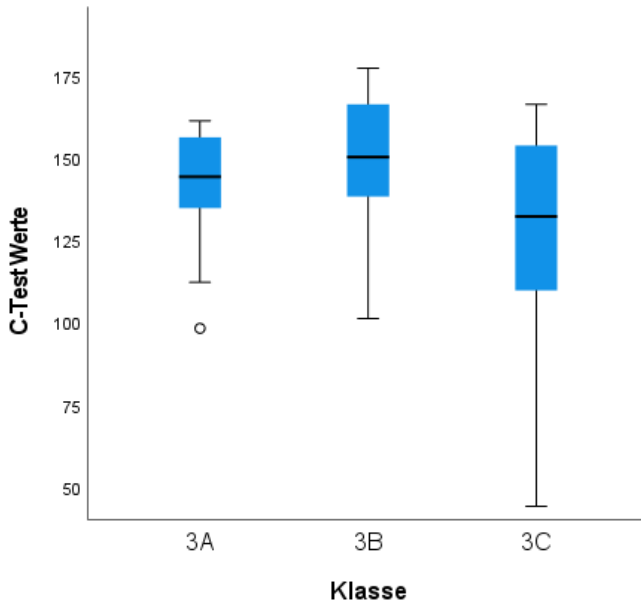


Abbildung 5: Verhältnis der Klassen im Sprachstand

In einer weiterführenden Analyse wurden die Performanz der Gruppen KI und A nur für die Klasse 3C betrachtet. Ziel war es, herauszufinden, ob die Experimentalgruppen bei Berücksichtigung von schwächeren Leistungen zu unterschiedlichen Ergebnissen führen. Da die Voraussetzungen entsprechend erfüllt waren,

wurde hier eine ANOVA (= Analysis von Variance) mit Messwiederholung berechnet. Die unabhängige Variable stellten einerseits der Zeitaspekt (Pre zu Post) und andererseits die Gruppe (A vs. KI) dar. Es konnten folgende Ergebnisse beobachtet werden:

Auch hier starteten jene Schüler/innen der Gruppe KI mit leicht schwächeren Werten im Schnitt ($M = 9,57$, $SD = 2,07$) im Vergleich zu jenen, die analog (A-Gruppe) unterrichtet wurden ($M = 10,32$, $SD = 2,61$). Im Post-Test zeigte sich dann allerdings eine scoreperformative Überlegenheit bei jenen Schülerinnen und Schülern, die das KI-Erklärvideo rezipierten ($M = 13,57$, $SD = 1,40$), im Vergleich zu den analog unterrichteten Schüler/innen ($M = 11,32$, $SD = 1,86$). Die statistische Überprüfung zeigte einen signifikanten Interaktionseffekt zwischen der Pre- und Post-Messung und der Experimentalgruppe, $F(1,24) = 7,18$, $p < 0,05$, $\eta^2_p = 0,23$. Somit konnte ein statistisch auffälliger Unterschied in der Entwicklung der beiden Experimentalgruppen über beide Messzeitpunkte hinweg beobachtet werden. Nachstehende Abbildung veranschaulicht das Verhältnis der jeweiligen Mittelwerte.

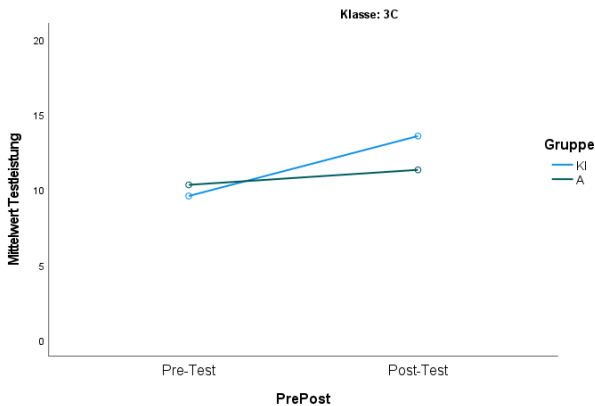


Abbildung 6: Signifikante Leistungssteigerung der KI-Gruppe innerhalb der Klasse 3C

Es muss jedoch erwähnt werden, dass die Experimentalgruppe KI mit $n = 7$ relativ klein ist, während Gruppe A (analog unterrichtete Schülerinnen und Schüler) mit $n = 19$ eine solide Größe aufweist. Ferner wurden die Varianzen der Gruppen mittels eines Levene-Tests (Prüfung der Stichproben; Gleichheit der Varianzen; Rasch et al., 2008) für beide Zeitpunkte statistisch untersucht. Es zeigten sich weder für den Pre-Test, $p = 0,896$, noch für den Post-Test, $p = 0,339$, signifikante Unterschiede, womit von homogenen Varianzen auszugehen ist, was sich günstig

auf die Vergleichbarkeit der Werte auswirkt. Somit kann Hypothese 2 (H2) angenommen werden, dass vor allem Lernende mit einem niedrigeren Sprachniveau (ausgehend von C-Test-Erhebungen) nach der KI-Erklärvideo-Exposition ein signifikant besseres Ergebnis bei den Post-Tests erzielen als jene Lernenden mit einem ähnlich niedrigem Sprachniveau, die den analogen Lehrendeninput erhielten.

Deskriptive Statistiken der Testleistung Pre und Post für die Klasse 3C.

	Gruppe	<i>M</i>	<i>SD</i>	<i>N</i>
Pre-Test (t1)	KI	9,57	2,07	7
	A	10,32	2,61	19
	Gesamt	10,12	2,46	26
Post-Test (t2)	KI	13,57	1,40	7
	A	11,32	1,86	19
	Gesamt	11,92	2,00	26

Tabelle 4: Mittelwertberechnung und Leistungssteigerungstendenzen (basierend auf t1 & t2) für Klasse 3C

5.4 Limitationen und Interpretationen

Aus einer messmethodischen Perspektive wurde versucht, die relativ kleine Stichprobe mit durchschnittlich $n = 76$ durch unterschiedliche Kalkulationsmodelle, jeweils adaptiert an das zu überprüfende Erhebungsitem, auf reliable Signifikanzkennzahlen zu prüfen. In der Regel wurden vor allem bei den Pre- und Post-Test-Auswertungen die Signifikanzen empirisch kohärent kalkuliert. Ausgehend davon war es Ziel dieser Studie, die Effekte der KI-gestützten Visualisierungstools (animierte Erklärvideos) vor allem in Hinblick auf eine punktemäßige Leistungssteigerung bei den vergleichbaren (aber bewusst nicht identischen) *cloze-tests* (Pre und Post) empirisch bedeutsam zu belegen. Die im vorigen Kapitel dargelegten Ergebnisse lassen darauf schließen, dass der Einsatz von KI-gestützten Videos nicht den signifikant-superioren Effekt im Vergleich zu analogen Lehrendeninputs im grammatikalischen Spektrum von *past simple* und *present perfect simple* erzielt. Dennoch zeigten sich Tendenzen einer Leistungssteigerung, die statistisch allerdings nicht auffällig höher waren. Was die Gründe für die Bestätigung der Nullhypothese (H0) und somit die Verwerfung der H1 sein könnten, müsste in Folgestudien eruiert werden, die folgende Aspekte intensiver behandeln und die hier vorliegenden Limitationen dieser Studie zukünftig beseitigen:

- Größere Stichprobe im altershomogenen Segment,
- Erweiterung der Stichprobe auf weitere Schulen und möglicherweise Schularten (soziodemografische, sozioökonomische, infrastrukturelle Vergleichbarkeit),

- Verstärkte Korrelierung der testspezifischen Messwertanalysen mit *delayed* Post-Tests und Post-Interviews mit Schülerinnen und Schülern bzw. Lehrerinnen und Lehrern (Auffassung von motivationalem Lernen, Rezeption der Erklärvideos im Kontext der Mayerschen *Cognitive Theory of Multimedia Learning*, Mindset der Lehrkräfte hinsichtlich Grammatikperformanzen etc.),
- Stärkere empirische Verortung des Phänomens der Verzerrung der Instruktionsqualität bei Lehrenden durch strikte Manuskriptvorgaben,
- Differenziertere Darlegung/Verortung möglicher sozialer Effekte im Kontext der Leistungssteigerung (Lehrendenpersönlichkeit, Leistungsstärke der Gruppen etc.).

Die Tatsache, dass vermeintlich leistungsschwächere Klassen durch den Einsatz KI-gestützter Erklärvideos signifikant und messeffektsrelevant besser abschnitten als jene Schülerinnen und Schüler aus dem gleichen punktetechnisch-reduzierten Segment, die analog instruiert wurden, legt den Grundstein für eine weitere empirische Exploration von KI-gestützten Werkzeugen mit leistungsschwächeren Schülerinnen und Schülern. In einem nächsten Schritt erscheint es als messmethodisch relevant, jene Schülerinnen und Schüler, die nach einem eher schwächeren C-Test-Resultat, aber vor allem nach der KI-Exposition signifikant zulegen, in Interviews zu befragen, was ihnen besonders gut an den Videos gefallen hat, um diese Antworten im Sinne einer kategorialen semantischen Reduktion auch auf wissenschaftliche Frameworks (vgl. Mayer 2014; Keller 2016) anzuwenden. Ferner bestünde auch Interesse daran, den Bildungsbackground dieser Schülerinnen- und Schülergruppe zu erheben, um mögliche Förderpotenziale KI-gestützter Tools auch im Sinne der Chancengleichheit empirisch zu belegen.

6. Conclusio und Ausblick

Ganz im Sinne der in Kapitel 1 kritisierten Positivkonnotation von Digitalisierung gibt diese kleine empirische Studie zumindest Grund zur Annahme, dass digitale Medien per se nicht immer automatisch den gewünschten Effekt bringen, den man sich erhofft. Auch oder gerade eben bei Künstlichen Intelligenzen wird es in Zukunft für den fremdsprachlichen Unterricht von großer Bedeutung sein, vermeintliche Positiveffekte empirisch zu verifizieren bzw. zu falsifizieren, da naturgemäß solche Technologien auch Ängste bei Sprachlehrkräften schüren. Was wird in Zukunft alles von der Maschine oder vom Algorithmus übernommen und wo braucht es dann noch die (Sprach-)Lehrkraft? Gerade im Bereich der schwachen KIs war

es nie das Narrativ, die Sprachlehrkraft zu ersetzen, sondern ihr vielmehr zu assistieren (Schmidt & Strasser 2022: 179). Schwache KIs, wie jene der gezeigten Erklärvideoanwendungen, wurden in erster Linie nicht dafür konzipiert, Lernerfolge und Punktestände bei Überprüfungen zu maximieren, sondern sind vielmehr dafür geeignet, die Lehrkraft bei der schnellen und KI-basierten bzw. automatisierten Produktion von motivationalen, multimedial-ansprechenden und vor allem für den Fremdsprachenunterricht so wichtigen mehrkanaligen (Audio, Video, Text etc.) Lehr-Lernartefakten zu unterstützen, damit der vermeintlich trockene, remediale Grammatik-Drill um eine dynamischere Facette reicher wird. Wenn sich dabei auch leistungsverbessernde Performanzen bei Tests und Schularbeiten ergeben, so ist dies explizit zu begrüßen.

Die vorliegende Studie zeigt, dass sich die leistungsschwächeren KI-exponierten Schüler/innen (im Vergleich zum analogen Input) signifikant verbessert haben. Dies kann für die Sprachlehrkraft Anlass genug sein, des Öfteren festigende bzw. remediale Grammatikvermittlungssequenzen mit KI-automatisierten Videos örtlich und zeitlich – ganz im Sinne des *Flipped-Classroom*-Ansatzes – vor allem auch für die Prüfungsvorbereitung auszulagern. Die Schülerinnen und Schüler schauen sich die Erklärvideos zuhause oder unterwegs an, sodass der Präsenzunterricht dafür verwendet wird, ein bestimmtes grammatikalisches Phänomen (wie z.B. den Unterschied zwischen *past simple* und *present perfect simple*) mit der methodisch-didaktischen und interaktionalen Kompetenz der Lehrkraft und dem Einsatz kommunikativ-explorativer Grammatikerarbeitungsmodelle in einem interpersonalen Lernraum (in der Regel dem Präsenzunterricht in der Klasse) zu behandeln. Die KI schafft somit Raum und zeitliche Ressourcen für die Lehrkraft, Grammatik induktiv und vor allem kommunikativ im Klassenzimmer zu unterrichten. Somit ist und bleibt die Lehrkraft die Expertin bzw. der Experte, die schwache Visualisierungs-KI die Assistentin.

Eingang des revidierten Manuskripts 13.02.2023

Literaturverzeichnis

- Arnold, Sebastian & Zech, Jonas (2019): *Kleine Didaktik des Erklärvideos: Erklärvideos für und mit Lerngruppen erstellen und nutzen*. Braunschweig: Westermann.
- Baker, Toby & Smith, Laurie (2019): *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Buckingham: Nesta.
- Bateman, John A. & Schmidt-Borcherding, Florian (2018): The Communicative Effectiveness of Education Videos: Towards an Empirically-Motivated Multimodal Account. *Multimodal Technologies and Interaction* 2, 1-27.
- Berendt, Bettina; Littlejohn, Allison & Blakemore, Mike (2020): AI in education: learner choice and fundamental rights. *Learning, Media and Technology* 45, 312-324.

- Berry, Roger (2021): *Doing English grammar: theory, description and practice*. Cambridge, UK; New York: Cambridge University Press.
- Bortz, Jürgen & Döring, Nicola (2016): *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.
- Bortz, Jürgen & Schuster, Christof (2016): *Statistik für Human- und Sozialwissenschaftler: extras online*. (7. vollst. überarb. u. erw. Aufl.). Berlin Heidelberg: Springer.
- Brehmer, Jana & Becker, Sebastian (2017): *Erklärvideos ... als eine andere und/oder unterstützende Form der Lehre*. [Online: https://www.uni-goettingen.de/de/document/download/5d0fa49e220547bded74a21f21d44fc0.pdf/03_Erklärvideos.pdf, 03.12.2022].
- Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences*. London: Routledge.
- Cunningham, Una; Rashid, Shaista & Le, Thanh (2019): The Effect of Learner Training on the Use of Digital Tools to Support English Writing Skills. *Asian EFL Journal* 21, 27-49.
- Glaser, Peter (2012): *Internjet – der digitale Untergang der Welt*. [Online: <https://futurezone.at/meinung/internet-der-digitale-untergang-der-welt/24.581.580>, 2.11.2022].
- Grimm, Nancy; Meyer, Michael & Volkmann, Laurenz (2015): *Teaching English*. Tübingen: Narr Francke Attempto.
- Hain, Johannes (2011): *Varianzanalyse*. [Online: https://www.uni-wuerzburg.de/fileadmin/10040800/user_upload/hain/SPSS/ANOVA.pdf, 4.10.2022].
- Haß, Frank (Hrsg.) (2006): *Fachdidaktik Englisch: Tradition, Innovation, Praxis*. Stuttgart: Klett Sprachen.
- Heil, Catherine Regina; Wu, Jason; Lee, Joey; Schmidt, Torben (2016): A review of mobile language learning applications: trends, challenges and opportunities. *The EUROCALL Review Volume* 24, 32-50.
- Hinkel, Eli & Fotos, Sandra (eds.) (2002): *New perspectives on grammar teaching in second language classrooms*. Mahwah, New Jersey, USA: L. Erlbaum.
- Jörissen, Benjamin (2019): *Jugend – YouTube – Kulturelle Bildung. Eine repräsentative Umfrage unter 12- bis 19jährigen zur Nutzung kultureller Bildungsangebote an digitalen Kulturrorten*. [Online: <https://www.bosch-stiftung.de/de/publikation/jugend-youtube-kulturelle-bildung-horizont-2019>, 8.2.2023].
- Keller, John M. (2016): Motivation, Learning, and Technology: Applying the ARCS-V Motivation Model. *Participatory Educational Research* 3, 1-15.
- Krämer, Andreas (2016): *International study on the use of explainer videos and effects of different video formats*. [Online: <https://simpleshows.com/de/press-releases/international-study-on-the-use-and-effects-of-different-explainer-video-formats-2016/>, 5.4.2022].
- Larsen-Freeman, Diane (2014): Teaching Grammar. In: Celce-Murcia, Marianne; Brinton, Donna M. & Snow, Marguerite Ann (eds.). *Teaching English as a second or foreign language* (4th ed.). Boston: Cengage Learning. 256-270.
- Lotze, Netaya (2018): *Präsenzunterricht ade? Künstliche Intelligenz fürs Sprachenlernen?* Magazin Sprache. [Online: <https://www.goethe.de/de/spr/eng/gls/21290629.html>, 14.05.2023, 2.6.2022].
- Mayer, Richard E. (2009): *Multimedia learning*. Cambridge: Cambridge University Press.
- Mayer, Richard E. (2014): Cognitive Theory of Multimedia Learning. In: Mayer, Richard E. (ed.): *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press, 43-71.
- Merkert, Pina (2017): *Maschinelle Übersetzer: DeepL macht Google Translate Konkurrenz*. [Online unter: <https://www.heise.de/newsticker/meldung/Maschinelle-Uebersetzer-DeepL-macht-Google-Translate-Konkurrenz-3813882.html>, 11.10.2022].

- Meurers, Detmar & De Kuthy, Kordula (2019): KI zur Lösung realer Schulherausforderungen: Interaktive und adaptive Materialien im Fach Englisch. In: *Schulmanagement-Handbuch 1/2019*. Berlin: Cornelsen, 65-84.
- Mitchell, Tom & Cohen, William (2018): Never-ending learning. *Communications of the ACM* 61, 103-115.
- Niemeier, Susanne (2017): *Task-based grammar teaching of English: where cognitive grammar and task-based language teaching meet*. Tübingen: Narr Francke Attempto.
- Pandarova, Irina & Schmidt, Torben (2019): Predicting the Difficulty of Exercise Items for Dynamic Difficulty Adaptation in Adaptive Language Tutoring. *International Journal of Artificial Intelligence in Education* 29, 342-367.
- Pokrivcakova, Silvia (2019): Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education* 7, 135-153.
- Puentedura, Ruben (2014): *Learning, technology, and the SAMR model: Goals, processes, and practice*. [Online: <http://www.hippasus.com/rpweblog/archives/2014/06/29/Learning-TechnologySAMRModel.pdf>, 12.7.2022].
- Raatz, Ulrich (2021): *C-Test. Der Sprachtest. Theorien des C-Tests*. [Online: http://www.c-test.de/deutsch/index.php?lang=de&content=beschreibung_theorie§ion=ctest, 3.3.2022].
- Rasch, Björn; Frieze, Malte & Hofmann, Wilhelm (2008): *Quantitative Methoden* (2., erw. Aufl., korr. Nachdr.). Heidelberg: Springer Medizin Verl.
- Roche, Jörg & Scheller, Julija (2004): Zur Effizienz von Grammatikanimationen beim Spracherwerb. Ein empirischer Beitrag zu einer kognitiven Theorie des multimedialen Fremdspracherwerbs. *ZIF* 9, 1-15.
- Schmidt, Torben & Strasser, Thomas (2022): Artificial Intelligence in Foreign Language Learning and Teaching. A CALL for Intelligent Practice. *Anglistik* 33, 165-184.
- Scrivener, Jim (2012): *Teaching english grammar: what to teach and how to teach it*. Ismaning: Hueber.
- Selke, Stefan & Klose, Philipp (Hrsg.) (2016): *Lifelogging: digitale Selbstvermessung und Lebensprotokollierung zwischen disruptiver Technologie und kulturellem Wandel*. Wiesbaden: Springer VS.
- Shifman, Limor (2014): *Memes in digital culture*. Cambridge, Massachusetts: The MIT Press.
- Spitzer, Manfred (2016): Smart Sheriff gegen Smombies. *Nervenheilkunde* 35, 205-212.
- Strasser, Thomas (2018a): *Mind the App 2.0. Inspiring internet tools and activities to engage your students*. Innsbruck, London: Helbling English.
- Strasser, Thomas (2018b): Plug into the digital zeitgeist. *EL Gazette* 42.
- Strasser, Thomas (2020): Explainer videos in the EFL-classroom. *Praxis Englisch* 2, 29-33.
- Thaler, Engelbert (2012): *Englisch Unterrichten: Grundlagen, Kompetenzen, Methoden*; Berlin: Cornelsen.
- Thornbury, Scott (2011): *How to teach grammar*. Harlow: Longman, Pearson Education.
- Voss, Tamar & Wittwer, Jörg (2020): Unterricht in Zeiten von Corona: Ein Blick auf die Herausforderungen aus der Sicht von Unterrichts- und Instruktionsforschung. *Unterrichtswissenschaft* 48, 601-627.
- Wolf, Karsten D. (2015): Bildungspotenziale von Erklärvideos und Tutorials auf YouTube. *Medien + Erziehung merz* 1, 30-36.