

4 Empirische Untersuchung der Lerneffekte

4.1 Grundlagen des Forschungsdesigns

4.1.1 Studie zur *Discovery Tour by Assassin's Creed: Ancient Egypt*

Als wichtigste Referenzstudie des vorliegenden Projekts kann die Arbeit des Geschichtsdidaktikers Marc-André Éthier bezeichnet werden. Als Professor an der Université de Montréal beschäftigt er sich seit mehreren Jahren mit dem Einsatz von populären Geschichtsdarstellungen im Unterricht. Sein Interesse an Digitalen Spielen rührt daher nicht aus eigenen Spielerfahrungen, sondern von der Popularität des Mediums, die er persönlich bei Schulbesuchen wahrgenommen hat.¹ Besonders das Franchise *Assassin's Creed* schien von Lehrpersonen unterschiedlicher Stufen und Niveaus eingesetzt zu werden. Éthier suchte den Kontakt zum federführenden Studio des *Ubisoft*-Konzerns, das sich ebenfalls im kanadischen Montreal befindet. Sein Ziel war der Kontakt zu den konsultierten HistorikerInnen des Franchise, der über Maxime Durand hergestellt wurde.² Éthier interessierte sich besonders für ihre Arbeit an und ihren Einfluss auf die Digitalen Spiele. Die Interviews wurden im Zeitraum von 2016 bis 2020 durchgeführt und anschließend unter dem Namen *Le jeu et l'histoire: Assassin's Creed vu par les historiens*³ veröffentlicht. Im Rahmen dieser Gespräche erfuhr er von dem geplanten Bildungsmodus, der den Kern seiner empirischen Studie bilden sollte. Erste Ergebnisse wurden 2018 auf der internationalen Konferenz *Games for Change* in New York vorgestellt und schließlich 2022 in einer kanadischen Fachzeitschrift publiziert.⁴ Forschungsgegenstand ist die 2018 erschienene *Discovery Tour by Assassin's Creed: Ancient Egypt* (Ubisoft Entertainment 2018), die auf dem Grundspiel *Assassin's Creed Origins* (Ubisoft Montreal 2017) basiert. Éthier wies darauf hin, dass der Konzern weder Einfluss auf das Forschungsdesign oder die Ergebnisse gehabt, noch Honorare angeboten habe. Nur die benötigte Hardware und die Software wurde im Vorfeld der Untersuchung zur Verfügung gestellt.⁵ Dies war einerseits erforderlich, da die Studie in den Räumlichkeiten der neun ausgewählten Schulen stattfand. Zudem war die *Discovery Tour by Assassin's Creed: Ancient Egypt* zum Zeit-

1 Éthier und Durand 2018. *Beyond Gaming*. 13:50–14:30.

2 Éthier und Lefrançois 2018. „Quelle lecture les historiens d'Assassin's Creed font-ils de ce jeu vidéo d'histoire?“

3 Marc-André Éthier und David Lefrançois. *Le jeu et l'histoire: Assassin's Creed vu par les historiens*. Montréal, Québec: Del Busso éditeur, 2020.

4 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“.

5 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 35.

punkt der Studie noch nicht im Handel erhältlich. Die Erweiterung erschien weltweit am 20. Februar 2018⁶, während erste Ergebnisse der Studie bereits fünf Tage vorher durch die Université de Montréal publik gemacht worden waren.⁷

Die Arbeit kann als exemplarische Fallstudie zu historischen Lerneffekten Digitaler Spiele bezeichnet werden. Éthier rekrutierte hierzu 321 Lernende aus neun verschiedenen Schulen (aus einer öffentlichen und acht privaten) der frankophonen Gebiete Kanadas. Die Rektoren oder Lehrpersonen übten keinerlei Druck auf die Lernenden hinsichtlich der Studienteilnahme aus, vielmehr geschah diese auf freiwilliger Basis.⁸ Dabei wurden unterschiedliche Stufen einbezogen, um einen möglichen Einfluss des Alters auf die Lerneffekte zu überprüfen. Betrachtet man die Daten Éthiers, fällt zudem auf, dass die Studie mehrheitlich von (männlichen) Schülern durchgeführt wurde.

Tab. 4-1: Demografische Angaben zu den Teilnehmenden der Studie Éthiers⁹

<i>Alter</i>		
12–13 (n = 194)	14–15 (n = 98)	16–17 (n = 29)
<i>Geschlecht</i>		
Jungen (n = 231)	Mädchen (n = 90)	

Die Methode basiert auf dem Vergleich der präsentierten Inhalte einer Tour des Digitalen Spiels mit dem angeleiteten Unterricht einer Lehrperson. Hierfür wurden jeweils Gruppen von 40 Teilnehmenden gebildet. Diese absolvierten gemeinsam einen *Pre-Test*, der zur Feststellung des Vorwissens der teilnehmenden SchülerInnen diente. Anschließend wurden die Teilnehmenden zufällig in zwei gleich große Gruppen eingeteilt. In dieser Lernphase erhielt die Untersuchungsgruppe zwölf Minuten Zeit, um die Tour *Die große Bibliothek von Alexandria* der *Discovery Tour by Assassin's Creed: Ancient Egypt* zu absolvieren.¹⁰ Dies geschah selbstständig und ohne die Begleitung oder Anleitung von Fachpersonen. Die zweite Hälfte – die Kontrollgruppe – erhielt die Präsentation einer Lehrperson. Der Lehrpersonenvortrag wurde ebenfalls auf eine maximale Dauer von zwölf Minuten be-

6 o. A. „Discovery Tour by Assassin's Creed Ancient Egypt – PC (Digital)“. Ubisoft Store, 2023. https://store.ubisoft.com/us/discovery-tour-by-assassins-creed-ancient-egypt/5a7d728d0c8ee45b54018f25.html?lang=en_US (aufgerufen am 23. Mai 2023).

7 Université de Montréal – Salle de Presse 2018. „Right on Target“.

8 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 18.

9 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 18.

10 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 21.

grenzt.¹¹ So wurde beiden Präsentationsformen die identische Zeit zur Verfügung gestellt.

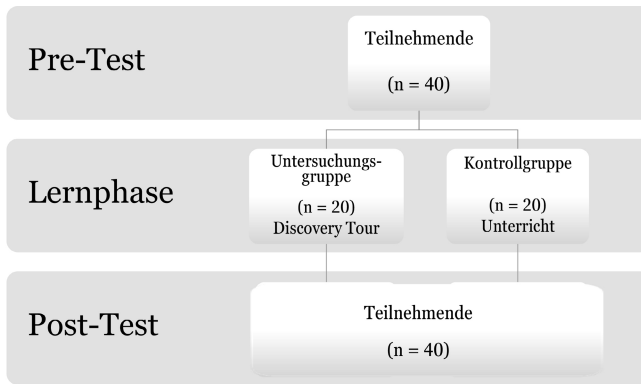


Abb. 4-1: Forschungsdesign der Fallstudie Marc-André Éthiers¹²

Laut Éthier waren die angebotenen Inhalte in allen Untersuchungs- und Kontrollgruppe identisch, da stets die gleiche Lektion und Lehrperson im Einsatz war.¹³ Hierzu wurden einige Screenshots und Texte des Digitalen Spiels in der Form von PowerPoint-Folien bereitgestellt. Diese wurden von der Lehrperson zur Visualisierung des Kurzunterrichts verwendet. Für den abschließenden *Post-Test* wurden die beiden Gruppen wieder zusammengeführt. Wie bereits bei der Durchführung des *Pre-Tests* existierte nur eine Version, die sämtliche Teilnehmenden ausfüllten. Um Auskunft über die subjektiven Lernerfahrungen zu erhalten, wurden zusätzlich qualitative Leitfaden-Interviews durchgeführt. Diese fanden nach Abschluss des *Post-Tests* mit Teilnehmenden beider Gruppen statt.¹⁴

Für die Auswertung des Lernzuwachses wurden die durchschnittlichen Resultate des *Pre-* und des *Post-Tests* miteinander verglichen. Statt einer Punkteskala entschied sich Éthier bei der Vorstellung auf der Konferenz für eine adressatengerechte Präsentation.¹⁵ Dazu wurde der Anteil der korrekten Antworten im Verhältnis zur Anzahl der Fragen berechnet. Mit dieser Variante konnte der durchschnittliche Lernzuwachs beider Gruppen nicht nur in Prozent ermittelt, sondern gleichzeitig öffentlichkeitswirksam gegenübergestellt werden.

¹¹ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 21.

¹² Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 20.

¹³ Éthier und Durand 2018. Beyond Gaming. 17:40–18:52.

¹⁴ Éthier und Durand 2018. Beyond Gaming. 19:20–20:10.

¹⁵ Éthier und Durand 2018. Beyond Gaming. 18:50–19:20.

Tab. 4-2: Deskriptive Analyse der Fallstudie Marc-André Éthiers¹⁶

Phase	Untersuchungsgruppe (Discovery Tour)	Kontrollgruppe (Unterricht mit Lehrperson)
Pre-Test	21 %	22 %
Post-Test	44 %	51 %

Auffallend ist, dass der durchschnittliche Wert der *Pre-Tests* beider Gruppen nahezu übereinstimmt. Daher ist anzunehmen, dass die Kontroll- und die Untersuchungsgruppen über ein ähnliches Vorwissen verfügten. Wenn man nun die Resultate der beiden Testphasen vergleicht, zeigt sich eine Steigerung von 23 Prozent bei der Gruppe der Spielenden, während die unterrichteten Teilnehmenden gar eine Steigerung von 29 Prozent aufweisen. Auf dieser Grundlage lässt sich argumentieren, dass die Lehrperson das historische Wissen effektiver vermittelt als das Digitale Spiel. Diese Erkenntnis lässt sich bereits anhand der deskriptiven Analyse gewinnen.

In der später publizierten Studie finden sich zudem detaillierte Angaben zu den gemessenen Mittelwerten der Untersuchungs- ($m_{\text{pre}} = 2,16$; $m_{\text{post}} = 4,97$) und der Kontrollgruppe ($m_{\text{pre}} = 2,07$; $m_{\text{post}} = 5,45$).¹⁷ Ein medienvergleichender Ansatz, wie Mayer ihn vorschlägt (siehe 4.1.2), wird jedoch nicht implementiert. Stattdessen nutzt Éthier die Methode der abhängigen *t*-Tests, um die durch die Lernphase bedingten Lerneffekte zu bestimmen (Tabelle 4-3). Somit werden die optisch beobachtbaren Veränderungen der Mittelwerte nach der Lernphase auf ihre statistische Signifikanz überprüft.¹⁸ Zudem sind Varianzanalysen enthalten, die Interesse, Alter und Schule der Teilnehmenden berücksichtigen.¹⁹ Das Folgende fokussiert jedoch auf jene Daten, die zentrale Forschungsfragen der vorliegenden Studie betreffen, das heißt sich ausschließlich auf die Resultate der *Pre*- und *Post-Tests* beziehen (siehe 4.2.1).

16 Éthier und Durand 2018. Beyond Gaming. 18:50–19:20.

17 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 26.

18 Jürg Schwarz. „t-Test für abhängige Stichproben“. Universität Zürich. Methodenberatung, 2023. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/ttestabh.html (aufgerufen am 31. Januar 2024).

19 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 36–44.

Tab. 4-3: t-Test für abhängige Stichproben der Studie Éthiers (Auswahl)²⁰

	Mittelwert	SD	Std.-Fehler Differenz	t	df	Cohens d
(Pre – Post)	-2,79	2,02	0,11	-24,76	320	-1,70
(Post-Gruppe)	4,71	1,72	0,09	47,58	303	4,02

Hinweis: Für beide Vergleichspaare lässt sich eine signifikante Differenz beobachten
($n = 321$; $p < 0.05$).

Laut Éthier lassen sich anhand der ausgewählten Vergleichspaare drei Erkenntnisse gewinnen. Erstens ist der bereits vermutete Lernzuwachs anhand des deskriptiven Mittelwertvergleichs nun statistisch belegt ($t = -24,76$; $d = -1,70$). Sowohl der Unterricht als auch das Spielen der Tour haben zu einem signifikanten Wissenserwerb geführt. Betrachtet man den Lernzuwachs beider Gruppen, zeigen sich im Vergleich der *Pre*- und *Post*-Tests signifikant positive Effekte.²¹ Zweitens konnte gezeigt werden, dass die Mittelwerte vor der Lernphase einen Einfluss auf die Mittelwerte nach der Lernphase ausüben (Tabelle 4-3: *Pre-Post*). Dies bedeutet, dass eine niedrigere Punktzahl im *Pre-Test* eine niedrigere Punktzahl im *Post-Test* zur Folge hatte.²² Demensprechend kann man die Behauptung aufstellen, dass vorhandenes Vorwissen den Wissenserwerb in der Lernphase erleichtert. Drittens konnten die gruppenbedingten Unterschiede im *Post-Test* bestätigt werden (Tabelle 4-3: *Post-Gruppe*). In Verbindung mit den Mittelwerten der Untersuchungs- und der Kontrollgruppe lässt sich darauf hinweisen, dass der Unterricht das Wissen besser vermittelt als die Tour. *Discovery Tour by Assassin's Creed: Ancient Egypt* kann den regulären Geschichtsunterricht somit nicht ersetzen. Éthier begründet dies mit der zwischenmenschlichen Interaktion, der strukturierteren Instruktion der Lehrperson und der Erwartungshaltung der Teilnehmenden.²³ Die Lernenden waren es gewohnt, Informationen durch Lehrervorträge aufzunehmen, zu verarbeiten und in Prüfungssituationen abzurufen. Digitale Spiele werden jedoch primär zu Unterhaltungszwecken verwendet und ohne fachliche Begleitung auch als solche behandelt. Die größeren Lerneffekte der Unterrichtssequenz gegenüber der Tour bestätigen die Vermutungen Éthiers.²⁴

²⁰ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 27.

²¹ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 25 f.

²² Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 31.

²³ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 32.

²⁴ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 34.

Wie bereits in der Keynote zu ersten Ergebnissen betont wird, waren der Fallstudie gewisse Grenzen gesetzt.²⁵ Die Fragen verlangten nur das Abrufen von deklarativem Wissen, wie es in den unteren Taxonomiestufen Blooms angesiedelt ist. Das Verständnis von historischen und politischen Konzepten und Prozessen wurde nicht überprüft.²⁶ Somit wurden auch keine eigenen Sachurteile oder Werturteile der Teilnehmenden verlangt, die zentrale Kompetenzen des historischen Lernens darstellen.²⁷ Zudem wurden die *Post-Tests* unmittelbar nach den Lernphasen durchgeführt, was eine Überprüfung der langfristigen Lerneffekte unmöglich machte. Neben den von Éthier angesprochenen Limitationen, könnten weitere Aspekte aufgeführt werden. So wurden die meisten Fragen des *Pre-Tests* für den *Post-Test* unverändert übernommen.²⁸ Ersterer überprüfte somit nicht nur das Vorwissen, sondern diente teilweise bereits als Lerngelegenheit. Die Teilnehmenden wurden somit auf die Inhalte der Lernphase vorbereitet. Dieser *testing-effect* kann sich auf die Resultate des *Pre-* und des *Post-Test-Vergleichs* auswirken und sollte bei der Diskussion berücksichtigt werden.²⁹ Gleiches gilt für das Vorwissen der Lehrpersonen. Wenn sie über die Fragen des *Post-Tests* informiert waren, konnten sie die Schwerpunkte Ihres Vortrags dementsprechend anpassen. Diese Fokussierung muss dabei keineswegs willentlich geschehen, sondern kann unbewusst stattfinden. Gerade die Interviews der Kontrollgruppe betonten die Wirkung der Lehrpersonen, da sie wichtige Inhalte herausgestellt haben.³⁰

Dennoch lässt sich abschließend festhalten, dass mit der *Discovery Tour by Assassin's Creed: Ancient Egypt* etwas über die Geschichte des alten Ägyptens gelernt werden kann. Diese Erkenntnis ist umso erstaunlicher, wenn man bedenkt, dass die Untersuchungsgruppe die Tour ohne zusätzliche Hilfe, Unterstützung oder Betreuung durchgeführt hat.³¹ Laut Éthier konnten die Teilnehmenden die zur Unterhaltung erschaffene Spielwelt zu Bildungszwecken einsetzen: „The research and detail that goes into a video game like Ubisoft's can be absorbed by students and help them get better marks.“³² Allerdings zeigt die Studie, dass auch moderne Digitale Spiele keinen gleichwertigen Ersatz für professionell angeleiteten Geschichtsunterricht darstellen. Die positiven Lerneffekte der Untersuchungsgruppe stimmen Éthier jedoch zuversichtlich, dass die neuen Medien in den Geschichtsunter-

25 Éthier und Durand 2018. *Beyond Gaming*. 20:10–22:45.

26 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 23.

27 Gautschi 2015. *Guter Geschichtsunterricht*. S. 50.

28 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 21 f.

29 Mayer 2014. *Computer games for learning*. S. 39.

30 Éthier und Durand 2018. *Beyond Gaming*. 19:30–19:50.

31 Université de Montréal – Salle de Presse 2018. „Right on Target“.

32 Université de Montréal – Salle de Presse 2018. „Right on Target“.

richt einfließen sollten. Seiner Meinung nach dürfte die *Discovery Tour* im hybriden Einsatz, das heißt mit Begleitung durch die Lehrperson, die größten Lerneffekte erzielen. Sein Ziel ist es, zukünftig weitere Studien zu Digitalen Spielen im Geschichtsunterricht durchzuführen, und er ruft die Fachdidaktik dazu auf, sich intensiver mit der Thematik auseinanderzusetzen.³³

4.1.2 Der Evidence-Based Approach nach Mayer

Richard E. Mayer ist Professor für Psychologie an der Universität von Kalifornien in Santa Barbara. Sein Forschungsschwerpunkt liegt auf dem Lehren und Lernen mit digitalen Anwendungen. Im Laufe seiner Arbeit sind die Digitalen Spiele zunehmend in den Fokus gerückt. In seiner Monografie *Computer games for learning: an evidence-based approach*³⁴ fasst er seine Erkenntnisse zur empirischen Beforschung des Mediums zusammen. Laut eigener Aussage begegneten ihm wiederholt gegensätzliche Aussagen zum Nutzen von Digitalen Spielen in Bildungskontexten. Einerseits werden sie von FürsprecherInnen über alle Maßen gelobt, während Gegenstimmen kaum positive Eigenschaften entdecken können. Laut Mayer basieren die meisten Aussagen auf Anekdoten, die sich mehr auf Wahrnehmungen als auf wissenschaftliche Grundlagen berufen.³⁵ Daher möchte er mit seinem Buch ein Basiswerk zur empirischen Beforschung der Lerneffekte Digitaler Spiele liefern. Enthalten ist auch der *media comparison approach*³⁶, der die Grundlage der vorliegenden Arbeit bildet. Dieser wird im nächsten Kapitel genauer vorgestellt (siehe 4.1.3).

Vorher müssen die theoretischen Grundlagen geklärt werden, die Mayer für die evidenzbasierte Forschung festlegt. Er fasst hierzu die aus seiner Sicht wichtigen Prinzipien erfolgreicher Studiendesigns zusammen:

33 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 34. Die ersten Ergebnisse seiner weiteren Arbeit liegen bereits vor, konnten jedoch für die vorliegende Studie nicht mehr berücksichtigt werden: Marc-André Éthier und David Lefrançois. „Chapter 1 Historical Video Games and Teaching Practices“. In „Assassin’s Creed“ in the Classroom, herausgegeben von Erik Champion und Juan Francisco Hiriart Vera, 11–34. Bd. 15. Video Games and the Humanities. München: De Gruyter Oldenbourg, 2023. <https://doi.org/10.1515/9783111253275-002> (aufgerufen am 26. Januar 2024).

34 Mayer 2014. *Computer games for learning*. S. 42 f.

35 Mayer 2014. *Computer games for learning*. S. xi f.

36 Mayer 2014. *Computer games for learning*. S. 225–249.

1. Signifikante Fragen formulieren, die empirisch untersucht werden können

- Um der Falle der anekdotischen Evidenz zu entgehen, empfiehlt Mayer die Ausarbeitung einer klaren Hypothese. Sie sollte sich auf die Lerneffekte fokussieren und einem der drei Genres der Beforschung Digitaler Spiele zuordnen lassen:
- Value-added: Hat die Anpassung eines Elementes des Digitalen Spiels Einfluss auf die Lerneffekte?
- Cognitive consequence: Kann das Digitale Spiel kognitive Kompetenzen fördern?
- Media comparison: Kann das Digitale Spiel größere Lerneffekte erzielen als andere Medien?

2. Die Studie auf relevante Theorie beziehen

- Das Forschungsdesign sollte sich auf theoretische Grundlagen stützen und über einen konzeptuellen Rahmen verfügen.

3. Methoden verwenden, die der Untersuchung der Forschungsfrage dienen

- Hier beschreibt Mayer die Passung von Forschungsfrage und Methoden. Für die vorgeschlagenen drei Genres empfiehlt er empirische Untersuchungen. Dabei muss beachtet werden, dass der Aufbau der Studie auf die Beantwortung der Forschungsfrage ausgerichtet ist.

4. Eine kohärente und schlüssige Argumentationskette aufstellen

- Die Schlussfolgerungen zu den Lerneffekten Digitaler Spiele müssen sich auf die gesammelten Daten beziehen. Hierzu sollten Auffälligkeiten und Schemen erkannt und analytisch verarbeitet werden.

5. Replizieren und Generalisieren der Studien

- Erst durch die wiederholte Ausführung können die Aussagen zu den Lerneffekten Digitaler Spiele allgemeine Gültigkeit erhalten. Diese müssen zu unterschiedlichen Zeitpunkten an unterschiedlichen Orten durchgeführt werden. Außerdem sollten auch bereits vorhandene Studien in die Schlussfolgerungen miteinbezogen werden.

6. Ergebnisse öffentlich zugänglich machen, um sie der Überprüfung und Kritik auszusetzen

- Die Lerneffekte Digitaler Spiele sind noch nicht breit erforscht worden. Gerade deshalb sollten entsprechende Studien veröffentlicht und der Peer-Review ausgesetzt werden. Nur so kann die weitere Entwicklung des Forschungsfeldes vorangetrieben werden.

Abb. 4-2: Mayers sechs Prinzipien der empirischen Forschung im Bildungswesen³⁷

Zusätzlich zu den sechs grundlegenden Prinzipien formuliert Mayer drei Charakteristika empirischer Versuchsanordnungen. Erstens müssen die Untersuchungs- und die Kontrollgruppe dieselbe Behandlung erfahren. Dieses Vorgehen wird von ihm als Versuchskontrolle (*Experimental Control*)³⁸ bezeichnet. Die einzige Variation stellt demnach jenes Element dar, dessen Lerneffekte überprüft werden sollen. Digitale Spiele bieten eine besondere Herausforderung für die Versuchskon-

³⁷ Eigene Zusammenstellung nach: Mayer 2014. Computer games for learning. S. 32–34.

³⁸ Mayer 2014. Computer games for learning. S. 35 f.

trolle. Das Medium ist interaktiv und non-linear, was die Vermittlung des identischen Informationsgehalts in einem bestimmten Zeitraum erschwert.³⁹ Diese Unsicherheit kann besonders bei der Methode des *media comparison approach* entstehen:

The results of media comparison studies must be interpreted in light of possible threats to experimental control, yet these threats also should be balanced against the practical usefulness of comparing two commonly used delivery modes.⁴⁰

Mayer beschreibt hier u. a. die Schwierigkeit, dass ein Lehrpersonenvortrag oder ein Video eine bestimmte Laufzeit aufweisen, während das Digitale Spiel durch bestimmte Interaktionen der Spielenden fortgesetzt wird. Das Zeitmanagement ist also den Spielenden überlassen und weitgehend individuell. Demnach sollten die Eigenschaften der eingesetzten Medien bereits in der Planungsphase der Studie berücksichtigt werden. Das zweite Charakteristikum ist die zufällige Zuteilung der Teilnehmenden zur Kontroll- bzw. zur Untersuchungsgruppe.⁴¹ Weder dürfen sie selbst eine Auswahl treffen, noch darf die Zuordnung anhand erhobener Daten des *Pre-Tests* oder bestimmter Eigenschaften vorgenommen werden. Drittens muss die Einheit der Messwerte der Forschungsfrage der Studie entsprechen. Für die Feststellung der Lerneffekte bedeutet dies, dass von sämtlichen Gruppen Mittelwert (m), Standardabweichung (SD) und Probengröße (n) angegeben werden müssen.⁴² Nur durch deren Erfassung erhält man valide Datensätze, die sich mit Resultaten anderer Studien vergleichen lassen. Zu diesem Zweck empfiehlt Mayer die Berechnung der Effektstärke (d), bei der die durchschnittliche Standardabweichung beider Gruppen (SD_{pooled}) miteinbezogen wird.⁴³

$$d = \frac{m_1 - m_2}{SD_{pooled}}$$

In seiner einflussreichen Meta-Studie *Visible Learning* definiert der Pädagoge John Hattie jene Lerneffekte als erwünscht, die über $d = 0,4$ liegen.⁴⁴ Mayer übernimmt diesen Schwellenwert und weist zudem auf die Berechnung der statistischen Signi-

³⁹ Mayer 2014. Computer games for learning. S. 36.

⁴⁰ Mayer 2014. Computer games for learning. S. 36.

⁴¹ Mayer 2014. Computer games for learning. S. 36.

⁴² Mayer 2014. Computer games for learning. S. 37.

⁴³ Mayer 2014. Computer games for learning. S. 45.

⁴⁴ Vgl.: John Hattie. Lernen sichtbar machen, herausgegeben von Wolfgang Beywl und Klaus Zierer, 2. Aufl. Baltmannsweiler: Schneider-Verl. Hohengehren, 2014. S. 293 f.

fikanz hin. Dieses Vorgehen entspricht einem t -Test mit unabhängigen Stichproben, wie er besonders in statistischen Anwendungen verbreitet ist.⁴⁵

Demzufolge müssen die Lerneffekte mithilfe dieser vergleichenden t -Tests für unabhängige Stichproben quantitativ ermittelt werden. Dabei können sie optional durch qualitative Analysen ergänzt werden.⁴⁶ Neben den soeben beschriebenen Prinzipien und Charakteristika der Best Practice, bietet Mayer ebenfalls die umgekehrte Perspektive an. Auf der Grundlage publizierter Studien definiert er acht Fehler, die bei der Untersuchung der Lerneffekte Digitaler Spiele auftreten können:

⁴⁵ Alexander Ly und Lotte Kehrer. „How to Conduct a Classical Independent Sample T-Test in JASP and Interpret the Results“. JASP – Free and User-Friendly Statistical Software, 18. April 2018. <https://jasp-stats.org/2018/04/18/how-to-conduct-a-classical-independent-sample-t-test-in-jasp-and-interpret-the-results/> (aufgerufen am 27. November 2023).

⁴⁶ Mayer 2014. Computer games for learning. S. 37.

1. Die falschen Merkmale messen

- Der wichtigste Faktor für das Gelingen der Studie ist die Entscheidung, was gemessen werden soll. Zur Feststellung der Lerneffekte ist es entscheidend, sich auf Veränderungen im Wissen und Können der ProbandInnen zu fokussieren. Hierzu können etwa die Post-Test-Resultate oder der Lernzuwachs von Pre- zu Post-Test verwendet werden.

2. Keine Kontrollgruppe einbinden

- Ohne eine Kontrollgruppe sind die Resultate der Studie nicht aussagekräftig. Im Zentrum stehen nicht die isolierten Lerneffekte der Untersuchungsgruppe, sondern deren Verhältnis zu den Ergebnissen der Kontrollgruppe.

3. Äpfel mit Birnen vergleichen

- Dieser Fehler bezieht sich auf die fehlende Versuchskontrolle. Bis auf die zu untersuchende Variable sollten die Bedingungen und die Durchführung für die Untersuchungs- und die Kontrollgruppe identisch sein.

4. Bereits existierende Gruppen verwenden

- Die zufällige Einteilung der ProbandInnen ist ein wichtiges Charakteristikum für das Gelingen der Studie. Daher sollte die Untersuchungs- und die Kontrollgruppe nicht aus bereits festgelegten Einheiten (bspw. Schulklassen) bestehen. Ebenso sollten weder die demografische Erhebung, noch die Resultate des Pre-Tests einen Einfluss auf die Zusammenstellung der Gruppen ausüben.

5. Mit der Anzahl der ProbandInnen knausern

- Eine empirische Studie kann nur über statistische Aussagekraft verfügen, wenn die Stichprobengröße einen bestimmten Wert erreicht. Idealerweise werden 100 ProbandInnen pro Gruppe vorgeschlagen. Ab einer Anzahl von 25 ProbandInnen pro Gruppe können Studien jedoch als aussagekräftig bezeichnet werden.

6. Sich auf Anekdoten stützen

- Das Medium der Digitalen Spiele wird in der öffentlichen Wahrnehmung mit unterschiedlichen Attributen versehen. Diese reichen von infantilen Killerspielen bis zu Heilsbringern des modernen Bildungswesens. Diese Vorstellungen beeinflussen ebenfalls die Spielerfahrung der ProbandInnen. Der Einbezug individueller Reaktionen und Wahrnehmungen der Spielenden kann Einblicke in die Rezeption der Digitalen Spiele bieten, lässt jedoch keine Rückschlüsse auf die tatsächlichen Lerneffekte zu. Grundlage der empirischen Untersuchung sollten daher stets die erhobenen Daten bilden.

7. Ein ineffektives Digitales Spiel verwenden

- Die Verwendungen von ungeeigneten Digitalen Spielen verletzt die Kriterien der Versuchskontrolle. Daher sollte das gewählte Digitale Spiel bereits auf seine Zweckmäßigkeit überprüft worden sein. Einerseits sollte es möglichst viele ProbandInnen ansprechen und motivieren. Gleichzeitig muss es Zugang zu Lerninhalten ermöglichen, deren fachliche Eignung bestätigt ist. Für die Untersuchung sollten die evaluierten Digitalen Spiele in den vorgesehenen Nutzungskontext eingebettet werden (etwa Unterricht oder regulärer Heimgebrauch).

8. Mittelwert, Standardabweichung und Probengröße nicht angeben

- Es genügt nicht die durchschnittlichen Ergebnisse der Untersuchungs- und der Kontrollgruppe miteinander zu vergleichen. Gerade bei divergierenden Gruppengrößen ist es unabdingbar, die Größe der Proben (n) und die Standardabweichung (SD) der Gruppen miteinzubeziehen. So kann die Effektgröße (d) berechnet und mit anderen empirischen Untersuchungen verglichen werden.

Abb. 4-3: Acht Untersuchungsfehler zu Lerneffekten Digitaler Spiele nach Mayer⁴⁷

⁴⁷ Eigene Zusammenstellung nach: Mayer 2014. Computer games for learning. S. 37–44.

4.1.3 Der Media Comparison Approach

Wie bereits im vorherigen Kapitel erläutert, schlägt Mayer drei verschiedene Ansätze für die empirische Beforschung Digitaler Spiele vor: *value-added*, *cognitive consequence* und *media comparison*.⁴⁸ Während die Erstgenannten sich auf das Digitale Spiel beschränken, werden bei Letzterem diverse Medien in das Forschungsdesign eingebunden. Die grundsätzliche Frage des *media comparison approach* ist, ob Digitale Spiele die Inhalte besser vermitteln können als andere, konventionellere Medien.⁴⁹

Die vermeintlichen Vorteile des neuen, interaktiven Mediums werden seit Jahrzehnten öffentlichkeitswirksam verbreitet. Sie würden Lernende besonders motivieren und seien aufgrund ihren spielerischen Zugangs wirksamer als Texte, Bilder oder Videos, die bereits im Unterrichtsalltag etabliert sind. Auf der Gegenseite finden sich jene kritischen Stimmen, die auf negative Auswirkungen wie Gewaltverherrlichung und das inhärente Suchtpotenzial hinweisen. Laut Mayer dient der *media comparison approach* in erster Linie dazu, diese Behauptungen mit wissenschaftlichen Methoden zu überprüfen:

Although strong claims have been made for the potential of games to revolutionize the way students learn, research evidence is needed to scientifically test those claims.⁵⁰

Der Ansatz birgt jedoch auch spezifische Herausforderungen, die sich in zwei Kategorien zusammenfassen lassen. Erstens ist es oftmals die mit dem Wechsel des Mediums einhergehende Veränderung der Methode, die einen bedeutenderen Einfluss auf die Lerneffekte ausüben kann.⁵¹ Betrachtet man etwa die Studie Éthiers (siehe 4.1.1), beschäftigt sich dort die Untersuchungsgruppe individuell mit dem Digitalen Spiel. Der Kontrollgruppe werden die Inhalte jedoch durch den Vortrag einer Lehrperson vermittelt. Somit wird nicht nur das eingesetzte Medium ausgetauscht, sondern ebenso der methodische Zugang modifiziert. Die zweite Herausforderung hängt unmittelbar mit dieser Problematik zusammen. Es ist anspruchsvoll, die identischen Inhalte in unterschiedliche Medien einzubinden.⁵² Diese Übereinstimmung ist jedoch eine unabdingbare Voraussetzung für die erfolgreiche Umsetzung des *media comparison approach*:

⁴⁸ Mayer 2014. Computer games for learning. S. 26–28.

⁴⁹ Mayer 2014. Computer games for learning. S. 28.

⁵⁰ Mayer 2014. Computer games for learning. S. 227.

⁵¹ Mayer 2014. Computer games for learning. S. 228.

⁵² Mayer 2014. Computer games for learning. S. 229.

Treatment group: People learn academic content by playing a game.

*Control group: People learn the same content with conventional media.*⁵³

Die inhaltlichen und methodischen Anforderungen verwendet Mayer als Filterkriterien seiner Meta-Studie. Hierzu recherchierte er in domänenspezifischen Online-Datenbanken nach Aufsätzen und Artikeln, die die Lerneffekte Digitaler Spiele mit denen anderer Medien vergleichen.⁵⁴ Aus seiner Zusammenstellung entfernte er anschließend jene Arbeiten, die nicht den von ihm definierten Prinzipien entsprechen. Es verblieben 32 Studien, die er in fünf inhaltliche Kategorien einteilte: Naturwissenschaft (16), Fremdsprachen (5), Mathematik (5), Sprache (3), Geistes- und Sozialwissenschaften (3).⁵⁵ Besonders für die beiden erstgenannten Kategorien zeigt sich eine signifikante Effektstärke. So lässt sich ein Median von $d = 0,69$ für die Untersuchungen mit naturwissenschaftlichen Inhalten feststellen, während jene zu Fremdsprachen einen Median von $d = 0,96$ aufweisen. Mayer hält den potenziellen Mehrwert der Digitalen Spiele in diesen Bereichen für vielversprechend.⁵⁶ Demgegenüber stehen die zusammengetragenen Resultate der Studien zu mathematischen Inhalten ($d = 0,03$). Hier lassen sich für die Untersuchungsgruppen (Digitale Spiele) im Vergleich zu den Kontrollgruppen (konventionelle Medien) keine signifikanten Effekte feststellen. Mayer zweifelt daher am Mehrwert der Digitalen Spiele für den Bildungsbereich Mathematik.⁵⁷ Für das Sprachlernen und die Geistes- und Sozialwissenschaften erlaubt er sich keine Beurteilung, da keine ausreichende Datenmenge vorhanden ist.

Er wünscht sich, dass gerade auf diesen Gebieten weitere Studien durchgeführt werden, die die Lerneffekte Digitaler Spiele mit denen anderer Medien vergleichen.⁵⁸ Dabei sollte unterschieden werden, ob sie als Selbstlernertools oder im Unterrichtskontext untersucht werden. Ähnlich wie Éthier sieht er hier das größte Potenzial für Digitale Spiele in der Bildung.⁵⁹ Das gilt ebenfalls, wenn sich herausstellen sollte, dass sie im isolierten Einsatz weniger effektiv sind als konventionelle Medien. Denn im Vergleich zu vielen dieser Medien werden Digitale Spiele von zahlreichen Lernenden als Freizeitbeschäftigung angesehen und dementsprechend freiwillig genutzt.⁶⁰

53 Mayer 2014. Computer games for learning. S. 28.

54 Mayer 2014. Computer games for learning. S. 229.

55 Mayer 2014. Computer games for learning. S. 232.

56 Mayer 2014. Computer games for learning. S. 232.

57 Mayer 2014. Computer games for learning. S. 238 f.

58 Mayer 2014. Computer games for learning. S. 244.

59 Université de Montréal – Salle de Presse 2018. „Right on Target“.

60 Mayer 2014. Computer games for learning. S. 244.

4.1.4 Historisches Lernen

In den letzten Jahren hat eine spezifische Auseinandersetzung mit Digitalen Spielen in der Geschichtsdidaktik eingesetzt (siehe 1.2). Der Zeitpunkt ist keineswegs zufällig, ist er doch eine Folge der zunehmenden Beschäftigung mit den digitalen Entwicklungen des 21. Jahrhunderts. Im interdisziplinären Vergleich kann man hingegen von einem Rückstand sprechen. Andere Fachgebiete haben sich bereits länger mit Fragen zu Digitalen Spielen und zum Lernen auseinandergesetzt.

„Unter Game-Based Learning versteht man die Verbindung von digitalen Spielen und Lernen.“⁶¹ Es wird mit gängigen Lerntheorien verknüpft, wobei besonders auf das intrinsische und extrinsische Motivationspotenzial sowie auf mögliche Selbstwirksamkeitserfahrungen der Spielenden hingewiesen wird.⁶² Verstehen wir die Geschichtsdidaktik als „Theorie und Praxis des historischen Lehrens und Lernens“⁶³, ist die Frage nach dem Nutzen für das historische Lernen jedoch zweifelsohne in dieser Disziplin zu verorten. Für die Geschichtsdidaktik existiert jedoch, anders als in anderen Disziplinen, kein einheitliches Kompetenzmodell. Einen Konsens in der Forschung bildet die Annahme, dass Geschichte nicht als Ansammlung von Ereignissen der Vergangenheit, sondern als „retrospektives Konstrukt, das sich jede Gegenwart und jeder Einzelne neu erarbeitet respektive aneignet“⁶⁴, verstanden wird. Die angesprochene Erarbeitung und Aneignung der Einzelnen werden durch das Geschichtsbewusstsein vollzogen. Dessen Entwicklung stellt das Ziel des historischen Lernens dar und macht somit ebenfalls die Kernaufgabe des Geschichtsunterrichts aus.⁶⁵ Es geht daher nicht um die bloße Akkumulation von reinem Faktenwissen, wie es dem Fach noch heute häufig vorgeworfen wird.⁶⁶ Ein ausgeprägtes Geschichtsbewusstsein kann den Zahlen, Daten und Fakten eine Bedeutung für die eigene Lebenswelt verleihen. Es ist ständigen Einflüssen ausgesetzt, die seine Beschaffenheit verändern können. Um sich weiter-

61 Michael G. Wagner und Sonja Gabriel. „Game-Based Learning – Computerspiele im Geschichtsunterricht“. In *Praxishandbuch Historisches Lernen und Medienbildung im digitalen Zeitalter*, herausgegeben von Daniel Bernsen und Ulf Kerber, 337–346. Opladen, Berlin, Toronto: Verlag Barbara Budrich, 2017. <https://doi.org/10.2307/j.ctvdf0ffd.32> (aufgerufen am 2. April 2020). Hier: S. 338.

62 Preisinger 2022. *Digitale Spiele in der historisch-politischen Bildung*. S. 91.

63 Baumgärtner 2019. *Wegweiser Geschichtsdidaktik*. S. 28.

64 Bernd Schönemann. „Geschichtsbewusstsein – Theorie“. In *Handbuch Praxis des Geschichtsunterrichts*, herausgegeben von Michele Barricelli und Martin Lücke, 98–111. 2. Aufl. Bd. 1, 2 Bde. Forum Historisches Lernen. Schwalbach: Wochenschau, 2017. Hier: S. 102.

65 Hans-Jürgen Pandel. „Geschichtsbewusstsein“. In *Wörterbuch Geschichtsdidaktik*, herausgegeben von Ulrich Mayer u. a., 80–81. 3. Aufl. Schwalbach/Ts: Wochenschau, 2014. Hier: S. 81.

66 Baumgärtner 2019. *Wegweiser Geschichtsdidaktik*. S. 76.

entwickeln zu können, muss das Geschichtsbewusstsein diese Einflüsse zielgerichtet verarbeiten. Jeismann definiert hierzu drei Dimensionen, als grundlegenden Dreischritt des historischen Lernprozesses: Analyse, Sachurteil und Wertung.⁶⁷

Die dabei getätigten kognitiven Prozesse des Geschichtsbewusstseins führen zur sogenannten historischen Sinnbildung.⁶⁸ Sie gilt als historisch, wenn in ihr verschiedene Zeitebenen gekoppelt werden. So erhalten die Erfahrungen der Vergangenheit Relevanz in der Gegenwart, um daraus Folgerungen für die Zukunft zu ziehen.⁶⁹ Die historische Sinnbildung ist somit gleichsam Aufgabe und aktive Leistung des individuellen Geschichtsbewusstseins. Doch wie lassen sich dessen Aktivitäten beschreiben, handelt es sich doch um den „psychischen Verarbeitungsmodus historischen Wissens“⁷⁰? Laut Rüsen ist es das Erzählen, durch das Ereignisse im Zeitverlauf Sinn erhalten.⁷¹ Dieses Argument wirkt nachvollziehbar, da das Erzählen sämtlichen Aspekten des Kollektivsingulars „Geschichte“ zugrunde liegt. Das gilt sowohl für den privaten Bericht über die letzten Sommerferien als auch für die wissenschaftlichen Darstellungen der Historiografie, die allesamt narrative Strukturen ausweisen.⁷² Das historische Erzählen wird dabei nicht nur als performativer Akt, also als tatsächliches Sprechen oder Schreiben, verstanden. Vielmehr handelt es sich um Denkopoperationen des Geschichtsbewusstseins, die „ebenfalls eine genuin narrative Funktion haben“⁷³. Rüsen definiert die vier mentalen Vorgänge der historischen Sinnbildung, die einzeln beschrieben werden können und dennoch miteinander verflochten sind: Wahrnehmung, Interpretation, Orientierung und Motivation.⁷⁴

Betrachtet man historisches Lernen vor diesem Hintergrund, wird deutlich, weshalb es nicht nur als Lern-, sondern auch als Denkmodell bezeichnet wird.⁷⁵ Im Zuge der Kompetenzorientierung im Bildungswesen kann diese Annahme, trotz des Fehlens eines einheitlichen Konzepts in der Geschichtsdidaktik, als Konsens gelten. In dieser Arbeit wird auf das Modell Peter Gautschis zurückgegriffen, da es sich an den vorgestellten Strukturen des Geschichtsbewusstseins orientiert

67 Baumgärtner 2019. Wegweiser Geschichtsdidaktik. S. 32.

68 Pandel 2014. „Geschichtsbewusstsein“. S. 81.

69 Rüsen 2020. Historische Sinnbildung. S. 76.

70 Pandel 2014. „Geschichtsbewusstsein“. S. 80.

71 Rüsen 2020. Historische Sinnbildung. S. 14.

72 Baumgärtner 2019. Wegweiser Geschichtsdidaktik. S. 18 f.

73 Jörn Rüsen. Historische Orientierung: Über die Arbeit des Geschichtsbewusstseins, sich in der Zeit zurechtzufinden. Köln, Weimar: Böhlau Verl., 1994. S. 217.

74 Rüsen 2020. Geschichte denken. S. 3.

75 Gautschi 2015. Guter Geschichtsunterricht. S. 42.

und sich zusätzlich auf die Prozesshaftigkeit des historischen Lernens fokussiert. Gautschi arbeitet hierzu die Operationen des Geschichtsbewusstseins in einen Zyklus ein:

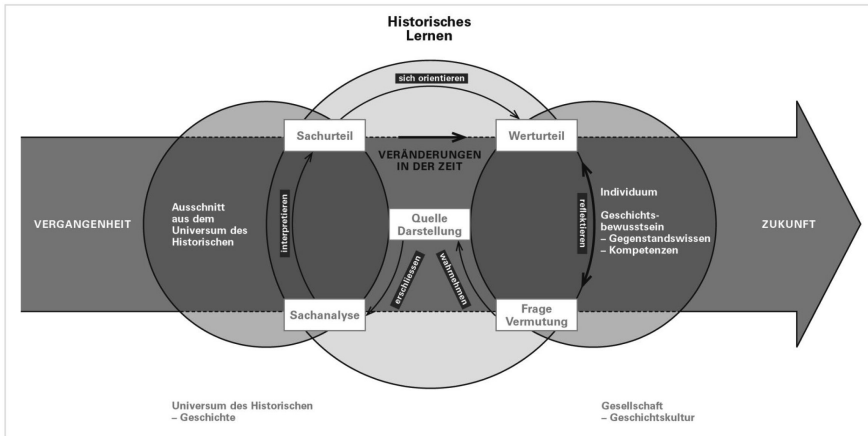


Abb. 4-4: Struktur- und Prozessmodell „Historisches Lernen“ nach Gautschi: Idealtypischer Ablauf⁷⁶

Betrachtet man die Elemente der Grafik, sind die Bezüge zur theoretischen Arbeit Jeismanns und Rüsens offensichtlich. Erstens ist hier die horizontale Zeitachse zu nennen, deren Verlauf durch den breiten Pfeil im Hintergrund dargestellt wird. Sie entspricht den drei zeitlichen Ebenen, die von Rüsen als essenziell für die historische Sinnbildung definiert wurden: Gegenwart, Vergangenheit und Zukunft.⁷⁷ Die zwei Letzteren sind durch die Beschriftung angegeben, während die Gegenwart im zentralen Kreis dargestellt ist (Abbildung 4-4). Die zyklische Anordnung des Modells im Uhrzeigersinn visualisiert den idealtypischen Prozess des historischen Lernens, indem

[...] von der Gegenwart aus Bezug auf die „andere“ Zeit der Vergangenheit genommen wird, wie in diesem Bezug der Zeitverlauf von der Vergangenheit über die Gegenwart in die Zukunft gedacht wird [...].⁷⁸

⁷⁶ Peter Gautschi, Jan Hodel und Hans Utz. „Kompetenzmodell für ‚Historisches Lernen‘ – eine Orientierungshilfe für Lehrerinnen und Lehrer“. Pädagogische Hochschule der FHNW, 2009. http://www.gesellschaftswissenschaften-phfhnw.ch/wp-content/uploads/2014/01/Gautschi_Hodel_Utz_2009_Kompetenzmodell.pdf (aufgerufen am 30. April 2020). S. 5.

⁷⁷ Rüsen 2020. Historische Sinnbildung. S. 76.

⁷⁸ Rüsen 2020. Historische Sinnbildung. S. 76.

Zweitens wurden die von Rüsen definierten mentalen Operationen der historischen Sinnbildung in das Modell eingearbeitet: Wahrnehmen, Interpretieren, Orientieren und Motivieren. Letzteres wird im idealtypischen Modell nicht explizit genannt, ist jedoch implizit vorhanden. Es kann als Anlass für Handlungen und Kommunikation in der Lebenswelt bezeichnet werden, die durch den Abschluss des Prozesses ausgelöst werden.⁷⁹ Die kleinen Pfeile des Kreises stellen somit die genannten Operationen Rüsens dar und bezeichnen allesamt Tätigkeiten des Geschichtsbewusstseins. Durch deren Vollzug können einzelne Teilschritte des historischen Lernprozesses abgeschlossen werden. Diese werden von Gautschi als Produkte bezeichnet und orientieren sich an den Dimensionen des Geschichtsbewusstseins nach Jeismann: Sachanalyse, Sachurteil und Werturteil.⁸⁰ Zudem fügt er dem Dreischritt noch einen weiteren hinzu, indem er die Aufmerksamkeit für den historischen Sachverhalt über die Formulierung einer Frage oder Vermutung einbezieht.⁸¹ Im idealtypischen Prozess werden durch die Reflexion des Werturteils neue Fragen und Vermutungen angestoßen. Wie jedoch Rüsen beschreibt, besteht zwischen den einzelnen Prozessen ein komplexer Zusammenhang.⁸² Gautschi bezieht diese Überlegungen in ein erweitertes Modell ein:

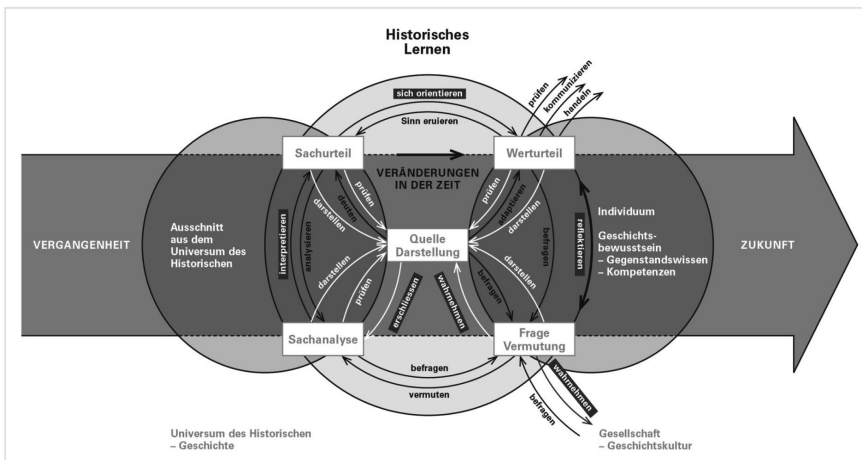


Abb. 4-5: Struktur- und Prozessmodell „Historisches Lernen“ nach Gautschi: Erweiterter Ablauf⁸³

⁷⁹ Gautschi 2015. Guter Geschichtsunterricht. S. 46: Im erweiterten Modell (Abbildung 4-5) ist das Motivieren als „prüfen, kommunizieren und handeln“ angegeben.

⁸⁰ Gautschi 2015. Guter Geschichtsunterricht. S. 43 f.

⁸¹ Gautschi 2015. Guter Geschichtsunterricht. S. 44.

⁸² Rüsen 2020. Geschichte denken. S. 3.

⁸³ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 6.

Die angesprochene Komplexität zeigt sich darin, dass historisches Lernen nicht nur vom idealtypischen Ausgangspunkt einer Frage oder einer Vermutung ausgehen kann, sondern in diverse Richtungen verlaufen kann.⁸⁴ Neben den gewünschten Prozessen werden im erweiterten Modell auch sogenannte „Kurzschlüsse“⁸⁵ dargestellt. Grundsätzlich gilt, dass für ein Gelingen des historischen Lernens, unabhängig vom Ausgangspunkt, der gesamte Zyklus durchlaufen und Abkürzungen vermieden werden sollten. Im Zentrum des Prozessmodells stehen die Quellen und Darstellungen, da sich sämtliche beschriebene Operationen des Geschichtsbewusstseins auf diese beziehen sollten (Abbildung 4-5). Durch diese Auseinandersetzung entsteht die Sinnbildung über das historische Erzählen, das den „Kern des Historischen Lernens“⁸⁶ ausmacht.

Gautschi überträgt diese Strukturen in sein Kompetenzmodell, wozu er vier Anforderungsbereiche definiert, um zu den Produkten des historischen Lernens zu gelangen: Die Wahrnehmungskompetenz führt zu Fragen und Vermutungen bezüglich der Quellen und Darstellungen. Die Erschließungskompetenz erlaubt die Überprüfung vorhandener Sachanalysen und befähigt dazu, eigene zu entwickeln. Die Interpretationskompetenz erlaubt es, Sachanalysen zu vergleichen, um zu einem historischen Sachurteil zu gelangen. Die Orientierungskompetenz zur Sinnbildung über Zeiterfahrung baut Einstellungen und Haltungen auf, die zu einem Werturteil führen.⁸⁷ Der oben definierte Kern der Sinnbildung, das historische Erzählen, steht im Zentrum des Kompetenzmodells. Daher spricht Gautschi auch von den „vier Teilbereichen der narrativen Kompetenz“⁸⁸.

84 Gautschi 2015. Guter Geschichtsunterricht. S. 44.

85 Gautschi 2015. Guter Geschichtsunterricht. S. 46.

86 Gautschi 2015. Guter Geschichtsunterricht. S. 46.

87 Gautschi 2015. Guter Geschichtsunterricht. S. 50–52.

88 Gautschi 2015. Guter Geschichtsunterricht. S. 50.

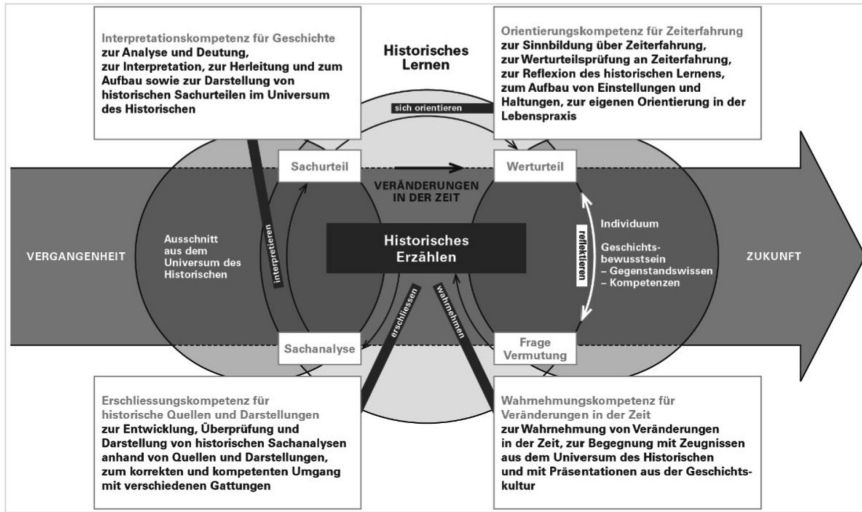


Abb. 4-6: Vier Kompetenzbereiche für „Historisches Lernen“⁸⁹

Diese vier Problemlösefähigkeiten sind entscheidend, damit der beschriebene Prozess des historischen Lernens gelingen kann, gleichzeitig sind sie jedoch dessen Ergebnis.⁹⁰ Bereits diese zusammenfassende Darstellung der Theorie macht deutlich, dass die Anforderungen an Lernende und Lehrende durch die Kompetenzorientierung gestiegen sind. Ein bloßes Auswendiglernen von Zahlen, Daten und Fakten, wie es über Jahrzehnte im Geschichtsunterricht praktiziert wurde, kann den Qualitätsansprüchen des historischen Lernens nicht genügen. KritikerInnen der Bildungsreformen bemängeln daher, dass kaum noch historisches Sachwissen vermittelt würde, was langfristig zu einer Schwächung des Fachs Geschichte führen müsse.⁹¹ Dem ist jedoch entgegenzusetzen, dass es sich stets um vergangenheitsbezogene Auseinandersetzungen handelt: „Alle Kompetenzbereiche müssen an historischen Inhalten ausgebildet, angewendet und ausdifferenziert werden.“⁹² Das oft geforderte historische Sachwissen ist somit ein willkommenes (Neben-)Produkt des historischen Lernens.

⁸⁹ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 10.

⁹⁰ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 11

⁹¹ Thomas Sandkühler. „Geschichtsunterricht im Widerstreit: Ein Blick in Presse und Onlinemedien“. Lernen aus der Geschichte, 14. September 2017. <http://lernen-aus-der-geschichte.de/Lernen-und-Lehren/content/13731> (aufgerufen am 31. Juli 2023).

⁹² Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 11.

4.1.5 Historisches Lernen und Digitale Spiele

Wie oben ausgeführt, stehen Quellen und Darstellungen im Zentrum des historischen Lernprozesses. Ist das Modell auch für interaktive, digitale Medien anwendbar oder gilt es nur für Texte, Bilder und Objekte, wie sie seit langem fester Bestandteil des Geschichtsunterrichts sind? Um mit den Worten Bettina Alavis zu fragen: „Lernen Schüler/innen Geschichte im Digitalen anders?“⁹³ Dies führt unweigerlich zur Kernfrage, wie die Geschichtsdidaktik, als „Theorie und Praxis des historischen Lehrens und Lernens“⁹⁴, mit diesen neuen Medien umgeht. Bedarf es zur Lösung dieser Herausforderungen gar eines neuen Fachgebiets, etwa einer eigenen Geschichtsdidaktik des Digitalen?

Anders als Digitale Spielen werden Texte, Bilder und Objekte in verschiedenen Ausprägungen seit Jahrzehnten in der grundlegenden fachdidaktischen Literatur behandelt. Häufig orientiert sich der Aufbau der Handbücher an der Unterscheidung von schriftlichen, bildlichen und gegenständlichen Medien.⁹⁵ In jüngeren Publikationen ist häufig Digitalität als ergänzende, alleinstehende Kategorie hinzugekommen. Grosch führt diese Ordnung auf die erziehungswissenschaftlichen Modelle zurück, die Unterrichtsmedien nach angesprochenen Sinnen strukturieren.⁹⁶ Aus der geschichtsdidaktischen Perspektive ist die Unterteilung nach äußerlichen Merkmalen jedoch nicht präzise genug. Zu vielfältig sind die Medien innerhalb der einzelnen Kategorien und dementsprechend auch ihr Wert für das historische Lernen.⁹⁷ Folglich existieren diverse Ansätze in der Geschichtsdidaktik, die den Medienbegriff praxisorientierter definieren wollen. Pandel schlägt vor, Medien als Präsentationsformen von Geschichte zu bezeichnen, die nicht nach ihrer sinnlichen Wahrnehmung, sondern nach den drei Kategorien Quelle, Darstellung und Imagination zu unterscheiden sind.⁹⁸ Dies entspricht nahezu der Differenzierung,

93 Bettina Alavi. „Lernen Schüler/innen Geschichte im Digitalen anders?“ In *Geschichte lernen im digitalen Wandel*, herausgegeben von Marko Demantowsky und Christoph Pallaske, 3–16. Berlin, München, Boston: De Gruyter Oldenbourg, 2015. <https://doi.org/10.1515/9783486858662-002> (aufgerufen am 7. Mai 2020).

94 Baumgärtner 2019. *Wegweiser Geschichtsdidaktik*. S. 28.

95 Vgl.: Hilke Günther-Arndt und Meik Zülsdorf-Kersting. *Geschichts-Didaktik: Handbuch für die Sekundarstufe I und II*. 6. Aufl. Berlin: Cornelsen, 2014.

96 Waldemar Grosch. „Der Einsatz digitaler Medien in historischen Lernprozessen“. In *Handbuch Praxis des Geschichtsunterrichts*, herausgegeben von Michele Barricelli und Martin Lücke, 125–145. 2. Aufl. Bd. 2, 2 Bde. Forum Historisches Lernen. Schwalbach: Wochenschau, 2017. Hier: S. 126.

97 Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“.

98 Hans-Jürgen Pandel. „Medien“. In *Wörterbuch Geschichtsdidaktik*, herausgegeben von Ulrich Mayer u. a., 135–136. 3. Aufl. Schwalbach/Ts: Wochenschau, 2014. Hier: S. 136.

wie sie im Modell des historischen Lernens nach Gautschi auftritt. Entscheidend für die Unterteilung ist demnach der Authentizitätsgrad der Präsentationsformen. Quellen sind möglichst zeitnah zu den Ereignissen entstanden, während es sich bei Darstellungen um spätere Beschreibungen handelt.⁹⁹ Die Imagination sieht Pandel als Gegenstück zu den Quellen, da es sich um Darstellungen handelt, die „sich dabei fiktionale[r] Mittel[...] bedienen“¹⁰⁰. Die zusätzliche Kategorie wird von vielen FachdidaktikerInnen kritisch gesehen. Gelten nur Produkte von professionellen HistorikerInnen und DokumentarfilmerInnen als Darstellungen? Sind diese nicht ebenfalls perspektivische Rekonstruktionen der Vergangenheit? Wie lassen sich jene Kunstwerke oder Romane systematisieren, die häufig mehr Erkenntnisgewinn für das historische Lernen bieten als so manche Autorentexte in Sachbüchern? Offensichtlich jedoch sind die Kategorien meist fließend, weshalb Grosch die Imaginationen als Unterform der Darstellungen bezeichnet.¹⁰¹ Diese können, je nach Applikationskontext und entsprechender Fragestellung, selbst zu Quellen werden. Fest steht jedoch, dass wir Medien benötigen, um uns Geschichte anzueignen und sie zu vermitteln.¹⁰² Daraus lässt sich folgern, dass es sich bei sämtlichen Quellen und Darstellungen um Medien handeln muss. Sie sind zentral für das historische Lernen, weshalb man Geschichte als *Schulfach des Umgangs mit vergangenheitsbezogenen Medien* beschreiben könnte.¹⁰³ In der heutigen Lebenswelt sind es vor allem die digitalen und interaktiven Medien, in denen die Vergangenheit thematisiert und Geschichtsbilder verbreitet werden. Das vielzitierte Bonmot Bergmanns „So viel Geschichte wie heute war nie ...“¹⁰⁴ ist gegenwärtig wohl zutreffender als jemals zuvor. Das historische Lernen soll die Verknüpfung von Vergangenheit, Gegenwart und Zukunft ermöglichen. Diese Gegenwart wird von digitalen Medien dominiert und wird dies voraussichtlich bleiben. So sieht Demantowsky die Geschichtsdidaktik zwar vor großen Herausforderungen, stellt jedoch fest, dass es keine eigenständige digitale Geschichtsdidaktik geben kann.¹⁰⁵ Sie gehört damit unweigerlich in den definierten Aufgabenbereich der „Wissenschaft vom Geschichtsbewusstsein“¹⁰⁶.

99 Pandel 2014. „Medien“.

100 Pandel 2014. „Medien“.

101 Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 127.

102 Pandel 2014. „Medien“. S. 135.

103 Preisinger 2022. Digitale Spiele in der historisch-politischen Bildung. S. 103.

104 Zitiert nach: Christian Kuchler. „Bespäßung vs. Behandlung: Kino und Geschichtsunterricht“. Public History Weekly, 2, Nr. 15 (2014). <https://dx.doi.org/10.1515/phw-2014-1765> (aufgerufen am 27. Juli 2023).

105 Demantowsky 2015. „Die Geschichtsdidaktik und die digitale Welt“. S. 157.

106 Baumgärtner 2019. Wegweiser Geschichtsdidaktik. S. 28.

Diese kurze Erörterung lässt folgenden Schluss zu: Digitale Spiele als Ausprägung der digitalen Medien müssen ein Bestandteil der Geschichtswissenschaft, der Fachdidaktik und des Unterrichts sein. Allgemein wird ihnen, gerade auch von Lehrpersonen, oft skeptisch begegnet. Sind sie tatsächlich weniger wertvoll für das historische Lernen als traditionelle Medien? Diese Pauschalisierung verbietet sich nicht nur aufgrund der schierer Vielfältigkeit des Angebots, sondern auch wegen des dargelegten Mediendiskurses in der Geschichtsdidaktik: „Historisch lernen kann man an allen Medien, die entweder selbst aus der Vergangenheit stammen oder sich auf Vergangenes beziehen.“¹⁰⁷ Wir sollten uns daher nicht die Frage stellen, *ob* historisches Lernen mit und durch die Digitalen Spiele möglich ist, sondern *wie* es ermöglicht werden kann. Dazu müssen wir uns zuerst bewusstwerden, wie sie sich von analogen Medien unterscheiden. Laut Winnerling handelt es sich bei Digitalen Spielen um eine Remedialisierungsform. Anders als Texte treffen sie ihre Aussagen nicht nur mittels ihrer fixierten Abbildungen von Zeichen auf dem Papier, sondern durch „alles, was das Spiel abzubilden vermag – Text, Bild, Ton, Animation, Interaktion“¹⁰⁸. Isoliert betrachtet sind diese Bestandteile bereits in statischer Form bekannt und ihr Nutzen für das historische Lernen ist durch zahlreiche Modelle und Raster beschrieben worden. In ihrem konzentrierten Auftreten erwecken sie jedoch den Eindruck eines gänzlich neuen Regelsystems geschichtsbezogener Darstellung. Dem widerspricht Winnerling, er weist jedoch zugleich darauf hin, dass gerade die Interaktivität es erschwere, Digitale Spiele in eine lineare Form zu überführen.¹⁰⁹ Sind die Unterschiede also nicht so gravierend, wie es der allgemeine Diskurs zum Digitalen Wandel suggeriert?

So sieht es u. a. Grosch, der bloß einen technischen Unterschied zwischen analogen und digitalen Medien feststellt.¹¹⁰ Entscheidender sind für ihn deren Funktionen im historischen Lernprozess. Der vorliegende Forschungsgegenstand *DTAG* kann der Funktion einer E-Learning-Anwendung zugeteilt werden (siehe 3.3): Sie setzt unterschiedliche Informationsträger ein (Multimedialität), verwendet unterschiedliche Kodierungen (Multikodalität), spricht verschiedene Sinne an (Multimodalität) und ist steuerbar (Interaktivität).¹¹¹ Sie vereint somit fast sämtliche Charakteristika digitaler Medien und ihrer analogen Vorgänger in einer Anwendung.

107 Bernsen 2017. „Medien im Geschichtsunterricht“. Hier: S. 39.

108 Tobias Winnerling. „[Wenn die Aussage lautet:] Spielerisch Geschichte lernen? [bezieht sich das worauf?]“ MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, Nr. 28, Bildung gemeinsam verändern: Diskussionsbeiträge und Impulse aus Forschung und Praxis (2017), 19–27. <https://doi.org/10.21240/mpaed/28/2017.02.22.X> (aufgerufen am 1. April 2020). Hier: S. 23.

109 Winnerling 2017. „Wenn die Aussage lautet“. S. 24.

110 Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 125.

111 Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 133.

Doch so sehr diese Fülle an Reizen und Eindrücken Lernende überfordern kann, bleiben die mentalen Prozess des historischen Lernens und die dafür erforderlichen Kompetenzen bestehen.¹¹² Geschichte wird mit Digitalen Spielen demnach nicht anders gelernt als mit traditionellen Medien, da die Gedankenoperationen, wie Wahrnehmen, Analysieren, Interpretieren und Orientieren, weiterhin Gültigkeit besitzen.¹¹³

Wir benötigen weder ein neues Verständnis noch ein neues Modell des historischen Lernens. Ebenso sieht es Demantowsky, der vorschlägt, „Herausforderungen wie den digitalen Wandel im Rahmen und mit den Mitteln ihrer disziplinären Eigenlogik zu bearbeiten“¹¹⁴. Diese Eigenlogik beinhaltet den charakteristischen Umgang mit Digitalen Medien, wie er in den Kompetenzmodellen der Fachdidaktik vorgestellt wird. Gautschi definiert vier Bereiche, die für das erfolgreiche historische Lernen entwickelt werden müssen. Eine Orientierungshilfe für Lehrpersonen enthält ein anschauliches Beispiel, das den Ablauf des historischen Lernprozesses anhand einer Fotografie beschreibt.¹¹⁵ Auf der Grundlage der bisherigen Überlegungen sollten die nötigen Operationen auf die *DTAG* übertragen werden können (Tabelle 4-4). Selbstverständlich sind die vorgestellten Beispiele stark idealisiert. Dennoch machen sie deutlich, dass historisches Lernen nicht nur durch Quellen, sondern auch durch neue Formen aus der Geschichtskultur, wie etwa Digitale Spiele, angestoßen werden kann.¹¹⁶

Tab. 4-4: Idealtypischer Ablauf des historischen Lernprozesses nach Gautschi: Adaptiert für Digitale Spiele

Kompetenzbereiche nach Gautschi ¹¹⁷	Beispiel Fotografie ¹¹⁸	Beispiel <i>Discovery Tour: Ancient Greece by Ubisoft</i>
Wahrnehmungskompetenz für Veränderungen in der Zeit: Begegnung mit Zeugnissen aus dem Universum des Historischen und Präsentationen aus der Geschichtskultur.	Urs findet eine Fotografie und merkt, dass sie bereits älter ist. Er stellt sich eine Reihe von Fragen, u. a. wer gezeigt wird und aus welcher Zeit sie stammt. Anhand der Bildunterschrift vermutet er, dass es sich um seinen Vater handeln könnte?	Urs spielt das Digitale Spiel und erkennt, dass hier nicht die Gegenwart dargestellt wird. Er stellt sich eine Reihe von Fragen, u. a. ob hier eine tatsächliche Vergangenheit dargestellt wird und, wenn ja, um welche Zeit es sich handelt. Anhand der Kleidung

112 Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 145.

113 Alavi 2015. „Lernen Schüler/innen Geschichte im Digitalen anders?“ S. 6.

114 Demantowsky 2015. „Die Geschichtsdidaktik und die digitale Welt“. S. 158 f.

115 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 3 f.

116 Gautschi 2015. Guter Geschichtsunterricht. S. 46.

117 Gautschi 2015. Guter Geschichtsunterricht. S. 50–52.

118 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 3 f.

Kompetenzbereiche nach Gautschi ¹¹⁹	Beispiel Fotografie ¹²⁰	Beispiel <i>Discovery Tour: Ancient Greece by Ubisoft</i>
Erschließungskompetenz für historische Quellen und Darstellungen: Entwicklung, Überprüfung und Darstellung von historischen Sachanalysen anhand von Quellen und Darstellungen.	Urs fragt seinen Vater, der sich auf dem Foto erkennt. Der Bildhintergrund und die Beschriftung weisen auf ein Kinderfest im Jahr 1961 hin. Für weitere Informationen fragt er seine Großmutter. Sie sagt ihm, dass das Mädchen eine Freundin war, die mit ihrer Familie aus Italien hergezogen war.	der Menschen und der Gebäude der Spielwelt vermutet er, dass es sich um die Antike handeln könnte. Obwohl die Statuen farbig sind?
Interpretationskompetenz für Geschichte: Analyse und Deutung, zur Interpretation, zur Herleitung und zum Aufbau sowie zur Darstellung von historischen Sachurteilen im Universum des Historischen.	Im Internet recherchiert Urs, dass in den 1960er Jahren viele ItalienerInnen zum Arbeiten in die Schweiz kamen. Doch sie waren nicht sehr beliebt. Die Statistik zeigt, dass sich die Schweiz vom Einwanderungs- zum Auswanderungsland entwickelt hatte.	Als Urs weiterspielt, wird ihm bewusst, dass die Gebäude zwar aufwändig und teuer waren, jedoch die Macht der attischen Demokratie repräsentieren sollten. Urs vergleicht die Informationen mit Wikipedia und seinem Schulgeschichtsbuch. Sie scheinen zu stimmen. Nur ist die Akropolis etwas anders dargestellt. Der Parthenon etwa scheint viel kleiner gewesen zu sein, als er im Spiel dargestellt wird.
Orientierungskompetenz für Zeiterfahrung: Sinnbildung über Zeiterfahrung, zur Werturteilsprüfung an Zeiterfahrung, zur Reflexion des historischen Lernens, zum Aufbau von Einstellungen und Haltungen, zur eigenen Orientierung.	Urs ist beeindruckt, dass seine Großmutter das Mädchen gehütet hatte. Denn auch heute haben es viele Kinder von EinwanderInnen schwer. Er erfährt es selber auf der Straße oder wenn er die Tagesschau guckt. Das obwohl die Schweiz, wie früher, heute von dieser	Die Tour und die monumentale Darstellung sollen ihm als Spielendem zeigen, wie wichtig den Athenern die Demokratie war und was sie geleistet haben mussten, um diese zu erhalten. Urs weiß, dass es auch in der Schweiz große Bauten gibt, die wichtig für die Demokratie des

¹¹⁹ Gautschi 2015. Guter Geschichtsunterricht. S. 50–52.

¹²⁰ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 3 f.

Kompetenzbereiche nach Gautschi ¹¹⁹	Beispiel Fotografie ¹²⁰	Beispiel <i>Discovery Tour: Ancient Greece by Ubisoft</i>
tierung in der gegenwärtigen Lebenspraxis.	Migration profitiert. Er überlegt sich wie die Spannungen und Konflikte in Zukunft bewältigt und überwunden werden könnten.	Landes sind und die diese repräsentieren. Denn sie ist auch hier aktuell noch wichtig, wie er an den Feierlichkeiten zum 175. Jahrestag der Bundesverfassung im Fernsehen gesehen hat. Die Geschichte der Akropolis von Athen hat ihm gezeigt, dass die Demokratie nicht selbstverständlich ist, sondern darauf geachtet werden sollte, sie auch in Zukunft zu erhalten.

So lassen sich etwa kompetenzorientierte Aufgabenstellungen anhand des Modells des historischen Lernens kreieren.¹¹⁹ Zudem zeigt sich, dass die Quellen und Darstellungen niemals isoliert betrachtet werden können. Sach- und Werturteile können nur durch Prüfung und Vergleich kompetent formuliert werden. Wie bereits Grosch feststellte, gilt dies auch für multimediale E-Learning-Anwendungen und Simulationen.¹²⁰ Die Erkenntnisse entsprechen der vorgelagerten Strukturanalyse der *DTAG* (siehe 3.3). Obwohl sie neben der virtuellen Umgebung noch weitere Mediengattungen wie etwa Fotografien von Objektquellen oder Autorentexte in ihre Präsentation einbindet, handelt es sich insgesamt nur um eine mögliche Darstellung der griechischen Antike. Folgt man der Terminologie Demantowskys, wird die *DTAG* durch das Spielen bereits zum *Lerngegenstand erster Ordnung*. Dabei wird *im* digitalen Angebot gelernt.¹²¹ Übertragen in das Modell Gautschis, können hiermit vor allem die Wahrnehmung angeregt sowie Fragen und Vermutungen aufgeworfen werden. Durch die gelieferten Informationen können, je nach Inhalt und Aufbau der Touren, weitere Kompetenzbereiche gefördert werden und zur Entwicklung einer Sachanalyse führen (Abbildung 4-6). Ohne ein bereits gut ausgebildetes Geschichtsbewusstsein lassen sich jedoch kaum reflektierte Sach- und Werturteile ausdifferenzieren, geschweige denn überprüfen. Die einzigen in die Tour integrierten Feedbackmöglichkeiten bilden daher die Quizfragen, die sich bloß auf das vermittelte historische Sachwissen beziehen. Wie oben beschrieben (siehe 4.1.4), kann dies im gesamten Prozess des historischen Lernens erworben

¹¹⁹ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

¹²⁰ Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 145.

¹¹⁹ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

¹²⁰ Grosch 2017. „Der Einsatz digitaler Medien in historischen Lernprozessen“. S. 145.

¹²¹ Demantowsky 2015. „Die Geschichtsdidaktik und die digitale Welt“. S. 161.

werden. Das gilt ebenso für das unbegleitete und unangeleitete Spielen der *DTAG*, als Lerngegenstand erster Ordnung. Ob es jedoch gelingt, „Wissensbestände mithilfe von Lernspielen besser zu speichern als mit anderen Methoden“¹²², ist eine zentrale Fragestellung der vorliegenden Studie.

So löblich dieser offensichtliche Nutzen ist, bringt er doch die Problematik des bei Lernenden äußerst beliebten *Information Snacking*¹²³ mit sich. Für sie werden digitale Medien zu Nachschlagewerken. Ähnlich wie es Hodel für *Wikipedia* feststellte, werden die Informationen häufig nur kurz gesichtet und unreflektiert übernommen.¹²⁴ Die Lernenden sind dabei meist geübt im raschen Sammeln und Zusammenstellen von Faktenwissen. Das ist auch für die *DTAG* anzunehmen. Denn ebenso wie die beliebte Online-Enzyklopädie, können Digitale Spiele als ein „für die Gegenwart prägendes Medium der Wissensspeicherung“¹²⁵ betrachtet werden. Ohne fachliche Begleitung verstärkt die narrative Struktur der Touren die Vorstellung einer Geschichte als linearer Abfolge von Ereignissen (siehe 3.3.4). Dieser Effekt wird durch den *realist simulation style* und das Immersive verstärkt (siehe 3.3.1). Zusätzlich dürften die Bekanntheit des Franchise *Assassin's Creed* und die marketingwirksame Legitimation durch FachexpertInnen die Annahme festigen, es werde „gesichertes allgemeingültiges Wissen“¹²⁶ präsentiert (siehe 2.7.4).

Doch um ganzheitliches historisches Lernen zu ermöglichen und einen kompetenten Umgang mit digitalen Medien zu erlangen, muss das Angebot vergleichend analysiert, interpretiert und kritisiert werden. Hierdurch wird das Digitale Spiel zum *Lerngegenstand zweiter Ordnung*, wobei weniger *in*, sondern *an* ihm gelernt wird.¹²⁷ Entscheidend ist somit weniger das einzelne Medium, sondern die angewendeten Methoden. Die verantwortlichen HistorikerInnen von *Ubisoft* scheinen sich den Limiten ihres Produkts bewusst zu sein. Laut offizieller Beschreibung soll die *DTAG* kein allumfassendes, autodidaktisches Lehrmittel darstellen, sondern einen niederschweligen Zugang zur Geschichte bieten, der sich idealerweise

122 Bernhardt 2018. Das Spiel im Geschichtsunterricht. S. 244.

123 Astrid Schwabe. „Das World Wide Web als historisches Informations-Medium?: Ausgewählte Ergebnisse zur Nutzung der historischen Website Vimu.info“. In *Geschichte lernen im digitalen Wandel*, herausgegeben von Marko Demantowsky und Christoph Pallaske, 35–58. Berlin, München, Boston: De Gruyter Oldenbourg, 2015. <https://doi.org/10.1515/9783486858662-004> (aufgerufen am 2. April 2020). Hier: S. 57.

124 Jan Hodel. Verkürzen und Verknüpfen: Geschichte als Netz narrativer Fragmente: wie jugendliche digitale Netzmedien für die Erstellung von Referaten im Geschichtsunterricht verwenden. Bd. 5. *Geschichtsdidaktik heute*. Bern: hep der bildungsverlag, 2013. S. 311.

125 Schwarz 2023. Geschichte in digitalen Spielen. S. 111.

126 Hodel 2013. Verkürzen und Verknüpfen.

127 Demantowsky 2015. „Die Geschichtsdidaktik und die digitale Welt“.

in einem gesteigerten Interesse für die Epoche zeigt.¹²⁸ Das Digitale Spiel soll im Geschichtsunterricht behandelt werden und das Angebot an eingesetzten Medien erweitern. Hierfür sprechen, neben den eigenen Behauptungen, zusätzlich in Zusammenarbeit mit der McGill-Universität entwickelte Unterrichtsmaterialien.¹²⁹

Diese Ausführungen zeigen erneut die Bedeutung des durch Lehrpersonen aufbereiteten Geschichtsunterrichts. Besonders für das historische Lernen *an* digitalen Medien und für die oft beschriebenen, überfachlichen Medienkompetenzen. Demzufolge sollte der Geschichtsunterricht, im Zuge der allgemeinen Kompetenzorientierung in der Bildung, an Relevanz gewinnen. Umso überraschender ist es, dass die entsprechenden Bildungsreformen teilweise zu einer gegenteiligen Entwicklung geführt haben.¹³⁰

4.2 Voraussetzungen der Studie

4.2.1 Forschungsfragen

Der Ausgangspunkt der empirischen Untersuchung ist die Annahme, dass Digitale Spiele einen Einfluss auf das Geschichtsbewusstsein Jugendlicher und junger Erwachsener ausüben (siehe 4.1.5).¹³¹ Wie oben beschrieben, ist die *Discovery Tour* ein Sonderfall in der Industrie, da sie explizit zur Vermittlung von historischen Inhalten entwickelt wurde (siehe 2.7.3). Die grundsätzlichsste Frage sollte demnach lauten, ob das Digitale Spiel diese Anforderungen für die Zielgruppe der Studie erfüllen kann. Besonders wenn die Inhalte der *Discovery Tour*, anders als bei der Referenzstudie Éthiers, im privaten Umfeld, das heißt ohne Begleitung einer Fachperson gespielt, gelesen oder betrachtet werden. Dies entspricht dem Forschungsdesiderat, wie es etwa von Bernhardt und Preisinger formuliert wird.¹³² Können durch den Einsatz Digitaler Spiele bei Jugendlichen und jungen Erwachsenen Lerneffekte, gemessen an historischem Wissen, festgestellt werden? Es handelt sich hierbei um eine der zentralen Forschungsfragen der vorliegenden Studie (ZF1). Wie bereits dargelegt, ist das Digitale Spiel ein interaktives und multikodales Medium, welches Texte, Bilder und Ton als Informationsträger verwendet. Somit

128 Delalande 2021. „Interviews“. Hier: S. 179.

129 Shankar 2022. „Ubisoft partnering with McGill University on Assassin’s Creed Discovery Tour curriculums“.

130 von Matt 2023. „Krise im Geschichtsunterricht?“

131 Giere 2019. Computerspiele – Medienbildung – historisches Lernen.

132 Bernhardt 2018. Das Spiel im Geschichtsunterricht. S. 244. Und: Preisinger 2022. Digitale Spiele in der historisch-politischen Bildung. S. 89.

unterscheidet es sich von anderen Medien, wie etwa dem Text oder gar dem Video. Gestützt auf die Referenzstudie Éthiers und den *media comparison approach* Mayers ergibt sich die zweite zentrale Forschungsfrage. Wie in Kapitel 4.1.2 erläutert, wird zu diesem Zweck ebenfalls die statistische Effektstärke (d) berechnet. Entsprechend gilt für die hierfür verwendeten t -Tests mit unabhängigen Stichproben die Alternativhypothese, dass Digitale Spiele die historischen Inhalte effektiver vermitteln als die bereits im Geschichtsunterricht etablierten Medien (ZF2).

Neben den beiden genannten Schwerpunkten gilt es weitere Aspekte zu beleuchten. Wie der Name bereits insinuiert, dienen die ergänzenden Forschungsfragen der weiteren Analyse der Befunde. Unter Einbezug der Ergebnisse der *Pre*- und der *Post-Tests* können etwa Voraussetzungen gelingender Lernprozesse thematisiert werden, wie etwa den Einfluss von Präkonzepten und medialen Vorprägungen. Welchen Einfluss übt das Vorwissen zur griechischen Antike auf die Lerneffekte der Teilnehmenden aus (EF1)?¹³³ Zudem kann davon ausgegangen werden, dass der Umgang mit Quellen und Darstellungen im Geschichtsunterricht geübt wurde. Lassen sich die Unterschiede zwischen den verschiedenen Bildungsabschlüssen der Teilnehmenden erkennen (EF2)? Analog dazu ist es eine verbreitete Annahme, dass die regelmäßige Nutzung des Mediums einen Einfluss auf das Lernen mit und an Digitalen Spielen ausüben kann.¹³⁴ Zeigen sich tatsächlich messbare Unterschiede zwischen den Ergebnissen der regelmäßig Spielenden und denen der weniger erfahrenen Teilnehmenden der Studie (EF3)? Die vorliegende Studie kann diese Angaben zum Nutzungsverhalten nicht nur einem Vergleich unterziehen, sondern auch ihren Einfluss auf die Lerneffekte bestimmen.

Wie im Kapitel 4.3.6 erläutert wird, sind die drei offenen Aufgabenblöcke des Fragebogens in der Punktwertung nicht berücksichtigt worden. Der Schwerpunkt der Arbeit liegt auf der quantitativen Erhebung der Lerneffekte. Dennoch können die Ergebnisse mit den offenen Aufgaben zielgerichtet ergänzt werden.¹³⁵ Passend zu den zentralen Forschungsfragen sind folgende zwei Aspekte zu beleuchten. Weisen jene Teilnehmenden, die die erweiterten Anforderungen erfüllen, gemessen am historischen Wissen, gleichzeitig die größten Lerneffekte auf (OF1)? Und kann ein möglicher Einfluss des Mediums auf die Erreichung jener Kompetenzziele festgestellt werden (OF2)? Bezogen auf die Erhebung der Präkonzepte, lässt sich zudem überprüfen, ob es die Teilnehmenden mit stark ausgeprägtem Vorwissen zur griechischen Geschichte sind, die die erweiterten Anforderungen der offenen Aufgaben erfüllen oder gar übertreffen können (OF3)?

133 Vgl.: Gautschi 2015. *Geschichte lehren*. S. 39; Metzger und Paxton 2016. „Gaming History“. S. 557.

134 Schwarz 2021. „Geschichte im digitalen Spiel. Ein interaktives Geschichtsbuch“. S. 582.

135 Mayer 2014. *Computer games for learning*. S. 37.

4.2.2 Auswahl des Forschungsgegenstands

Bei der Beschäftigung mit der *DTAG* wurde bereits nach kurzer Zeit ersichtlich, dass das Digitale Spiel in seinem vollen Umfang nicht für die empirische Untersuchung geeignet ist. Bereits die angegebenen Durchschnittswerte der 30 verfügbaren Touren belaufen sich auf knapp zehn Stunden. Die Bewegungsmöglichkeiten innerhalb der Spielwelt und die zahlreichen Entdeckungsorte erhöhen die benötigte Spielzeit noch. Zudem lässt sich ein freier Spielprozess kaum kontrollieren. Laut Mayer ist dies jedoch unabdingbar, um das Digitale Spiel mit anderen Mediengattungen vergleichen zu können.¹³⁶ Da das gesamte Digitale Spiel inhaltlich und zeitlich zu umfangreich ist, um in einer empirischen Studie als Forschungsgegenstand behandelt zu werden, musste der Untersuchungsgegenstand somit auf eine Tour begrenzt werden. Dies entspricht ebenfalls dem Vorgehen der Referenzstudie Éthiers, die sich auf eine zwölfminütige Sequenz fokussiert.¹³⁷ Nach dem Absolvieren sämtlicher Inhalte der *DTAG* wurde *Thermopylen* aus den 30 vorhandenen Touren ausgewählt. Sie findet am Ort der gleichnamigen Schlacht aus den Perserkriegen statt und besteht aus acht Stationen. Als Guide führt Leonidas durch die Tour, der bei der Verteidigung der Passage im Jahr 480 v. u. Z. gefallen ist.¹³⁸ Das Digitale Spiel macht die Anwesenheit 49 Jahre nach seinem Tod dennoch möglich.

Für die Auswahl dieser Tour lassen sich drei Gründe nennen. Erstens weist sie eine ideale Länge für die Einbindung in die Studie auf. Von den Entwickelnden wird ihre durchschnittliche Spieldauer mit 20 Minuten angegeben. Durch mehrfache Durchgänge in verschiedenen Tempi ließ sich diese Einstufung bestätigen. Zweitens handelt es sich um eine Sequenz aus dem Kapitel *Schlachten und Kriege*. Die Strukturanalyse der *DTAG* hat gezeigt, dass die Touren dieser Kategorie einen thematischen und zeitlichen Fokus aufweisen, der eine kohärente Geschichtserzählung ermöglicht (siehe 3.3.4). Die Tour *Thermopylen* ist geografisch angebunden und stellt die Ursachen, die Entwicklung und die Folgen der Schlacht vor. Zudem wird im letzten Drittel auf die Wirkungsgeschichte der Schlacht hingewiesen, was das historische Ereignis in einen größeren Kontext einbettet. Sie entspricht somit den Anforderungen Mayers, der eine fachliche Eignung der Spielinhalte voraussetzt.¹³⁹ Daraus ergibt sich der dritte Grund der Auswahl. Es gibt wohl kaum ein Ereignis der griechischen Antike, welches eine derartige Verbreitung in der Geschichtskultur aufweisen kann. Bereits im antiken Griechenland wurde die Erzäh-

¹³⁶ Mayer 2014. Computer games for learning. S. 36.

¹³⁷ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 21.

¹³⁸ Will 2019. Athen oder Sparta. S. 37.

¹³⁹ Mayer 2014. Computer games for learning. S. 43.

lung tradiert und mythisiert. Spätestens seit der Frühen Neuzeit wurden die Ereignisse und Protagonisten aus Herodots *Historien*¹⁴⁰ in Gedichten und Gemälden verarbeitet. In den folgenden Jahrhunderten wurde der Kampf der Spartiaten von Politikern instrumentalisiert und in den Altertumswissenschaften idealisiert. Diese Vorstellungen halten sich bis in die Gegenwart und erreichten ihren Höhepunkt in der öffentlichen Wahrnehmung durch Frank Millers Comic *300*¹⁴¹ und dessen gleichnamige Verfilmung¹⁴². Spätestens im 21. Jahrhundert waren die Perserkriege somit auch einem Teil des nicht genuin geschichtsinteressierten Publikums bekannt. Auch das Digitale Spiel *Assassin's Creed Odyssey* nutzt die Schlacht als wirkmächtigen Einstieg in seine virtuelle Darstellung des antiken Griechenlands. Doch nicht nur die künstlerische Verarbeitung des Stoffes, sondern auch die gezielte Instrumentalisierung für politische Zwecke ist gegenwärtig weiterhin zu beobachten. Bei den gewalttätigen Unruhen am Kapitol im Jahr 2021 trugen einige Protestierende Verkleidungen, die an Spartiaten erinnern sollten.¹⁴³

Die Schlacht scheint auch nach knapp 2500 Jahren noch rezipiert zu werden, wobei sich ein breites Spektrum der möglichen Deutungen beobachten lässt.¹⁴⁴ Es wird daher angenommen, dass die Studienteilnehmenden über Präkonzepte und Alltagstheorien bezüglich der angebotenen Inhalte verfügen. Die Vielzahl der vorhandenen Narrative und die schiere Verbreitung in der Geschichtskultur machen die Tour *Thermopylen* zum idealen Untersuchungsgegenstand dieser Studie.

4.2.3 Rekrutierung der Teilnehmenden

Die Umfrage stand über einen Zeitraum vom 1. September 2021 bis zum 7. Januar 2022 zur Verfügung. Die Entscheidung, die Studie nicht in Präsenz durchzuführen, wurde bereits vor der einsetzenden Pandemie getroffen. Die Online-Umfrage ermöglichte eine standortunabhängige Rekrutierung der Teilnehmenden. So konnte die zeitliche Flexibilität der Teilnehmenden berücksichtigt und eine mögliche Laborsituation vermieden werden. Die Untersuchung sollte idealtypisch dem alltägli-

140 Herodotus. *Historien*. Übersetzt von Josef Feix, Bibliothek der alten Welt. Düsseldorf: Artemis & Winkler, 2004.

141 Frank Miller. *300*. Milwaukie, Or: Dark Horse Comics, 1999.

142 Snyder 2007. 300.

143 Simon Mallory und Sara Sidner. „Decoding the Extremist Symbols and Groups at the Capitol Hill Insurrection“. CNN, 11. Januar 2021, Abschn. Edition. <https://www.cnn.com/2021/01/09/us/capitol-hill-insurrection-extremist-flags-soh/index.html> (aufgerufen am 2. August 2023).

144 Mischa Meier. „Die Thermopylen-,Wanderer, kommst du nach Sparta?““. In *Die griechische Welt: Erinnerungsorte der Antike*, herausgegeben von Elke Stein-Hölkeskamp und Karl-Joachim Hölkeskamp, 98–113. 2. Aufl. München: C. H. Beck, 2019.

chen Nutzungsverhalten der Jugendlichen und jungen Erwachsenen entsprechen.¹⁴⁵ Hierzu wurde die Verwendung privater Geräte im häuslichen Umfeld zugelassen. Zudem ermöglichte diese Art der Durchführung eine geografische Öffnung der Zielgruppe, die sich auf die deutschsprachigen Länder Schweiz, Deutschland und Österreich fokussierte.

Durch die Nutzung des Anbieters *Qualtrics* konnte ein sicherer Zugangslink zur Umfrage erstellt werden, der beliebig oft und funktional verbreitet werden konnte. Zur Rekrutierung der Teilnehmenden wurden mehrere Kanäle genutzt. Zwangsläufig wurde der Link auf Social Media wie *Twitter*, *Instagram* und *LinkedIn* veröffentlicht. Hierzu wurde der Aufruf zur Teilnahme von mir persönlich, aber auch von den deutlich reichweitenstärkeren Konten meiner Betreuer Gerhard Lauer und Marko Demantowsky geteilt.¹⁴⁶ Letztere haben in ihren Lehrveranstaltungen an der Johannes-Gutenberg-Universität Mainz und der Universität Wien ebenfalls auf die Umfrage hingewiesen. Zusätzlich wurden weitere Bildungseinrichtungen angefragt. So wurde der Link während der Öffnung des Untersuchungsfensters auf dem Basler Bildungsserver angezeigt. An dieser Stelle sei erneut Alexandra Binnenkade für ihre Kooperation gedankt. Weiter wurden Gymnasiallehrpersonen des Kantons Aargau von Daniel Siegenthaler informiert. Als einzige direkt kontaktierte Schule wurde das Freie Gymnasium Basel ausgewählt. Aufgrund meiner dort absolvierten Praktika bestanden hier gute Kontakte zum damaligen Qualitätsverantwortlichen Michael Holzer. Zum Zweck der Randomisierung wurde jedoch bewusst darauf verzichtet, bereits bestehende Gruppen zu rekrutieren (beispielsweise Schulklassen).¹⁴⁷ Neben der Verbreitung über Social Media und dem Kontakt zu Bildungseinrichtungen in der Nordwestschweiz wurden Kanäle genutzt, die spezifisch zur Rekrutierung von Teilnehmenden für wissenschaftliche Arbeiten kreiert wurden. Namentlich sind dies die Plattformen *surveycircle.com*, *poll-pool.com* und *surveyswap.io*. Hier wurde die Umfrage anhand der geforderten Voraussetzungen kategorisiert und nach den entsprechenden Kriterien eingebettet.

Als zusätzlicher Anreiz bei der Rekrutierung erhielten die Teilnehmenden die Chance auf Gewinn eines von zehn Gutscheinen. Konkret handelte es sich um Guthabekarten der Anbieter *Sony*, *Nintendo* und *Microsoft*. Die Teilnehmenden hat-

145 Mayer 2014. Computer games for learning. S. 43.

146 Rees, Marvin [@rees_marvin]. „Spielend Geschichte lernen? Kann das Digitale Spiel ‚Discovery Tour by Ubisoft‘ Historisches Lernen ermöglichen? Die Umfrage zu meiner Studie an der Universität Basel ist ab jetzt online: https://acsurvey.qualtrics.com/jfe/form/SV_2uzifCMcq9BQBE #DiscoveryTour #AssasinsCreed #GameStudies“. Tweet. Twitter, 9. September 2021. https://twitter.com/rees_marvin/status/1435903790734090241 (aufgerufen am 13. Oktober 2023).

147 Mayer 2014. Computer games for learning. S. 41.

ten am Ende des Fragebogens die Möglichkeit, sich für die Verlosung anzumelden. Sie wurde über die Funktion „anonymisierte Auslosungen“¹⁴⁸ der Plattform *qualtrics XM* kreiert. Aus Sicherheits- und Datenschutzgründen erfolgte die Zustellung der Guthabenkarten über die eigens eingerichtete Funktionsmail der Universität Basel. Die Anmeldung erfolgte auf freiwilliger Basis und hatte keinen Einfluss auf den Abschluss der Umfrage.

4.3 Forschungsdesign und Fragebogen

4.3.1 Disposition

Auf Grundlage der zentralen Forschungsfragen, der Erkenntnisse der Referenzstudien und der eigenen Analyse des Gegenstands ergibt sich folgende Basisstruktur des Forschungsdesigns (Abbildung 4-7). Grundsätzlich folgt die Struktur somit dem *media comparison approach* Mayers:

Treatment group: People learn academic content by playing a game.

Control group: People learn the same content with conventional media.¹⁴⁹

Im Unterschied hierzu besteht die vorliegende Studie aus zwei Kontrollgruppen. So werden die Lerneffekte des Digitalen Spiels nicht nur mit denjenigen eines Textes, sondern auch mit denen einer Videoaufzeichnung der Tour (*Let's Play*) verglichen. Die wichtigsten Faktoren für das Gelingen dieses empirischen Ansatzes sind die vollständige Randomisierung der Gruppen und die inhaltliche Übereinstimmung der präsentierten Medien. Um die Umsetzung der Forschungserkenntnisse und deren Implementierung in die Studie zu erläutern, wird im Folgenden der Fragebogen vorgestellt. Er ist in sechs Blöcke unterteilt: Einführung, Mediennutzung, Demografie, *Pre-Test*, Lernphase, *Post-Test*, Abschluss. Die optionale Verlosung wird unabhängig von der Umfrage betrieben und deshalb hier nicht gesondert aufgeführt. Der vollständige Fragebogen findet sich im Anhang (II). Es gilt zu beachten, dass die angegebenen Nummerierungen der Fragen nicht der exakten Reihenfolge der Items oder Indikatoren entsprechen.

¹⁴⁸ o. A. „Erstellen anonymisierter Auslosungen“. Qualtrics Support, 2023. <https://www.qualtrics.com/support/de/survey-platform/common-use-cases-rc/creating-an-anonymized-affle/?rid=lang-Match&prevsite=en&newsite=de&geo=CH&geomatch=> (aufgerufen am 19. Oktober 2023).

¹⁴⁹ Mayer 2014. Computer games for learning. S. 28.

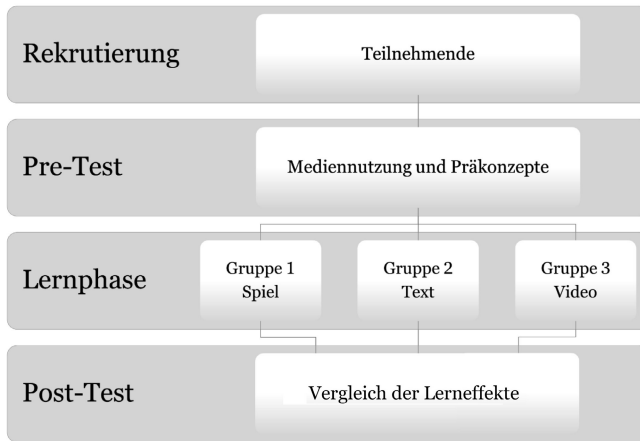


Abb. 4-7: Schematischer Aufbau des Forschungsdesigns (eigene Darstellung)

4.3.2 Einführung

Anders als die folgenden Blöcke besteht die Einführung nur aus einer Seite. Sie wird den Teilnehmenden angezeigt, wenn sie den Link zur Umfrage in einem der vorgestellten Rekrutierungskanäle öffnen (siehe 4.2.3). Die Einführung besteht aus einem Informationstext und einer abschließenden Frage. Zuerst werden grundsätzliche Aspekte der Studie angegeben und es wird auf die Anonymisierung und Löschung der Daten hingewiesen. Dann werden Zweck, Aufbau und Dauer der Umfrage kurz und prägnant beschrieben. Dabei wird nicht versprochen, dass die Teilnehmenden das Digitale Spiel erhalten oder spielen werden, sondern bloß, dass sie sich mit dessen Inhalten auseinandersetzen. Zudem wird auf die Verlosung der Guthabenkarte nach Abschluss des Fragebogens hingewiesen und eine Mailadresse für Fragen zur Studie angegeben. Zuletzt wird das Alter der Teilnehmenden überprüft, um sicherzustellen, dass sie der Zielgruppe der Studie angehören (E1). Sollten die Teilnehmenden sich nicht in der festgelegten Altersspanne von 16 bis 21 Jahren befinden, wird die Umfrage direkt beendet und der Versuch wird nicht für die Ergebnisse der Untersuchung berücksichtigt. Aufgrund der implementierten Sperrung mehrmaliger Durchführungen lässt sich der Umfragelink nach dem erzwungenen Abbruch nicht erneut öffnen. Dies verhindert einen großen Teil der unsachgemäßen Nutzung des Fragebogens (siehe 4.4).

4.3.3 Mediennutzung und Demografie

Wie in den Forschungsfragen dargelegt, ist die Erhebung des Nutzungsverhaltens hinsichtlich Digitaler Medien (*M3*) von erheblichem Wert für die Studie. Neben der Häufigkeit der Nutzung Digitaler Spiele (*M1*) wird der Fokus daher explizit auf die Verbindungen zum Fach Geschichte gelegt (*M6–M12*). Von besonderem Interesse kann hier die Angabe „über die Hinwendungsmotive zu digitalen Spielformaten durch genuin geschichtliches Interesse“¹⁵⁰ gelten (*M6*). Anschließend folgt eine Fragebatterie zu den Erfahrungen mit dem Franchise *Assassin's Creed* (*A1–A5*). Diese sogenannte Verzweigung wird jedoch nur angezeigt, wenn die Teilnehmenden angeben, bereits Teile der Reihe gespielt zu haben (*A1*). Besonderes Interesse gilt hier der Verbreitung der zuletzt veröffentlichten *Discovery Tours*, deren Popularität deutlich geringer scheint als diejenige der Haupttitel des Franchise (*A5*).¹⁵¹ So lassen sich individuelle Voraussetzungen und Wahrnehmungen der Teilnehmenden feststellen, die Informationen zur Rezeption Digitaler Spiele bereitstellen.

So wertvoll diese persönlichen Vorstellungen und Erfahrungen sind, können sie jedoch die Erhebung der Lerneffekte nicht ersetzen.¹⁵² Letztere bilden bekanntlich die primäre Forschungsfrage der Studie. Ähnlich verhält es sich mit dem Block zur Demografie (*D1+D2*). Dieser beschränkt sich neben der Frage zum Geschlecht auf die Abfrage des momentanen Bildungsstatus, um Angaben für die ergänzende Forschungsfrage (*EF2*) zu erheben.

4.3.4 Pre-Test

Der *Pre-Test* hat im Kontext der Studie zwei wichtige Funktionen. Er dient einerseits zur Erhebung des Vorwissens der Teilnehmenden. Andererseits bildet er somit die Grundlage der zentralen Forschungsfragen, die den Lernzuwachs der Teilnehmenden feststellen. Wie in der Referenzstudie Éthiers werden hierzu die Ergebnisse des *Pre-Tests* mit denjenigen des *Post-Tests* verglichen (siehe 4.1.1).

Der *Pre-Test* besteht somit aus allgemeinen *Multiple-Choice*-Fragen zur griechischen Geschichte. Dies wird in der Einführung des Blocks deklariert und somit transparent gemacht (*G0*). Die Aufgaben beinhalten neben Fragen zu Politik und Religion ebenso Elemente der Alltags- und Wissenschaftsgeschichte der griechi-

¹⁵⁰ Milch, Heinze und Friedrich 2020. „Digitale Spiele“. Hier: S. 270.

¹⁵¹ Vedran Radić. „The Highest-Selling Games Developed By Ubisoft Ranked (& How Much They Sold)“. Game Rant, 24. August 2021. <https://gamerant.com/ubisoft-highest-selling-games-ranked-how-much/> (aufgerufen am 5. Juli 2022).

¹⁵² Vgl. Abbildung 4-3.

schen Welt. In der ersten Aufgabe wird das Wirklichkeitsbewusstsein der Teilnehmenden diagnostiziert, indem sie historische Persönlichkeiten von mythischen und fiktiven Figuren unterscheiden (G1).¹⁵³ Für jede korrekte Persönlichkeit wird ein Punkt verliehen, während falsche Antworten einen Abzug zur Folge haben. Dies ergibt ein Maximum von 17 Punkten. Die nächste Kategorie verwendet wiederum die übliche Variante geschlossener Fragen, die jeweils vier Antwortmöglichkeiten bieten, aus denen eine korrekte ausgewählt wird (G2.1.–G2.10.). Pro richtige Antwort werden zwei Punkte verteilt. Aufgrund der höheren Anforderungen werden die folgenden fünf Aufgaben hingegen mit jeweils drei Punkten bewertet. Sie zeigen unterschiedliche Themenbereiche, die jeweils mit vier Aussagen beschrieben werden. Hier sollen die Teilnehmenden jeweils die falsche Aussage identifizieren (G3.1.–G3.5.). Zum Abschluss des *Pre-Tests* werden vier Entscheidungsfragen gestellt, die mit *richtig* oder *falsch* beantwortet werden können (G4.1.–G4.4.). Da hier nur zwei Optionen zur Auswahl stehen, werden diese jeweils nur mit einem Punkt bewertet.

Der *Pre-Test*-Block besteht somit aus insgesamt 20 Fragen, die eine maximale Punktzahl von 56 aufweisen. Er fungiert dabei als Diagnoseinstrument, wie es häufig von Lehrpersonen in einer prognostischen Evaluation zur zukünftigen Unterrichtsplanung genutzt wird. Insgesamt wurde darauf geachtet, inhaltliche Fragen zum Forschungsgegenstand *Thermopylen* zu vermeiden. Aufbauend auf der Diskussion der Studie Éthiers, wurde dabei gänzlich auf identische Fragen in *Pre*- und *Post-Test* verzichtet (siehe 4.1.1). Dies verhinderte Lerneffekte aus der Durchführung des *Pre-Tests* für den *Post-Test*, und damit den sogenannten *testing effect*¹⁵⁴.

4.3.5 Lernphase

Durch die Auseinandersetzung mit der Tour *Thermopylen* werden die Teilnehmenden in dieser Phase mit einer Darstellung aus der Geschichtskultur konfrontiert.¹⁵⁵ Der Aufbau der Lernphase ist auf die beiden zentralen Forschungsfragen ausgelegt. Können durch den Einsatz Digitaler Spiele bei Jugendlichen und jungen Erwachsenen überhaupt Lerneffekte, gemessen an historischem Wissen, festgestellt werden (ZF1)? Und ist es das Digitale Spiel, dass die historischen Inhalte effektiver vermittelt als bereits im Geschichtsunterricht etablierte Medien (ZF2)?

¹⁵³ Gautschi 2015. Geschichte lehren. S. 26 f.

¹⁵⁴ Mayer 2014. Computer games for learning. S. 39.

¹⁵⁵ Gautschi 2015. Guter Geschichtsunterricht. S. 50.

Daraus ergibt sich die Anordnung einer Untersuchungsgruppe (*S*) und der beiden Kontrollgruppen (*T/V*).¹⁵⁶ Erstere konsumiert die Inhalte der Tour *Thermopylen*, wie es von den Entwickelnden intendiert wurde. Um den intermedialen Vergleich zu ermöglichen, liest die Kontrollgruppe *T* einen Text, während *V* eine Videoaufzeichnung der Tour betrachtet. Es wird darauf geachtet, dass allen drei Gruppen möglichst identische Inhalte präsentiert werden, was aufgrund der bereits beschriebenen Eigenschaften von Digitalen Spielen jedoch nicht vollumfänglich möglich ist (siehe 3.2.1). Als Text wird ein bebildertes Transkript der Erzählstimme erstellt, die die Spielenden während der Tour informiert (siehe Anhang I). Die acht verwendeten Grafiken sind Screenshots aus der *DTAG*, die den Teilnehmenden als visuelle Unterstützung dienen. Der Text ähnelt somit einem AutorInnen-Text, wie er etwa in gängigen Lehrmitteln zu finden ist. Der Gruppe *V* hingegen wird ein eigens erstelltes *Let's Play* der Tour präsentiert.¹⁵⁷ Ebenso wie das Material der Gruppe *T* werden die aufrufbaren schriftlichen Zusatzinformationen hier nicht verwendet, da ihre Inhalte im *Post-Test* unberücksichtigt bleiben. Bis zur Lernphase ist die Umfrage für sämtliche Gruppen identisch. Wie oben beschrieben, wird versucht, die angebotenen Inhalte der Medien ebenfalls deckungsgleich zu gestalten, um die Bedingungen und die Durchführung der Untersuchungs- und der Kontrollgruppen zu vereinheitlichen (siehe 4.1.2).

Die Einteilung der Teilnehmenden in die Gruppen, und somit die erste Verzweigung der Umfrage, erfolgt nach dem Zufallsprinzip. Dabei wird die Gliederung weder von den Angaben zur Mediennutzung noch von demografischen Faktoren beeinflusst. So kann die Bildung bereits bestehender Gruppen, beispielsweise täglich Spielender, verhindert werden.¹⁵⁸ Den Teilnehmenden werden je nach Zuteilung unterschiedliche Blöcke in der Lernphase angeboten. Der eigentliche Auftrag ist jedoch identisch. Die Teilnehmenden sollen die Medieninhalte aufmerksam spielen, lesen oder betrachten. Entscheidend ist, dass jede Lernphase mit einer Anleitung versehen ist, die auf die anschließenden Fragen zu den präsentierten Inhalten hinweist (*S2.2/S2.3*, *T1*, *V1*). Somit werden die Teilnehmenden bereits vor der eigentlichen Auseinandersetzung mit den Medien über deren Zweck informiert und für die anstehenden Aufgaben aktiviert. Denn nur wenn sie ihre Aufmerksamkeit dem Thema widmen und sich dazu bereit zeigen, sich neues

156 Mayer 2014. Computer games for learning. S. 40.

157 AC Survey 2021. Entdeckungstour Thermopylen. Eigene Aufnahme des Autors der Entdeckungstour Thermopylen aus Discovery Tour: Ancient Greece by Ubisoft (Ubisoft Montreal 2019). 14:03, YouTube, 2021. <https://www.youtube.com/watch?v=8losE9RDzsU> (aufgerufen am 25. Januar 2024).

158 Mayer 2014. Computer games for learning. S. 41 f.

Wissen anzueignen, können entsprechende Lernprozesse initiiert werden.¹⁵⁹ Die zielgerichtete Verarbeitung der historischen Inhalte aus der Tour *Thermopylen* betrifft somit die Kompetenzbereiche der Wahrnehmung und der Erschließung anhand einer Darstellung aus der Geschichtskultur (siehe 4.1.5).

Der Gruppe *T* wird hierzu das bebilderte Transkript als PDF-Dokument angezeigt (siehe Anhang I). Dieses wurde durch die Betätigung eines integrierten Links geöffnet. In seiner Struktur richtet es sich nach den acht Stationen der Tour, die von Ansprachen des Leonidas gerahmt werden. Zusätzlich sind die drei Fragen und Antworten des abschließenden Quiz integriert, damit die Inhalte denen der Entdeckungstour entsprechen. Nach demselben Prinzip wird mit dem Medienformat der Gruppe *V* verfahren. Die Aufnahme besteht aus acht Stationen, Einführung und Abschluss. Die Sprecher des Leonidas und der Erzähler tragen hier jenen Text vor, der Gruppe *T* schriftlich präsentiert wird. Das Video ist auf der Auftragsseite *VI* eingebettet und kann von den Teilnehmenden durch Betätigung der angezeigten Play-Taste abgespielt werden. Beide Gruppen bestätigten im Anschluss, dass sie die Medieninhalte aufmerksam und vollständig betrachtet hätten. Die Implementierung der Medieninhalte der Gruppe *S* erfordert mehrere Schritte, da drei Herausforderungen für eine gelungene Umsetzung existieren. Erstens kann weder das Digitale Spiel *DTAG* noch die *einzelne* spielbare Tour in das Online-Umfrage-Tool integriert werden. Zweitens besitzen viele Teilnehmende weder das Digitale Spiel noch einen PC mit den entsprechenden Hardwareanforderungen. Drittens ist die Einzelversion der *DTAG* nur über den hauseigenen Dienst *Ubisoft Connect* erhältlich, für den sich keine spezifischen Geschenkgutscheine oder Guthabekarten erwerben ließen.¹⁶⁰ So kann das Digitale Spiel den Teilnehmenden nicht unmittelbar zur Verfügung gestellt werden, was die festgelegte Versuchsprämisse, das heißt die möglichst alltagsnahe Nutzung auf Privatgeräten, erschwert.¹⁶¹ Diese drei Problemfelder der Gruppe *S* wurden im Vorfeld der Umfrage antizipiert und erforderten dementsprechende Maßnahmen. Sie führten zu folgendem Aufbau der Umfrage für jenes Drittel der Teilnehmenden, welches dieser Lernphase zugewiesen wurde.

Zuerst werden die Teilnehmenden gefragt, ob sie aktuell auf das Digitale Spiel zugreifen können, das heißt die *DTAG* (*S1*) oder den Haupttitel *Assassin's Creed Odyssey* besitzen. Letztgenannte Software ermöglicht ebenfalls, auf die Tour *Thermopylen* (*S1.1.*) zuzugreifen. Sollte eine dieser Items mit *Ja* bestätigt werden, werden die Teilnehmenden direkt zur eigentlichen Lernphase weitergeleitet (*S2*). Falls dies nicht der Fall ist, folgt die Verteilung der zur Verfügung gestellten Codes. Hier-

¹⁵⁹ Gautschi 2015. Guter Geschichtsunterricht. S. 48.

¹⁶⁰ Abgeklärt im persönlichen Mailverkehr mit „acdcoverytour@ubisoft.com“.

¹⁶¹ Mayer 2014. Computer games for learning. S. 43.

zu müssen jedoch vorab die Voraussetzungen der Hardware und die Verfügbarkeit geklärt werden. Daher werden die Mindestanforderungen der *Discovery Tour* präsentiert und können von den Teilnehmenden mit den Spezifikationen ihrer Geräte abgeglichen werden (S3.1.). Anschließend wird auf das erforderliche Konto bei *Ubisoft Connect* hingewiesen (S3.2.). Für jene Teilnehmenden, die noch kein Konto haben, kann hier die Anmeldung über den eingebetteten Link vollzogen werden. Nach der erfolgreichen Registrierung wird den Teilnehmenden der Zugang zum Digitalen Spiel bereitgestellt. Dies geschieht durch Codes, die über eine eigens eingerichtete Funktionsmail versendet werden (S3.4.). Die benötigten 50 Einheiten der *DTAG* wurden der Studie von *Ubisoft* zu Forschungszwecken zur Verfügung gestellt. Der Kontakt wurde dabei über eine Mailadresse hergestellt, die ausdrücklich für interessierte Lehrpersonen eingerichtet wurde.¹⁶² Die Gültigkeit der Codes beschränkte sich auf den Durchführungszeitraum der Umfrage. Zur Unterstützung der Teilnehmenden wird ein bebildeter Leitfaden angezeigt, der den Installationsvorgang beschreibt (S3.5.) und anschließend den Zugriff auf die *Tour Thermopylen* anleitet (S2.1.–S.3.). Da es sich bei dem Digitalen Spiel um eine externe Software handelt, kann der Browser mit der aktiven Umfrage während der *Tour* minimiert werden. Nach der Rückkehr zum Browser muss, wie auch bei den Gruppen *T* und *V*, der abgeschlossene Spielvorgang bestätigt werden.

Der Prozess der Gruppe *S* ist wegen der oben beschriebenen Gesichtspunkte komplexer als der der anderen Kontrollgruppen und bietet den Teilnehmenden daher drei Ausweichmöglichkeiten an. Sollte eines der Items zu Systemanforderungen (S3.1.), *Ubisoft Connect* (S3.2.) oder Installation (S3.5./S3.6.) mit *Nein* beantwortet werden, wird der Block *S* beendet. Anders als bei der Einführungsfrage (*E1*) wird die Umfrage in diesem Fall jedoch nicht abgebrochen, da es sich hierbei um technische, das heißt nicht studienrelevante, Faktoren handelt. Folglich findet eine weitere Randomisierung statt, die die Teilnehmenden einer der Gruppen *T* oder *V* zuweist. So kann die Umfrage regulär abgeschlossen werden, ohne die technischen Voraussetzungen der Hardware als Ausschlusskriterium einzubeziehen.

4.3.6 Post-Test

Die oben beschriebenen Randomisierungen und Verzweigungen werden im *Post-Test* wieder zusammengeführt (siehe Abbildung 4-7). In der Lernphase wurde die *Tour Thermopylen*, unabhängig von der medialen Form, von den Teilnehmenden

¹⁶² „acdcoverytour@ubisoft.com“, in: o. A. 2022. „Die Discovery Tour von Ubisoft“. Ubisoft Entertainment verlangte keinen vorzeitigen Einblick in den Forschungsprozess und beeinflusste in keiner Weise die Ergebnisse der vorliegenden Studie.

als „Begegnung mit einer Präsentation aus der Geschichtskultur“¹⁶³ wahrgenommen. Laut Gautschi führt dies zu eigenen Fragen und Vermutungen.¹⁶⁴ Diese werden jedoch im Rahmen des *Post-Tests* weder aufgezeichnet noch überprüft, da nicht die Rezeption, sondern die quantitative Bestimmung der Lerneffekte im Vordergrund steht. Wie Mayer für deren Untersuchung vorschlägt, fokussiert sich der *Post-Test* daher auf die Veränderungen des Wissens und Könnens der Teilnehmenden.¹⁶⁵ Die einzelnen Blöcke folgen jedoch dem vorgestellten Modell des historischen Lernens (siehe 4.1.4) und den daraus resultierenden Aufgabenstellungen.¹⁶⁶ Die fünf Fragetypen des *Post-Tests* unterscheiden sich damit nicht nur in Bezug auf die zu prüfenden Kompetenzen, sondern auch im Format. Zwei Batterien bestehen aus geschlossenen *Multiple-Choice*-Fragen (*P1.1.–P1.8./P4.1.–P4.8.*), während es sich bei den übrigen drei Blöcken um offene Aufgabenformate (*P2, P3, P5*) handelt.

Der *Post-Test* beginnt mit einem kurzen Infotext (*P0*). Hier werden die Teilnehmende darauf hingewiesen, dass die vorliegende Umfrage keine Prüfungssituation darstellt und sie die Fragen daher möglichst spontan und direkt beantworten sollen. Anschließend folgt der Block (*P1*), der insgesamt aus acht Fragen des Typus *richtig oder falsch* besteht. Sämtliche Items behandeln Inhalte, die in der soeben präsentierten Tour vorkommen. Es handelt sich somit um eine Wissensaufgabe mit niedriger Anforderung, da die bloße „Wiedergabe von Erlerntem in der erlernten oder in einer transformierten Form“¹⁶⁷ verlangt wird. Pro korrekte Antwort werden in diesem Block (*P1.1.–P1.8.*) zwei Punkte vergeben, sodass ein Maximum von 16 Punkten erreicht werden kann. Die zweite Fragebatterie mit geschlossenen Aufgaben folgt als vorletzter Block der Umfrage (*P4*). Hier werden den Teilnehmenden vier Auswahlmöglichkeiten zur Verfügung gestellt. Sie beziehen sich ebenfalls unmittelbar auf die Tour *Thermopylen* und sind nur durch deren Inhalte zu beantworten. Aus dieser Geschichtsdarstellung sollen verschiedene Phänomene, Sachverhalte und Personen identifiziert werden.¹⁶⁸ Die Fragen überprüfen somit, ob die Teilnehmenden die Sachverhalte der Tour und die dort behandelten Ereignisse verstanden und durchdrungen haben. Die eingesetzten *Multiple-Choice*-Fragen sind hierfür geeignet, da sie nicht nur den Wissens-, sondern auch den Erschließungsaufgaben zuzuordnen sind.¹⁶⁹ Neben der erhöhten Anzahl an Distrak-

163 Gautschi 2015. Guter Geschichtsunterricht. S. 51.

164 Gautschi 2015. Guter Geschichtsunterricht. S. 50.

165 Mayer 2014. Computer games for learning. S. 33.

166 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

167 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

168 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 12.

169 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

toren, im Vergleich zum Block *P.1*, ist eine höhere Punktevergabe somit ebenfalls aufgrund der gesteigerten Anforderungen gerechtfertigt. Für die acht Aufgaben (*P4.1-P4.8.*) werden daher jeweils fünf Punkte verliehen, was ein Maximum von 40 Punkten ergibt.

Im Folgenden werden nun die offenen Aufgabenformate vorgestellt. In der Reflexion seiner Studie kritisierte Éthier bereits, dass die genutzten Fragen bloß deklaratives Wissen, und damit keine weiteren Kompetenzen des historischen Lernens überprüfen. Er wünsche sich für kommende Untersuchungen zu Lerneffekten Digitaler Spiele jene Fragen, die höheren Anforderungsbereichen zuzuordnen sind.¹⁷⁰ In Block *P2* werden die Teilnehmenden mit zwei Aussagen des griechischen Historikers Herodot konfrontiert. Sie sind seinen *Historien* entnommen und werden in der Tour nicht explizit zitiert. Bei den Textstellen handelt es sich um Aussagen über die Spartiaten¹⁷¹ und die Perser¹⁷², die von den Teilnehmenden geordnet und in Beziehung gesetzt werden sollen. Durch Auseinandersetzung mit zusätzlichen Quellen verlassen die Teilnehmenden das Material der Tour und gelangen zu einem historischen Sachurteil (siehe 4.1.4).¹⁷³ Die Aufgabe *P2* kann damit dem Kompetenzbereich der Interpretation zugeordnet werden.¹⁷⁴ Im anschließenden Block *P3* wird wiederum eine Aussage der Tour vorgegeben. Leonidas lobt im Digitalen Spiel die Rolle der 300 Spartiaten und ihre Bedeutung für die griechische Welt. Die Teilnehmenden beziehen das Zitat auf die Vergangenheit, ohne die Perspektive der eigentlichen AutorInnen, das heißt der Entwickelnden, zu vernachlässigen. Obwohl sie Vermutungen zur Darstellung des Leonidas anstellen, handelt es sich nicht nur um eine Wahrnehmungsaufgabe.¹⁷⁵ Leonidas' Aussage kann als Sachurteil gesehen werden, das nun durch die Teilnehmenden geprüft wird. Somit handelt es sich ebenfalls um eine Orientierungsaufgabe, indem ein Bezug zwischen Phänomenen der Vergangenheit und der Gegenwart hergestellt wird.¹⁷⁶ Ersteres Phänomen betrifft die eigentliche Rolle der 300 Spartiaten in den Perserkriegen, während Letzteres die gegenwärtige Beurteilung in der populären Wahrnehmung widerspiegelt. Der abschließende Block des *Post-Tests (P5)* erfordert eine noch anspruchsvollere Transferleistung der Teilnehmenden. In dieser Aufgabe müssen zwei unterschiedliche Zeitebenen der Vergangenheit miteinander in Beziehung gesetzt werden. So verglich Hermann Göring nach der Schlacht von Sta-

170 Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 34.

171 Hdt.: Hist. VII, 104,4.

172 Hdt.: Hist. VII, 210,2.

173 Gautschi 2015. Guter Geschichtsunterricht. S. 52.

174 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

175 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

176 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

lingrad im Winter 1943, die deutschen Soldaten mit den 300 Spartiaten an den Thermopylen.¹⁷⁷ Die Kontextualisierung und Interpretation dieser Aussage weist bereits hohe Anforderungen auf. Als einzige Aufgabe des *Post-Tests* setzt sie zudem Kontextwissen voraus, welches nicht explizit in der Lernphase behandelt wurde. Denn das vorgegebene Zitat stellt ein Sachurteil dar, welches nicht aus der Tour übernommen, sondern als externe Quelle eingebaut ist. Um diese Aufgabe erfolgreich zu bewerkstelligen, müssen die Teilnehmenden Sinnbildungsmuster erkennen, reflektieren und schließlich erläutern. Dabei müssen sie abwechselnd in den drei Zeitkontexten Antike, Zweiter Weltkrieg und Gegenwart operieren. Die Auseinandersetzung kann somit mehrere Kompetenzbereiche abdecken und verlangt die Formulierung eines eigenen Werturteils, wie Gautschi es für die „Sinnbildung über Zeiterfahrung“¹⁷⁸ beschreibt.

Die soeben vorgestellten offenen Aufgaben (*P.2/P.3/P.5*) werden in der quantitativen Untersuchung nicht berücksichtigt. Für die Berechnung der Lerneffekte werden nur die 16 geschlossenen Frageformate (*P.1/P.4*) verwendet. Der *Post-Test* zur quantitativen Analyse der Lerneffekte umfasst somit insgesamt 56 Punkte (*P.1/P.4*). Wie oben erläutert, werden damit Wahrnehmungs- und Erschließungskompetenz sowie die Wiedergabe von Erlerntem, das heißt historische Wissensbestände, überprüft. Zusätzlich werden mit den offenen Aufgaben sämtliche Denkopoperationen des historischen Lernens berücksichtigt (siehe 4.1.4). Der Ausschluss dieser Aufgaben aus der summativen Bewertung hat praktische und theoretische Gründe. So verfügt die Software *qualtrics XM* über die integrierte Funktion *ExpertReview*¹⁷⁹. Hierdurch werden die Durchführbarkeit, die Barrierefreiheit und die Dauer der Umfrage bewertet. Der vorliegende Fragebogen wurde als akzeptabel eingeschätzt, wobei die Länge der Umfrage und die Anzahl der Texteingabefelder beanstandet wurden. So übersteigt er die empfohlene Länge von zwölf Minuten und verfügt über mehr als drei offene Antwortmöglichkeiten. Laut der *ExpertReview* können diese Faktoren zur Ermüdung oder Verstimmung der Teilnehmenden führen. Neben sinkenden Abschlussraten kann dies auch Folgen für die Auswertung der Ergebnisse haben, da grundsätzlich „in Webumfragen häufig keine sehr wertvollen Textantworten“¹⁸⁰ zu erwarten sind. Die Einschätzung der Software wurde bereits durch die Rückmeldungen zu einer Beta-Test-Version des Fragebo-

177 Meier 2019. „Die Thermopylen – ‚Wanderer, kommst du nach Spa(rta)‘“. S. 98–99.

178 Gautschi 2015. Guter Geschichtsunterricht. S. 52.

179 David L. Vanette. „Umfragemethodik und bewährte Vorgehensweisen rund um die Einhaltung von Vorschriften“. Qualtrics Support, 2023. <https://www.qualtrics.com/support/de/survey-platform/survey-module/survey-checker/survey-methodology-compliance-best-practices/?rid=langMatch&prevsite=en&newsite=de&geo=CH&geomatch=> (aufgerufen am 21. November 2023).

180 Vanette 2023. „Umfragemethodik“.

gens bestätigt. Um den Teilnehmenden eine gewisse Orientierung über den maximalen Umfang der Antworten zu bieten, wurde die Eingabe daher auf 800 (*P2*), 600 (*P3*) und 1000 (*P5*) Zeichen beschränkt.

Neben den praktischen Gründen der Durchführbarkeit, fußt die Isolation der offenen Aufgaben ebenso auf einer theoretischen Grundlage. Der Fokus liegt auf den geschlossenen Fragen, weil die Lerneffekte laut Mayer im datenbasierten Zugang nur quantitativ bestimmt werden können.¹⁸¹ Das schließt jedoch die Einbindung offener Aufgaben nicht grundsätzlich aus. Einerseits besteht die Möglichkeit einer Quantifizierung der Textantworten durch eine Bewertungsskala, wie sie etwa bei summativen Geschichtsprüfungen eingesetzt wird. Andererseits können zusätzliche qualitative Aussagen, die Datenbasis der Studie unterstützen.¹⁸² Hierzu muss festgestellt werden, ob die erforderlichen Kompetenzziele der drei Aufgaben erreicht wurden. In der vorliegenden Studie werden die Antworten in die vier Kategorien *nicht erreicht*, *teilweise erreicht*, *erreicht* und *übertrroffen* eingeteilt. Anhand dieses Schemas werden die Aufgaben zuerst separat bewertet und anschließend zu einer Gesamtbeurteilung der einzelnen Teilnehmenden zusammengefasst:

nicht erreicht (ne):

- Die Kompetenzziele werden bei sämtlichen Aufgaben nicht erreicht
- Die Kompetenzziele werden nur bei einer Aufgabe teilweise erreicht

teilweise erreicht (te):

- Die Kompetenzziele werden bei mindestens zwei Aufgaben teilweise erreicht

oder

- Die Kompetenzziele werden nur bei einer Aufgabe erreicht und bei mindestens einer weiteren teilweise erreicht

erreicht (e):

- Die Kompetenzziele werden bei mindestens zwei Aufgaben erreicht und bei der dritten mindestens teilweise erreicht

übertrroffen (ü):

- Die Kompetenzziele werden bei sämtlichen Aufgaben erreicht und bei mindestens zwei übertrroffen

181 Mayer 2014. Computer games for learning. S. 37.

182 Mayer 2014. Computer games for learning. S. 37.

Bei der Beurteilung hinsichtlich dieser vier Kategorien werden weder demografische Angaben noch die quantitativen Resultate des *Pre-* oder des *Post-Tests* berücksichtigt.

4.3.7 Abschluss

Nach dem Ende des *Post-Tests* wird den Teilnehmenden eine kurze Danksagung angezeigt (*GC1*). Bei bestehendem Interesse an der Studie, kann die eingerichtete Funktionsmailadresse kontaktiert werden. Die Verabschiedung wird mit den Hinweisen auf die Kanäle zur Rekrutierung abgeschlossen (siehe 4.2.3). Die Verlosung der Gutscheine wird über eine externe Funktion des Umfragetools realisiert und geschieht auf freiwilliger Basis. Bei beiden Varianten, das heißt auch bei einer Entscheidung gegen das Gewinnspiel, wird die Umfrage an dieser Stelle (*GC2*) beendet und ein Hinweis angezeigt, der die anonymisierte Sicherung der Daten bestätigt. Somit gilt die Teilnahme an der Studie offiziell als abgeschlossen. Die übrigen Teilnehmenden werden durch das Annehmen der Frage *GC1* zur entsprechenden Seite weitergeleitet (siehe Anhang II). Aufgrund dieser gesonderten Bearbeitung der Verlosung können keine Rückschlüsse auf die Ergebnisse Teilnehmenden gezogen werden. Nach dem Losverfahren erfolgte der Versand der Gutscheine über eine separat eingerichtete Funktionsmail der Universität Basel.

4.4 Sicherheit und Datenaufbereitung

Wichtige Aspekte der datenschutzrechtlichen und sicherheitstechnischen Anforderungen einer empirischen Studie wurden bereits in den vorherigen Kapiteln beschrieben. Vor der Durchführung der Umfrage wurde daher ein Antrag bei der Ethikkommission der Universität Basel gestellt, um das Forschungsprojekt freizugeben. Darin waren sämtliche Informationen zu Ablauf, Inhalt und verwendeten Materialien enthalten. Die mittlerweile abgeschlossene Erhebung war dabei mit den gestellten Anforderungen konform. So wurde die komplette Studie anonym durchgeführt und dabei wurden keine personenbezogenen Konten oder IP- oder Mailadressen erhoben. Zusätzlich konnten die Teilnehmenden die Umfrage jederzeit abbrechen und wurden zu keinem Zeitpunkt der Umfrage getäuscht. Über diese Sicherheitsmaßnahmen wurden die Teilnehmenden im Informationsschreiben zu Beginn der Umfrage unterrichtet (siehe 4.3.2). Der Zugang zur eigentlichen Studie war jedoch erst nach vorheriger Bestätigung der Altersüberprüfung gestattet. Dies entspricht den Anforderungen der Ethikkommission der Universität Basel und der Freigabe des gewählten Forschungsgegenstands (USK 6 bzw. Pegi 7). Die

Teilnahme am Gewinnspiel erfolgte ebenfalls auf freiwilliger Basis. Da die Anmeldung erst im Anschluss an die Umfrage stattfand, konnte sie keinen Einfluss auf deren Ergebnisse nehmen. Ebenso wie für Rückfragen und das Versenden der Spielcodes wurde hierfür eine spezifische Funktionsmailadresse eingerichtet. So erfolgte die Zustellung der Gewinne unabhängig von der anonymen Umfrage und erlaubt keine Rückschlüsse auf die Teilnehmenden.¹⁸³ Die drei Funktionsmailadressen werden nach Abschluss des Promotionsstudiums von den IT-Services der Universität Basel gelöscht. Das Forschungsprojekt entsprach somit den datenschutzrechtlichen und sicherheitstechnischen Anforderungen der Ethikkommission.

Überdies reduzierten die weiteren in der Umfrage implementierten Maßnahmen den Aufwand der Datenaufbereitung. Erstens konnten durch die automatisierte Bot-Erkennung verdächtige Teilnahmen gefiltert und gegebenenfalls entfernt werden. Beim sogenannten *reCaptcha* handelt es sich um ein Programm, das die Wahrscheinlichkeit von Bot-Teilnahmen senken kann.¹⁸⁴ Es weist den abgeschlossenen Umfragen einen Wert von 0 bis 1 zu, wobei ein Ergebnis von 0,5 und höher für eine menschliche Teilnahme spricht, während die Wahrscheinlichkeit einer Bot-Bearbeitung unter einem Wert von 0,5 steigt.¹⁸⁵ Bei der vorliegenden Umfrage wurden jene Einreichungen entfernt, die den *reCaptchaScore* von 0,4 unterschritten. Die zweite Betrugserkennung wurde über die *Captcha-Fragen*¹⁸⁶ integriert, die an vier Stellen der Umfrage integriert wurden. Diese sind nicht mit der automatisierten Bot-Erkennung zu verwechseln, da sie ein aktives Handeln der Teilnehmenden erfordern.¹⁸⁷ Neben dem unsachgemäßen Zugriff automatisierter Programme kann das Gewinnspiel auch reguläre Teilnehmende zu einer mehrfachen Einreichung animieren. Die Umfrage wurde so gestaltet, dass ein erneutes Öffnen des Links zu einem direkten Abbruch führt.¹⁸⁸ Über ein eingeblendetes Textfeld wird auf die bereits erfolgte Abgabe hingewiesen. Die technischen Verfahren fanden begleitend zur laufenden Umfrage statt. Doch auch nach deren Abschluss wurden die Ergebnisse weiterführend bereinigt. Im Besonderen betrifft dies die Antworten der offenen Textaufgaben (P2/P3/P5). Um ein bloßes Weiterklicken zu verhindern, wurde eine Zeicheneingabe in den Textfeldern verlangt.

183 o. A. 2023. „Erstellen anonymisierter Auslosungen“.

184 o. A. „Betrugserkennung“. Qualtrics Support, 2023. <https://www.qualtrics.com/support/de/survey-platform/survey-module/survey-checker/fraud-detection/> (aufgerufen am 21. November 2023).

185 o. A. 2023. „Betrugserkennung“.

186 o. A. „Captcha-Prüfung Frage“. Qualtrics Support, 2023. <https://www.qualtrics.com/support/de/survey-platform/survey-module/editing-questions/question-types-guide/advanced/captcha-verification/#RespondentExperience> (aufgerufen am 19. Oktober 2023).

187 o. A. 2023. „Betrugserkennung“.

188 o. A. 2023. „Betrugserkennung“.

Platzhalter oder andere zweckfremde Antworten wurden jedoch nicht entfernt. Nachträglich gelöscht wurden nur Einreichungen, die anstößige oder menschenverachtende Beiträge enthielten.

Die oben beschriebenen Maßnahmen zu Sicherheit und Datenschutz wurden allesamt in der Browseranwendung *qualtrics XM* realisiert.¹⁸⁹ Sie waren im Paket für den Bildungssektor enthalten, welches vom Juli 2021 bis Juni 2022 lizenziert wurde.¹⁹⁰ Für die weitere Verarbeitung der Daten standen verschiedene Programme zur Auswahl. Die Wahl fiel schließlich auf das Open-Source-Angebot der Universität Amsterdam. Das Programm *JASP* bietet Datenanalyse auf dem neusten Stand, ohne die hierfür üblichen Gebühren zu verlangen. Der Zugang ist zudem durch ein möglichst intuitives Interface gewährleistet, welches auch unerfahrene Nutzende anspricht.¹⁹¹ Daher wird das Programm bereits an vielen Hochschulen eingesetzt. Der entscheidende Faktor, der für *JASP* sprach, waren jedoch die Anwendungsmöglichkeiten für die vorgestellten theoretischen Grundlagen der Studie. So lassen sich sowohl die deskriptive Analyse Éthiers (siehe 4.1.1) als auch die Mayer'sche Bestimmung der medienvergleichenden Lerneffekte (siehe 4.1.2) mit dem Programm durchführen. Hierbei werden die geforderten Mittelwerte, die Standardabweichungen und die Stichprobengrößen berücksichtigt und dokumentiert.¹⁹² Des Weiteren können sämtliche gängigen Dateiformate integriert werden. So wurde der Datensatz aus *qualtrics XM* als *comma-separated values* (.csvs) exportiert und konnte abschließend in *JASP* aufgerufen werden. Er erschien dort in einer tabellarischen Darstellung, die bereits metrische, ordinale und nominale Variablen unterschied. Hingegen fand die Punktwertung des *Pre-* und des *Post-Tests* sowie die entsprechende Gewichtung der einzelnen Aufgaben bereits im Umfragetool *qualtrics XM* statt. Nachträglich in *JASP* wurden einzig die Abzüge aufgrund falscher Antworten bei der Aufgabe *G1* vorgenommen, um die finalen Resultate der *Pre-Tests* zu bestimmen. Da die Antwortmöglichkeiten der geschlossenen Aufgaben in zufälliger Reihenfolge präsentiert wurden, war jedem wählbaren Item ein Zahlenwert zugeteilt (siehe Anhang II). Die offenen Aufgaben hingegen wurden nicht automatisiert bewertet, sondern anhand der vorgestellten Skala beurteilt und mit dem entsprechenden Label ergänzt (siehe 4.3.6).

189 o. A. „Online-Umfragesoftware: Top Umfrage-Plattform testen“. Qualtrics Core XM, 2022. <https://www.qualtrics.com/de/core-xm/umfrage-plattform/> (aufgerufen am 29. August 2022).

190 Das Digital Humanities Lab der Universität Basel hat die Kosten für das Abonnement dankenswerterweise übernommen.

191 Eric-Jan Wagenmakers. „A Fresh Way to Do Statistics“. Departement of Psychological Method, University of Amsterdams JASP, 2022. <https://jasp-stats.org/> (aufgerufen am 31. August 2022).

192 Mayer 2014. Computer games for learning. S. 44.

Die Umfrage stand dabei im Zeitraum vom 1. September 2021 bis zum 7. Januar 2022 zur Verfügung. Abschließend wurden 94 valide Teilnahmen in die bereinigte Datei von *JASP* integriert (Tabelle 4-5).

Tab. 4-5: Demografische Angaben der Teilnehmenden

	Häufigkeit	Prozent
Alter 16–21 Jahre	94	100
Männlich	45	47,87
Weiblich	47	50,00
Nichtbinär/drittes Geschlecht	1	1,06
Keine Angabe	1	1,06

Die Altersspanne bildete ein Ausschlusskriterium der Teilnahme und wurde bereits zu Beginn der Umfrage überprüft (*E1*). Auf die Erhebung des Geschlechts hingegen sollte ursprünglich bewusst verzichtet werden. Denn häufig werden diese Angaben in empirischen Studien beiläufig erfasst, ohne in den Fragestellungen genderspezifische Perspektiven abzubilden.¹⁹³ Dies gilt ebenso für die formulierten Forschungsfragen der vorliegenden Untersuchung. Um jedoch sicherzustellen, dass das Teilnehmendenfeld nicht eindimensional besetzt ist, das heißt, dass die Ergebnisse nicht gegebenenfalls zur „Geschlechtsblindheit“¹⁹⁴ führen, wurde ein dreigliedriges Item integriert (*D2*). Die demografischen Angaben in Tabelle 4-5 verdeutlichen, dass zumindest die Geschlechter männlich und weiblich gleichmäßig im Datensatz vertreten sind. Auf Basis dieser Erkenntnis und aufgrund der fehlenden Passung zu den formulierten Forschungsfragen wurden Angaben zum Geschlecht der Teilnehmenden in den Ergebnissen und der Analyse jedoch nicht gesondert thematisiert.

Entscheidender für die Fragestellung der vorliegenden Studie ist die Zusammensetzung der Lernphasengruppen ($n_S = 20$; $n_V = 35$ und $n_T = 40$). Die beiden Letzteren erfüllen die erforderliche Stichprobengröße ($n \geq 25$), während die erstge-

193 Nora Berner, Lena Rosenkranz und Julia Schütz. „Geschlecht als Variable in der quantitativen Bildungsforschung“. In *Methoden und Methodologien in der erziehungswissenschaftlichen Geschlechterforschung*, herausgegeben von Melanie Kubandt und Julia Schütz, 26–44. Opladen, Berlin, Toronto: Verlag Barbara Budrich, 2020.

194 Nicola Döring. „Zur Operationalisierung von Geschlecht im Fragebogen: Probleme und Lösungsansätze aus Sicht von Mess-, Umfrage- und Queer-Theorie“. *GENDER – Zeitschrift für Geschlecht, Kultur und Gesellschaft*, 5, Nr. 2 (2013), 94–113. <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-396603> (aufgerufen am 5. März 2024).

nannte Untersuchungsgruppe dieses Kriterium knapp unterschreitet.¹⁹⁵ Dieser Aspekt sollte in der Diskussion der Ergebnisse berücksichtigt werden. Die unterschiedliche Stichprobengröße kann hingegen vernachlässigt werden, da der *t*-Test in *JASP* diesen Umstand bereits einbezieht. So gewichtet das Statistikprogramm die Anzahl der Teilnehmenden pro Gruppe für die Berechnung der zusammengefassten Standardabweichung (SD_{pooled}).¹⁹⁶ Zur grundlegenden Bewertung des Lernzuwachses nach Éthier (siehe 4.1.1) wurden die deskriptiven Statistiken und ein *t*-Test mit abhängigen Stichproben verwendet. Für die Feststellung der Lerneffekte nach Mayer (siehe 4.1.2) konnte der *t*-Test mit unabhängigen Stichproben durchgeführt werden. Auf Basis der unterschiedlichen Zusatzfunktionen, wie der Auswahl der Visualisierung und der Filter, können die Daten den Forschungsfragen entsprechend verarbeitet werden. Der Fokus liegt auf den zentralen Forschungsfragen (*ZF1*, *ZF2*), die sich mit Lernzuwachs und Effektstärke der eingesetzten Medien beschäftigen. Daher wird auf Varianzanalysen (ANOVA), wie sie etwa Éthier verwendet, verzichtet.¹⁹⁷ Stattdessen werden entsprechende deskriptive Statistiken zu den erweiterten Forschungsfragen (*EF1–EF3*) präsentiert. Hierzu können entsprechende Variablen im Statistikprogramm *JASP* berechnet werden. Diese Verschmelzung von Zugänglichkeit, Anwendbarkeit und Zweckmäßigkeit erfüllt sämtliche Anforderungen der Studie.

4.5 Ergebnisse

4.5.1 Analyse des Lernzuwachses nach Éthier

Wie im vorherigen Kapitel vorgestellt, existieren mehrere Methoden, die Lerneffekte Digitaler Spiele empirisch zu belegen. Durch den Forschungsaufbau können so etwa die Ergebnisse des *Pre-Tests* mit denjenigen des *Post-Tests* verglichen werden. Tabelle 4-6 zeigt die Ergebnisse sämtlicher abgeschlossenen Teilnahmen der Studie ($n = 94$). Diese Auswertung geschieht unabhängig vom präsentierten Medienformat der Lernphase. Die Ergebnisse sind in Prozenten angegeben und stellen das Verhältnis zwischen der erreichten und der maximalen Punktzahl in den zwei Testformaten dar. Die deskriptive Analyse bietet erste Einblicke in die Frage, ob die Inhalte Digitaler Spiele bei Jugendlichen und jungen Erwachsenen historischen Wissenszuwachs generieren können (*ZF1*).

¹⁹⁵ Mayer 2014. Computer games for learning. S. 43 f.

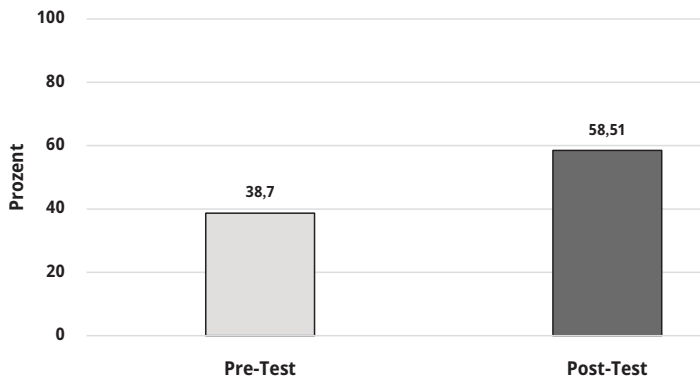
¹⁹⁶ Ly und Kehler 2018. „How to Conduct a Classical Independent Sample T-Test“.

¹⁹⁷ Éthier, Lefrançois und Déry 2022. „Utilisation en classe d’histoire“. S. 28–32.

Tab. 4-6: Deskriptive Statistik der Lerneffekte (n = 94)

	% Pre-Test	% Post-Test
Gültig	94	94
Median	37,50	58,03
Mittelwert	38,70	58,51
Minimum	12,50	19,64
Maximum	71,42	100,00

Der Mittelwert des *Pre-Tests* liegt bei 38,7 Prozent. Bei einem Punktemaximum von 56 entspricht dies einer durchschnittlich erreichten Punktzahl von 21,67. Der *Post-Test* wurde mit 58,51 Prozent und damit durchschnittlich elf Punkte besser abgeschlossen. Dies entspricht einer Steigerung von fast 20 Prozent nach der Lernphase, wie sie in Abbildung 4-8 dargestellt ist.

**Abb. 4-8:** Mittelwert der erreichten Punktzahl in Prozent (n = 94)

Die Differenz zwischen dem maximalen und dem minimalen Ergebnis der Teilnehmenden beträgt für den *Pre-Test* 58,92 Prozent und für den *Post-Test* 80,36 Prozent. Eine teilnehmende Person hat den *Post-Test* ohne Fehler absolviert, was für die Höhe dieses Werts verantwortlich ist. Betrachtet man jedoch die geringe Differenz zwischen Mittelwert und Median, ist davon auszugehen, dass solche Extreme die vorliegende Datenbasis nicht beeinflussen.¹⁹⁸ Somit lässt sich die Steigerung nicht durch einige Ausreißer erklären. Hierzu sollte ebenfalls festgehalten wer-

198 Björn Christensen, Sören Christensen und Martin Missong. Statistik klipp & klar. WiWi klipp & klar. Wiesbaden: Springer Gabler, 2019. S. 23.

den, dass 13 Teilnehmende sich im Verhältnis von *Pre-* zu *Post-Test* nicht verbessern konnten. Im Umkehrschluss lässt sich jedoch feststellen, dass bei 81 von 94 Teilnehmenden nach der Lernphase ein gesteigertes Ergebnis zu beobachten war. So weist bereits die deskriptive Statistik auf einen bedeutenden Einfluss der Auseinandersetzung mit den Tourinhalten hin (Tabelle 4-7).

Tab. 4-7: Deskriptive Statistik (Pre- und Post-Test)

	Gültig	Mittelwert %	Mittelwert	SD	Standardfehler
Pre-Test	94	38,70	21,67	7,72	0,80
Post-Test	94	58,51	32,77	12,54	1,29

Wie in der Referenzstudie Éthiers kann nun der *t*-Test für abhängige Stichproben durchgeführt werden (Tabelle 4-8). So kann festgestellt werden, ob sich die beobachteten Unterschiede der erreichten Punktzahlen des *Pre-* und des *Post-Tests* auch statistisch belegen lassen.¹⁹⁹

Tab. 4-8: *t*-Test für abhängige Stichproben

	Mittelwert	SD	Std.-Fehler Differenz	t	df	Cohens d
(Pre – Post)	-11,10	10,41	1,10	-10,04	93	-1,04

Hinweis: Für das Vergleichspaar lässt sich eine signifikante Differenz beobachten ($n = 94$; $p < 0.05$).

Es kann festgestellt werden, dass sich die Lernphase signifikant auf die Ergebnisse im *Post-Test* auswirkt ($t = -10,04$; $d = -1,04$). Die durchschnittliche Steigerung der Teilnehmenden um fast 20 Prozent ($n = 94$; $\Delta = 11,10$) ist somit direkt auf die Beschäftigung mit der *DTAG* zurückzuführen. Dies gilt unabhängig vom Spielen, Lesen oder Betrachten der Tour.

Folgerichtig wird im nächsten Schritt der Einfluss der in der Lernphase verwendeten Medien einbezogen. Sind die Lerneffekte der Tour *Thermopylen*, gemessen am historischen Wissen, unabhängig von der medialen Präsentationsform der Inhalte identisch (*ZF2*)? Dies kann ebenfalls mit den Methoden der deskriptiven Statistik untersucht werden. Sie lassen einen direkten Vergleich der drei Gruppen zu, ohne jedoch Standardabweichung und Größenverhältnisse der Stichproben zu berücksichtigen. Ob die beobachteten Unterschiede ebenfalls statistisch signifikant sind, wird hingegen anhand des *media comparison approach* überprüft (siehe 4.5.2). Zur Erinnerung sei erneut erwähnt, dass die Kontrollgruppe *V* eine unkom-

¹⁹⁹ Schwarz 2023. „*t*-Test für abhängige Stichproben“.

mentierte Aufzeichnung betrachtet hat, während *T* ein bebildertes Transkript zur Verfügung gestellt bekam. Die Untersuchungsgruppe *S* hingegen hat die Tour *Thermopylen* selbstständig gespielt. Um der Methodik Éthiers zu folgen, wird nun der Lernzuwachs anhand der Differenz zwischen *Pre-* und *Post-Test* gemessen. Anders als in der Referenzstudie Éthiers sind die Umfrageaufgaben in der vorliegenden Arbeit gewichtet. Daher wird nicht die Anzahl richtiger Antworten, sondern das Verhältnis zwischen der erreichten und der maximalen Punktzahl als Referenzwert verwendet. Sämtliche hierfür nötigen Angaben des *Pre-Tests* sind in Tabelle 4-9 aufgeführt, während Tabelle 4-10 die entsprechenden Daten des *Post-Tests* enthält.

Tab. 4-9: Deskriptive Statistik des Pre-Tests (Spiel – Text – Video)

	Pre-Test			% Pre-Test		
	S	T	V	S	T	V
Gültig	20	39	35	20	39	35
Median	23,00	23,00	20,00	41,07	41,07	35,71
Mittelwert	23,25	21,51	20,94	41,52	38,42	37,40
Minimum	9,00	7,00	9,00	16,07	12,50	16,07
Maximum	40,00	39,00	35,00	71,43	69,64	62,50

Tab. 4-10: Deskriptive Statistik des Post-Tests (Spiel – Text – Video)

	Post-Test			% Post-Test		
	S	T	V	S	T	V
Gültig	20	39	35	20	39	35
Median	29,50	30,00	33,00	52,68	53,57	58,93
Mittelwert	30,70	31,13	35,77	54,82	55,59	63,88
Minimum	11,00	11,00	11,00	19,64	19,64	19,64
Maximum	56,00	52,00	54,00	100,00	92,86	96,43

Die Gruppe *S* erreicht im *Post-Test* einen Mittelwert von 54,82 Prozent, was einer durchschnittlichen Punktzahl von 30,7 entspricht. Auf vergleichbarem Niveau bewegt sich Gruppe *T*, die eine mittlere Punktzahl von 31,13 erreichen konnte. Dies entspricht 55,59 Prozent des Maximums von 56. Das höchste Ergebnis lässt sich in der Gruppe *V* beobachten. Hier wurden mit 35,77 durchschnittlich fast zwei Drittel der Punkte (63,88 Prozent) erzielt. Abbildung 4-9 zeigt die Ergebnisse des *Pre-* und

des *Post-Tests* in gegliederter Form. Mithilfe der grafischen Darstellung lassen sich diese Werte mit denjenigen des *Pre-Tests* vergleichen.

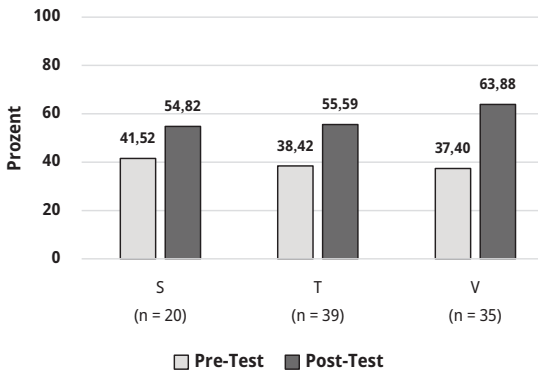


Abb. 4-9: Mittelwert der erreichten Punktzahl in Prozent (Spiel – Text – Video)

So können erste Tendenzen bestätigt werden, die bereits in der deskriptiven Analyse der gesamten Stichprobe ($n = 94$) erkennbar sind. Jede der drei Gruppen konnte ihr Ergebnis nach der Lernphase steigern. Analog zur Analyse des gesamten Teilnehmerfelds bleiben die Verhältnisse bei der Berechnung der Mediandifferenz bestehen. Die in Abbildung 4-9 ersichtlichen Steigerungen sind somit nicht auf eine geringe Anzahl abweichender Ergebnisse zurückzuführen. Zudem ist auffällig, dass der Mittelwert der *Post-Tests* stets die 50 Prozent überschreitet.

In der deskriptiven Analyse dient die Differenz zwischen den Resultaten des *Pre-* und des *Post-Tests* als zentraler Referenzwert zur Beschreibung des Lernzuwachses. Diesen kann man in einem einfachen Balkendiagramm visualisieren (Abbildung 4-10). Demnach lässt sich bei der Gruppe V die größte Steigerung beobachten (26,48 Prozent), während T (17,17 Prozent) und S (13,3 Prozent) weniger Zunahme aufweisen. Auf den ersten Blick kann somit die Videoaufnahme die Inhalte der DTAG am besten vermitteln (ZF2). Der Vergleich zwischen *Pre-* und *Post-Tests* weist hier den größten Lernzuwachs auf.

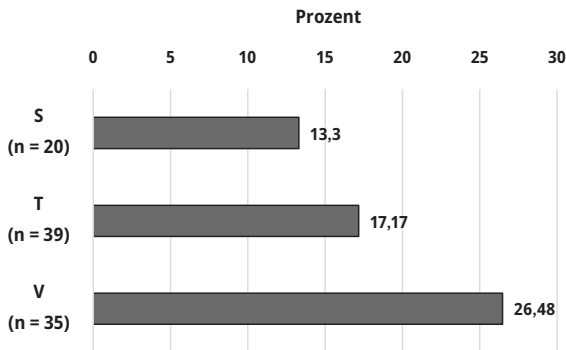


Abb. 4-10: Differenz der Mittelwerte des Pre- und Post-Tests in Prozent (Spiel – Text – Video)

Die große Differenz erklärt sich jedoch auch durch das Ergebnis des *Pre-Tests*, das bei der V-Gruppe unter demjenigen der anderen Gruppen lag. Dies zeigt sich besonders im Median der *Pre-Tests*, wo ein Unterschied von sechs Prozentpunkten zu verzeichnen ist ($T/S = 41,07$ Prozent; $V = 35,71$ Prozent). Diese Diskrepanz ist dem Aufbau der Studie entsprechend ausgefallen. Denn wie bereits erläutert, wurde die Einteilung der Gruppen nach dem Zufallsprinzip vorgenommen und war somit keine Folge der in der laufenden Umfrage getätigten Antworten. Neben diesen Faktoren wird die deskriptive Analyse durch die unterschiedliche Anzahl der validen Einreichungen und der daraus resultierenden Stichprobengrößen erschwert ($n_v = 35$; $n_t = 39$; $n_s = 20$). Die Berücksichtigung der Standardabweichung und eine Gewichtung der Gruppengrößen kann mit der deskriptiven Statistik nur unzureichend beschrieben werden. Hierzu müssen spezifische Methoden zur Berechnung der medienvergleichenden Lerneffekte angewandt werden, wie sie Mayer für den *media comparison approach* (siehe 4.1.2) vorschlägt.

4.5.2 Media Comparison Approach nach Mayer

Neben dem Vergleich der *Pre-* und der *Post-Tests* bietet der *media comparison approach* einen differenzierten Zugang, um die zentrale Forschungsfrage zu beantworten (ZF2). Dazu werden jedoch nur die Resultate des *Post-Tests* beachtet. Anders als bei der deskriptiven Analyse ist hier nicht das Verhältnis zwischen der erreichten und der maximalen Punktzahl in den zwei Testformaten maßgebend. Um der Methodik Mayers zu entsprechen, werden die Mittelwerte der tatsächlich

erreichten Punktzahl des *Post-Tests* verwendet.²⁰⁰ Wie in Abbildung 4-11 ersichtlich, hat dieser Wechsel keinen Einfluss auf das Verhältnis der Gruppen, da nur die zu messende Einheit angepasst wurde. Statt eines Maximums von 100 Prozent wird nun der absolute Wert von 56 Punkten festgelegt.

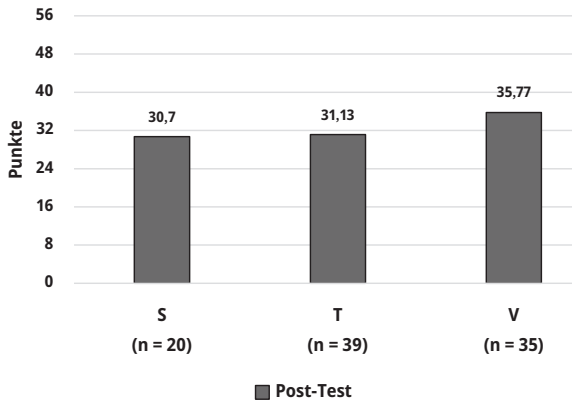


Abb. 4-11: Mittelwerte der Punktzahlen des Post-Test (Spiel – Text – Video)

Gruppe *V* hat durchschnittlich das beste Resultat im *Post-Test* erreicht, während *S* und *T* einen ähnlichen Wert aufweisen. Entscheidender sind hingegen zwei Ansätze, die den *media comparison approach* von der Analyse in Kapitel 4.5.1 unterscheiden. Erstens wird das erhobene Vorwissen des *Pre-Tests* in diesem Verfahren nicht beachtet. Zweitens basiert der Ansatz nicht auf dem beschreibenden Vergleich der Mittelwerte der Untersuchungs- und der Kontrollgruppe. Denn zur Bewertung der statistischen Signifikanz dient hier die Berechnung der Effektstärke. Als Referenzwert wird hier die Effektstärke *d* verwendet.

Nach der Methodik Cohens können jene Lerneffekte als signifikant bewertet werden, die $d = 0,4$ übersteigen.²⁰¹ Die Berechnung erfolgt anhand der bereits vorgestellten Formel (siehe 4.1.2):

$$d = \frac{m_1 - m_2}{SD_{pooled}}$$

²⁰⁰ Mayer 2014. Computer games for learning. S. 45.

²⁰¹ Mayer 2014. Computer games for learning. S. 45.

Zudem sind für sämtliche Gruppen Mittelwert (m), Standardabweichung (SD) und gültige Probengröße (n) zu berücksichtigen, wie sie in Tabelle 4-11 angegeben werden.²⁰²

Tab. 4-11: Mittelwert der Punktzahl des Post-Tests (m) und Standardabweichung (SD) (Spiel – Text – Video)

	Post-Test		
	S	T	V
Gültig	20	39	35
Mittelwert	30,70	31,13	35,77
Standardabweichung	14,20	11,61	12,32

Anders als bei Varianzanalysen (ANOVA) bedingt die Berechnung der Effektstärke nach Mayer jedoch, dass die Mittelwerte der Untersuchungsgruppe jeweils nur im Vergleich mit einer der Kontrollgruppen berechnet werden können. Demnach muss die zentrale Forschungsfrage präzisiert werden: Kann das Digitale Spiel die Inhalte tatsächlich effektiver vermitteln als der bebilderte Text (ZF2)? Die Differenz der durchschnittlich erreichten Punktzahl von 0,43 ($m_S = 30,70$; $m_T = 31,13$) deutet auf keinen erheblichen Unterschied zwischen den beiden Gruppen hin. Um jedoch die Signifikanz der Lerneffekte zu überprüfen, also die Standardabweichung (SD) und die Gruppengröße (n) zu berücksichtigen, wird die Effektstärke (d) berechnet. Die Tabelle 4-12 beinhaltet die Ergebnisse des t -Tests für unabhängige Stichproben nach Mayer für die Gruppen S und T .

Tab. 4-12: t -Test für unabhängige Stichproben (Spiel – Text)

	t	df	p	Cohens d	Std.-Fehler Cohens d
Post-Test	-0,12	57	0,55	-0,03	0,28

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Gruppe S größer ist als Gruppe T .

Basierend auf der theoretischen Grundlage ist besonders die Effektstärke (d) von Interesse. Laut Mayer ist sie der wichtigste Referenzwert für die Bestimmung der Lerneffekte. Wenn man davon ausgeht, dass bei der Gruppe S höhere Lerneffekte auftreten, wird ein negativer Wert ($d = -0,03$) gemessen. Das Digitale Spiel scheint daher die Inhalte der Tour nicht besser vermitteln zu können als der bebilderte Text. Allerdings können keine signifikanten Unterschiede zwischen den Medien

²⁰² Mayer 2014. Computer games for learning. S. 37.

festgestellt werden, da die Effektstärke den dafür nötigen Schwellenwert von $d = 0,2$ deutlich unterschreitet.²⁰³ Ein ähnliches Bild zeigt sich, wenn wir die übrige Teststatistik betrachten. So ist der t -Wert ebenfalls negativ. Er wird in der Methode Mayers zwar nur beiläufig erwähnt, kann jedoch in Kombination mit den Freiheitsgraden ($df = n_S + n_T - 2$) weiteren Aufschluss über eine eventuell signifikante Differenz der Mittelwerte bieten.²⁰⁴ Wie ein Blick auf die entsprechende Verteilungstabelle zeigt, kann der vorliegende t -Wert bei einem geläufigen Signifikanzniveau nicht als kritischer Wert eingestuft werden ($t = -0,12$; $df = 57$; $\alpha = 0,05$).²⁰⁵ Auch der p -Wert von 0,55 gibt keinen Anlass dazu anzunehmen, dass sich die Mittelwerte der Gruppe S grundlegend von denjenigen der Gruppe T unterscheiden.

Anhand der Ergebnisse des Tests lässt sich kein Einfluss der medialen Präsentationsformen feststellen. Das Digitale Spiel kann die Inhalte der Tour demnach nicht erkennbar besser vermitteln als der bebilderte Text ($d = -0,03$; $t = -0,12$; $p = 0,55$). Diese Aussage basiert auf der Grundlage, dass Varianzhomogenität vorliegt, dass also die Streuung der *Post-Test*-Ergebnisse beider Gruppen sich nicht voneinander unterscheidet.²⁰⁶

Tab. 4-13: Annahmeprüfungen zum t -Test für unabhängige Stichproben (Spiel – Text)

Prüfung auf Normalverteilung (Shapiro-Wilk)				
	Gruppe	W	p	
Post-Test	S	0,95	0,32	
	T	0,97	0,28	
Prüfung auf Gleichheit der Varianzen (Levenes)				
	F	df ₁	df ₂	p
Post-Test	1,68	1	57	0,20

Hinweis. Ein signifikantes Ergebnis deutet auf Abweichung von einer Normalverteilung hin.

Der Levenes-Test in Tabelle 4-13 zeigt, dass Varianzhomogenität besteht ($F = 1,68$; $p = 0,20$; $n = 59$), da kein signifikantes Ergebnis vorliegt.²⁰⁷ Eine weitere Annahme

203 Mayer 2014. Computer games for learning. S. 45.

204 Jürg Schwarz. „t-Test für unabhängige Stichproben“. Universität Zürich Methodenberatung, 2023. http://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/ttestunabh.html (aufgerufen am 27. November 2023).

205 Springer Verlag. „Lehrbuch Psychologie: Tabellen statistischer Kennziffern“. Springer, 2023. https://lehrbuch-psychologie.springer.com/sites/default/files/atoms/files/tabellen_statistischer_kennziffern.pdf (aufgerufen am 11. Dezember 2023). S. 1.

206 Schwarz 2023. „t-Test für unabhängige Stichproben“.

207 Springer Verlag 2023. „Lehrbuch Psychologie: Tabellen statistischer Kennziffern“. S. 3 f.

ist die nicht nur StatistikerInnen bekannte Normalverteilung der Resultate des *Post-Tests*.²⁰⁸ Wie im Verfahren nach Shapiro-Wilk ersichtlich (Tabelle 4-13), kann für keine der beiden Gruppen eine Abweichung von der Normalverteilung festgestellt werden ($p_s = 0,32$; $p_T = 0,28$). Dies wird umso deutlicher, wenn man die grafische Darstellung der Mittelwerte des *Post-Tests* betrachtet (Abbildung 4-12). Die Kurven der beiden Gruppen ($n_s = 20$; $n_T = 39$) sind auf der rechten Seite der Grafik dargestellt. Sie weisen Ähnlichkeiten zur Normalverteilungsdichte auf, die mit der Glockenkurve nach Gauß dargestellt wird.²⁰⁹ Obwohl die Dichte der Gruppe *T* im Bereich des Mittelwerts eine kleine Einprägung aufweist, kann die statistisch festgelegte Normalverteilung bestätigt werden.

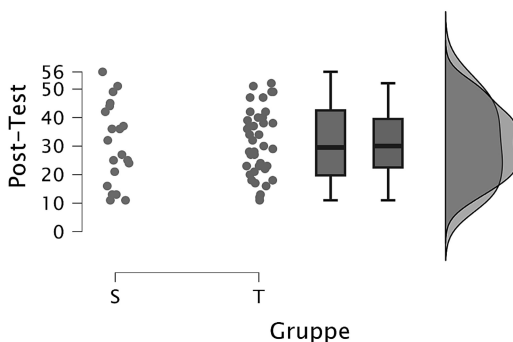


Abb. 4-12: Grafische Darstellung der Post-Test-Resultate (Spiel – Text)

Neben diesen Angaben visualisiert Abbildung 4-12 noch weitere Resultate der des *t*-Tests. Die Punkte des *Jitters* auf der linken Seite stellen die Ergebnisse der Teilnehmenden beider Gruppen dar ($n_s = 20$; $n_T = 39$). Sie bestätigen die Erkenntnisse der numerischen Auswertungen der Effektstärke und Annahmen. Zwischen den *Jitters* und der Verteilungskurve sind die Box-Plots angelegt. Sie orientieren sich am Median der Ergebnisse des *Post-Tests*, die jeweils durch den schwarzen Balken in der Grafik markiert sind ($x_s = 29,50$; $x_T = 30,00$). Ebenso wie bei Mittelwert weisen die Resultate der beiden Gruppen nur eine geringe Differenz auf. Gleiches gilt ebenso für die schwarzen Linien, sogenannte *Whiskers*, die das Minimum und das Maximum der erreichten Punktzahl darstellen ($\wedge_s = 11,00$; $\wedge_T = 11,00$; $\vee_s = 56,00$; $\vee_T = 52,00$).²¹⁰ Ein Unterschied zeigt sich in der Box, die jeweils das obere und das untere Viertel der Ergebnisse ausschließt ($Q1_s = 19,75$; $Q3_s = 42,50$; $Q1_T = 22,50$;

²⁰⁸ Ly und Kehrler 2018. „How to Conduct a Classical Independent Sample T-Test“.

²⁰⁹ Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 103.

²¹⁰ Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 24.

$Q3_T = 39,50$). Hier entsteht der optische Eindruck, dass die Spannweite bei der Gruppe S ($R_S = 22,75$) bedeutend größer ausfällt als bei der Gruppe V ($R_V = 17,00$). Dies lässt sich jedoch anhand der berechneten Standardabweichungen, der Normalverteilung und der Varianzhomogenität nicht bestätigen. Es ist davon auszugehen, dass die unterschiedlichen Teilnehmendenzahlen der Gruppen für diesen optischen Eindruck verantwortlich sind. Vermutlich würden die Ausmaße der Boxen sich angleichen, wenn die Stichprobengröße identisch wäre. Abbildung 4-12 bietet somit eine visuelle Zusammenfassung der vorliegenden Daten. Auch optisch ist weder in den Resultaten noch in der Verteilung eine bedeutende Differenz der Lernphasengruppen zu erkennen.

Analog zur vorangegangenen Methodik kann nun der t -Test der Gruppen S ($n_S = 20$) und V ($n_V = 35$) durchgeführt werden. Dieser kann der Beantwortung der Frage dienen, ob das Digitale Spiel die Inhalte tatsächlich effektiver vermitteln kann als die aufgezeichnete Videosequenz der Tour (ZF2). Die hierzu nötigen Daten sind in Tabelle 4-11 abzulesen. Beginnend mit dem direkten Vergleich der Mittelwerte fällt auf, dass die Differenz 5,07 Punkte beträgt ($m_S = 30,70$; $m_V = 35,77$). Somit kann ein bedeutender Unterschied zwischen der medialen Vermittlung des Digitalen Spiels und des Videos vermutet werden. Ob dieser tatsächlich statistisch signifikant ist, zeigt sich erst anhand der gemessenen Effektstärke (Tabelle 4-14).

Tab. 4-14: t -Test für unabhängige Stichproben (Spiel – Video)

	t	df	p	Cohens d	Std.-Fehler Cohens d
Post-Test	-1,39	53	0,91	-0,39	0,29

Hinweis. Bei allen Tests gibt die Alternativhypothese an, dass Gruppe S größer ist als Gruppe V .

Die Effektstärke ist negativ und grenzt an den signifikanten Schwellenwert an. Mayer spricht bei $d \geq 0,4$ von einem Wert, der als „educationally important“²¹¹ angesehen werden kann. Demnach kann dieses Ergebnis als mittlerer Effekt bewertet werden. Somit kann die Alternativhypothese (ZF2), dass das Digitale Spiel die Inhalte effektiver vermittelt als das Video, nicht bestätigt werden ($d = -0,39$; $p = 0,91$). Eher zeigen sich gar bedeutend größere Lerneffekte bei der Kontrollgruppe V . Der bei Mayer nicht näher erwähnte t -Wert zeichnet ein ähnliches Bild. Auch hier ist ein negatives Ergebnis, das heißt ein optisch besseres Resultat für die Gruppe V , zu beobachten. Allerdings wird der kritische Wert nicht erreicht, wes-

²¹¹ Mayer 2014. Computer games for learning. S. 45.

halb kein statistisch signifikanter Unterschied vorliegt ($t = -1,39$; $df = 53$; $\alpha = 0,05$).²¹² Gerade die gemessenen Lerneffekte lassen jedoch eine Tendenz hin zum Medium der aufgezeichneten Tour erkennen. Diese Annahmen müssen wiederum hinsichtlich Normalverteilung und Varianzhomogenität überprüft werden (Tabelle 4-15). Letzteres geschieht wiederum durch den Test nach Levene, der ein nicht signifikantes Ergebnis aufweist ($F = 0,66$; $p = 0,42$; $n = 55$). Es kann somit von Varianzhomogenität ausgegangen werden.

Tab. 4-15: Annahmeprüfungen zum t-Test für unabhängige Stichproben (Spiel – Video)

<i>Prüfung auf Normalverteilung (Shapiro-Wilk)</i>			
	Gruppe	W	p
Post-Test	S	0,95	0,32
	V	0,93	0,04
<i>Prüfung auf Gleichheit der Varianzen (Levenes)</i>			
	F	df ₁	df ₂ p
Post-Test	0,67	1	53 0,42

Hinweis. Ein signifikantes Ergebnis deutet auf Abweichung von einer Normalverteilung hin.

Die Normalverteilung für die Gruppe S ist bereits oben überprüft worden und kann nicht abgelehnt werden. Es zeigt sich jedoch ein signifikantes Ergebnis für die Gruppe V ($p_s = 0,32$; $p_v = 0,04$). Dies deutet auf eine Abweichung von der Normalverteilung hin.²¹³ Dazu muss jedoch erwähnt werden, dass in der Statistik eine Faustregel existiert, nach der eine Abweichung von der Normalverteilung bei größeren Stichproben ($n \geq 30$) keinen Einfluss auf die Resultate ausübt.²¹⁴ Bei der Gruppe V ist diese Bedingung erfüllt ($n_v = 35$). Dennoch lohnt es sich, visuelle Darstellungen des Datensatzes zu präsentieren, da sie zur weiteren Orientierung dienen können.

212 Springer Verlag 2023. „Lehrbuch Psychologie: Tabellen statistischer Kennziffern“. S. 1.

213 Ly und Kehrer 2018. „How to Conduct a Classical Independent Sample T-Test“.

214 Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 137.

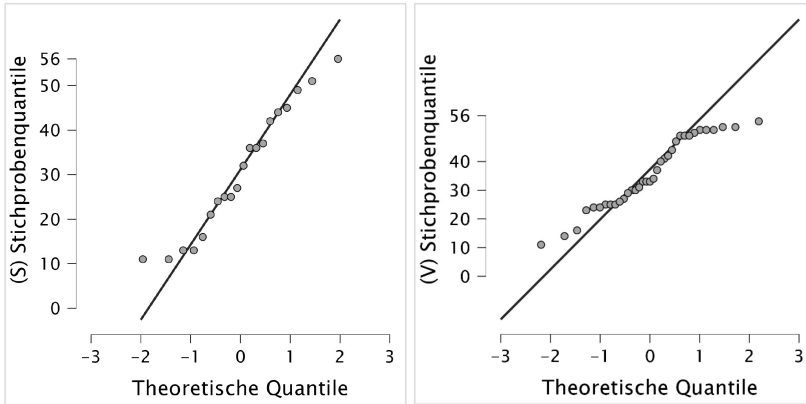


Abb. 4-13: Quantil-Quantil-Diagramm der Post-Test-Resultate (Spiel – Video)

Die Darstellung in Abbildung 4-13 ermöglicht einen direkten Vergleich der theoretischen und der tatsächlichen Verteilung der Ergebnisse. Im Idealfall würden die einzelnen Datenpunkte sich auf der Linie der Normalverteilung befinden. Während das Diagramm der Gruppe *S* sich dem Idealzustand annähert, weisen die Resultate der Gruppe *V* eine sichtbare Schiefe auf. Für Letztere Gruppe kann die Normalverteilung somit auch optisch nicht bestätigt werden. Zudem ist dies an der Einkerbung in der grafischen Darstellung der Dichte erkennbar (Abbildung 4-14). Sie zeigt die Konzentration in den mittleren und hohen Punktzahlen, während die Werte für das untere und obere Mittelfeld nicht entsprechend vertreten sind.

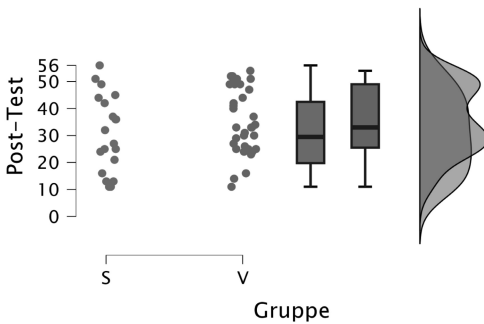


Abb. 4-14: Grafische Darstellung der Post-Test-Resultate (Spiel – Video)

Wie in den Box-Plots ersichtlich, weisen die Mediane ($x_S = 29,50$; $x_V = 33,00$) der beiden Gruppen eine geringere Differenz auf als ihre Mittelwerte ($m_S = 30,70$; $m_V =$

35,77). Der Median teilt den Datensatz in zwei Hälften und ist somit weniger von extremen Ergebnissen beeinflusst.²¹⁵ Der mittlere Effekt, der durch den *t*-Test gemessen wird ($d = -0,39$), kann somit nicht statistisch belegt werden.

4.5.3 Präkonzepte und mediale Vorprägungen

4.5.3.1 Vorhandenes Vorwissen

Da der Fokus auf der Beantwortung der zentralen Forschungsfragen liegt, wird an dieser Stelle bewusst auf die Methoden des *t*-Tests mit abhängigen Stichproben (Éthier), der unabhängigen Stichproben (Mayer) sowie der geeigneten Varianzanalysen (ANOVA) verzichtet. Um die einzelnen Aspekte der erweiterten Forschungsfragen zu beleuchten, werden stattdessen deskriptive Statistiken verwendet.

So kann etwa überprüft werden, ob das Vorwissen zur griechischen Antike einen Einfluss auf die Resultate des *Post-Tests* hat (*EF1*). Dieses Vorwissen wurde durch den *Pre-Test* erhoben. Um die Teilnehmenden in zwei Gruppen zu unterteilen, ist der Median der Gesamtstichprobe herangezogen worden ($x = 21$; $n = 94$). Er gilt als Referenzgröße der Unterscheidung des Vorwissens. So wurden jene Ergebnisse des *Pre-Tests*, die den Wert von 21 Punkten unterschreiten, der Gruppen *B* ($m_B < 21$; $n = 45$) zugewiesen, während die übrigen Teilnehmenden die Gruppe *A* ($m_A \geq 21$; $n = 49$) bilden.

Tab. 4-16: Deskriptive Statistik nach Vorwissen der Teilnehmenden [*A* ($m_A \geq 21$); *B* ($m_B < 21$)]

	Pre-Test		Post-Test	
	A	B	A	B
Gültig	49	45	49	45
Median	26,00	15,00	42,00	25,00
Mittelwert	27,63	15,18	38,76	26,24
Minimum	21,00	7,00	12,00	11,00
Maximum	40,00	20,00	56,00	51,00

Selbstverständlich lässt sich die Differenz im Mittelwert des *Pre-Tests* mit dem Selektionsverfahren der Gruppen erklären ($m_A = 27,63$; $m_B = 15,18$). Bedeutender ist jedoch die Beobachtung der unterschiedlichen Resultate des *Post-Tests*. Hier zeigt sich ein deutlich stärkeres Ergebnis bei der Gruppe *A* ($m_A = 38,76$; $m_B = 26,24$). In Abbildung 4-15 wird dieses Verhältnis noch deutlicher dargestellt.

215 Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 23.

Interessanter für die Frage nach dem Einfluss des Vorwissens bei der Vermittlung der Inhalte der DTAG ist jedoch der festgestellte Lernzuwachs. Nach Éthier kann dieser in der Differenz zwischen *Pre-* und *Post-Test* ermittelt werden (siehe 4.1.1). Während sich für Gruppe A ein Lernzuwachs von 11,13 Punkten beobachten lässt, weist B eine Steigerung von 11,06 Punkten auf. Es zeigt sich, dass die Unterschiede hier sehr gering ausfallen. Um den unterschiedlichen Medien Rechnung zu tragen, können die Ergebnisse des *Post-Tests* ebenfalls nach diesen Kriterien untersucht werden. Dies wird in Abbildung 4-16 aufgearbeitet.

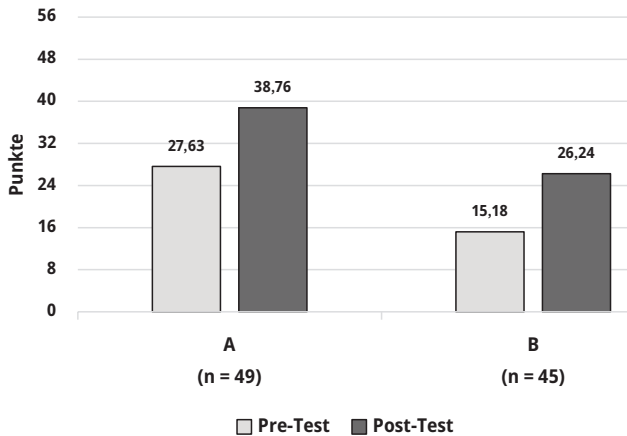


Abb. 4-15: Mittelwerte nach Vorwissen der Teilnehmenden [A ($m_A \geq 21$); B ($m_B < 21$)]

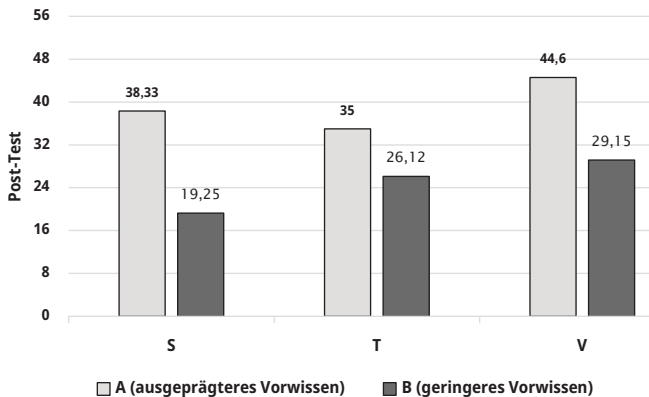


Abb. 4-16: Mittelwerte der Punktzahlen des *Post-Tests* nach Vorwissen ($n_A = 49$; $n_B = 45$) der Teilnehmenden in den Lernphase (Spiel – Text – Video)

Auf den ersten Blick lässt sich feststellen, dass die Gruppe A in jeder Form der medialen Präsentation der Lernphase bessere Resultate aufweist. Besonders die aufgezeichnete Tour scheint hier zu hohen Werten geführt zu haben ($m_{A-V} = 44,60$). Gleiches gilt für die Teilnehmenden mit weniger vorhandenem Vorwissen ($m_{B-V} = 29,15$). Gehen wir vom durchschnittlichen Resultat des *Pre-Tests* aus (Tabelle 4-16) konnten Erstere ihr Ergebnis um 16,97 Punkte steigern, während bei wenig vorhandenem Vorwissen im Post-Test 13,97 Punkte im Vergleich zum *Pre-Test* hinzugewonnen werden konnten. Ein ähnliches Verhältnis lässt sich auch für den bebilderten Text feststellen ($m_{A-T} = 35$; $m_{B-T} = 26,12$). Hier erzielte hingegen die Gruppe A nur 7,37 Punkte mehr im Vergleich zum *Pre-Test*. Gruppe B konnte hingegen mit einer Steigerung von 10,94 Punkten von der Präsentation des Textes in der Lernphase profitieren.

Der gravierendste Unterschied zeigt sich bei den *Post-Tests*, die nach dem tatsächlichen Spielen der Tour absolviert wurden. Dieser kann nicht nur im direkten Vergleich der Vorwissensgruppen festgestellt werden ($m_{A-S} = 38,33$; $m_{B-S} = 19,25$). Während jene Teilnehmende mit umfangreicherem Vorwissen durch das Spielen der Tour 10,7 Punkte hinzugewannen, konnte Gruppe B kaum von der Lernphase profitieren. Mit nur 4,07 Punkten mehr im Vergleich zum durchschnittlichen *Pre-Test*-Ergebnis ($m_B = 15,18$) kann nicht von einem bedeutenden Lernzuwachs gesprochen werden (EF1). Dieses Resultat muss jedoch mit Vorsicht betrachtet werden, da nur ein kleiner Teil der Gruppe B die Tour gespielt hat ($n_B = 45$; $n_{B-S} = 8$).

4.5.3.2 Bildungsabschluss

In der Umfrage wurde weder nach dem Alter der Teilnehmenden noch nach ihrem höchsten Bildungsabschluss gefragt. Stattdessen wurde die aktuell laufende Ausbildung (D2) erhoben. Die festgelegte Altersspanne von 16 bis 21 Jahren führte zu einer enormen Bandbreite der beruflichen und schulischen Bildungssituationen. Um die Angaben für eine deskriptive Analyse verwenden zu können, wurden die Bildungsabschlüsse in drei Kategorien eingeteilt. Erstens jene Teilnehmenden, die zum Zeitpunkt der Umfrage bereits an einer Universität oder Hochschule studieren ($n_{UH} = 20$). Zweitens die Lernenden einer gymnasialen Maturitätsschule oder äquivalenter Angebote im deutschsprachigen Raum ($n_{GM} = 63$). Drittens jene Gruppe, die sich in der beruflichen Grundbildung, in berufsbegleitenden schulischen Angeboten oder momentan nicht in Ausbildung befindet ($n_{BG} = 11$). Abbildung 4-17 zeigt die Ergebnisse des *Pre*- und des *Post-Tests* anhand dieser drei Kategorien.

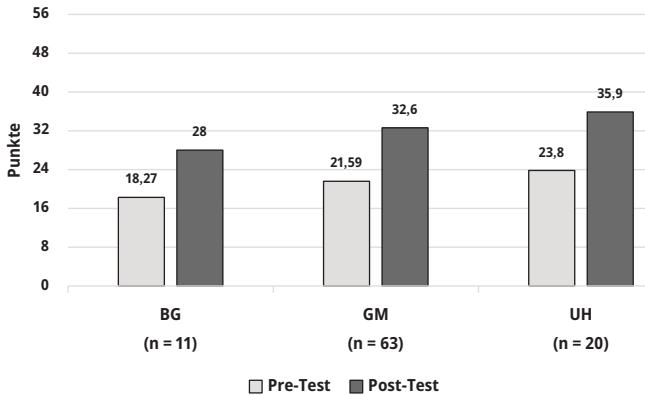


Abb. 4-17: Mittelwerte der Punktzahlen nach aktueller Ausbildung der Teilnehmenden [UH (Universität/Hochschule), GM (Gymnasiale Maturitätsschule) und BG (Berufliche Grundbildung, Fach- und Berufsmaturitätsschule, andere)]

Die Größe der Stichprobe *GM* zeigt, dass rund zwei Drittel des gesamten Datensatzes von Lernenden der Sekundarstufe II stammt. Eine Erkenntnis, die auf der Grundlage des Rekrutierungsverfahrens kaum überrascht (siehe 4.2.3). Im *Pre-Test* lassen sich Unterschiede zwischen den aktuellen Ausbildungssituationen erkennen. So kann bei den Studierenden ein umfangreicheres Vorwissen beobachtet werden als bei den übrigen Teilnehmenden ($m_{UH} = 23,8$; $m_{GM} = 21,59$; $m_{BG} = 18,27$). Erwähnenswert ist jedoch, dass sich bei den drei unterschiedlichen Ausbildungssituationen wiederum eine ähnliche Steigerung vom *Pre-* zum *Post-Test* feststellen lässt. So konnten die Lernenden der Maturitätsschulen einen Lernzuwachs von 11,01 Punkten verzeichnen, während vergleichbare Differenzen bei Studierenden ($\Delta_{UH} = 12,1$) und Teilnehmenden in der Berufsbildung ($\Delta_{BG} = 9,73$) erkennbar sind. Das bedeutet, dass sämtliche Gruppen von den Inhalten der Lernphase profitieren konnten (EF2).

Ob nun aber die mediale Präsentationsform einen Einfluss auf den Lernzuwachs ausübt, kann durch die Aufteilung der Ergebnisse des *Post-Tests* in die bereits bekannten Verzweigungen Spiel (*S*), Text (*T*) und Video (*V*) beobachtet werden. Diese ist in Abbildung 4-18 grafisch dargestellt.

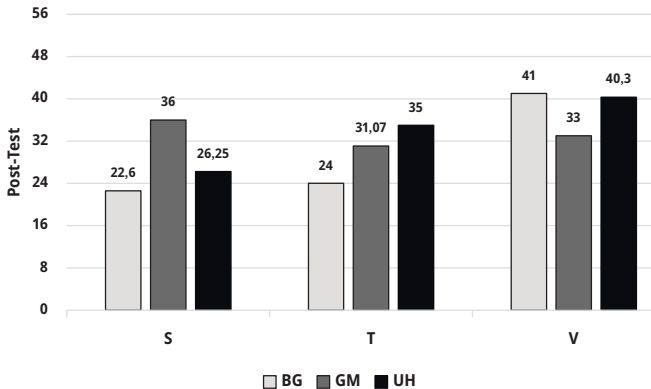


Abb. 4-18: Mittelwerte der Punktzahlen des Post-Tests nach aktueller Ausbildung der Teilnehmenden [UH (Universität/Hochschule), GM (Gymnasiale Maturitätsschule), BG (Berufliche Grundbildung, Fach- und Berufsmaturitätsschule, andere)] in den Lernphasen (Spiel – Text – Video)

Betrachtet man nur die Teilnehmenden, die das Digitale Spiel in der Lernphase gespielt haben, ist auffällig, dass die Lernenden von Maturitätsschulen hier ein besonders gutes Ergebnis aufweisen ($m_{GM-S} = 36$; $n_{GM-S} = 11$). Aufgrund der geringen Stichprobengröße lassen sich keine signifikanten Aussagen treffen. Dies betrifft insbesondere die Vergleichsgruppen mit identischer Lernphase, die sich aus wenigen Teilnehmenden zusammensetzen ($n_{BG-S} = 5$; $n_{UH-S} = 4$). Ähnliche Verhältnisse lassen sich für die übrigen Medienformate feststellen, weshalb keine signifikanten Aussagen zulässig sind ($n_{BG-T} = 3$; $n_{BG-V} = 3$; $n_{UH-T} = 6$; $n_{UH-V} = 10$). Auch innerhalb der Kategorie der Lernenden von Maturitätsschulen können Unterschiede beobachtet werden ($n_{GM-S} = 11$; $n_{GM-T} = 30$; $n_{GM-V} = 22$). An dieser Stelle genügt die Anzahl der Teilnehmenden jedoch, um die obige Tendenz zu bestätigen. Verglichen mit dem gepoolten *Pre-Test*-Ergebnis ($m_{GM} = 21,59$; $n_{GM} = 63$), lässt sich ein Lernzuwachs von 14,41 Punkten beobachten. Gleichzeitig führten die Aufzeichnung und das bebilderte Transkript in der Lernphase zu geringeren Steigerungen von *Pre*- zu *Post-Test* ($\Delta_{GM-T} = 9,48$; $\Delta_{GM-V} = 11,41$).

4.5.3.3 Mediale Vorprägungen

Die allgemeine Mediennutzung der Teilnehmenden zeugt bereits von klaren Präferenzen der Teilnehmenden. Hierzu konnten fünf Medien aus einer Auswahl von 15 Möglichkeiten gewählt werden (M3). Die fünf häufigsten Nennungen werden in Abbildung 4-19 dargestellt.

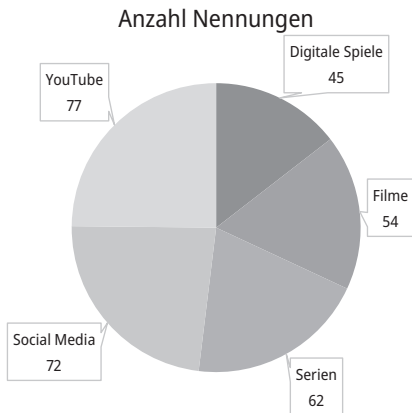


Abb. 4-19: Die fünf am häufigsten verwendeten Medien der Teilnehmenden (n = 94)

Es zeigt sich, dass die Teilnehmenden besonders visuelle Medien, wie etwa Videoportale und Social Media nutzen. Digitale Spiele sind ebenfalls vertreten, wobei das Forschungsdesign der Studie Spielende unweigerlich anspricht. Doch selbst in digitaler Form werden schriftliche Erzeugnisse deutlich seltener genutzt. So erhielten die Newsportale bloß elf Nennungen, während Wikipedia immerhin 30-mal in den häufig genutzten Medien auftauchte.

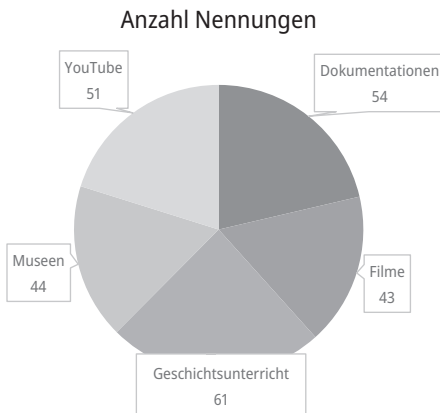


Abb. 4-20: Die fünf meistgenannten Medien, durch die die Teilnehmenden in Kontakt mit historischen Themen kommen (n = 94)

Bevor hier der Fokus auf die Erfahrung mit Digitalen Spielen gelegt wird, betrachten wir, wie die Teilnehmenden dem „Historischen Universum“²¹⁶ begegnen. Die anschließende Frage beschäftigte sich mit den Berührungspunkten der Teilnehmenden mit historischen Themen (M4). Die wählbaren Items waren dabei identisch zur vorgelagerten Frage zur Mediennutzung.

Aufgrund der hohen Anzahl von Lernenden der gymnasialen Stufe, ist die häufige Nennung des Geschichtsunterrichts wenig überraschend. Neben dem außerschulischen Lernort des Museums dominieren auch in dieser Kategorisierung die visuellen Medien. Wobei Filme, Dokumentationen und YouTube sicherlich nicht immer klar voneinander zu trennen sind. Somit ist eine deskriptive Analyse, die nach medialen Vorprägungen durch Videos unterscheidet, nicht zu leisten. Es gab nicht genügend Teilnehmende, die keine Videoformate als Kontaktmedium zu historischen Themen favorisiert haben ($n = 2$). Digitale Spiele werden hingegen nur von 20 Teilnehmenden als häufiger Kontakt zu historischen Themen angegeben, während sie jedoch 45 Nennungen bei den meistgenutzten Medien erhielten (Abbildung 4-19). Um nun den Einfluss der medialen Vorprägungen zu berücksichtigen, können diese Angaben als Referenzwert genutzt werden (Tabelle 4-17).

Tab. 4-17: Deskriptive Statistik der Punktzahlen nach häufiger Nutzung Digitaler Spiele

	Pre-Test		Post-Test	
	häufiger	seltener	häufiger	seltener
Gültig	45	49	45	49
Median	21,00	21,00	34,00	32,00
Mittelwert	21,58	21,76	33,38	32,20
Minimum	7,00	9,00	11,00	11,00
Maximum	40,00	39,00	54,00	56,00

Fast die Hälfte der Teilnehmenden hat Digitale Spiele als eines der fünf meistgenutzten Medien angegeben ($n_{\text{häufiger}} = 45$; $n_{\text{seltener}} = 49$). Die Spielerfahrung hat dabei keinen Einfluss auf die Ergebnisse des *Pre-Tests*, da sich weder Mittelwert noch Median bedeutend unterscheiden. Ähnliches lässt sich für den *Post-Test* beobachten (Tabelle 4-17). Somit kann der Fokus nun auf die Frage gelegt werden, ob die Vermittlung der historischen Inhalte durch vorhandene Spielerfahrung beeinflusst wird (EF3). Obwohl das bebilderte Transkript und die aufgezeichnete Tour ebenfalls aus dem Digitalen Spiel stammen, hat nur die Gruppe S die DTAG in der Lernphase tatsächlich gespielt (Tabelle 4-18).

²¹⁶ Gautschi 2015. Geschichte lehren. S. 6.

Tab. 4-18: Deskriptive Statistik der Gruppe S (Spiel) nach Häufigkeit der Nutzung Digitaler Spiele

	Pre-Test		Post-Test	
	häufiger	seltener	häufiger	seltener
Gültig	14	6	14	6
Median	24,00	20,00	36,00	23,00
Mittelwert	25,29	18,50	32,21	27,17
Minimum	11,00	9,00	11,00	13,00
Maximum	40,00	28,00	51,00	56,00

Auf den ersten Blick scheint die Spielerfahrung keinen Einfluss auf den Lernzuwachs der Tour auszuüben. Vergleicht man die Mittelwerte des *Pre-* und des *Post-Tests*, können die häufig Spielenden ihr Ergebnis nach der Lernphase gar weniger steigern als die andere Gruppe ($\Delta_{\text{häufiger}} = 6,92$; $\Delta_{\text{seltener}} = 8,67$). Bei beiden Gruppen lassen sich jedoch erhebliche Differenzen zwischen Minimum und Maximum erkennen. Diese haben einen Einfluss auf den gemessenen Mittelwert, da aufgrund der Größe der Stichprobengruppen extreme Ergebnisse im Datensatz stärker ins Gewicht fallen. Hier rückt der Median in den Fokus, da er weniger von sogenannten „Ausreißern“²¹⁷ beeinflusst wird. Während die Resultate des *Pre-Tests* beider Stichproben hier enger zusammenliegen, zeigt sich eine bedeutende Differenz im *Post-Test* (EF3). Hier kann ein Unterschied von 13 Punkten festgestellt werden ($x_{\text{häufiger-S}} = 36$; $x_{\text{seltener-S}} = 23$).

4.5.3.4 Offene Aufgaben

Wie im Forschungsdesign beschrieben, folgt die vorliegende Untersuchung keinem *Mixed-Methods*-Ansatz (Siehe 4.2.1). Die drei offenen Aufgabenblöcke des Fragebogens (*P3*, *P4*, *P5*) werden außerhalb der Punktwertung geführt. Dennoch können Sie mit der quantitativen Erhebung verglichen werden und diese ergänzen.²¹⁸ Hierzu musste festgestellt werden, ob die erforderlichen Kompetenzziele erreicht werden konnten (siehe 4.1.5). Dabei wurden weder demografische Angaben noch die Resultate des *Pre-* oder des *Post-Tests* berücksichtigt. Stattdessen wurden die drei offenen Aufgaben anhand einer vierstufigen Skala beurteilt: *nicht erreicht*, *teilweise erreicht*, *erreicht*, *übertrroffen*. Maßgebend für die Beurteilung der Zielerreichung war die zusammengefasste Leistung in allen drei offenen Aufgaben (siehe 4.3.6).

²¹⁷ Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 23.

²¹⁸ Mayer 2014. Computer games for learning. S. 37.

Tab. 4-19: Deskriptive Statistik nach den erreichten Kompetenzzielen der offenen Aufgaben [nicht erreicht (ne), teilweise erreicht (te), erreicht (e), übertroffen (ü)]

	Pre-Test				Post-Test			
	ne	te	e	ü	ne	te	e	ü
Gültig	50	26	10	8	50	26	10	8
Median	17,00	22,00	26,50	28,00	25,00	33,00	44,50	49,00
Mittelwert	19,18	22,19	26,70	29,25	27,38	34,81	41,70	48,63
Minimum	7,00	12,00	17,00	21,00	11,00	11,00	25,00	40,00
Maximum	39,00	33,00	35,00	40,00	56,00	51,00	54,00	52,00

Anhand des vorgestellten Beurteilungsverfahrens konnten die offenen Aufgaben kategorisiert werden (Tabelle 4-19). Auffallend ist die hohe Anzahl an Teilnehmenden, die die geforderten Kompetenzziele nicht erreichen konnten ($n_{ne} = 50$). Somit ist diese Gruppe stärker vertreten als die übrigen drei zusammen ($n_{te} = 26$; $n_e = 10$; $n_{\ddot{u}} = 8$). Zudem wird die Anzahl der Teilnehmenden geringer, je höher die Anforderungen ausfallen. Die Verteilung entspricht somit eher einer Pyramide als der Gauss'schen Kurve. Es finden sich höhere Mittelwerte bei jenen *Pre-Tests* der oberen Kategorie ($m_e = 26,70$; $m_{\ddot{u}} = 29,25$), während bei den unteren geringeres Vorwissen festgestellt wurde ($m_{ne} = 19,18$; $m_{te} = 22,19$). So zeigt sich, dass es die Teilnehmenden mit stark ausgeprägtem Vorwissen sind, die die erweiterten Anforderungen der Interpretations- und Orientierungskompetenzen erfüllen oder gar übertreffen (OF3). Noch auffälliger ist die Differenz der Mittelwerte des *Post-Tests*. Obwohl sämtliche Gruppen ihr Ergebnis steigern konnten, zeigt sich ein Zusammenhang zwischen der Zielerreichung bei den offenen Aufgaben und dem Mittelwert der punktebasierten Auswertung (OF1). So sind es jene Teilnehmende, die erhöhte Anforderungen bei Interpretations- und Orientierungsaufgaben mindestens erfüllen, die gleichzeitig die besten Resultate bei den bepunkteten Wahrnehmungs- und Erschließungsaufgaben des *Post-Tests* aufweisen ($m_{ne} = 27,38$; $m_{te} = 34,81$; $m_e = 41,70$; $m_{\ddot{u}} = 48,63$). Die Daten des Mittelwerts unterscheiden sich dabei kaum von jenen des Medians, was die Aussagekraft der Ergebnisse stützt.

Um den Effekt der dort eingesetzten Medien auf die Erreichung der erweiterten Kompetenzziele zu überprüfen, werden die Häufigkeiten untersucht. In Abbildung 4-21 sind die Teilnehmenden daher nicht nur nach den präsentierten Medien der Lernphase, sondern ebenfalls nach der Kompetenzzielerrreichung der offenen Aufgaben des *Post-Tests* sortiert. Anhand der Häufigkeit der Zielerreichung in den einzelnen Gruppen kann ein möglicher Einfluss des Mediums auf die Erreichung der Kompetenzziele festgestellt werden (OF2).

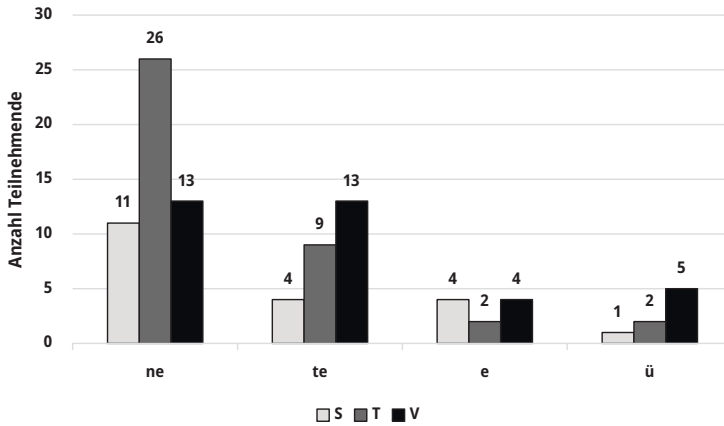


Abb. 4-21: Anzahl der Teilnehmenden nach den erreichten Kompetenzzielen der offenen Aufgaben [nicht erreicht (ne), teilweise erreicht (te), erreicht (e) und übertroffen (ü)] in den Lernphasen (Spiel – Text – Video)

Der größten Anzahl der Teilnehmenden, die die Anforderungen der offenen Aufgaben nicht erfüllen konnte, wurde das bebilderte Transkript präsentiert ($n_{ne-T} = 26$). Ein vergleichbares Bild zeichnet sich bei den übrigen Gruppen ab. Mit der identischen Lernphase konnten nur vier Teilnehmende die erweiterten Kompetenzziele erreichen oder übertreffen ($n_{e-T} = 2$; $n_{ü-T} = 2$). Wenn die Tour tatsächlich gespielt wurde, können bessere Resultate festgestellt werden. So haben fast die Hälfte der 20 Spielenden die Anforderungen der offenen Aufgaben zumindest teilweise erreichen können ($n_{te-S} = 4$; $n_{e-S} = 4$; $n_{ü-S} = 1$). Das aufgezeichnete Video der Tour ist jedoch das einzige Medium der Lernphase, das mehr Teilnehmende mindestens zum teilweisen Erreichen der Kompetenzziele führte. So ist die Lernphase V in den Gruppen mit erweiterten Anforderungen überproportional vertreten ($n_{te-V} = 13$; $n_{e-V} = 4$; $n_{ü-V} = 5$). Sie enthält die größte Anzahl jener Teilnehmenden, die ein angemessenes Sach- und Werturteil in den offenen Aufgaben formulierten (OF2).

4.6 Diskussion

4.6.1 Zentrale Forschungsfragen

Der *Pre-Test* bestand aus allgemeinen *Multiple-Choice*-Fragen zur Geschichte des antiken Griechenlands. Sie überprüften Aspekte, die vermutlich von Teilen der Ge-

sellschaft als Basis- oder Allgemeinwissen bezeichnet werden würde.²¹⁹ Der *Post-Test* hingegen stellte spezifische Fragen zur Schlacht an den Thermopylen, die ohne vorherige Beschäftigung mit dem Thema kaum zu beantworten sind (siehe 4.3.6, Anhang II). Betrachten wir das Verhältnis von *Pre-* zu *Post-Test* mittels deskriptiver Analyse, zeigt sich, dass der Großteil der Teilnehmenden ihre Resultate steigern konnte (Tabelle 4-6). Fasst man sämtliche Gruppen zusammen, kann eine durchschnittliche Verbesserung von 20 Prozent beobachtet werden ($n = 94$). Dieser optische Ersteindruck konnte zudem als statistisch signifikant bestätigt werden (Tabelle 4-8). Folglich hat der Einsatz der *DTAG* in der Lernphase einen messbaren Effekt auf das Wissen der Teilnehmenden. Es ist daher anzunehmen, dass die Ergebnisse des *Post-Test* ohne das Spielen, Lesen oder Betrachten der Tourinhalte nicht in dieser Form aufgetreten wären. Somit lässt sich die Referenzstudie Éthiers bestätigen, dessen Fazit er selbst treffend auf den Punkt brachte: „Everybody learns something!“²²⁰ Diese Aussage ist stets vor dem Hintergrund zu betrachten, dass in der vorliegenden Studie nur Materialien aus dem Digitalen Spiel zum Einsatz kamen. Sämtliche Medien wurden unverändert übernommen und die Lernphase wurde ohne Anleitung einer Fachperson durchgeführt. Bei Jugendlichen und jungen Erwachsenen können demnach Lerneffekte, gemessen an historischem Wissen, festgestellt werden (*ZF1*). Aufgrund des aufwändigen Entstehungsprozesses und der intensiven Zusammenarbeit mit FachexpertInnen wird das Digitale Spiel als „Speicher für (populär-)geschichtliche Wissensbestände“²²¹ wahrgenommen. Wie jedoch die Analyse der *DTAG* gezeigt hat, ist der Forschungsgegenstand *Thermopylen* eine jener Touren, die strukturierte Ereignisgeschichte vermitteln (siehe 3.3.4). Das Urteil lässt sich daher nicht unmittelbar auf Touren ohne klaren geografischen oder temporalen Bezug übertragen.

Ist das Digitale Spiel jedoch in der Lage, die Inhalte der Tour am effektivsten zu vermitteln (*ZF2*)? Die Alternativhypothese der Studie ist, dass das Spielen des Digitalen Spiels bei Jugendlichen und jungen Erwachsenen zu den größten Lerneffekten führt. Anders als das Video oder der Text ist das Spiel das Format, das die Entwickelnden im Sinn hatten. Zudem wurde im Vorfeld der Studie davon ausgegangen, dass eine Mehrheit der Teilnehmenden mit dem Medium vertraut ist. Diese Annahme basiert auf der Referenzstudie Éthiers, bei der mehr als 80 Prozent der freiwilligen Teilnehmenden angaben, häufig Digitale Spiele zu nutzen.²²² Daher würde der Link einer Umfrage zu *Assassin's Creed* sowie die Aussicht auf eine thematisch passende Guthabenkarte potenziell eher jene ansprechen, die bereits

²¹⁹ Vgl.: „Basic Knowledge“ in: Demantowsky 2018. „What Is Public History“. S. 18 f.

²²⁰ Éthier und Durand 2018. *Beyond Gaming*. 18:55–19:00.

²²¹ Schwarz 2023. *Geschichte in digitalen Spielen*. S. 108.

²²² Éthier, Lefrançois und Déry 2022. „Utilisation en classe d'histoire“. S. 19.

mit dem Medium vertraut sind. Die Randomisierung würde anschließend dazu führen, dass sie sich gleichmäßig auf die Untersuchungs- und die Kontrollgruppen verteilen. Beides konnte mithilfe der Erhebung der medialen Vorprägungen bestätigt werden (siehe 4.5.3.3).

Folgt man dem Vergleich der drei Lernphasengruppen muss festgestellt werden, dass kein empirischer Beweis dafür vorliegt, dass das Digitale Spiel die Inhalte der Tour besser vermittelt als die Videoaufzeichnung oder das bebilderte Transkript. Gegen die ursprüngliche These sprechen sowohl die deskriptive Analyse (Abbildung 4-9) als auch die paarweise durchgeführten *t*-Tests für unabhängige Stichproben (siehe Tabelle 4-12 bzw. Tabelle 4-14). Besonders bei Ersterer ist jedoch die Differenz der Stichprobengrößen zu beachten, die das Bild verzerren könnten. Dennoch zeigt sich, dass die Teilnehmenden in den Lernphasen Spiel ($m_S = 54,82$ Prozent) und Text ($m_T = 55,59$ Prozent) vergleichbare *Post-Test*-Resultate erreicht haben, während die Gruppe Video ($m_V = 63,88$ Prozent) durchschnittlich fast zwei Drittel der Fragen korrekt beantworten konnte (ZF2). Verglichen mit den Ergebnissen des *Pre-Tests* lässt sich festhalten, dass Letztere auch den größten Lernzuwachs zu verzeichnen hat ($\Delta_S = 13,3$ Prozent; $\Delta_T = 17,17$ Prozent; $\Delta_V = 26,48$ Prozent). Aus methodischer Sicht könnte man das Ergebnis der Gruppe S mit der Untersuchungsgruppe der Studie Éthiers vergleichen. So wurde jeweils eine Tour ohne Anleitung einer Fachperson gespielt und von einem *Pre*- und einem *Post-Test* gerahmt. Hier konnten die Spielenden sich nach der Lernphase um ganze 23 Prozent steigern, was eine Differenz von fast zehn Prozentpunkten darstellt (Tabelle 4-2). Dies kann wohl am ehesten mit dem unterschiedlichen Ansatz begründet werden. So fanden sich bei Éthier teilweise identische Fragen im *Pre*- und im *Post-Test*, was in der vorliegenden Untersuchung bewusst vermieden wurde (siehe 4.3.4, 4.3.6). Analog zur Referenzstudie lässt sich jedoch ebenfalls festhalten, dass auch für die Lernphase des Digitalen Spiels ein positiver Effekt zu beobachten ist. Die These Éthiers, wonach das Spielen der *Discovery Tour* Lernen auch im unbeleiteten, privaten Gebrauch ermöglicht, kann demnach bestätigt werden.²²³

Wie oben beschrieben, weist jedoch die Gruppe V im Vergleich von *Pre*- und *Post-Test* einen erheblichen Lernzuwachs auf. Für den *media comparison approach*, der Stichprobengröße und Standardabweichung der Gruppen berücksichtigt, zeichnet sich ein ähnliches Bild (siehe 4.5.2). Bei dieser Methode dienen jeweils die *Post-Test*-Resultate zweier Gruppen zur Berechnung der Lerneffekte. Die vergleichenden Effektstärken lassen weder auf einen Vorteil des Digitalen Spiels gegenüber dem Text ($d = -0,03$) noch gegenüber der Videoaufzeichnung ($d = -0,39$) schließen (ZF2). Letztere nähert sich dabei gar dem Grenzwert der Signifikanz von

223 Université de Montréal – Salle de Presse 2018. „Right on Target“.

$d = 0,4$ an. Unweigerlich stellt man sich die Frage nach möglichen Erklärungen für dieses deutliche Ergebnis. So hat die Annahmeprüfung der Gruppe *V* gezeigt, dass eine Abweichung von der Normalverteilung vorliegt (Tabelle 4-15, Abbildung 4-13). Da jedoch eine relativ große Stichprobe vorhanden ist ($n_V = 35$), kann deren Einfluss auf die Resultate nicht endgültig bestätigt werden (siehe 4.5.2). Zudem verringern sich jedoch die Differenzen der *Post-Test*-Resultate erheblich, wenn statt des Mittelwerts ($m_S = 30,7$; $m_T = 31,13$; $m_V = 35,77$) der Median verwendet wird ($x_S = 29,50$; $x_T = 30,00$; $x_V = 33,00$). Die Gruppe *V* ist demnach am stärksten von extremen Ergebnissen beeinflusst. Diese beiden Faktoren sollten bei der Betrachtung der großen Effektstärke ($d = -0,39$) des *t*-Tests der Gruppen *S* und *V* berücksichtigt werden. Trotz dieser statistischen Feinheiten lassen sich jedoch klare Tendenzen erkennen, die auf Videoaufzeichnungen als besonders effektives Medium hinweisen. Hierfür können vier mögliche Interpretationen vorgebracht werden, die meines Erachtens nicht getrennt voneinander zu beurteilen sind. Erstens die unvermeidlichen Nachteile der Lernphasen der Gruppen *S* und *T*. Zweitens nehmen Videoformate eine zentrale Rolle im Medienkonsum der Teilnehmenden ein. Dabei besteht ein nicht zu vernachlässigender Teil aus Aufzeichnungen von Digitalen Spielen (*Let's Plays*). Drittens die nachweisliche Nutzung des Mediums Video in der Aneignung von historischem Wissen durch Jugendliche und junge Erwachsene. Und schließlich die Beschaffenheit der Tour *Thermopylen* im Kontext des Digitalen Spiels *DTAG*.

Das erste Argument betrifft die Erwartungshaltung der Spielenden und die technischen Anforderungen des Forschungsdesigns. Wie die Auswertung der medialen Vorprägungen zeigt, spielt eine bedeutende Anzahl der Teilnehmenden regelmäßig Digitale Spiele. In der Einleitung wurden sie zudem darauf hingewiesen, dass die Umfrage sich mit den Inhalten der *DTAG* beschäftigen wird (siehe 4.3.2). Wie bereits erwähnt, wurden sie in keiner Weise getäuscht. Dennoch erwarteten die Teilnehmenden der Gruppe *T* wohl nicht, dass sie sich mit einem bebilderten Transkript auseinandersetzen würden. Somit ist anzunehmen, dass zumindest bei einigen die Motivation in der Lernphase gesunken ist. Bei der Gruppe *S* sind es hingegen eher technische Hürden, die die Erfahrung der Umfrage beeinflussen konnten. Wenn die Teilnehmenden noch nicht über das Digitale Spiel verfügten, verlängerte sich die Dauer der Umfrage um einige Minuten. Der Anmelde- und Downloadprozess erforderte, gerade von den nicht regelmäßig Spielenden, ein gewisses Maß an Konzentration, Geduld und Entgegenkommen (siehe 4.3.5). Somit verlangte die Lernphase der Gruppe *S* zumindest den Wechsel zu Drittprogrammen, um das Digitale Spiel erhalten. Insbesondere die nötige Registrierung bei *Ubisoft Connect* brachte neun Teilnehmende dazu, den Installationsprozess abzubrechen. Zusätzlich kann nicht sicher geklärt werden, ob die *Discovery Tour* auf sämtlichen Endgeräten in angemessener Qualität gelaufen ist. Es handelt sich hier um

eben jene Hürden, die viele Lehrpersonen vom Einsatz Digitaler Spiele im Geschichtsunterricht abhalten.²²⁴ Ein Faktor, der bei Texten und Videos, aufgrund ihrer ständigen Verfügbarkeit, keine Rolle mehr spielt.

Allerdings kann die Effektstärke der Videoaufzeichnung nicht nur mit den eventuellen Nachteilen der anderen Lernphasen erklärt werden. Hierfür müssen die medialen Vorprägungen der Jugendlichen und jungen Erwachsenen einbezogen werden. Bei dieser Zielgruppe sind Videos, in jeglicher Form, das mit Abstand am weitesten verbreitete Medium. Das lässt sich für die Stichprobe der Studie statistisch belegen und kann wohl als These auch für den gesamten deutschsprachigen Raum angenommen werden. Als am häufigsten verwendete Medien wurden von den Teilnehmenden *YouTube*, *Social Media*, *Serien* und *Filme* genannt (siehe Abbildung 4-19). Alle diese Medien bieten Videos in verschiedenen Kategorien, Formaten und Längen an. Man kann zusätzlich annehmen, dass bei den Social Media vor allem jene Plattformen genutzt werden, die visuell ansprechende Clips präsentieren (*Instagram*, *TikTok* etc.). Für die Teilnehmenden sind Videoaufzeichnungen wohl die alltäglichste und gleichzeitig zugänglichste Medienform. Das Lesen von längeren Texten scheint hingegen im Konsumverhalten der Zielgruppe nur eine geringe Rolle zu spielen. Welche Inhalte die Jugendlichen und jungen Erwachsenen auf *YouTube* und anderen Videoplattformen betrachten, wurde nicht erhoben. Von besonderem Interesse für die Studie ist jedoch, dass die Beschäftigung mit Digitalen Spielen häufig nicht durch das tatsächliche Spielen geschieht. Wenn man einer repräsentativen Studie des deutschen Games-Verbands folgt, kann davon ausgegangen werden, dass rund zwei Drittel der Spielenden zwischen 16 und 24 Jahren Aufzeichnungen von Digitalen Spielen (*Let's Plays*) auf *YouTube* und anderen Videoplattformen betrachten.²²⁵ Dieses Erkenntnis lässt sich auf die vorliegende Untersuchung übertragen, da die Mehrheit der Teilnehmenden nachweislich an Digitalen Spielen interessiert ist (siehe 4.5.3.3). Somit ist davon auszugehen, dass sie im Umgang mit aufgezeichneten Spielsessionen geübt sind. Dies muss, neben den oben erwähnten Nachteilen der Lernphasen *S* und *T*, als weiterer Faktor für die große Effektstärke der Gruppe *V* betrachtet werden.

Der dritte Erklärungsansatz betrifft den wahrnehmbaren Kontakt der Teilnehmenden mit dem Universum des Historischen, also die „unendliche Fülle von In-

224 Über den dafür nötigen technischen und organisatorischen Aufwand kann ich aus der eigenen Lehrtätigkeit berichten. Doch das würde den Rahmen der vorliegenden Arbeit sprengen und muss daher auf zukünftige Projekte verschoben werden.

225 Martin Puppe. „Bei Gaming-Livestreams und Let's Plays schalten täglich Millionen Menschen in Deutschland ein“. game – Verband der deutschen Games-Branche, 27. März 2023. <https://www.game.de/bei-gaming-livestreams-und-lets-plays-schalten-taeglich-millionen-menschen-in-deutschland-ein/> (aufgerufen am 8. Januar 2024).

halten und Themen, von Personen und Ereignissen der Vergangenheit“²²⁶, wie es von Gautschi beschrieben wird. Die medialen Vorprägungen wurden im Vorfeld der Studie bereits antizipiert, weshalb die Frage nach dem Kontakt zu historischen Themen integriert wurde (siehe Abbildung 4-20). Zusammen mit dem außerschulischen Lernort des *Museums* wird auch der *Geschichtsunterricht* genannt. Als Lehrperson ist es beruhigend zu wissen, dass diese Angebote tatsächlich als Kontakt mit historischen Themen wahrgenommen werden. Drei der fünf häufigsten Nennungen sind wiederum den Videoformaten zuzuordnen: *Dokumentationen*, *YouTube* und *Filme*. Obwohl die Formate nicht immer klar zu trennen sind, ist die Auswahl bewusst erstellt worden. Denn es wird noch ersichtlicher, dass der Kontakt zu historischen Themen, außerhalb des schulischen Kontextes, größtenteils über Videoformate stattfindet. Textbasierte Medien und Digitale Spiele wurden hingegen deutlich seltener genannt. Es ist anzunehmen, dass Erstere zwar zu einer kurzen Recherche auf *Google* oder *Wikipedia* dienen, jedoch nicht zu einer längerfristigen Beschäftigung führen. Zweitere werden hingegen häufiger genutzt (siehe Abbildung 4-19), allerdings meist nicht bewusst mit dem Kontakt zu historischen Inhalten verbunden (Abbildung 4-20). Selbstverständlich dienen auch Videoformate vor allem der Unterhaltung. Auf Grundlage der erhobenen Daten werden einige jedoch gleichwohl mit der Intention betrachtet, sich historisches Wissen anzueignen. Das deckt sich mit den Erkenntnissen diverser Studien, nach denen Schülerinnen und Schüler zu einem großen Teil regelmäßig *YouTube* nutzen, um sich für Prüfungen oder Präsentationen vorzubereiten.²²⁷ Das Angebot der Plattform deckt dabei den gesamten Fächerkanon ab und wird dementsprechend häufig auch für historische Inhalte genutzt. Solche Videoangebote scheinen sich im schulischen Alltag vieler Lernender etabliert zu haben. Wie die demografische Erhebung zeigt, besuchen zwei Drittel der Teilnehmenden eine allgemeinbildende Schule der Sekundarstufe II ($n_{GM} = 63$). Zudem ist fast ein Viertel an Hochschulen oder Universitäten eingeschrieben und hat erst kürzlich eine Maturitätsschule abgeschlossen ($n_{UH} = 20$). Die meisten Teilnehmenden der Studie waren darin geübt, zur Prüfungsvorbereitung historische Inhalte in sogenannten Erklär-Videos auf *YouTube*

²²⁶ Gautschi 2015. Guter Geschichtsunterricht. S. 43.

²²⁷ Andreas Cord. „Sesamstraße und Telekolleg als Vorbilder? Erklärvideos auf YouTube als Fortsetzung des traditionellen Schul- und Bildungsfernsehens ‚mit anderen Mitteln‘“. In *Geschichte auf YouTube: neue Herausforderungen für Geschichtsvermittlung und historische Bildung*, herausgegeben von Christian Bunnenberg und Nils Steffen, 27–60. *Medien der Geschichte*, Band 2. Berlin: De Gruyter Oldenbourg, 2019. <https://doi.org/10.1515/9783110599497-002> (aufgerufen am 19. Mai 2020). Hier: S. 46–48.

zu betrachten.²²⁸ Die Lernphase der Gruppe V mit anschließendem *Post-Test* bildet somit ihre alltägliche Lebenswelt ab.

Als vierter Faktor ist schließlich die Struktur der Tour *Thermopylen* im Kontext der *DTAG* zu erwähnen. Aufgrund des dominanten *framing narrative* und des untergeordneten *historical ludonarrative* (siehe 3.3.4), besitzt das tatsächliche Spielen kaum erkennbaren Mehrwert gegenüber dem passiven Betrachten der Videoaufzeichnung (ZF2). Das gilt insbesondere für die fehlenden *exploratory* und *performative challenges* innerhalb der Tour. Hier wird das interaktive Motivationspotenzial des Mediums nicht ausgeschöpft und mögliche Selbstwirksamkeitserfahrungen der Spielenden kaum gefördert.²²⁹ Um dies zu erreichen, müssten die Entwickelnden bereit sein, ihr Konzept zu überdenken und spielerische Freiheiten anzubieten. In diesem Fall wäre es unausweichlich, das Schema der unveränderlichen Erzählungen zu verlassen und die Erinnerungsumgebung (*heritage environment*) in eine Vergangenheitsumgebung (*reenactment*) zu transformieren, in der den Spielenden weitreichendere Handlungsmöglichkeiten angeboten werden (siehe 3.3.5). Dieser Ansatz wird im neusten Teil der *Discovery Tour* bereits verfolgt.²³⁰ Es bleibt abzuwarten, ob aus den *Spielwiesen der Geschichte* der Grundspiele tatsächlich interaktive Lernlandschaften entstehen können.

Wie oben beschrieben, kann besonders die jahrelange Übung und Erfahrung mit dem Medium als historischem Wissensspeicher als zentraler Faktor der Lerneffekte gesehen werden. Ohne Anleitung und Begleitung einer Fachperson ist es somit nicht das Digitale Spiel, das die Inhalte der Touren effektiver vermitteln konnte, sondern deren Videoaufzeichnung (ZF2). Dafür spricht sowohl der deskriptive Vergleich des *Pre-* und des *Post-Tests* nach Éthier als auch – nach Mayer – die gemessenen Effektstärken der Untersuchungsgruppe S: gegenüber der Kontrollgruppen T ($d = -0,03$) und gegenüber der Gruppe V ($d = -0,39$). Zusammenfassend muss dennoch erwähnt werden, dass die Inhalte der drei Lernphasen allesamt aus der *DTAG* entnommen wurden (siehe 4.3.5). Daher bestätigt sich die These, dass durch den Einsatz Digitaler Spiele bei Jugendlichen und jungen Erwachsenen Lerneffekte, gemessen an historischem Wissen, festgestellt werden können (ZF1). Laut der erhobenen Daten gilt dies sowohl für die Präsentation der Tour in Spielform als auch in Video- oder Textform.

228 Anke John. „Wissen2go – Frontalunterricht YouTube“. *Public History Weekly*, 5, Nr. 25 (2017). <https://dx.doi.org/10.1515/phw-2017-9584> (aufgerufen am 8. Januar 2024).

229 Preisinger 2022. Digitale Spiele in der historisch-politischen Bildung. S. 91.

230 AKGWDS Redaktion. „Ein Valhalla für Erinnerungen – Stimmen aus dem AKGWDS zum Release der ‚Discovery Tour: Viking Age‘“. Billet. gespielt, 8. Dezember 2021. <https://gespielt.hypothesen.org/4440> (aufgerufen am 4. März 2022).

4.6.2 Ergänzende Forschungsfragen

Für die Beantwortung der beiden zentralen Forschungsfragen wurden bereits einige Aspekte der Präkonzepte und medialer Vorprägungen einbezogen. Dennoch lohnt sich der Blick auf weitere Erkenntnisse der Studie. Diejenigen Teilnehmenden, die bereits über ein breiteres Vorwissen zur griechischen Geschichte verfügten, konnten ein deutlich besseres Resultat im *Post-Test* erreichen (Tabelle 4-16). Dies gilt unabhängig von der absolvierten Lernphase und entspricht somit der festgestellten Abhängigkeit zwischen den *Pre*- und den *Post-Test*-Resultaten in der Referenzstudie Éthiers (siehe 4.1.1). Aus geschichtsdidaktischer Perspektive waren diese Resultate ebenfalls zu erwarten, da Vorwissen als einer der wichtigsten Faktoren für das Gelingen von Lernprozessen betrachtet werden kann.²³¹ Allgemein fällt es Lernenden leichter, neue Lerninhalte zu verstehen, wenn sie sie mit bereits vorhandenen verknüpfen können. Da es sich hierbei um individuelle Entwicklungen handelt, bilden die Punktzahlen des *Post-Tests* nur einen Teil des Lernpotenzials ab. Umso entscheidender ist der Lernzuwachs, der vom *Pre*- zum *Post-Test* beobachtet werden kann. Denn die 49 Teilnehmenden mit geringerem Vorwissen weisen durchschnittlich eine nahezu identische Steigerung auf ($\Delta_B = 11,06$) wie die mit umfangreicherem Vorwissen ($\Delta_A = 11,13$). Beide konnten gleichermaßen von den Inhalten der Tour profitieren (*EF1*).

Betrachtet man die einzelnen Medienformate separat, zeigt sich, dass die Videoaufzeichnung diesen Gleichschritt passend abbildet. Bei den Lernphasen *S* und *T* hingegen lassen sich Unterschiede erkennen. Interessanterweise konnten die Teilnehmenden mit geringerem Vorwissen ihre Punktzahl im *Post-Test* besonders durch das bebilderte Transkript steigern. Gleichzeitig war es die Lernphase des Spielens der Tour, die sich bei vorhandenem Vorwissen deutlich effektiver zeigte (siehe Abbildung 4-16). Laut diesem Ergebnis können besonders jene von der interaktiven Beschäftigung mit dem Digitalen Spiel profitieren, die bereits über gewisse fachliche Grundlagen verfügen. Die Erfahrung mit dem Medium scheint hingegen kaum Auswirkungen zu haben. Denn beide Kategorien bestanden rund zur Hälfte aus den Teilnehmenden, die häufig Digitale Spiele nutzen. Es ist anzunehmen, dass die Orientierung in der neuartigen, ungewohnten Form der Geschichtsdarstellung besser gelingt, wenn bereits eine gewisse fachliche Grundlage vorhanden ist (*EF1*). So können die präsentierten Lerninhalte der virtuellen Spielwelt direkter als solche wahrgenommen, erschlossen und aufgenommen werden, wenn sie an das Vorwissen der Spielenden anknüpfen.²³² Dies entspricht dem Zweck der mit Punkten bewerteten Aufgaben des *Post-Tests*, die die Wahrnehmungs- und Er-

²³¹ Gautschi 2015. Geschichte lehren. S. 39.

²³² Gautschi 2015. Geschichte lehren. S. 39.

schließungskompetenzen überprüfen (siehe 4.1.5, 4.3.6). Selbstverständlich sollten diese unabhängig vom gewählten Medium gelten. Es zeigte sich jedoch, dass die Teilnehmenden mit geringerem Vorwissen im Umgang mit dem Video und dem Text profitieren konnten, während sie das tatsächliche Digitale Spiel als Informationsträger teilweise zu überfordern schien.

Bei der Aufstellung der Forschungsthese wurde davon ausgegangen, dass die Erfahrung und Übung mit Quellen und Darstellungen im Geschichtsunterricht einen Einfluss auf die Ergebnisse des *Pre-* und des *Post-Tests* ausüben (EF2). Zwar lässt die Halbierung der Stichprobe durch die Erhebung des Vorwissens keine Rückschlüsse auf die momentane Ausbildungssituation der Teilnehmenden zu. Umgekehrt lässt sich bei differenzierter Betrachtung dennoch eine klare Tendenz erkennen. Medienunabhängig weist die Gruppe der Studierenden im *Pre-* und im *Post-Test* die höchsten Mittelwerte auf (Abbildung 4-17). Dies ist sicherlich mit der allgemeinbildenden Laufbahn und der bereits abgeschlossenen Sekundarstufe II zu erklären, die sie von Lernenden in der Berufsbildung oder am Gymnasium unterscheidet. Dies bestätigt die Annahme im Vorfeld der Untersuchung und spricht für die Zweckmäßigkeit der Maturitätsschulen. Ähnlich wie bei der Analyse des Vorwissens kann jedoch beobachtet werden, dass sämtliche Gruppen von den Inhalten der Tour profitieren konnten (EF2). Denn unabhängig vom eingesetzten Medium der Lernphase steigerten sich die Resultate vom *Pre-* zum *Post-Test*. Aus didaktischer Sicht bieten sich demnach Möglichkeiten für den Einsatz im Geschichtsunterricht, da das Medium Imagination und Handlungsorientierung fördern kann. Es ist anzunehmen, dass sich diese Erkenntnis auch auf sämtliche Leistungszüge der obligatorischen Schulzeit übertragen lässt. Denn hier ist der Lebensweltbezug noch stärker hervorzuheben, da mehr als die Hälfte der Lernenden des unteren Niveaus der Sekundarstufe I häufig Digitale Spiele nutzen.²³³ Selbstverständlich wäre es an dieser Stelle angebracht, die Resultate der medialen Präsentationsformen in Betracht zu ziehen. Wie jedoch bereits bei der Ergebnisdarstellung ersichtlich, sind für Hochschulen und Berufsbildung spezifische Stichproben nur im tiefen einstelligen Bereich vorhanden (siehe 4.5.3.2). Daher lassen sich keine Zusammenhänge zwischen Lerneffekten einzelner Medien und der Ausbildungssituation der Teilnehmenden erkennen, geschweige denn Aussagen dazu zu treffen.

Auf den ersten Blick erlaubt es der Datensatz hingegen, Rückschlüsse auf die Lernenden an Maturitätsschulen zu ziehen ($n_{GM-S} = 11$; $n_{GM-T} = 30$; $n_{GM-V} = 22$). Wie in Abbildung 4-18 ersichtlich, scheinen sie im besonderen Maße vom tatsächlichen Spielen der Tour profitiert zu haben. Diese Aussage ist jedoch unter dem Aspekt zu betrachten, dass zehn dieser elf GymnasiastInnen Digitale Spiele zu ihren

233 Bernath u. a. 2020. „JAMES“. S. 21.

meistgenutzten Medien zählten. Somit berichten die Daten weniger über den Einfluss der aktuellen Ausbildungssituation als über die Bedeutung der Spielerfahrung. Die medialen Vorprägungen wurden bereits zur Ergänzung der zentralen Forschungsfragen herangezogen (siehe 4.6.1). Zudem hat die Umfrage ergeben, dass etwa die Hälfte der Teilnehmenden Digitale Spiele als eines der fünf meistgenutzten Medien bezeichnen. Diese Angabe wird als Referenz für die Erfahrung mit dem Medium verwendet. Unabhängig von der medialen Präsentationsform der Lernphase können in beiden Kategorien nahezu identische Ergebnisse festgestellt werden (siehe Tabelle 4-17). Dies gilt ebenso für die Differenz zwischen *Pre-* und *Post-Test*. So kann festgestellt werden, dass vorhandene Spielerfahrung zumindest keine negativen Auswirkungen auf den Lernzuwachs ausübt (*EF3*). Selbstverständlich muss der Fokus nun auf die Lernphase *S* gelegt werden, die dem ursprünglichen Verwendungszweck der *DTAG* entsprach. Die dort zu beobachtende Stichprobengröße der unerfahrenen Spielenden entspricht den Erwartungen der Studie ($n_{\text{häufiger-S}} = 14$; $n_{\text{seltener-S}} = 6$). Dies ist durch die vorhandene Hardware und die Verfügbarkeit des Digitalen Spiels bei den häufiger spielenden Teilnehmenden bedingt (siehe 4.3.5). Obwohl die geringe Anzahl der Stichprobe keine statistisch signifikante Aussage zulässt, darf aufgrund der Datenlage eine Annahme gewagt werden (Tabelle 4-18). So kann die Gruppe der häufig spielenden einen deutlich größeren Lernzuwachs vom *Pre-* ($x_{\text{häufiger-S}} = 24$; $x_{\text{seltener-S}} = 20$) zum *Post-Test* ($x_{\text{häufiger-S}} = 36$; $x_{\text{seltener-S}} = 23$) aufweisen. Demnach übt die vorhandene Erfahrung mit dem Medium einen positiven Effekt auf die Vermittlung historischer Inhalte durch das Spielen Digitaler Spiele aus (*EF3*). Diese Interpretation ist zulässig, obwohl in den Mittelwerten keine bedeutenden Differenzen festzustellen sind. Grund dafür ist ein registriertes Maximalergebnis der seltener spielenden Teilnehmenden im *Post-Test*. Dieses wird auch von der Statistik-Software *JASP* als „Ausreisser“²³⁴ erkannt und dementsprechend markiert (ID = 11). Somit lassen sich die stark divergierenden Resultate des Mittelwerts und des Medians erklären (siehe Tabelle 4-18).

Nach dem Modell des historischen Lernens lässt sich der mit Punkten bewertete *Post-Test* den Wahrnehmungs- und Erschließungsaufgaben zuordnen, wobei der Fokus hier auf die „Wiedergabe von Erlerntem in der erlernten Form“²³⁵ gelegt wurde. Obwohl die Studie keinen *Mixed-Methods*-Ansatz verfolgt, können die offenen Aufgaben der Ergänzung der quantitativen Ergebnisse dienen.²³⁶ Sie erforderten von den Teilnehmenden anspruchsvollere Transferleistungen und die Verschriftlichung eines eigenen Sach- oder Werturteils (siehe 4.1.5). Aufgrund des Ein-

²³⁴ Christensen, Christensen und Missong 2019. Statistik klipp & klar. S. 23.

²³⁵ Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 13.

²³⁶ Mayer 2014. Computer games for learning. S. 37.

bezugs externer Zitate mussten sie sich zudem argumentativ mit Quellen auseinandersetzen, die nicht den Inhalten der *DTAG* entnommen waren. So wurden die Interpretations- und Orientierungskompetenzen der Teilnehmenden überprüft, die zusätzlich für das Gelingen des historischen Lernprozesses benötigt werden.²³⁷ Maßgebend für die Beurteilung der erweiterten Kompetenzzieleerreichung war die zusammengefasste Leistung in den drei offenen Aufgaben (*P2*, *P3*, *P5*). Sie wurde normorientiert und unabhängig von den Ergebnissen des Pre- und des *Post-Tests* vorgenommen (siehe 4.3.6).

In Tabelle 4-19 ist die Kategorisierung der Teilnehmenden nach *nicht erreicht* ($n_{ne} = 50$), *teilweise erreicht* ($n_{te} = 26$), *erreicht* ($n_e = 10$) und *übertrifft* ($n_{ü} = 8$) dargestellt. Bereits die unterschiedliche Anzahl der Teilnehmenden bestätigt die Analyse der *Expert Review* in *qualtrics XM* (siehe 4.4). So wurde ein bedeutender Teil der offenen Aufgaben mit Lückenfüllern vervollständigt, um die Texteingabe zu überspringen. Dennoch kann die Kategorisierung der offenen Antworten interessante Ergänzungen zur quantitativen Erhebung bieten. Wie bereits anhand der zentralen Forschungsfrage (*ZF1*) beschrieben, können durch den Einsatz Digitaler Spiele bei Jugendlichen und jungen Erwachsenen Lerneffekte, gemessen an historischem Wissen, festgestellt werden. So ist es nicht von der Hand zu weisen, dass es jene Teilnehmenden sind, die erhöhte Anforderungen des Interpretierens und Orientierens erfüllen, die gleichzeitig die größten Lerneffekte, gemessen an historischem Wissen, aufweisen (*OF1*). Die Datenlage ist sowohl für die Verwendung des Mittelwerts als auch auf Grundlage des Medians eindeutig. So verringert sich die Anzahl der Teilnehmenden in den oberen Kategorien, während gleichzeitig die durchschnittlichen Ergebnisse des *Post-Tests* kontinuierlich steigen ($m_{ne} = 27,38$; $m_{te} = 34,81$; $m_e = 41,70$; $m_{ü} = 48,63$). Auf Grundlage der Theorie konnte dieses Ergebnis bereits antizipiert werden (siehe 4.1.5). Besonders wenn man den idealtypischen Ablauf des historischen Lernens berücksichtigt, wie er etwa im zyklischen Modell Gautschis dargestellt wird (Abbildung 4-4). Denn die einzelnen Prozesse sollten in einer festgelegten Reihenfolge stattfinden, die sich jeweils auf die vorherigen beziehen. Nur so können „unerwünschte Kurzschlüsse“²³⁸ vermieden werden. Um hingegen zu reflektierten Sach- und schließlich zu Werturteilen zu gelangen, müssen zuvor die Schritte des Wahrnehmens und des Erschließens gelingen. Gleiches gilt für den Prozess in entgegengesetzter Richtung. Da die mit Punkten bewerteten Aufgaben die dafür nötigen Kompetenzen überprüften (siehe 4.3.6), decken sich die beobachteten Ergebnisse mit den zuvor getätigten Annahmen.

Dennoch ist hier eine weitere Voraussetzung für das Erreichen der Kompetenzziele der offenen Aufgaben zu erwähnen. So hat die Analyse der *DTAG* gezeigt,

²³⁷ Gautschi 2015. Guter Geschichtsunterricht. S. 48.

²³⁸ Gautschi 2015. Guter Geschichtsunterricht. S. 48.

dass die Touren zu einem bestimmten Ereignis, zu denen etwa der vorliegende Untersuchungsgegenstand *Thermopylen* zählt, besonders für die Vermittlung von deklarativem Wissen geeignet sind (siehe 3.3). In isolierter Form, das heißt ohne professionelle Einbettung und Begleitung einer Fachperson, werden weiter gefasste Konzepte und Strukturen nur marginal thematisiert. Neben der Bereitschaft, sich aktiv mit der Darstellung auseinanderzusetzen, ist vorhandenes Kontextwissen zur griechischen Geschichte nötig, um die Anforderungen der offenen Aufgaben des *Post-Tests* zu erreichen. Wie es bereits die Daten für die geschlossenen Wissens- und Erschließungsaufgaben belegen, können die präsentierten historischen Inhalte direkter als solche erkannt, verarbeitet und gespeichert werden, wenn sie an das Vorwissen der Spielenden anknüpfen (*EF1*). Die Beurteilung der offenen Aufgaben (*P2*, *P3*, *P5*) und die Einteilung der Antworten der Teilnehmenden in die vier vorgestellten Kategorien ergänzen diesen Befund. Vorhandenes historisches Kontextwissen begünstigt nicht nur weiteren Wissenserwerb, sondern es kann auch das Gelingen der Prozesse des historischen Lernens unterstützen und fördern (*OF3*). Dafür spricht die Beobachtung, dass jene Teilnehmenden, die bereits im *Pre-Test* hohe Ergebnisse erzielten, die Anforderungen der offenen Aufgaben erreichen oder übertreffen konnten ($m_{ne} = 19,18$; $m_{te} = 22,19$; $m_e = 26,70$; $m_{ü} = 29,25$). Da die hierzu erforderlichen Kompetenzen „an historischen Inhalten ausgebildet, angewendet und ausdifferenziert werden“²³⁹, ist Sachwissen somit gleichzeitig Folge und Voraussetzung des historischen Lernens. Dies ist als Kernaussage der erhobenen Daten in Tabelle 4-19 zu bezeichnen. Die Fähigkeiten zur Lösung der offenen Aufgaben zeigten sich insbesondere bei jenen Teilnehmenden, die bereits über entsprechendes Vorwissen verfügten und denen es gleichzeitig besser gelang, sich die Inhalte der Lernphase anzueignen (*OF3*, *OF1*).

Ob nun die mediale Präsentationsform einen Einfluss auf die Erreichung der Kompetenzziele hat, ist mit den bisher verwendeten Methoden nicht zu bestimmen. Das betrifft besonders die Verwendung der durchschnittlichen Ergebnisse des *Pre-* oder des *Post-Tests*, die als Referenzwerte zur Beantwortung der zentralen Forschungsfragen dienten (siehe 4.6.1). Denn die Beurteilung der offenen Aufgaben fand unabhängig von den Resultaten der quantitativen Erhebung statt. Vielmehr geht es darum herauszufinden, ob die drei Lernphasen sich gleichmäßig innerhalb der vier Beurteilungsstufen *nicht erreicht*, *teilweise erreicht*, *erreicht* und *übertroffen* verteilen (Abbildung 4-21). In der untersten Kategorie ist auffällig, dass mehr als die Hälfte der Teilnehmenden das bebilderte Transkript gelesen hat. Demnach lässt sich festhalten, dass der Text nur bei wenigen Teilnehmenden zur Erreichung der Kompetenzziele verhelfen konnte (*OF2*). Grund dafür sind vor al-

239 Gautschi, Hodel und Utz 2009. „Kompetenzmodell für ‚Historisches Lernen‘“. S. 11.

lem die zahlreichen Antworten, die nur aus Lückenfüllern oder Nonsens-Eingaben bestehen (~60 Prozent). Besonders erwähnt sei an dieser Stelle eine Einreichung, die die drei Fragetexte kopierte, um sie als Antwort einzufügen. Für solche Aktionen ist wohl die fehlende Motivation der Gruppe *T* nach der Lernphase verantwortlich. Diese wurde bereits erläutert (siehe Kapitel 4.6.1) und kann ebenso für die Gruppe *S* herangezogen werden. Erst recht, da die üblichen Hürden beim Einsatz von Digitalen Spielen bei den ungenügenden Antworten nach der spielenden Lernphase ($n_{\text{ne-S}} = 11$) zu vernachlässigen sind. So gaben neun der Teilnehmenden in der Umfrage an, bereits Teile des Franchise *Assassin's Creed* gespielt zu haben (*A2*), während sechs täglich am Computer spielen (*M1*, *M2*). Mangelnde technische Voraussetzungen oder Schwierigkeiten im Umgang mit Digitalen Spielen können damit als Argument ausgeschlossen werden. Vielmehr ist es der Tatsache geschuldet, dass „in Webumfragen häufig keine sehr wertvollen Textantworten“²⁴⁰ zu erwarten sind. Somit stützt es eher die verbreitete These des Motivationspotenzials Digitaler Spiele, wenn fast die Hälfte der Teilnehmenden die Kompetenzziele zumindest teilweise erreichen konnte.²⁴¹ Indessen lässt sich dies nur für ein Drittel derjenigen beobachten, die das bebilderte Transkript gelesen haben (siehe Abbildung 4-21). Bei der Videoaufzeichnung hingegen kann ein diametral gegensätzliches Resultat zur Lernphase *T* festgestellt werden. Hier befinden sich nur ein Drittel der Teilnehmenden auf der untersten Stufe der Kompetenzzieleerreichung ($n_V = 35$; $n_{\text{ne-V}} = 13$). Zudem ist diese Gruppe auch in der obersten Kategorie am stärksten vertreten. Dies deckt sich mit den Befunden des medienvergleichenden Lernzuwachses zwischen *Pre-* und *Post-Test* (siehe 4.6.1). Somit weist die Gruppe *V* nicht nur die höchsten Mittelwerte im bewerteten *Post-Test* auf, sondern enthält auch die größte Anzahl jener Teilnehmenden, die ein angemessenes Sach- und Werturteil in den offenen Aufgaben formulierten (*OF2*). Individuell und ohne Begleitung einer Fachperson sind historische Lernprozesse demnach effektiver mittels Videoaufzeichnung zu initiieren als durch das Lesen oder Spielen der Tourinhalte (*OF2*).

Neben der fehlenden Motivation für Textantworten bei Online-Umfragen sind die Ursachen für die oft mangelnde Kompetenzzieleerreichung im Angebot des Forschungsgegenstands zu finden. Zahlreiche Teilnehmende der Kategorien *nicht erreicht* und *teilweise erreicht* haben keine Lückenfüller genutzt, sondern versucht, sich aktiv mit den offenen Aufgaben auseinanderzusetzen. Wie die Analyse der *DTAG* jedoch zeigt, funktionieren besonders Touren wie die *Thermopylen* als chronologische Berichte von Ereignissen (siehe 3.3.4). Die statistische Auswertung belegt, dass diese tatsächlich zu einem Lernzuwachs, gemessen an historischem Wissen, beitragen können (*ZF1*). Eine Auseinandersetzung mit weiter gefassten Kon-

240 Vanette 2023. „Umfragemethodik“.

241 Vgl.: Preisinger 2022. Digitale Spiele in der historisch-politischen Bildung. S. 89–92.

zepten und Strukturen, die auf der Grundlage historischer Fragestellungen untersucht werden, findet kaum statt. Wichtige geschichtsdidaktische Prinzipien wie Kontroversität und Multiperspektivität wurden gar bewusst vermieden, indem die Texte der FachexpertInnen von den Entwickelnden angepasst und gekürzt wurden (siehe 2.7.3). Daher kann die *DTAG* im privaten Gebrauch zwar Wissen vermitteln, ohne jedoch die die Ausbildung der für das historische Lernen charakteristischen Problemlösefähigkeiten zu fokussieren. Es kann angenommen werden, dass die guten Ergebnisse der offenen Aufgaben weniger auf die mediale Präsentationsform der Tour, sondern eher auf bereits vorhandene Kompetenzen der Teilnehmenden zurückzuführen sind (*OF2*).

4.6.3 Limitationen und Forschungsdesiderata

Bereits bei der Interpretation der Ergebnisse wurden Limitationen berücksichtigt und auf deren Auswirkungen eingegangen. Sie betreffen die Versuchskontrolle, den Forschungsgegenstand und die Auswertung der Ergebnisse.

So müssen zuvorderst die Limitationen der Online-Umfrage thematisiert werden. Diese betreffen Aspekte der Versuchskontrolle, wie sie von Mayer definiert werden.²⁴² Die Entscheidung für die asynchrone Form der Untersuchung wurde unabhängig von der pandemischen Lage getroffen. Die private Nutzung an heimischen Geräten ließ nicht nur eine höhere Anzahl an Teilnahmen erwarten, sondern entsprach auch dem Forschungsdesiderat der Referenzstudie Éthiers.²⁴³ Im Endeffekt muss diese frühe Festlegung als Glücksfall betrachtet werden, da eine Präsenzdurchführung im Winter 2021/22 nicht möglich gewesen wäre. Daraus ergaben sich jedoch spezifische Herausforderungen, die auf die Einbindung des Digitalen Spiels zurückzuführen sind. Denn dieses war, anders als Texte und Videos, nicht unmittelbar in das Umfragetool integrierbar. Zudem war die *DTAG* nicht kostenlos abrufbar. Dank der Bereitstellung von Codes durch *Ubisoft* konnte zumindest dieses Problem gelöst werden. Neben der Registrierung in der externen Anwendung *Ubisoft Connect* mussten die Teilnehmenden zudem gewisse Hardwarevoraussetzungen erfüllen, um auf das Digitale Spiel zugreifen zu können. Daher wurde bereits im Vorfeld mit einer hohen Abbruchrate in der Lernphase *S* gerechnet. Um dennoch ausreichende Einreichungen zu erhalten, wurden die jeweiligen Teilnehmenden stattdessen einer der beiden Kontrollgruppen zugeteilt. Folglich führte dies, trotz zuvor gleichmäßiger Randomisierung durch die Software, zu einer geringeren Stichprobengröße in der Untersuchungsgruppe ($n_S = 20$; $n_T = 39$;

²⁴² Mayer 2014. Computer games for learning. S. 34–37.

²⁴³ Université de Montréal – Salle de Presse 2018. „Right on Target“.

$n_V = 35$). Neben dem erschwerten Zugang zum Digitalen Spiel ergaben sich aus der asynchronen Durchführung weitere Limitationen. So konnte nicht überprüft werden, ob die Teilnehmenden sich weiterer Hilfsmittel bedienten. Im Verlauf der Umfrage wurde mehrfach darauf hingewiesen, dass es sich nicht um eine Prüfung handelt, und darum gebeten, auf externe Hilfsmittel zu verzichten (*G0*, *P0*). Allerdings kann deren Einsatz dennoch weder für den *Pre*- noch für den *Post-Test* ausgeschlossen werden. Eine geschlossene Browserumgebung war bei der Verwendung des Programms *Ubisoft Connect* ebenfalls nicht umzusetzen. Zudem wäre jeder technische Aufwand wohl kaum praktikabel gewesen, da man den Zugriff der Teilnehmenden auf ihre Smartphones ohnehin nicht hätte verhindern können. Der Datensatz lässt jedoch darauf schließen, dass die meisten Teilnehmenden den Aufruf zum Verzicht auf externe Hilfsmittel beherzigt haben.

Eine bedeutend größere Herausforderung stellten die Anforderungen der Versuchskontrolle an den Forschungsgegenstand dar. Der *media comparison approach* verlangt die Übertragung in verschiedene Medienformate, die jedoch identische Inhalte präsentieren. Für das Video war das Aufzeichnen eines *Let's Play* die offensichtlichste und zuverlässigste Variante. Allerdings stellte die Bereitstellung identischer Informationen in Textform eine große Herausforderung dar. Ursprünglich sollte ein populärwissenschaftlicher Aufsatz aus einem Geschichtsmagazin für die Lernphase Text verwendet werden.²⁴⁴ Allerdings konnten die Anforderungen damit keinesfalls erfüllt werden, da sie unweigerlich andere Informationen präsentierten als die AutorInnentexte der *DTAG*. Daher wurde die Tour eigenhändig transkribiert und mit passenden Screenshots versehen. Doch wie wohl auch Regisseure einer Romanverfilmung wissen, ist eine Übertragung auf ein anderes Medium nicht ohne Kompromisse zu leisten. Bezogen auf interaktive Digitale Spiele, betrifft dies besonders die Entscheidungsfreiheit und die Gliederung der Informationen.²⁴⁵ So lässt das aufgezeichnete Video keine eigenen Richtungsänderungen oder Blickrichtungswechsel zu. Trotz der eingefügten Bilder kann der Text zudem weder die Umgebung der Spielwelt darstellen noch die Tonspur wiedergeben. Diese Limitation ist unvermeidlich, da sie die Kerneigenschaften der drei Medien betrifft.

Unterschiedliche Stichprobengrößen oder die aufwändige Aneignung des Digitalen Spiels können sich auf die Resultate auswirken. Auch wenn sie, bis auf die *Post-Tests* der Gruppe V, nicht verdächtig sind, von der statistischen Normalverteilung abzuweichen. Daher erfolgte die Auswertung der quantitativen Daten auf Grundlage der bereits erwähnten Limitationen. Insofern dies möglich war, wur-

²⁴⁴ Lars Abromeit. „Ein Leben für den Krieg“. Geo Epoche Kollektion, Nr. 8, Das antike Griechenland. (2017), 42–50.

²⁴⁵ Mayer 2014. Computer games for learning. S. 36.

den sie jedoch bei der Vorstellung (4.5) und der Interpretation der Ergebnisse (4.6.1) berücksichtigt. Im Gegensatz dazu wurden die offenen Aufgaben bereits im Vorhinein aus dem bewerteten Teil der Studie ausgeklammert. Es zeigte sich, dass der Einschätzung der *Expert Review* vertraut werden konnte, die von ihrem Einsatz in Online-Umfragen abriet.²⁴⁶ Noch dazu wiesen die Fragen einen hohen Schwierigkeitsgrad auf, da sie erweiterte Anforderungsbereiche überprüfen sollten. Die Limitationen betreffen somit die allgemeine Bereitschaft, Texteingaben in Online-Umfragen zu tätigen, wie auch die vorgenommene Kategorisierung der Antworten. Da es sich nicht um eine qualitative Untersuchung handelt, wurden sie anhand einer vierstufigen Skala beurteilt. Dies entspricht dem Vorgehen einer Geschichtslehrperson, die die Textantworten der Lernenden anhand von bestimmten Kriterien bewertet und zwecks Notenfindung quantifizieren muss. Eine vollkommen objektive Bewertung ist, anders als bei *Multiple-Choice*-Aufgaben, hier nicht zu erreichen.²⁴⁷ Gleiches gilt selbstverständlich auch für die Einstufung der offenen Aufgaben der vorliegenden Studie. Damit sind die grundlegenden Limitationen von Textaufgaben in Online-Umfragen beschrieben. Ein vielversprechender Ansatz, kompetenzorientierte Aufgabenstellungen in quantitative Studien zu integrieren, ist mir leider erst nach Ende der Datenerhebung zur Kenntnis gelangt. Die Rede ist von den *umgekehrten Multiple-Choice*-Aufgaben, wie sie bereits in der amerikanischen Geschichtsdidaktik eingesetzt werden.²⁴⁸ Umgekehrt deshalb, da nur eine der vier Antworten tatsächlich falsch ist und die anderen mit einem bis drei Punkten bewertet werden. Laut Manuel Köster können diese tatsächlich „historische Denkleistungen“²⁴⁹ abbilden, was üblichen Mehrfachwahlaufgaben nicht gelingt. Potenziell wären sie daher in der Lage, die Interpretations- und Orientierungskompetenz zu überprüfen, ohne eine längere Texteingabe zu erfordern. Zudem wären die Sach- oder Werturteile bereits mit Punkten versehen, was die Integration in eine quantitative Studie erleichtert.

Damit lassen sich die Forschungsdesiderata unmittelbar aus den Limitationen der Studie ableiten. Denn gerade in den Geistes- und Sozialwissenschaften sind auf diesem Gebiet kaum empirische Arbeiten vorhanden. Es fehlt somit schlicht an Vergleichsstudien, anhand derer man die Erkenntnisse einordnen könnte. Um diese Verhältnisse zu ändern, werden zwei methodische Konzepte vorgeschlagen. Erstens birgt der medienvergleichende Ansatz anhand von Online-Umfragen wei-

²⁴⁶ Vanette 2023. „Umfragemethodik“.

²⁴⁷ Klaus Fieberg. „Test – Lernkontrollen als Instrumente der Leistungsbeurteilung und der Diagnose“. *Praxis Geschichte*, 1 (2010), I–VI. Hier: S. II.

²⁴⁸ Manuel Köster. *Aufgabenkultur im Geschichtsunterricht. Methoden historischen Lernens*. Frankfurt/M: Wochenschau, 2021. S. 73.

²⁴⁹ Köster 2021. *Aufgabenkultur im Geschichtsunterricht*. S. 73.

terhin großes Potenzial für die Erhebung der Lerneffekte Digitaler Spiele. Neben dem Wissenserwerb könnte zudem spezifischer auf die Kompetenzbereiche des historischen Lernens eingegangen werden. Da Textantworten sich bekanntlich nicht für ein solches Setting eignen, sind Überlegungen zu alternativen Aufgabenformaten erwünscht. Interessant wäre etwa die Integration der oben vorgestellten Variante der *umgekehrten Multiple-Choice*-Aufgaben. Sie könnten die Vorteile der geschlossenen und der offenen Aufgaben vereinen und somit einen Mehrwert für Online-Umfragen bieten. Selbstverständlich müsste vorher überprüft werden, ob die formulierten Fragen tatsächlich die erwünschten Denkleistungen abbilden.

Ein weiteres Anliegen betrifft den Forschungsgegenstand selbst. So wäre eine vergleichbare Untersuchung zu anderen Touren der *DTAG* zu begrüßen. Damit könnte die Strukturanalyse des Digitalen Spiels auch empirisch belegt werden (siehe 3.3). Noch interessanter wäre jedoch ein Blick auf die stetige Weiterentwicklung der Reihe. So unterscheidet sich bereits der Nachfolgetitel *Discovery Tour: Viking Age* deutlich vom vorliegenden Forschungsgegenstand. Es wäre äußerst interessant zu überprüfen, ob die vorgenommenen Anpassungen sich auch auf die messbaren Lerneffekte auswirken. Das gilt selbstverständlich auch für kommende Veröffentlichungen der Reihe, in der Hoffnung, dass diese weiterhin von *Ubisoft* unterstützt wird. Das führt unweigerlich zu einem Wunsch, der nur indirekt an die Geschichtsdidaktik gerichtet ist. So lassen sich Lehrmittel, Sachbücher oder Lernvideos miteinander vergleichen. Die *DTAG* kann hingegen nur ihren Vor- und Nachfolgetiteln gegenübergestellt werden. Es liegt großes Potenzial darin, dass die bekannten Studios mit FachexpertInnen zusammenarbeiten, um vergleichbare Angebote zu entwickeln. Konkurrenz belebt bekanntlich das Geschäft, was der Entwicklung des gesamten Genres zugutekäme. Umgekehrt sollte die Geschichtsdidaktik sich dazu bereit erklären, entsprechende Projekte zu realisieren. Wie Demantowsky es für die Forschung und Entwicklung zur Public History vorschlägt, ist hier die Kooperation mit nicht-akademischen Institutionen erforderlich.²⁵⁰ Nur so kann deren Qualität gesichert werden und ein tatsächlicher Mehrwert für den Unterricht entstehen. Von den so entstehenden geschichtskulturellen Angeboten, sofern sie sowohl technisch als auch fachlich hochwertig produziert sind, könnten Lehrpersonen und Lernende profitieren. Einen vielversprechenden Ansatz verfolgt etwa das Projekt *When we disappear*, welches in Zusammenarbeit der Pädagogischen Hochschule Luzern und des Zürcher Studios *Inlusio Interactive* entwickelt wird.²⁵¹

²⁵⁰ Demantowsky 2018. „What Is Public History“. S. 31.

²⁵¹ Institut für Geschichtsdidaktik und Erinnerungskulturen der Pädagogische Hochschule Luzern und Inlusio Interactive. „Schülerprobung: Ein Videogame für historische Bildung“. *When We*

Als letztes und wichtigstes Desiderat sollten künftige Studien den ursprünglichen Verwendungszweck der Reihe *Discovery Tour* berücksichtigen. So ist sie gleichzeitig für die individuelle Nutzung geeignet, wurde allerdings spezifisch für den Einsatz in Bildungskontexten konzipiert (siehe 2.7.3). Ziel der Entwickelnden war es dabei nicht, die Lehrperson zu ersetzen, sondern ihr ein geeignetes Medium für den Unterricht zur Verfügung zu stellen.²⁵² Während die vorliegende Studie den privaten Gebrauch beobachtet, hat Éthier diesen Aspekt bereits untersucht und wird diesen auch in einem Folgeprojekt fokussieren.²⁵³ Dieser Ansatz sollte weiterverfolgt werden, um Erkenntnisse zur optimalen Einbindung der *Discovery Tour* und anderer Digitaler Spiele in den Geschichtsunterricht zu gewinnen. So könnten etwa unterschiedliche Lehr-Lern-Settings miteinander verglichen werden. Eignet sie sich besonders als Einstieg oder zur Ergebnissicherung und Repetition? Mit welchen konkreten Aufträgen kann ein Mehrwert des Digitalen Spiels gegenüber etablierten Medien und Methoden festgestellt werden? Diese Fragen sollten auch für kommerzielle Unterhaltungsspiele geklärt werden. Denn der Zweck dieser und weiterer Forschungsarbeit sollte die möglichst sinnvolle Einbettung der geschichtskulturellen Angebote in den Unterricht sein. Nicht als Pauschallösung oder Wundermittel, sondern als weitere „Präsentationsform der Geschichte“²⁵⁴, die historische Lernprozesse initiieren und fördern kann.

Disappear – Remember the Past, 2023. <https://www.whenwedisappear.com/de/edu> (aufgerufen am 2. März 2024).

252 Picard 2019. „Discovery Tour Ancient Greece – a behind the Scene Look with Maxime Durand“.

253 Mittlerweile erschienen: Éthier und Lefrançois 2023. „Chapter 1 Historical Video Games and Teaching Practices“.

254 Pandel 2014. „Medien“. S. 136.