Digitale Souveränität in der Lehrkräftebildung

Eine empirische Untersuchung professioneller Entwicklungsprozesse im Kontext generativer Künstlicher Intelligenz und des Physikunterrichts

Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Jannik Henze

Angenommen im Jahr 2026

Inhaltsverzeichnis

Danksagung

Auch wenn diese Arbeit ein Ergebnis individueller Forschung ist, wäre sie ohne das Netzwerk aus Menschen, die mich auf diesem Weg begleitet, unterstützt und inspiriert haben, nicht möglich gewesen. Ihnen allen gilt mein aufrichtiger Dank.

Ich danke Professor Dr. André Bresges dafür, dass er mir in meiner Forschung so große Freiheiten eingeräumt hat. Er hat mich in sein Team aufgenommen und mir damit ein Umfeld geschaffen, in dem ich meinen eigenen Weg finden konnte. Seine Offenheit für neue Ideen, sein Vertrauen in meine Arbeit und die vielen anregenden Gespräche haben meine Denk- und Arbeitsweisen, fachlich wie persönlich, nachhaltig geprägt. Seine Fähigkeit, Begeisterung für Forschung zu wecken und gleichzeitig einen Blick für das Wesentliche zu behalten, war und ist mir stets eine große Hilfe.

Professor Dr. Andreas Schadschneider danke ich für seine verlässliche Begleitung durch alle Phasen meiner Promotion. Mit seiner ruhigen Art und seinem Humor lockerte er so manche Situation auf und wahrte eine angenehme Gelassenheit, die oft genau das Richtige war. Seine Ratschläge waren immer treffend und haben mir in vielen Momenten Orientierung gegeben. Für seine Zeit, seine Geduld und seine kluge, wohlwollende Art bin ich ihm sehr dankbar.

Ich danke auch Prof. Dr. Benjamin Rott für die Übernahme des Vorsitzes und Dr. Cristal Schult für die Übernahme des Beisitzes der Disputation und ihre Unterstützung im gesamten Promotionsprozess.

Professor Dr. Sebastian Becker-Genschow möchte ich besonders für seine inhaltliche und strukturelle Unterstützung danken. Durch seine Analysen und Rückmeldungen hat er entscheidend dazu beigetragen, dass meine Arbeit an Tiefe und Struktur gewann.

Julia Lademann danke ich von Herzen für ihre fachliche, strukturelle und persönliche Unterstützung. Ihre klugen Anregungen, ihr offenes Ohr und ihre ehrliche Art haben mir stets sehr geholfen. Ebenso möchte ich Sascha Therolf, Tobias Solymosi und Simon Höfting für die gemeinsame Zeit als Kollegen und Freunde danken. Ihre Hilfsbereitschaft und Zuverlässigkeit machten den Arbeitsalltag oft leichter. Auch Christine Quambusch danke ich für ihre regelmäßige Unterstützung. Besonders dankbar bin ich ihnen allen dafür, dass ich mich immer auf sie verlassen konnte und kann.

Allen weiteren Kolleg:innen, auch den ehemaligen, im Institut für Physikdidaktik danke ich herzlich für ihre Geduld und die vielen Gespräche, die meinen Blick erweitert haben. Sie trugen mit ihren Ideen, Kritik und Unterstützung dazu bei, dass mein Weg klarer werden konnte.

Von Herzen danke ich meiner Verlobten Maike für ihre Liebe, Geduld und dafür, dass sie immer an mich geglaubt hat. Sie hat mir den Rücken freigehalten, mich in schwierigen Momenten aufgefangen und unterstützt. Ohne sie wäre dies alles nicht möglich gewesen.

Zusammenfassung

Diese Dissertation untersucht, wie Fortbildungskonzepte gestaltet werden müssen, um die digitale Souveränität von Lehrkräften im Umgang mit modernen Technologien zu stärken und für den Fachunterricht nutzbar zu machen. Die Arbeit analysiert in drei aufeinander aufbauenden Studien, unter welchen Bedingungen sich digitale Souveränität bei Lehrkräften entwickeln lässt und wie diese in die fachdidaktische Praxis integriert werden kann.

Die erste Studie fokussiert die motivationalen und einstellungsbezogenen Grundlagen bei der Implementierung digitaler Technologien im schulischen Kontext. Mit 14 Lehrpersonen wurde im Pre-Post-Test-Design untersucht, wie digitale Kreativitätswerkzeuge die digitale Souveränität von Lehrkräften beeinflussen. Die Ergebnisse zeigen ein Spannungsverhältnis zwischen der durch die Lehrpersonen wahrgenommenen Schüler:innenkompetenz im Umgang mit technischen Werkzeugen und der eigenen Anwendungssouveränität. Die Studie identifizierte die Notwendigkeit einer reflexiven Austauschphase zum etablierten 5E-Modell.

Die zweite Studie untersucht den Aufbau von AI Literacy bei angehenden Physiklehrkräften. Studierende erhielten in zwei Iterationen mit jeweils zwei Sitzungen eine Einführung in die Arbeit mit KI-Werkzeugen. Die Ergebnisse zeigen nicht signifikante Zuwächse in der AI Literacy, steigerten jedoch das Gefühl der professionellen Handlungsfähigkeit. Der Vergleich der Iterationen verdeutlicht, dass kollektiver Austausch essenziell ist, da ohne diesen Fortschritte in anspruchsvollen Bereichen limitiert bleiben.

Die dritte Studie fokussiert den didaktischen Mehrwert von KI bei der Auswertung physikalischer Pendelexperimente im Vergleich zur Auswertung mit Excel. Der Chatbot ExperiMentor unterstützte Lernende durch Hinweise und Visualisierungen, ohne fachliche Lösungen vorzugeben. In einem randomisierten Pre-Post-Kontrollgruppendesign mit 50 Lehramtsstudierenden zeigten beide Gruppen signifikante Wissenszuwächse ohne Unterschiede zwischen den Gruppen. Die KI-Gruppe erreichte jedoch signifikant höhere Werte in Motivation und positiven emotionalen Lernerfahrungen bei gleichzeitig reduzierter Frustration und geringerer kognitiver Belastung.

Die Studien belegen, dass digitale Werkzeuge und KI-Systeme zwar als Scaffolding kognitive Entlastung bieten und Motivation fördern können, jedoch ohne menschliche Bezugspersonen keine umfassende Erweiterung digitaler Souveränität ermöglichen. Reflexion, kritische Bewertung und die Entwicklung einer professionellen Haltung bleiben an soziale Aushandlungsprozesse gebunden. Digitale Souveränität entsteht durch die systematische Kombination beider Scaffolding-Formen. Die Ergebnisse zeigen, dass erfolgreiche Professionalisierung handlungsorientierte Zugänge zu Technologie erfordert, um Berührungsängste abzubauen. Formate müssen über Einzelinterventionen hinausgehen und Strukturen etablieren, in denen wiederholt zwischen praktischer Anwendung und reflexivem Austausch gewechselt wird.

Abstract

This dissertation examines how continuing education concepts must be designed in order to strengthen teachers' digital sovereignty in dealing with modern technologies and make it usable for subject teaching. In three interrelated studies, the work analyses the conditions under which digital sovereignty can be developed among teachers and how it can be integrated into subject-specific teaching practice.

The first study focuses on the motivational and attitudinal foundations for implementing digital technologies in a school context. Using a pre-post test design, 14 teachers were examined to determine how digital creativity tools influence teachers' digital sovereignty. The results show a tension between the teachers' perception of their students' competence in using technical tools and their own confidence in using them. The study identified the need for a reflective exchange phase on the established 5E model.

The second study examines the development of AI literacy among prospective physics teachers. Students received an introduction to working with AI tools in two iterations, each consisting of two sessions. The results show insignificant increases in AI literacy, but did increase the feeling of professional competence. A comparison of the iterations illustrates that collective exchange is essential, as without it, progress in challenging areas remains limited.

The third study focuses on the didactic added value of AI in the evaluation of physical pendulum experiments compared to evaluation with Excel. The chatbot ExperiMentor supported learners with hints and visualisations without prescribing technical solutions. In a randomised pre-post control group design with 50 teacher training students, both groups showed significant knowledge gains with no differences between the groups. However, the AI group achieved significantly higher scores in motivation and positive emotional learning experiences, while experiencing reduced frustration and lower cognitive load.

The studies show that digital tools and AI systems can offer cognitive relief and promote motivation as scaffolding, but without human reference persons, they cannot enable a comprehensive expansion of digital sovereignty. Reflection, critical evaluation and the development of a professional attitude remain tied to social negotiation processes. Digital sovereignty arises from the systematic combination of both forms of scaffolding. The results show that successful professionalisation requires action-oriented approaches to technology in order to reduce reservations. Formats must go beyond individual interventions and establish structures in which there is a repeated shift between practical application and reflective exchange.

1. Einleitung

Der Begriff *KI-Ära* wurde von der Gesellschaft für deutsche Sprache zum Wort des Jahres 2025 gewählt (Gesellschaft für deutsche Sprache e. V., 2025) und steht damit für den Wandel, den Künstliche Intelligenz (KI)[1] in nahezu allen Lebensbereichen ausgelöst hat. Digitale Werkzeuge prägen den Alltag, indem sie Informations- und Kommunikationsstrukturen verändern und damit neue Formen sozialer Interaktion ermöglichen (Döbeli Honegger, 2017; Pfiffner et al., 2021). Der souveräne Umgang mit digitalen Medien und Künstlicher Intelligenz ist somit zu einem wichtigen und festen Bestandteil des gesellschaftlichen Lebens geworden. Unter den Bedingungen einer digital geprägten Gesellschaft gewinnt in diesem Zusammenhang die reflektierte Auseinandersetzung mit KI-Systemen besondere Bedeutung (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025).

Ein angemessenes Verständnis der heutigen Welt erfordert sowohl bei Lernenden als auch bei Lehrenden grundlegende digitale Kompetenzen (Döbeli Honegger, 2017). Eine von Zawacki-Richter et al. (2019) durchgeführte Metaanalyse zeigt das zunehmende Interesse an KI im Bildungsbereich und verdeutlicht das didaktische Potenzial. Die Autor:innen betonen, dass Lehrkräfte ein fundiertes Verständnis von KI entwickeln müssen, um die Technologie gewinnbringend für Lehr- und Lernprozesse einsetzen zu können. Die Verschränkung digitaler Transformationsprozesse mit soziokulturellen Lebensbereichen macht die technologische Urteilskraft zu einer wichtigen Voraussetzung demokratischer Partizipation (Thimm, 2023). Gleichzeitig entstehen mit dem Einsatz digitaler Technologien neue Formen kollaborativer Wissens- und Innovationsprozesse, die bisherige Strategien der Wissensvermittlung grundlegend verändern (Ferdinand et al., 2017).

Schulen und Bildungseinrichtungen im Allgemeinen übernehmen in diesem Kontext eine Schlüsselrolle, denn sie haben die Möglichkeit, digitale Kompetenzen zu fördern und damit Voraussetzungen für autonome Entscheidungs- und Gestaltungsprozesse zu schaffen (Initiative D21 e.V., 2025). Lehrende stehen somit vor der Aufgabe, auf technologische Entwicklungen flexibel zu reagieren und ihr professionelles Wissen kontinuierlich anzupassen (Huwer et al., 2019). Digitale Souveränität entsteht dabei nicht beiläufig, sondern setzt den langfristigen Aufbau individueller Kompetenzen im Umgang mit digitalen Technologien voraus. Dieser Entwicklungsprozess muss durch gezielte strukturelle und curriculare Maßnahmen innerhalb von Bildungseinrichtungen aktiv begleitet und ermöglicht werden (Blossfeld et al., 2018; Loroff et al., 2017).

[1] In dieser Arbeit wird der Begriff KI für generative KI in Form von Large Language Models (LLM) verwendet, die als Chatbots eingesetzt werden, zur Vereinfachung durchgängig mit KI abgekürzt.

Wie wirksam Lehrkräfte digitale Technologien wie KI in den Unterricht integrieren, hängt maßgeblich von ihrem fachlichen, didaktischen und technologischen Wissen sowie ihren Einstellungen ab (Backfisch et al., 2021). Empirische Befunde zeigen, dass insbesondere im Umgang mit KI viele Unsicherheiten bestehen (Jude et al., 2025). Hier setzen Aus- und Fortbildungsangebote an, die darauf zielen, digitale Souveränität gezielt zu fördern und Lehrkräften handlungsfähige Kompetenzen im Kontext von KI zu vermitteln.

Aus dem Anspruch einer reflektierten und theoretisch fundierten Integration digitaler Technologien rückt die fachliche Perspektive in den Mittelpunkt (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024). Das Fach Physik bietet durch seine enge Verbindung von Natur, Technik und gesellschaftlicher Reflexion besondere Ansatzpunkte, um digitale und KI-bezogene Kompetenzen aufzubauen (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025). Physikunterricht kann Lernenden ermöglichen, technologische Entwicklungen kritisch einzuordnen, ihren Einfluss auf das eigene Leben zu verstehen und digitale Informationen verantwortungsvoll zu nutzen (Schröter et al., 2021). Damit bietet der Physikunterricht eine ideale Grundlage, um Konzepte zu entwickeln und zu erproben, die die digitale Souveränität von Lehrkräften gezielt stärken.

1.1. Zielsetzung der Arbeit

Ziel dieser Dissertation ist es, zu analysieren, wie sich digitale Souveränität von Lehrkräften im Umgang mit modernen Technologien wie KI durch gestaltungsorientierte Aus- und Fortbildungskonzepte schrittweise entwickelt und welche Implikationen sich daraus für die fachdidaktische Gestaltung von Lehr Lernprozessen ergeben. Dabei verbindet die Arbeit eine theoretische Fundierung des Konzeptes der digitalen Souveränität mit der empirischen Analyse von Aus- und Fortbildungskonzepten. Die fachdidaktische Perspektive des Physikunterrichts dient als Rahmen, um allgemeine Erkenntnisse zur Professionalisierung von Lehrkräften mit spezifischen Anforderungen an eine reflexive Auseinandersetzung mit Technologien zu verknüpfen.

1.2. Struktur der Arbeit

Die vorliegende Dissertation befasst sich mit der Frage, wie digitale Souveränität in der physikalischen Lehrkräftebildung im Kontext generativer Künstlicher Intelligenz entwickelt werden kann. Ausgangspunkt ist die zunehmende Bedeutung digitaler und KI-gestützter Werkzeuge für Unterricht, Ausbildung und professionelle Entscheidungsprozesse von Lehrkräften.

In Kapitel 1 werden die Problemstellung eingeführt, die Forschungsfragen formuliert und der Aufbau der Arbeit skizziert. In Kapitel 2 werden die theoretische Fundierung dargelegt und lernpsychologische Perspektiven, didaktische Modelle sowie zentrale Konzepte zu generativer

KI, AI Literacy und digitaler Souveränität in einen gemeinsamen Bezugsrahmen gestellt. Die Methodologie der kumulativen Dissertation wird in Kapitel 3 beschrieben. Drei empirische Studien bilden die Grundlage der Arbeit und beleuchten digitale Souveränität aus unterschiedlichen Perspektiven. In Kapitel 4 werden diese Studien in einer Synopse eingeordnet und ihr jeweiliger Beitrag zum Gesamtzusammenhang verdeutlicht. Kapitel 5 umfasst die drei Studien, während in Kapitel 6 die Ergebnisse diskutiert und synthetisiert werden. Mit dem siebten Kapitel wird die Arbeit mit einem Fazit und einem Ausblick auf zukünftige Forschungs- und Entwicklungsbedarfe abgeschlossen.

2. Theoretische Fundierung

Mehrere der folgenden Konzepte wurden bereits in den der Dissertation zugrunde liegenden Einzelpublikationen entwickelt und dort erläutert. Die folgende Darstellung konzentriert sich daher auf eine einordnende Zusammenfassung dieser Ansätze, während übergreifende Konzepte, die für das Gesamtverständnis der Arbeit zentral sind, ausführlicher behandelt werden.

2.1. Digitale Souveränität und AI Literacy

Digitale Souveränität bezeichnet die Fähigkeit mit modernen digitalen Technologien verantwortungsvoll, kompetent und autonom das eigene Handeln zu gestalten (Thimm, 2023). Der Begriff beschreibt also neben den technischen Kompetenzen auch eine soziale Handlungsfähigkeit und demnach einen reflektierten Umgang mit digitalen Technologien, die kritische Einordnung potenzieller gesellschaftlicher Folgen und die verantwortungsvolle Mitgestaltung der damit verbunden Möglichkeiten (Krupka, 2020; Stubbe, 2017; Thimm, 2023). Dies gilt auch für den Umgang mit KI. Mit digitaler Souveränität verbunden ist ein grundlegendes Verständnis der Funktionsweisen und Wirkmechanismen KI-basierter Systeme sowie die Fähigkeit, deren ethischen und rechtlichen Implikationen systematisch in pädagogische Entscheidungen einzubeziehen (Sekretariat der Kultusministerkonferenz, 2024). Ein kompetenter Umgang mit digitalen Medien bildet die Grundlage für eine dauerhaft verankerte Medienbildung, die sich darin ausdrückt, Verantwortung für das eigene digitale Informationshandeln zu übernehmen und dessen Auswirkungen reflektiert zu beurteilen (Blossfeld et al., 2018).

Digitale Souveränität entsteht nicht durch isolierten Kompetenzerwerb, sondern durch Interaktionen, die digitales und soziales Handeln verknüpfen. Ein sicherer und reflektierter Umgang mit Technologie wird insbesondere durch Lerngelegenheiten unterstützt, in denen digitale Werkzeuge aktiv erprobt, hinterfragt und eingeordnet werden (Stubbe, 2017). Für die Entwicklung digitaler Souveränität ist es daher erforderlich, Lerngelegenheiten zu schaffen, die über Wissensvermittlung hinausgehen und erfahrungsbasierte Auseinandersetzungen mit digitalen Technologien ermöglichen (Loroff et al., 2017).
Eine Notwendigkeit dieser pädagogischen Ausrichtung wird deutlich, denn in Deutschland verfügt weniger als die Hälfte der Bevölkerung über grundlegende digitale Kompetenzen (Initiative D21 e.V., 2025). Des Weiteren ist bezüglich der Nutzung von KI eine Abhängigkeit vom Bildungsniveau erkennbar. KI-Anwendungen werden vor allem von Personen mit höherem Bildungsabschluss genutzt. Dies verweist auf einen Bedarf an niedrigschwelligen und inklusiven Fördermaßnahmen im Bereich KI-bezogener Kompetenzen (Initiative D21 e.V., 2025). Für die Bereitschaft zur Nutzung digitaler und KI gestützter Technologien erweist sich dabei insbesondere der wahrgenommene praktische Nutzen im Alltag und im beruflichen Kontext als entscheidend (Initiative D21 e.V., 2025).

Im internationalen Kontext zeigt sich ein ähnliches Bild. Drei Viertel der befragten Pädagog:innen einer US-amerikanischen Studie äußerten den Bedarf nach Weiterbildungsmaßnahmen zu den grundlegenden Funktionsweisen und Potenzialen generativer KI (Esbenshade et al., 2025). Dies wird durch eine Vergleichsstudie gestützt, die auf ein ausgeprägtes Defizit in der KI-bezogenen Professionalisierung von Lehrkräften hinweist, da bislang nur ein geringer Anteil über entsprechende Fortbildungserfahrungen verfügt, während gleichzeitig ein hoher Bedarf an strukturierten Weiterbildungsangeboten besteht (Traga Philippakos & Rocconi, 2025). Ergänzend zeigt die US-amerikanische Befragung, dass sich mehr als zwei Drittel der US-Lehrpersonen wünschten, Weiterbildungen zur Automatisierung zeitintensiver Planungs- und Förderprozesse zu erhalten, welche durch KI geschehen können. Weiterhin sahen 65 % der Befragten zudem die Notwendigkeit professioneller Hilfestellungen, um Schüler:innen kompetent im reflektierten Umgang mit generativer KI anleiten zu können (Esbenshade et al., 2025).

Damit KI sinnvoll in unterrichtliche Kontexte integriert werden kann, benötigen Lehrkräfte ein konkretes Orientierungswissen, das realistische Einschätzungen von Einsatzmöglichkeiten erlaubt (Traga Philippakos & Rocconi, 2025). Dies wird im internationalen Diskurs als AI Literacy verstanden, einem Sammelbegriff für grundlegende Kompetenzen, um mit Künstlicher Intelligenz reflektiert umgehen zu können (Chiu, 2025; EU Artificial Intelligence Act, 2024; D. Long & Magerko, 2020; Traga Philippakos & Rocconi, 2025). In dieser Arbeit wird der etablierte Begriff AI Literacy im Sinne der offiziellen deutschen Fassung der EU KI Verordnung verwendet: Das Europäische Parlament übersetzt den Bergiff AI Literacy in Artikel 3 Nummer 56 als KI-Kompetenz und definiert ihn als "*Fähigkeiten, Kenntnisse und Verständnis, die es [...] Anwendern [...] ermöglichen, KI-Systeme […] in Kenntnis der Sachlage einzusetzen und sich über die Chancen und Risiken von KI und mögliche Schäden, die sie verursachen kann, bewusst zu werden*" (EU Artificial Intelligence Act, 2024). Dies deckt sich mit den Definitionen von Long und Magerko (2020) sowie Traga Philippakos und Rocconi (2025), aus welchen hervorgeht, dass eine ausgeprägte AI Literacy Personen in die Lage versetzt, den Einsatz von KI kritisch einzuordnen, mit KI zielgerichtet zusammenzuarbeiten und sie verantwortungsvoll und zweckorientiert zu nutzen, ohne sich auf die technische Fähigkeit des Umgangs mit KI zu beschränken. Empirisch zeigte sich im Rahmen einer KI-Fortbildung für Lehrkräfte ein signifikanter Zuwachs der allgemeinen AI Literacy sowie signifikante Verbesserungen, insbesondere in schulpraxisnahen Kompetenzen (Lademann et al., 2026). Weitere Befunde weisen darauf hin, dass Personen mit höherem Bildungsniveau in der Regel über eine ausgeprägtere AI Literacy verfügen (Zhao et al., 2022) und dass insbesondere die Fähigkeit Lehrender, kompetent mit KI umzugehen, entscheidend dafür ist, ob die Integration dieser in den Unterricht erfolgreich verläuft (Bauer et al., 2025; Traga Philippakos & Rocconi, 2025). Die professionelle Kompetenz der Lehrkräfte im Umgang mit KI ist somit eine zentrale Voraussetzung, um die Zusammenarbeit mit KI-Systemen sinnvoll und dauerhaft gewinnbringend zu gestalten.

2.2. Lernpsychologische Perspektiven

2.2.1. Vygotskys Zone der proximalen Entwicklung

Der russische Psychologe Lev Vygotsky entwickelte eine Theorie, die Lernen als soziales Phänomen versteht: Vygotsky betont, dass Wissen nicht isoliert entsteht, sondern durch soziale Interaktionen geprägt ist (Leong & Bodrova, 1996). Höhere kognitive Fähigkeiten sind seinem Verständnis nach eng in soziale Prozesse eingebunden und entwickeln sich im Austausch mit anderen (Bliss, 1996).

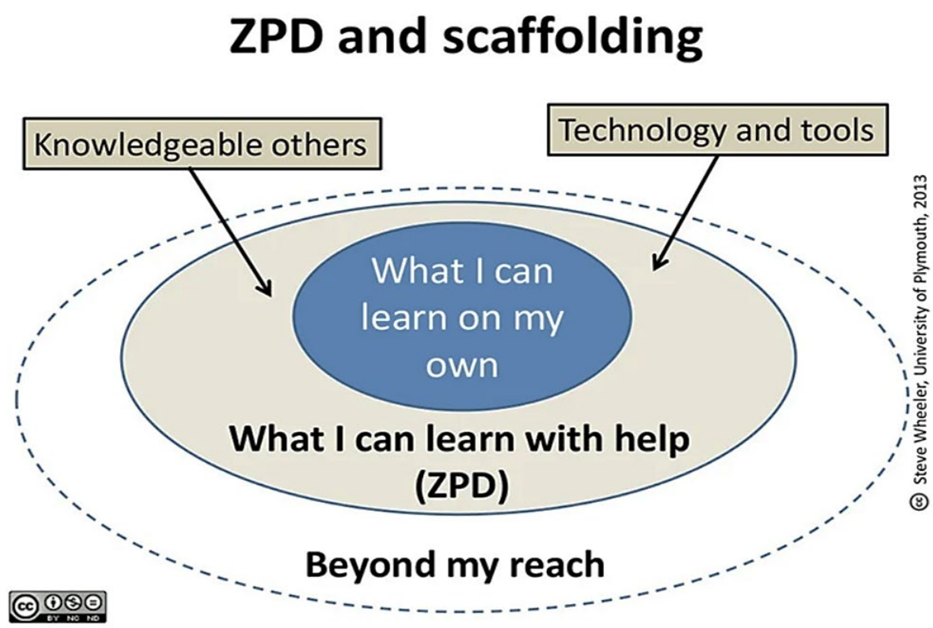

Abbildung 1 Die Zone der proximalen Entwicklung nach Vygotsky. © 2013 von Steve Wheeler, lizenziert unter CC BY-NC-ND nach McLeod (2025).

Darauf aufbauend beschreibt Vygotsky die Unterscheidung zwischen verschiedenen Kompetenzstufen (siehe Abbildung 1). Die erste Stufe stellt das eigenständige Vermögen einer lernenden Person dar (*What I can learn on my own* in Abbildung 1), eine Aufgabe ohne Hilfe zu bewältigen. Die zweite Stufe beschreibt, was diese Person mit Unterstützung erreichen kann (*What I can learn with help* in Abbildung 1). Diese Stufe wird als Zone of proximal development (ZPD, dt. Zone der proximalen Entwicklung) (Bliss, 1996) bezeichnet, und gibt an, in welchem Kontext Entwicklung stattfinden kann. Je kompetenter eine Person wird, desto weiter dehnt sich ihre individuelle ZPD aus und ermöglicht die Bewältigung schwierigerer Herausforderungen, die zuvor außerhalb der eigenen Lernreichweite lagen (*Beyond my reach* in Abbildung 1). Dies ermöglicht es ihr wiederum, anderen dabei zu helfen, ihre eigenen Kompetenzen zu erweitern (Leong & Bodrova, 1996). Abbildung 2 veranschaulicht die Zone der proximalen Entwicklung als dynamisches Verhältnis von Kompetenzen und Herausforderungen. Sie macht deutlich, dass Lernen besonders wirksam dort stattfindet, wo Anforderungen und Fähigkeiten in Balance stehen und damit ein Flow-Zustand erreicht wird, während Überforderung zu Angst und Unterforderung zu Langeweile führt (Repenning, 2012).

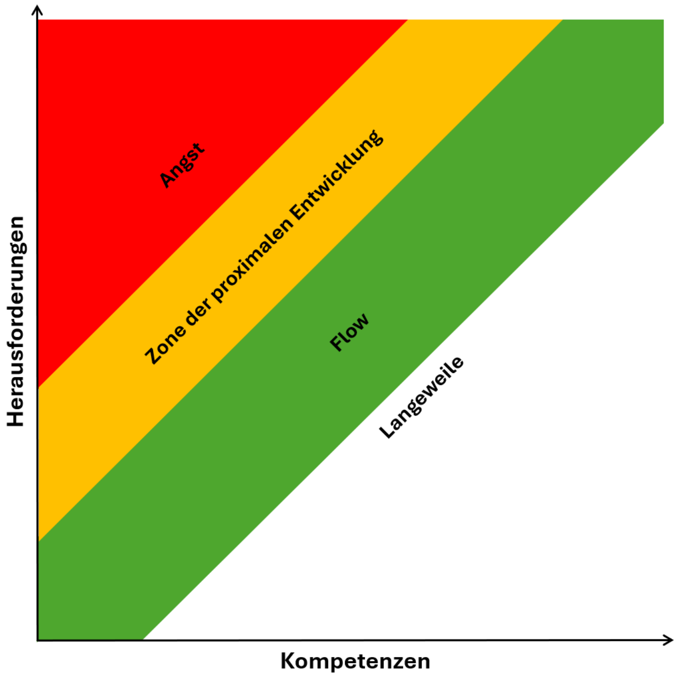

Abbildung 2 Zone der proximalen Entwicklung nach Vygotsky: Verhältnis von Kompetenzen und Herausforderungen sowie Einordnung von Angst, Flow und Langeweile. Übersetzt nach Repenning (2012).

Vygotsky beschreibt in diesem Zusammenhang das Konstrukt des More Knowledgeable Other (MKO), also einer Person mit einem Wissensvorsprung gegenüber der lernenden Person. Diese unterstützt Lernende dabei, ihre aktuellen Möglichkeiten zu überschreiten und Aufgaben innerhalb der Zone der proximalen Entwicklung zu bewältigen. Diese Rolle muss nicht zwingend durch eine Lehrperson erfüllt sein, sondern kann auch von Gleichaltrigen oder durch eine Technologie wie KI (siehe Abbildung 1) ausgefüllt werden (McLeod, 2025; Rigopouli et al., 2025).

In der Praxis wird für die Umsetzung das Prinzip des Scaffoldings angewandt. Dieser Begriff beschreibt eine pädagogische Praxis, bei der Lernende ein strukturierendes Gerüst erhalten, welches im Verlauf des Lernprozesses schrittweise abgebaut wird, damit die Lernenden selbstständig agieren können (Bliss, 1996). Durch Scaffolding wird demnach die Lücke zwischen Aufgaben, die eigenständig zu bewältigen sind und solchen, die nur mit Unterstützung zu bewältigen sind, geschlossen (Rigopouli et al., 2025).

2.2.2. Cognitive Load Theory

Die Cognitive Load Theory (CLT, Theorie der kognitiven Belastung (Seidel & Krapp, 2014)) beschreibt Wissenserwerb als Prozess, der durch optimierte kognitive Informationsverarbeitung unterstützt wird. Sie bietet einen instruktionspsychologischen Rahmen, um Lernumgebungen unter Berücksichtigung der Kapazität des Arbeitsgedächtnisses so zu konzipieren, dass kognitive Ressourcen optimal eingesetzt werden können. Die namensgebende kognitive Belastung beschreibt dabei die Summe aller Anforderungen, die durch die Verarbeitung von Lerninhalten an das Arbeitsgedächtnis gestellt werden (Sweller, 2012).

In der Theorie wird zwischen drei Formen kognitiver Belastung unterschieden (siehe Abbildung 3). Die inhaltsbedingte kognitive Belastung (*intrinsic cognitive load*) bezeichnet den Teil der kognitiven Belastung, der direkt aus dem Anspruch und der Komplexität des zu lernenden Stoffes resultiert (Seidel & Krapp, 2014). Sachfremde kognitive Belastung (*extraneous cognitive load*) wird unter anderem durch ungünstige didaktische Aufbereitung der Inhalte oder störende Rahmenbedingungen verursacht (Seidel & Krapp, 2014). Ergänzend dazu beschreibt die lernrelevante kognitive Belastung (*germane cognitive load*) den Ressourcenanteil, der für den Lernprozess, also das Konstruieren und Integrieren von Wissen ohne die inhaltliche Ebene, aufgewendet wird (Seidel & Krapp, 2014).

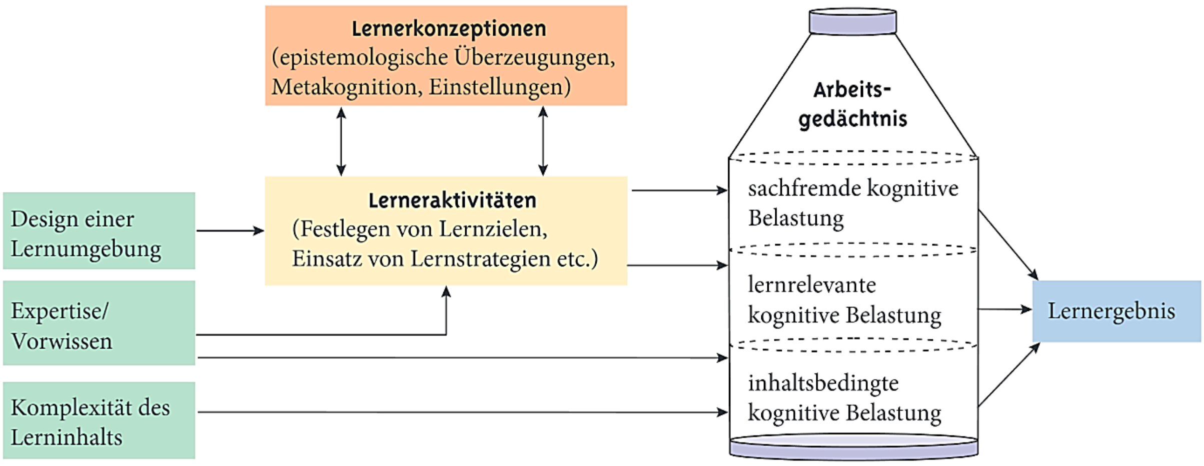

Abbildung 3 Theorie der kognitiven Belastung (Seidel & Krapp, 2014) in Anlehnung an Gerjets & Hesse (2004).

Positive Emotionen erweisen sich als maßgebliche Einflussfaktoren auf die kognitive Leistung, da sie die wahrgenommene intrinsische Komplexität eines Lerngegenstandes reduzieren und damit zusätzliche Ressourcen im Arbeitsgedächtnis freisetzen können. Diese stehen für eine vertiefte Auseinandersetzung mit den Lerninhalten zur Verfügung und bewirken eine Steigerung des Lernerfolgs (Young et al., 2021). Dieser Effekt geht nicht nur mit einer Reduktion der inhaltsbedingten, sondern auch mit einer Reduktion der sachfremden kognitiven Belastung einher. Bei positiven Emotionen wurden zusätzlich auch geringere Werte lernrelevanter kognitiver Belastung beobachtet. Dies kann darauf hindeuten, dass eine durch emotionale

Faktoren begünstigte kognitive Entlastung dazu führt, dass Lernprozesse effizienter ablaufen und damit für denselben Lernzuwachs ein geringerer kognitiver Konstruktionsaufwand erforderlich ist (Young et al., 2021). In Anlehnung an Abbildung 2 besagt die Cognitive Load Theory außerdem, dass eine ungünstige kognitive Belastung durch ein unausgewogenes Verhältnis zwischen Herausforderung und Kompetenzen sowie Angst oder Langeweile einhergeht, während ein lernförderliches Belastungsniveau im Flow-Bereich liegt (Repenning, 2012).
Für die Gestaltung von Lernumgebungen folgt daraus, dass neben strukturellen und inhaltlichen Aspekten auch die affektive Dimension berücksichtigt werden muss. Lernumgebungen, die positive Emotionen fördern, können die kognitive Belastung verringern, indem sie sachfremde und inhaltsbedingte Beanspruchung reduzieren und zugleich mehr Ressourcen für konstruktive Lernprozesse verfügbar machen.

2.2.3. Self-Determination Theory

Die Self-Determination Theory (SDT, Selbstbestimmungstheorie (Deci & Ryan, 1993)) besagt, dass Menschen sich in einem stetigen Entwicklungs- und Lernprozess befinden. Dafür werden spezifische Bedingungen benötigt, da diese Prozesse nicht automatisch geschehen sondern auf äußere Unterstützung angewiesen sind (Ryan & Deci, 2020). Zentral dafür sind grundlegende psychologische Bedürfnisse, deren Befriedigung entscheidend dafür ist, ob Motivation, Lernen und Persönlichkeitsentwicklung gelingen (Ryan & Deci, 2000).

Eines dieser Grundbedürfnisse ist Autonomie, was das Erleben von persönlicher Verantwortlichkeit und selbstbestimmtem Handeln beschreibt (Deci & Ryan, 1993). Eng damit verknüpft ist das Grundbedürfnis der Kompetenz, welches sich darauf bezieht, Aufgaben wirksam bewältigen und die eigenen Fähigkeiten erweitern zu können (Deci & Ryan, 1993). Das Grundbedürfnis der sozialen Eingebundenheit beschreibt das Gefühl von Zugehörigkeit und Verbundenheit, das in Kontexten entsteht, in denen Anerkennung und Wertschätzung erfahrbar sind (Ryan & Deci, 2020).

Der Begriff der Motivation ist zentraler Bestandteil der SDT, wobei Motivation nicht nur nach ihrer Ausprägung, sondern zwischen intrinsisch und extrinsisch unterschieden wird. Intrinsische Motivation bedeutet, dass Personen sich einer Tätigkeit zuwenden, weil sie diese als interessant oder gewinnbringend empfinden und dadurch Neugierde geweckt wird (Ryan & Deci, 2000). Extrinsische Motivation liegt hingegen vor, wenn eine Tätigkeit mit äußeren Bedingungen und Konsequenzen, wie etwa Belohnungen, Bewertungen oder auch negativen Sanktionen verknüpft ist (Ryan & Deci, 2000). Diese beiden Arten der Motivation lassen sich nicht additiv verknüpfen, da stark kontrollierende äußere Zwänge (extrinsische Motivation) die intrinsische Motivation durch Einschränkung der Autonomie und damit auch die Qualität des Lernprozesses verringern können (Deci, 1975; Ryan & Deci, 2017). Lernende erleben sich unter Bedingungen affektiver Ermutigung und Anerkennung eher als fähig, ihre Potenziale auszuschöpfen, während eine dauerhafte Frustration der Grundbedürfnisse mit reduzierter Motivation, oberflächlicherem Lernen und Beeinträchtigungen des psychischen Wohlbefindens einhergeht (La Ossa et al., 2024; Ryan & Deci, 2020). Aus der SDT lässt sich ableiten, Lernumgebungen so zu gestalten, dass Autonomie, Kompetenzerleben und soziale Eingebundenheit systematisch unterstützt werden, um Lernprozesse bestmöglich zu fördern (Ryan & Deci, 2000).

2.3. Didaktische Konzepte

2.3.1. STEAM und 4K

Der Begriff STEAM steht für ***S**cience*, ***T**echnology*, ***E**ngineering*, ***A**rt* sowie ***M**athematics* und erweitert damit das klassische STEM um einen künstlerisch-kreativen Bereich (Yakman, 2008). Diese gestalterischen Zugänge für das Verständnis naturwissenschaftlicher Phänomene gewannen zunehmend an Bedeutung und wird daher im STEAM-Kontext systematisch in Bildungsprozesse integriert (Yakman, 2008). In STEAM-orientierten Lernumgebungen stehen vor allem kooperative und problemlösungsorientierte Arbeitsformen im Vordergrund, wobei fachliche Inhalte mit kreativen Gestaltungsprozessen verknüpft werden (Jackson et al., 2020). Die Umsetzung dieser integrativen Lernansätze erfordert ein breites Spektrum an Kompetenzen, die über rein fachliches Wissen hinausgehen. Daher können 4K-Kompetenzen – Kritisches Denken und Problemlösen, Kommunikation, Kooperation sowie Kreativität und Innovation (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025) – eng mit STEAM verknüpft werden. Diese Kompetenzen werden als zentral für die Bearbeitung komplexer Zukunftsaufgaben und der Anknüpfung von Bildungsprozessen an gesellschaftlichen Herausforderungen erachtet (Pfiffner et al., 2021). Auch der Physikunterricht trägt neben fachlichem Wissenserwerb zur Ausbildung überfachlicher Kompetenzen wie den 4K bei. Er bietet Lerngelegenheiten, in denen Kreativität, kooperative Problemlösung, diskursiver Austausch und kritisches Denken systematisch gefördert werden können (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025).

2.3.2. 5E Modell & Erweiterung zum 6E Modell

Das 5E Modell, das die Phasen von *Engage* bis *Evaluate* (siehe Abbildung 4) umfasst, wurde bereits in den 1980er Jahren im Rahmen der Biological Sciences Curriculum Study entwickelt und seither in unterschiedlichen Bildungskontexten eingesetzt (Bybee, 2009; Bybee et al., 2006). Empirische Befunde zeigen, dass Lehrkräfte nach einer gezielten Fortbildung zu dem 5E Modell eine höhere Sicherheit bei der Vermittlung naturwissenschaftlicher Inhalte wahrnehmen. Dieses gestärkte professionelle Selbstverständnis wirkt sich zugleich positiv auf die

Lernenden aus und unterstützt deren fachliche Auseinandersetzung im Unterricht (Ballone Duran & Duran, 2004).

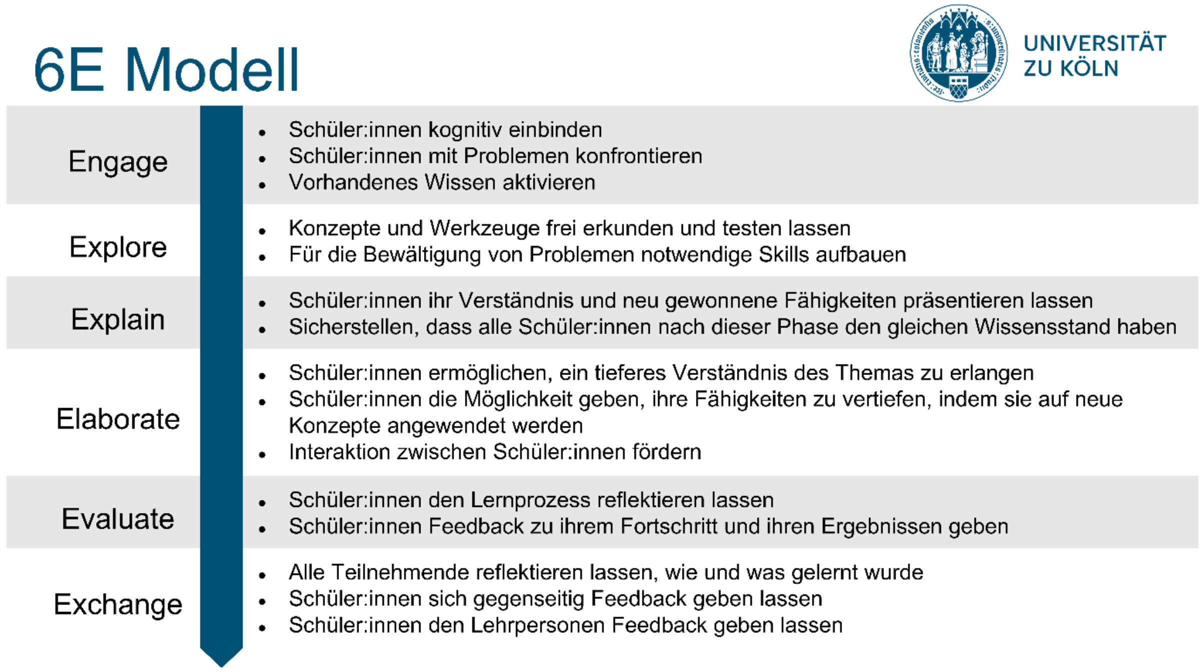

Abbildung 4: Phasen und Beschreibung des 6E-Modells. Eigene deutschsprachige Übersetzung und Überarbeitung auf Grundlage des 5E Modells von Bybee et al. (2006) sowie der Weiterentwicklung durch Henze et al. (2022). Die zugrunde liegende englischsprachige Fassung wurde in Henze et al. (2025) veröffentlicht.

Im Rahmen der in Henze et al. (2022) beschriebenen schulischen Studie, entstand in Kooperation mit der *Heliosschule – Inklusive Universitätsschule der Stadt Köln* die Idee zur konzeptionellen Weiterentwicklung des 5E Modells zum 6E Modell. Dies erweitert das 5E Modell mit der *Exchange*-Phase um eine zusätzliche Reflexions- und Auswertungsphase. Ziel dieser Erweiterung ist es, Lernprozesse nicht ausschließlich anhand fachlicher Ergebnisse zu beurteilen, sondern auch die Perspektiven und Erfahrungen der Lernenden sowie der Lehrenden systematisch einzubeziehen. Diese Phase des Modells ist stärker auf metakognitive Prozesse ausgerichtet und ermöglicht es, Lernwege, Entscheidungsprozesse und subjektive Einschätzungen bewusst zu reflektieren (Henze et al., 2022).

Das erweiterte Modell fördert außerdem den kontinuierlichen Austausch zwischen Lernenden sowie zwischen Lernenden und Lehrenden. Dieser dialogische Charakter unterstützt nicht nur kooperative Lernprozesse, sondern eröffnet auch vertiefte Einblicke in die Lernstrategien, Denkweisen und Entwicklungsverläufe der Lernenden. Das 6E Modell eignet sich damit sowohl als didaktische Struktur für Unterrichtsentwicklung, als auch als analytisches Instrument zur Untersuchung von Lehr-Lernprozessen (Henze et al., 2022).

2.4. Design-Based Research

Design-Based Research (DBR) ist ein forschungsmethodischer Ansatz, der auf die Entwicklung wirksamer und kontextspezifischer Innovationen zielt. Charakteristisch ist ein iterativ-zyklisches Vorgehen, bei dem nach Erprobungsphasen systematisch Rückmeldungen in die nächste Entwicklungsstufe integriert werden (van Zyl & Karsten, 2022).
Besonders im Kontext technologiebasierter Lehr-Lernsituationen erweist sich DBR als geeignet, da Gestaltung und Forschung eng miteinander verflochten sind und so praxisnahe Lösungen für komplexe Bildungsprobleme entstehen können (Anderson & Shattuck, 2012; van Zyl & Karsten, 2022). Auf diese Weise trägt DBR dazu bei, konkrete Bildungsinterventionen weiterzuentwickeln und zugleich theoretisch anschlussfähige Einsichten zur Verbesserung von Lernumgebungen zu gewinnen (Minichiello & Caldwell, 2021).

2.5. Generative KI in der Bildung

Kompetenzen im Umgang mit KI sollen nach der Kultusministerkonferenz (2024) verbindlich in allen drei Phasen der Lehrkräftebildung verankert werden. Neben spezifischen KI-bezogenen Kompetenzen sind dafür allgemeine digitale Kompetenzen, Medienkompetenz und die Fähigkeit zu kritisch-reflexivem Denken erforderlich. Mit diesen Voraussetzungen wird generative KI verstärkt lernförderlich und nicht nur als Informationsquelle genutzt (Chiu, 2024). Souveränität im Umgang mit generativer KI umfasst damit nicht nur die Verwendung einer KI, sondern vor allem die Fähigkeit, die generierten Inhalte reflektiert einzuordnen und produktiv zu verwenden. Gleichzeitig zeigt sich in der schulischen Praxis, dass dieser Anspruch mit einem hohen Aufwand der Professionalisierung einhergeht. Nach Jude et al. (2025) standen im Jahr 2024 Angebote zum Einsatz digitaler Medien im Unterricht besonders häufig bei Fortbildungen im Mittelpunkt. Gleichzeitig berichteten Lehrkräfte von einem deutlichen Unterstützungsbedarf bei der pädagogisch sinnvollen Nutzung digitaler Medien (Jude et al., 2025).

Diese Unsicherheit spiegelt sich auch in der Einschätzung möglicher Auswirkungen wider. Lehrkräfte, die regelmäßig KI nutzen, bewerten deren Folgen in schulischen Kontexten etwas weniger kritisch als Nicht-Nutzende (Jude et al., 2025). Dennoch überwiegt insgesamt die Sorge vor negativen Effekten, woraus ein Fortbildungsbedarf sichtbar wird, der sowohl die fachdidaktische und pädagogische Gestaltung von Unterricht mit und über digitale Medien wie KI, als auch die gezielte Förderung kritischen Denkens umfasst (Blossfeld et al., 2018; Jude et al., 2025; Schröter et al., 2021). Ergänzend zeigt sich, dass der Aufbau von Wissen über KI sowie entsprechender Kompetenzen in der Lehrkräftebildung im Hochschulbereich bislang unzureichend verankert ist (Huwer et al., 2024). Die beschriebenen Unsicherheiten lassen sich zugleich als Ausdruck eines wiederkehrenden Musters im Umgang mit neuen Technologien verstehen. Häufig zeigt sich hier der sogenannte Hype Cycle, bei dem auf eine Phase ausgeprägter Euphorie zunächst eine kritische Fokussierung auf Probleme folgt, bevor sich schrittweise eine ausgewogenere Bewertung zwischen Nutzen und Risiken herausbildet (Döbeli Honegger, 2017).

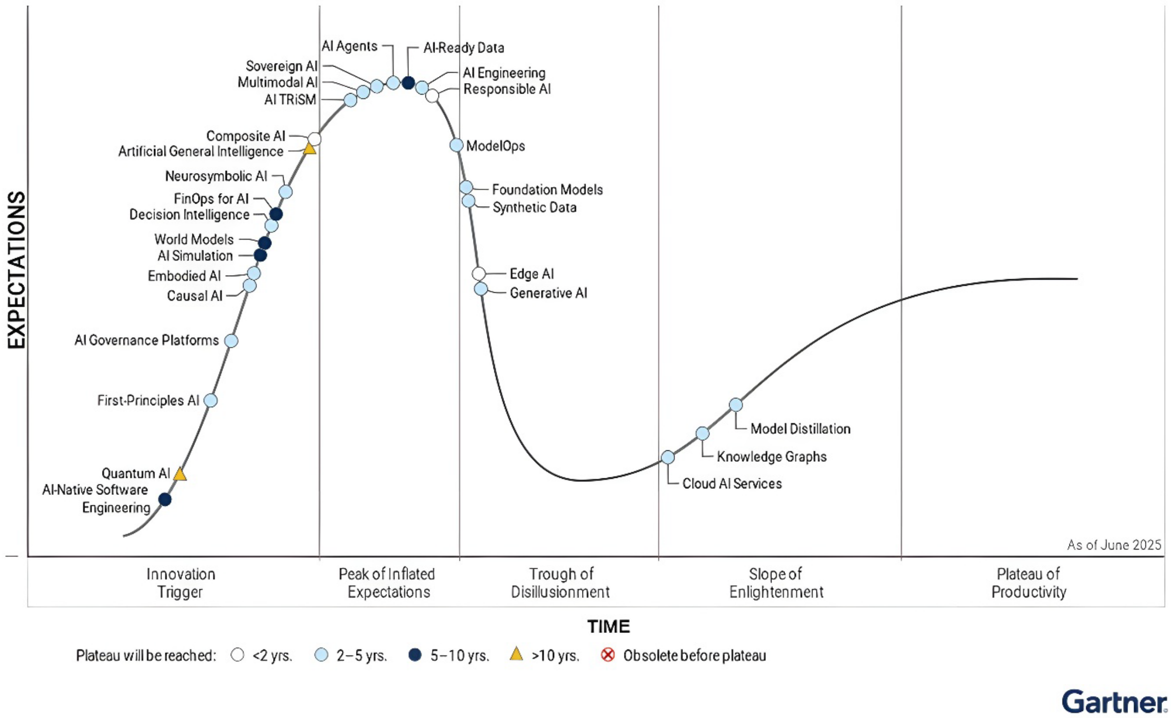

Abbildung 5: AI Hype Cycle (Gartner, Inc., 2025)

Die aktuelle Einordnung von Gartner verortet generative KI nach dem Überschreiten des Gipfels übersteigerter Erwartungen (*Peak of Inflated Expectations*) nun im Tal der Enttäuschungen (*Trough of Disillusionment*) (siehe Abbildung 5). In dieser Phase zeigt sich, dass viele Anwendungen den hohen Erwartungen nicht gerecht werden. Gleichzeitig führt die zunehmende Erfahrung mit der Technologie zu einem differenzierteren Umgang. Ein vertieftes Verständnis für ihre tatsächlichen Stärken und Grenzen bildet die Grundlage zuverlässiger Nutzungskonzepte (Dedehayir & Steinert, 2016). Das hellblaue Kreissymbol im AI Hype Cycle in Abbildung 5 weist darauf hin, dass das Plateau der Produktivität für generative KI zwischen 2027 und 2030 erreicht werden könnte (Gartner, Inc., 2025). Generative KI befindet sich damit in einem Übergang von einem überwiegend experimentellen Einsatz hin zu einem Einsatz, der im Sinne des Plateaus der produktiven Nutzung (*Plateau of Productivity*) vor allem durch eine zunehmende wahrgenommene Sinnhaftigkeit und Nutzbarkeit in Lehr-Lernkontexten geprägt ist. Die Integration digitaler Ressourcen in schulische Kontexte erfordert eine kontinuierliche wissenschaftliche Begleitung sowie eine theoretische Fundierung, um ihren Einsatz pädagogisch und didaktisch verantwortungsvoll zu gestalten (Schröter et al., 2021).

2.5.1. Bildungspotenziale generativer KI

In der schulischen Praxis zeigt sich ein ambivalentes Bild im Umgang mit Künstlicher Intelligenz. Lehrkräfte betonen das Potenzial von KI für eine stärker individualisierte Lernunterstützung. Gleichzeitig überwiegt jedoch Skepsis gegenüber ihren Wirkungen (Jude et al., 2025). Die Erhebung des Deutschen Schulbarometers 2025 zeigt diese Spannung deutlich auf. Etwa ein Drittel der Lehrkräfte hatte im vorangegangenen Jahr keine KI eingesetzt, während ein ähnlich großer Anteil entsprechende Werkzeuge gelegentlich und ein ebenso großer Anteil regelmäßig, teilweise täglich, nutzte (Jude et al., 2025). Die Nutzung von KI durch Lehrpersonen konzentriert sich bislang vor allem auf vorbereitende Tätigkeiten wie Aufgabenentwicklung und Unterrichtsplanung, während prüfungsbezogene Anwendungen oder diagnostische Einsätze seltener realisiert werden (Jude et al., 2025). Dabei liegt ein zentrales Potenzial von KI in der umfassenden Unterstützung von Lehrkräften über diese vorbereitenden Tätigkeiten hinaus. Sie kann bei der Organisation und Durchführung von Unterricht entlasten, Freiräume für pädagogische Kernaufgaben schaffen und insbesondere im Umgang mit heterogenen Lerngruppen durch differenzierte Zugänge und individuelle Lernunterstützung zur Förderung kooperativer Prozesse beitragen (Sekretariat der Kultusministerkonferenz, 2024).

Trotz Zurückhaltung in der Praxis verweisen verschiedene Studien auf substanzielle didaktische Potenziale. Das disruptive Potenzial von KI im Bildungskontext zeigt sich in der Möglichkeit, Lernaufgaben neu zu gestalten und damit Lernumgebungen zu entwickeln, die konstruktive und interaktive Wissensprozesse unterstützen (Bauer et al., 2025; Huwer et al., 2024). Der didaktische Mehrwert generativer KI hängt dabei maßgeblich davon ab, inwieweit ihre spezifischen Nutzungsmöglichkeiten mit einem durchdachten pädagogischen Konzept in Einklang gebracht werden (Guo et al., 2025).

Auch im Bereich der adaptiven Lernunterstützung können positive Effekte aufgezeigt werden. Tutorielle KI Unterstützung konnte Lernenden helfen in kürzerer Zeit deutlich höhere Lernzuwächse zu erzielen und sich im Vergleich zu Präsenzformaten stärker eingebunden und motiviert zu fühlen (Kestin et al., 2025). Parallel dazu entwickelt sich die naturwissenschaftsdidaktische Forschung in Richtung multipler Einsatzmöglichkeiten, multimodaler Ansätze und generativer Werkzeuge für Forschung und Unterricht, sodass mit einer zunehmenden Verbreitung großer multimodaler Modelle in Praxis und Forschung zu rechnen ist (Lee et al., 2025). Auf multimodalen großen Sprachmodellen (MLLMs) basierende KI-Chatbots wie ChatGPT, Gemini und auch der in Studie III dieser Dissertation genutzte ExperiMentor (Henze et al., 2026) können sowohl darstellungs- und medienübergreifende Lernmaterialien und Präsentationsformen erschließen als auch personalisierte Lernerfahrungen und Inhalte barriereärmer sowie motivierend aufbereiten (Bewersdorff et al., 2025).

Gleichzeitig bleibt die Rolle menschlicher Verantwortung zentral. KI kann Lehr- und Lernprozesse wirkungsvoll unterstützen, doch Bewertungsergebnisse und daraus resultierende Entscheidungen müssen dem aktuellen Entwicklungsstand entsprechend weiterhin bei den verantwortlichen Personen bleiben. Eine ungeprüfte Übernahme KI-generierter Inhalte kann zu Fehlentscheidungen und Ungerechtigkeiten führen und langfristig einen Abbau von Fachkompetenz begünstigen (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024).

2.5.2. Grenzen und Herausforderungen generativer KI im Bildungskontext

Empirische Befunde zeigen, dass angehende Lehrkräfte KI in der naturwissenschaftlichen Bildung zugleich sowohl mit hohen Erwartungen als auch mit Vorbehalten begegnen. Einerseits wird ihr Potenzial zur Unterstützung des Problemlösens, des Verständnisses komplexer Inhalte und zur Förderung von Teilhabe betont. Andererseits bestehen Bedenken hinsichtlich des Datenschutzes, der Möglichkeit des Plagiats, der Veränderung von Lehrrollen und der möglichen Verstärkung sozialer Ungleichheiten (Ishmuradova et al., 2025). Auch im Deutschen Schulbarometer 2025 zeigt sich, dass viele Lehrkräfte den Einsatz von KI mit Blick auf die Kompetenzentwicklung der Schüler:innen, insbesondere hinsichtlich sozialer und kommunikativer Fähigkeiten sowie kritischen Denkens, kritisch beurteilen. Zusätzlich wurde von einem Großteil der Befragten die eigene Handlungssicherheit im Hinblick auf den Einsatz von KI Anwendungen als gering eingeschätzt (Jude et al., 2025).

Aktuelle generative KI bietet bislang nur eingeschränkte Möglichkeiten, um Lernprozesse dauerhaft und umfassend zu unterstützen. Daraus ergibt sich ein deutlicher Entwicklungsbedarf hinsichtlich didaktischer und gestalterischer Konzepte, um KI-Systeme künftig gezielt auf unterschiedliche Lernvoraussetzungen und Bedürfnisse abzustimmen (Guo et al., 2025). Gleichzeitig zeigen Studien zu traditionellen Lehr-Lern Konzepten, dass der Umgang mit heterogenen Wissensständen bislang nur unzureichend gelingt (Kestin et al., 2025). Aus diesem Grund der Einsatz von KI insbesondere dann kritisch zu bewerten, wenn sie primär dazu genutzt wird, eigenständige gedankliche Auseinandersetzung zu umgehen und damit zentrale Denkprozesse zu unterlaufen(Kestin et al., 2025).

Neben den zuvor angeführten Problematiken bestehen zudem potenzielle Qualitätsrisiken. Analysen von KI-generierten Nachrichteninhalten zeigen, dass ein erheblicher Teil der Ausgaben fehlerbehaftet ist und die Mehrheit der Texte zumindest leichte Qualitätsdefizite enthält (European Broadcasting Union, 2025). Solche Fehler oder inhaltliche Schwächen lassen sich häufig nur schwer identifizieren oder überprüfen, da Antworten mit großer sprachlicher Sicherheit formuliert werden und dadurch den Eindruck von Verlässlichkeit erwecken (Bender et al., 2021; European Broadcasting Union, 2025; Kortemeyer, 2023). Dies kann bei ungeprüfter Übernahme zu möglichen Fehlvorstellungen führen, die sich verfestigen und den weiteren Lernprozess nachhaltig beeinträchtigen können (Ding et al., 2023). Aus technischer Perspektive beruht dies darauf, dass bei maschinellem Lernen, auf dem bisherige KI-Systeme basieren, Parameter einer Modellfunktion an statistische Zusammenhänge in den Trainingsdaten angepasst werden. Dabei entwickelt eine KI jedoch kein inhaltliches Verständnis für die zugrunde liegenden Problemstellungen, sondern generiert Ausgaben auf Grundlage von Mustererkennung (Tiemann et al., 2024). Die Ausgaben basieren somit auf Basis statistischer Wahrscheinlichkeiten für Wortfolgen im gegebenen Kontext (Bender et al., 2021).

Im Rahmen der Leistungsbewertung im Fach Physik sind diese Grenzen ebenfalls von Relevanz. Im Kernlehrplan Physik für Gymnasien und Gesamtschulen in Nordrhein-Westfalen wird eine Dimension hilfsmittel- und werkzeuggestützter Leistungen aufgeführt. So sind auch unter Nutzung geeigneter digitaler Werkzeuge erbrachte Leistungen bewertbar, sofern ihre fachliche Qualität gesichert bleibt (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025). Daraus ergibt sich ein Rahmen, in dem KI-gestützte Werkzeuge grundsätzlich als legitime Unterstützung in Leistungsprozessen berücksichtigt werden können und damit ein Teil zeitgemäßer Erkenntnisprozesse sind. Bewertungsentscheidungen müssen sich jedoch weiterhin an fachlichen Kriterien orientieren und der Anteil der Lernenden am Ergebnis muss transparent sein. Damit wird die kritisch-reflektierte Prüfung KI-generierter Inhalte zu einem zentralen Bestandteil zeitgemäßer Leistungsbewertung (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2025). Voraussetzung hierfür ist eine hinreichend ausgeprägte AI Literacy der Lehrkräfte, die es ihnen erlaubt, KI-generierte Inhalte hinsichtlich ihrer fachlichen Korrektheit, des Entstehungsprozesses sowie möglicher Verzerrungen einzuordnen und ihren didaktischen Einsatz sowie ihre Nutzung in Lernprozessen reflektiert abzuwägen.

Zuletzt sind auch Probleme im Kontext von Bildungsgerechtigkeit zu betrachten. KI-Anwendungen schaffen nicht automatisch Barrierefreiheit für alle Lernenden. Unterschiedliche Voraussetzungen, wie etwa begrenzte Sprachkompetenzen, Schwierigkeiten im Bereich des Lesens und Schreibens oder visuelle Beeinträchtigungen können den Zugang erheblich einschränken. Soziale Ungleichheiten verstärken diese Effekte, beispielsweise wenn kostenpflichtige Angebote oder unzureichende Unterstützung den Zugang begrenzen. Ohne gezielte kompensatorische Maßnahmen besteht somit die Gefahr, dass der Einsatz von KI bestehende Bildungsungleichheiten eher reproduziert als zu deren Abbau beiträgt (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024).

2.5.3. Emotionale & Motivationale Wirkungen generativer KI

Studien weisen darauf hin, dass sich der Einsatz von Chatbots wie ChatGPT in didaktisch angeleiteten Lernsituationen unterstützend auf kritische, kreative und reflexive Denkprozesse von Lernenden auswirken kann (Essel et al., 2024; Nikolopoulou, 2024). Zudem zeigen sich positive Effekte die Lernleistung und die subjektive Wahrnehmung des Lernprozesses betreffend, wobei diese Effekte nicht durchgängig auftreten, sondern entscheidend von der jeweiligen Nutzungssituation und der didaktischen Gestaltung abhängen (Wang & Fan, 2025).

Studien zu affektiven Reaktionen verdeutlichen, dass Zufriedenheit und Engagement bei der Arbeit mit ChatGPT von insbesondere drei Aspekten beeinflusst werden: der wahrgenommenen Nützlichkeit des Systems, den an seine Leistung gerichteten Erwartungen sowie der wahrgenommenen Verlässlichkeit der bereitgestellten Informationen. Weitere Einflussfaktoren sind die empfundene Freude bei der Nutzung, die wahrgenommene Autonomie, die Einfachheit der Handhabung die Übereinstimmung zwischen der eigenen Erwartung und dem, was tatsächlich erreichbar ist (Lo, 2023).

Über spezifische Anwendungssituationen hinaus zeigen Forschungsergebnisse, dass der Einsatz von KI in Bildungskontexten die Neugier, das kreative Problemlösen sowie das selbstregulierte Lernen stärken kann (Essel et al., 2024). Hierbei heben Lernende vor allem die Möglichkeit der individuellen Anpassung und des dialogischen Austauschs hervor (Gada & Chudasana, 2024). Eine gezielte und didaktisch reflektierte Einbindung von KI in Lernkontexte kann demnach dazu beitragen, Lernprozesse positiv zu unterstützen und günstigere Einstellungen der Lernenden gegenüber dem Lernen zu fördern (La Ossa et al., 2024). Darüber hinaus geben Lernende an, sich sowohl kognitiv als auch emotional stärker in das Unterrichtsgeschehen eingebunden zu fühlen (Guo et al., 2025).
Diese Befunde lassen sich eng mit dem Kompetenzniveau der Lernenden im Umgang mit KI verknüpfen. Lernende mit höher ausgeprägter AI Literacy berichten häufiger von einem gesteigerten Vertrauen in die eigene Handlungsfähigkeit und größerem Interesse an KI-gestützten Arbeitsformen. Sie haben insgesamt positiveren Einstellungen, während ablehnende Haltungen deutlich seltener auftreten (Hornberger et al., 2023). Damit wird deutlich, dass die emotionalen und motivationalen Effekte generativer KI maßgeblich durch die individuellen Voraussetzungen der Nutzenden sowie durch die didaktische Gestaltung der Lernumgebung geprägt werden.

3. Forschungsfragen

Basierend auf dem identifizierten Bedarf an qualitativen Fortbildungsmaßnahmen richtet sich die zentrale Forschungsfrage dieser Arbeit darauf, wie Fortbildungskonzepte gestaltet sein müssen, um gezielt die digitale Souveränität von Lehrkräften im Umgang mit KI-Technologien zu stärken und für den Fachunterricht nutzbar zu machen. Diese übergeordnete Fragestellung wird in drei Studien aus unterschiedlichen Perspektiven beleuchtet, die jeweils untersuchen, unter welchen Bedingungen sich digitale Souveränität bei Lehrkräften entwickeln lässt und wie diese in die fachdidaktische Praxis integriert werden kann.

In Studie I werden motivationale und einstellungsbezogene Voraussetzungen von Lehrkräften in das Zentrum gerückt. Bevor digitale Souveränität entstehen kann, bedarf es einer grundsätzlichen Offenheit gegenüber Technologie sowie positiver Erfahrungen im Umgang mit neuen Werkzeugen. Im Rahmen der folgenden Forschungsfragen wird analysiert, inwiefern sich im Kontext universitärer und beruflicher Lehrer:innenbildung Interesse an Robotik, Programmierung, KI, STEAM und nachhaltiger Entwicklung systematisch entwickeln lässt:

> *Wie kann STEAM-Bildung auf Basis des 5E-Modells in Schulen eingeführt werden?*
>
> *Wie stehen Lehrkräfte zur Einführung von modernen Kreativitätswerkzeugen im Rahmen des STEM-Unterrichts und wie verändert sich ihre Einstellung im Laufe eines siebenwöchigen Fortbildungsprogramms?*

Aufbauend auf den motivationalen und akzeptanzbezogenen Aspekten aus Studie I, wird in Studie II der Fokus auf den gezielten Aufbau fachspezifischer KI-Kompetenzen im Rahmen der universitären Lehrer:innenbildung gelegt. Die Studie folgt dem Ansatz des Design-Based-Research und untersucht, wie das iterative Design einer KI-Intervention über zwei Sitzungen die Kompetenzentwicklung angehender Physiklehrkräfte unterstützt. Neben der Erfassung fachlicher und technologischer Kompetenzen werden auch professionsbezogene Überzeugungen und Einstellungen in den Blick genommen, um mögliche Wechselwirkungen zwischen Kompetenzaufbau und Haltung zu analysieren:

> *Wie unterstützt das iterative Design einer KI-Intervention über zwei Sitzungen die Entwicklung von KI-Kompetenzen bei angehenden Physiklehrkräften?*
>
> *Wie stehen angehende Lehrkräfte zu KI im Bildungsbereich und wie beeinflusst die Intervention ihre Einstellung?*

Mit Studie III wird die Analyse auf die konkrete Anwendungsebene im physikalischen Kontext ausgeweitet. Während die vorherigen Studien Voraussetzungen und Kompetenzentwicklungen untersuchen, steht hier die Frage im Zentrum, inwiefern KI als funktionales Werkzeug zu veränderten Lernergebnissen und Nutzungserfahrungen führt. In einem kontrollierten

Vergleich zwischen einer Experimentalgruppe mit KI-Unterstützung und einer Kontrollgruppe mit traditionellen Werkzeugen wird die Wirksamkeit der eingesetzten Technologien empirisch überprüft. Neben kognitiven Lernergebnissen werden auch emotionale und motivationale Dimensionen erfasst, um auch affektive Aspekte digitaler Souveränität zu erfassen:

> *Wie verändern sich die Lernergebnisse der Studierenden innerhalb der Experimental- und Kontrollgruppe und wie unterscheiden sich die Gruppen voneinander?*
>
> *Wie erleben die Studierenden das jeweilige Werkzeug emotional und motivational, in Bezug auf Engagement, Frustration und wahrgenommene Effektivität?*

In ihrer Gesamtheit ermöglichen die drei Studien eine multiperspektivische Annäherung an die Frage, wie digitale Souveränität von Lehrkräften im Umgang mit KI aufgebaut werden kann. Sie verbinden motivationale Voraussetzungen, gestaltungsorientierte Kompetenzentwicklung und konkrete Anwendungserfahrungen im Fachkontext.

4. Methodologie

In diesem Kapitel werden die methodischen Zugänge der Dissertation beschrieben und dargelegt, wie die drei empirischen Studien geplant, durchgeführt und ausgewertet wurden. Über alle drei Studien hinweg folgt die Dissertation einem multimethodischen Forschungsansatz, der unterschiedliche Erhebungs- und Auswertungsmethoden kombiniert. Während Studie I explorativ ausgerichtet ist und vor allem deskriptive Einsichten in Einstellungen und Wahrnehmungen liefert, erweitern Studie II und Studie III die Perspektive um inferenzstatistische Analysen unter kontrollierten Bedingungen.

4.1. Studie I

Studie I wurde als Mixed-Methods-Design angelegt und erfasst Einstellungen, Interesse und subjektive Einschätzungen von Lehrkräften im Kontext einer siebenwöchigen schulbegleitenden Fortbildung zu digitalen Technologien. Ergänzend zu der quantitativen Erhebung wurden qualitative Daten genutzt, um die Perspektiven der 14 beteiligten Lehrkräfte zu dem Verlauf und den Rahmenbedingungen der Fortbildung zu dokumentieren. Untersucht wurde die Einführung von STEAM-Bildung entlang des 6E-Modells sowie dessen Einfluss auf Interesse und Einstellungen von Lehrkräften zu modernen digitalen Technologien im Kontext einer siebenwöchigen Fortbildung. Die Studie wurde an der *Heliosschule – Inklusive Universitätsschule der Stadt Köln* durchgeführt. Die Schule ist als kooperative Einrichtung der Universität zu Köln konzipiert und arbeitet projektorientiert mit einer engen Verbindung von schulischer Praxis und universitärer Lehrer:innenbildung (Henze et al., 2022). Die Fortbildung wurde in den regulären Unterricht der siebten Jahrgangsstufe eingebettet und unterrichtsbegleitend durchgeführt.
Beteiligt waren Lehrkräfte unterschiedlicher Fächer, Studierende in der Lehramtsausbildung sowie Schüler:innen. Die Studierenden unterstützten die Durchführung der praktischen Einheiten und begleiteten die Arbeitsphasen. Inhaltlich wurden digitale Werkzeuge aus dem STEAM-Kontext, unter anderem in Verbindung mit gesellschaftlichen Fragestellungen im Rahmen der Ziele für nachhaltige Entwicklung, eingesetzt (United Nations Department of Economic and Social Affairs, 2021).

Die Fortbildung wurde entlang des 5E-Modells (Bybee, 2009; Bybee et al., 2006) strukturiert und um eine zusätzliche sechste Phase erweitert (Henze et al., 2022). In der Engage-Phase erfolgte die thematische Einführung durch multimediale Impulse. Darauf aufbauend wurden in der Explore-Phase die eingesetzten digitalen Werkzeuge zunächst offen erkundet. Anschließend wurden in der Explain-Phase die gemachten Erfahrungen und Funktionen der Werkzeuge beschrieben und zusammengeführt. Daran anknüpfend war die Elaborate-Phase projektorientiert angelegt und erstreckte sich über mehrere Wochen mit dem Ziel, die erprobten Werkzeuge zur eigenständigen Entwicklung eines Projektprodukts einzusetzen. Abschließend umfasste die Evaluate-Phase die Vorbereitung und Durchführung einer öffentlichen Präsentation der Ergebnisse. Ergänzend wurde eine Exchange-Phase umgesetzt, die dem Austausch über Erfahrungen und den Verlauf des Projektes diente.

Zur Datenerhebung wurden Pre-Post-Fragebögen eingesetzt, die sich an Instrumenten der Interessensforschung orientierten und thematische, kontextuelle sowie aktivitätsbezogene Dimensionen abbildeten. Die quantitative Auswertung erfolgte deskriptiv. Hierzu wurden Häufigkeiten, Mittelwerte und Veränderungen zwischen Pre- und Post-Test dargestellt. Zusätzlich wurden leitfadengestützte Interviews mit beteiligten Lehrkräften und Expert:innen zu mehreren Zeitpunkten durchgeführt, transkribiert und anonymisiert. Die Auswertung der Interviews erfolgte mittels der qualitativen Inhaltsanalyse nach Kuckartz (2018). Die Zusammenführung der quantitativen und qualitativen Ergebnisse erfolgte im Sinne eines multiperspektivischen Mixed-Methods-Ansatzes durch eine gemeinsame, beschreibende Betrachtung beider Datenquellen.

4.2. Studie II

Studie II folgte dem Ansatz des Design-Based Research und diente der iterativen Entwicklung und Erprobung einer KI-gestützten Lehr-Lern-Einheit in der Lehrkräftebildung im Fach Physik. Das Ziel war die iterative Weiterentwicklung und Dokumentation der Intervention im Sinne des Design-Based-Research. Im Fokus stand dabei die Frage, wie unterschiedliche Gestaltungsvarianten die KI-Kompetenz angehender Physiklehrkräfte fördern und ihre Einstellungen zu KI im Bildungsbereich beeinflussen.

Die Intervention wurde in zwei aufeinanderfolgenden Zyklen umgesetzt und systematisch variiert, um Gestaltungsentscheidungen unter unterschiedlichen Rahmenbedingungen zu untersuchen. An der Studie nahmen insgesamt 31 Studierende der Physik im Lehramtsstudium der Universität zu Köln teil. Die Lehr-Lern-Einheit war in ein Modul der Lehramtsausbildung eingebettet und umfasste zwei Sitzungen mit jeweils 90 Minuten Dauer, die im Abstand von einer Woche stattfanden. Die beiden Zyklen wurden in zwei aufeinanderfolgenden Semestern durchgeführt. Während der erste Zyklus als Präsenzformat mit gemeinsamer Reflexionsphase umgesetzt wurde, erfolgte der zweite Zyklus als stärker selbstgesteuerten Variante, wodurch sich Unterschiede im Ablauf der Intervention ergaben (Henze et al., 2025).

Den theoretischen Bezugsrahmen der Studie bildete Vygotskys Konzept der Zone der proximalen Entwicklung. Dementsprechend wurden Unterstützungsangebote, wie die Nutzung generativer KI-Tools als Werkzeuge, Austausch in Peer-Konstellationen sowie strukturierende Impulse durch die Seminarleitung in die Aufgabenbearbeitung integriert. Die didaktische Struktur der Intervention orientierte sich am 5E-Modell und wurde ebenfalls um eine zusätzliche Exchange-Phase ergänzt (Henze et al., 2022). Diese Struktur bildete den organisatorischen

Rahmen beider Zyklen, wobei einzelne Phasen in Abhängigkeit vom jeweiligen Durchführungsformat unterschiedlich ausgestaltet waren.

Zur Datenerhebung wurden quantitative Instrumente eingesetzt. Die Erfassung der KI-Kompetenz erfolgte mittels eines standardisierten Tests mit insgesamt 30 Items nach Hornberger et al. (2023), der verschiedene Kompetenzbereiche abbildet. Ergänzend wurden Einstellungen gegenüber KI im Bildungsbereich mithilfe selbst entwickelter Fragebögen erhoben. Abhängig vom jeweiligen DBR-Zyklus wurden die Einstellungen entweder querschnittlich (Iteration 1 des DBR-Zyklus) oder im Pre-Post-Design (Iteration 2 des DBR-Zyklus) erfasst. Die KI-Kompetenz wurde in beiden Zyklen als Pre-Post-Messung erhoben.

Die statistische Auswertung erfolgte in mehreren Schritten. Zunächst wurden für alle Variablen deskriptive Kennwerte der gemessenen Items berechnet, getrennt nach Erhebungszeitpunkt und Zyklus. Für die inferenzstatistische Analyse der AI Literacy wurden Differenzwerte zwischen Pre- und Post-Test gebildet. Die Normalverteilung dieser Differenzwerte wurde mittels Shapiro-Wilk-Test überprüft. Bei erfüllter Normalverteilungsannahme wurden gepaarte t-Tests zur Prüfung von Mittelwertsunterschieden zwischen Pre- und Post-Test eingesetzt, anderenfalls Wilcoxon-Vorzeichen-Rang-Tests. Diese Vorgehensweise wurde sowohl auf Gesamtscores als auch auf Ebene einzelner Kompetenzbereiche angewendet. Für die Einstellungsskalen wurden analoge Verfahren genutzt, wobei bei querschnittlich erhobenen Daten Vergleiche mit einem definierten Skalenmittelpunkt vorgenommen wurden. Alle statistischen Analysen wurden mit einem einheitlichen Signifikanzniveau durchgeführt. Die inferenzstatistische Auswertung diente der formalen Prüfung von Veränderungen zwischen den Erhebungszeitpunkten innerhalb der jeweiligen Zyklen sowie der dokumentierenden Gegenüberstellung beider Iterationen.

4.3. Studie III

Studie III wurde als randomisiertes Pre-Post-Kontrollgruppendesign angelegt und untersuchte den Einfluss unterschiedlicher Werkzeuge zur Datenauswertung auf fachliche Lernzuwächse sowie auf emotionale und motivationale Variablen. An der Studie nahmen insgesamt $n = 50$ Lehramtsstudierende der Physik teil, die zufällig einer Versuchsgruppe oder einer Kontrollgruppe zugewiesen wurden.

Alle Teilnehmenden werteten identische Datensätze aus einem Fadenpendel- und einem Federpendelversuch aus. Ziel der Auswertung war die Bestimmung physikalischer Kenngrößen, insbesondere der Fallbeschleunigung g sowie der Federkonstante k. Die Kontrollgruppe nutzte zur Datenauswertung das Tabellenkalkulationsprogramm Microsoft Excel, während die Versuchsgruppe mit dem für diese Studie entwickelten Chatbot ExperiMentor arbeitete. ExperiMentor übernahm die Datenverarbeitung mithilfe im Hintergrund agierender Python-basierter

Skripte und unterstützte die Auswertung dialogisch. Beide Gruppen bearbeiteten identische Aufgabenstellungen unter vergleichbaren instruktionalen Rahmenbedingungen, um beobachtete Unterschiede möglichst auf die eingesetzten Werkzeuge zurückführen zu können.

Fachliche Lernzuwächse wurden mithilfe von Pre- und Post-Tests zu physikalischen Konzepten und Auswertungsmethoden erfasst. Ergänzend wurden emotionale und motivationale Konstrukte wie Motivation, wahrgenommene Methodenwirksamkeit, positive und negative emotionale Lernerfahrungen sowie Frustration und kognitive Belastung mittels Likert-Skalen erhoben.

Die statistische Auswertung erfolgte in mehreren aufeinander aufbauenden Schritten. Zu Beginn wurden für alle Variablen, getrennt nach Gruppe und Messzeitpunkt, deskriptive Kennwerte berechnet. Anschließend wurden für die leistungsbezogenen Variablen individuelle Differenzwerte zwischen Pre- und Post-Test gebildet. Die Verteilung dieser Differenzwerte wurde mithilfe von Normalverteilungstests überprüft und diente als Entscheidungsgrundlage für die Auswahl geeigneter inferenzstatistischer Verfahren.

Da die Normalverteilungsannahme für einen Großteil der Variablen nicht erfüllt war, kamen überwiegend nicht-parametrische Testverfahren zum Einsatz. Lernzuwächse innerhalb der jeweiligen Gruppen wurden mithilfe von Wilcoxon-Vorzeichen-Rang-Tests analysiert. Zur Untersuchung von Gruppenunterschieden unter Berücksichtigung unterschiedlicher Ausgangsniveaus wurde zusätzlich eine Kovarianzanalyse (ANCOVA) durchgeführt, wobei die jeweiligen Pretestwerte als Kovariate berücksichtigt wurden. Diese gilt als verlässlich bei Verletzung der Normalverteilung bei gleich großen Gruppengrößen (Bortz & Schuster, 2010).

Ergänzend wurden normalisierte Lernzuwächse nach Hake berechnet, um individuelle Lernfortschritte relativ zu der maximal möglichen Verbesserung darzustellen. Diese Kennwerte wurden deskriptiv ausgewertet und für gruppenbezogene Vergleiche herangezogen. Die emotional-motivationalen Variablen wurden ebenfalls mithilfe des Wilcoxon-Vorzeichen-Rang-Tests statistisch analysiert.

5. Synopse der eingebundenen Publikationen

Im Rahmen dieser Dissertation werden professionelle Entwicklungsprozesse von Lehrkräften im Kontext digitaler Transformation untersucht. Der Schwerpunkt liegt auf der digitalen Souveränität im Umgang mit Künstlicher Intelligenz sowie darauf, wie entsprechende Kompetenzen durch Aus- und Fortbildung aufgebaut und fachdidaktisch wirksam im Physikunterricht umgesetzt werden können.

Die Untersuchung erfolgt in einem konsekutiv angelegten Forschungsprozess (siehe Abbildung 6), der auf eine progressive Professionalisierung ausgerichtet ist. Ausgehend von der initialen Akzeptanz digitaler Werkzeuge und den damit verbundenen Überzeugungen von Lehrkräften und Lernenden (Studie I) wird der Fokus in Studie II auf den kognitiven und reflexiven Kompetenzaufbau im Umgang mit Künstlicher Intelligenz gelegt. Mit Studie III wird dieser Prozess abgeschlossen, indem die entwickelte Kompetenzbasis in einer fachdidaktischen Anwendungssituation überprüft und hinsichtlich ihrer Wirkung auf Lernprozesse sowie auf affektiv-motivationale Faktoren evaluiert wird. Dabei wird professionelle Souveränität nicht eindimensional verstanden, sondern als Zusammenspiel kognitiver, affektiv-motivationaler und didaktischer Dimensionen konzeptualisiert. Die drei Teilstudien ermöglichen in ihrer Gesamtheit eine differenzierte Betrachtung der Bedingungen, unter denen KI-gestützte Werkzeuge im Physikunterricht sinnvoll eingesetzt werden können. Anstelle eines isolierten Wirkungsnachweises liefert die Arbeit eine evidenzbasierte Analyse, die aufzeigt, wie technologische, kognitive und emotionale Faktoren ineinandergreifen, um digitale Technologien nachhaltig und verantwortungsvoll in professionelle Lehrhandlungen zu integrieren.

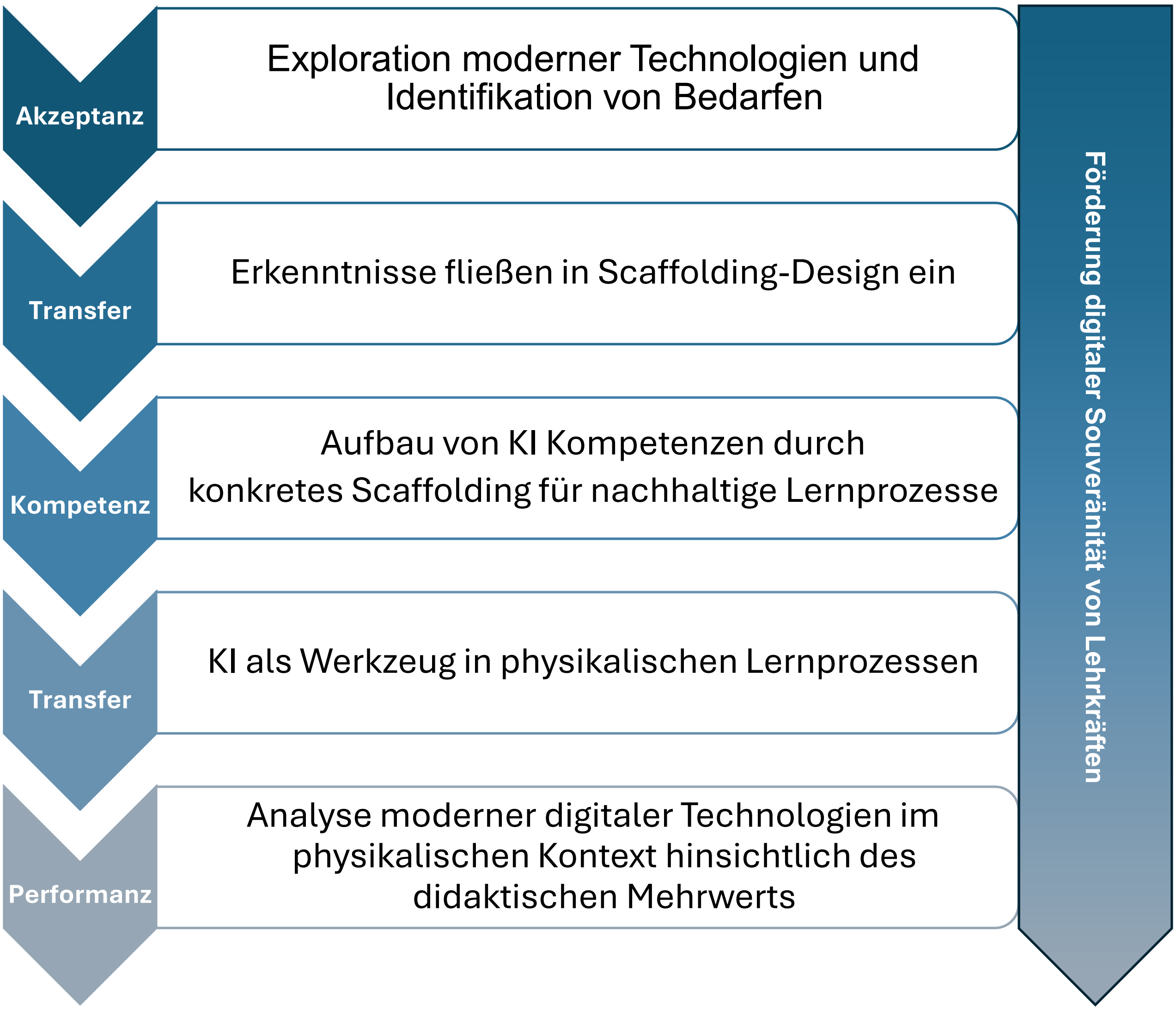

Abbildung 6 Phasenmodell der Untersuchung zur Förderung professioneller Souveränität von Lehrkräften. Die vertikale Struktur verdeutlicht die Entwicklungslinie von der technologischen Exploration über die Vermittlung von KI-Kompetenzen bis hin zur funktionalen Analyse und Evaluation des didaktischen Mehrwerts im Physikunterricht. Eigene Darstellung.

5.1. Studie I

Henze, J., Schatz, C., Malik, S. & Bresges, A. (2022). How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training? Frontiers in Education, 7, Artikel 872637. https://doi.org/10.3389/feduc.2022.872637

In der ersten Studie steht die Exploration der motivationalen und einstellungsbezogenen Grundlagen bei der Implementierung von digitalen Technologien im schulischen Kontext im Fokus. Der Schwerpunkt liegt hier auf der Identifikation von Gelingensbedingungen für die Interessenentwicklung. Ergänzend werden die Überzeugungen von Lehrkräften im Umgang mit modernen digitalen Technologien sowie Barrieren analysiert, die den Abbau von Hemmschwellen im edukativen Kontext beeinflussen. Im Fokus steht, wie der Einsatz von digitalen Kreativitätswerkzeugen die professionelle Souveränität von Lehrkräften beeinflusst und

welche Anforderungen diese Werkzeuge erfüllen müssen, um in der fachdidaktischen Praxis nachhaltig verankert zu werden.
Die Teilnehmenden an diesem Programm setzten sich aus 14 Lehrpersonen, 40 Studierenden und 116 Schüler:innen der Jahrgangsstufe 7 zusammen. Die Datenerhebung erfolgte im Rahmen eines Pre-Post-Test-Designs. Während die quantitativen Erhebungsinstrumente auf die Erfassung der Selbstwirksamkeitserwartung und der Einstellung zu digitalen Werkzeugen zielten, dienten qualitative offene Fragen der Analyse individueller Überzeugungen. Ergänzend wurden strukturierte Interviews mit Expert:innen aus den Bereichen Inklusion und Bildungstechnologie geführt, um die erhobenen Daten vor dem Hintergrund edukativer Anforderungen zu kontextualisieren.
Hinsichtlich der technologischen Akzeptanz der Lehrkräfte kann ein Spannungsverhältnis zwischen wahrgenommener Schüler:innenkompetenz im Umgang mit den technischen Werkzeugen und der eigenen Anwendungssouveränität identifiziert werden. Vor der Fortbildung hatte keine der beteiligten Lehrkräfte digitale Kreativitätswerkzeuge im Unterricht eingesetzt. Nach Abschluss der Intervention zeigt sich eine positive Tendenz bezüglich der Bereitschaft zur Adaption. Während im Pre-Test Unsicherheit und mangelnde didaktische Konzepte dominieren, befürwortet im Post-Test die Mehrheit der Teilnehmenden den Einsatz spezifischer Werkzeuge.
Ein wesentliches Ergebnis ist der Fokus auf Intuitivität und Robustheit. Werkzeuge mit variabler Komplexität, die sowohl einen niederschwelligen, spielerischen Zugang als auch tiefergehende Individualisierung und Programmierung ermöglichen, werden als besonders förderlich für die fachdidaktische Souveränität bewertet. Kritisch reflektiert wird die Rolle digitaler Kommunikation: Während die Lernenden durch das explorative Arbeiten hohe Autonomie und die 4K-Kompetenzen (Pfiffner et al., 2021) entwickeln, führt die pandemiebedingte Distanzlehre zu Frustration und Kommunikationsbarrieren, was die Notwendigkeit physischer Präsenzphasen für den sozialen Austausch unterstreicht.

Studie I zeigt damit, dass Lehrkräfte für die Integration neuer Technologien weniger eine umfassende technische Qualifizierung als vielmehr didaktische Orientierung, explorative Freiräume und niedrigschwellige Werkzeuge benötigen. Die Erweiterung des 5E-Modells um eine Austauschphase fördert dabei durch Transparenz der Prozesse die professionelle Souveränität und bildet die Grundlage für Studie II.

5.2. Studie II

Henze, J., Lademann, J., Bresges, A. & Becker-Genschow, S. (2025). Iterative development of an AI intervention for pre-service physics teachers from a Vygotskian perspective. Frontiers in Education, 10, Artikel 1707534. https://doi.org/10.3389/feduc.2025.1707534

Im Rahmen der zweiten Studie werden die Erkenntnisse der ersten Studie auf den Aufbau von AI Literacy im Bereich der physikalischen Lehramtsausbildung übertragen. Das Ziel der Untersuchung ist die iterative Entwicklung und Evaluation einer kompakten Interventionsform die angehende Physiklehrkräfte dazu befähigt, generative KI-Tools fachdidaktisch begründet einzusetzen. Der Fokus liegt dabei auf der Frage, wie ein strukturiertes Training das Verständnis der technologischen Funktionsweise sowie die Reflexionsfähigkeit im Umgang mit Künstlicher Intelligenz beeinflusst und inwiefern sich dies auf die wahrgenommene Handlungsfähigkeit im Einsatz von KI im späteren Schuldienst auswirkt.
Die didaktische Struktur der Intervention basiert, analog zur ersten Studie (Henze et al., 2022), auf dem 6E-Modell. Die konkrete Umsetzung umfasste zwei 90-minütige Sitzungen pro Iteration, in denen die Studierenden unter anderem Bildklassifikatoren mittels *Teachable Machine*[2] trainierten, um KI-Konzepte handlungsorientiert zu erarbeiten. Des Weiteren setzten sie KI-Tools zur Planung von Unterrichtssequenzen zu physikalischen Themen ein, indem die KI Vorschläge generierte und alternative didaktische Perspektiven aufzeigte. Die fachliche Bewertung und Entscheidung über den sinnvollen Einsatz der vorgeschlagenen Unterrichtseinheiten wurde diskursiv besprochen, so dass eine finale Entscheidung zur Unterrichtsgestaltung stets auf Seite der angehenden Lehrpersonen lag. Auf diese Weise konnte KI Impulse für eine kreative Unterrichtsgestaltung geben, ohne jedoch eigenständige Entscheidungen zu treffen. Die Evaluation erfolgte über einen validierten 30-Item-Test zur *AI Literacy* nach Hornberger et al. (2023) sowie einen Likert-basierten Fragebogen zu Einstellungen und motivationalen Variablen.

Hinsichtlich der AI Literacy zeigen beide Iterationen lediglich geringe, nicht signifikante Zuwächse im Gesamttestwert. Dies deutet darauf hin, dass Kurzinterventionen zwar spezifische Teilkompetenzen aktivieren können, jedoch für einen umfassenden und dauerhaften Kompetenzaufbau unzureichend sind. Punktuelle Verbesserungen ließen sich primär in Bereichen beobachten, die durch explizite Hands-on-Aktivitäten und gezieltes Scaffolding unterstützt wurden, wie etwa das Verständnis von maschinellen Lernprozessen oder die Interdisziplinarität von KI.
Trotz der geringen Wissenszuwächse zeigen die Ergebnisse, dass bereits kurze, aufgabenbasierte Interventionen das Gefühl der professionellen Handlungsfähigkeit sowie die regelmäßige Nutzung von KI im Alltag signifikant steigern können. Die Studierenden akzeptieren KI als

[2] https://teachablemachine.withgoogle.com/

Werkzeug zur Unterrichtsvorbereitung, lehnen jedoch die Auslagerung pädagogischer Kernaufgaben, wie automatisierte Leistungsbewertung, konsequent ab. Diese Haltung zeigt, dass die Teilnehmenden KI als kreatives Werkzeug wahrnehmen, die Notwendigkeit menschlicher Kontrolle jedoch betonen. Die Akzeptanz bezieht sich demnach vor allem auf KI zur Generierung von Ideen, während Bereiche, die pädagogisches Urteilsvermögen und persönliche Beziehungsarbeit erfordern, menschlicher Expertise bedürfen. Zudem erweist sich die gemeinsame Reflexion in der Exchange-Phase auch in dieser Studie als essenziell. Der Vergleich der Iterationen zeigt, dass ohne den kollektiven Austausch Fortschritte in ethisch anspruchsvollen und konzeptionell komplexen Bereichen limitiert bleiben. Inhaltsbereiche, die eine tiefer gehende Reflexion, insbesondere in Bezug auf ethische Fragestellungen, betreffen, benötigen zwingend den diskursiven Raum, um kognitive Prozesse erfolgreich abzuschließen.

Studie II zeigt, dass die Entwicklung von AI Literacy ein multidimensionaler Prozess ist, der über technisches Wissen hinausgeht und durch kurze, fachspezifisch eingebettete Interventionen mit sozialen Aushandlungsprozessen gefördert werden kann. Diese kognitive Fundierung ermöglicht es angehenden Lehrkräften, KI als reflektiertes fachdidaktisches Werkzeug zur kognitiven Entlastung im Physikunterricht einzusetzen, wie in Studie III gezeigt wird.

5.3. Studie III

Henze, J., Lademann, J., Becker-Genschow, S. & Bresges, A. (2026). AI-supported data analysis boosts student motivation and reduces stress in physics education. Frontiers in Education, 11, Artikel 1719670. https://doi.org/10.3389/feduc.2026.1719670

Die dritte Teilstudie fokussiert die fachdidaktische Performanz und Evaluation generativer KI im Physikunterricht anhand eines funktionalen Vergleichs zwischen einer KI-gestützten Datenauswertung und einer etablierten, tabellenbasierten Vorgehensweise mit Excel. Aufbauend auf den Erkenntnissen zur Akzeptanz digitaler Werkzeuge aus Studie I, sowie zur Kompetenzentwicklung im Umgang mit KI aus Studie II, wird in dieser Studie der konkrete didaktische Mehrwert von KI als Unterstützungswerkzeug bei der Auswertung physikalischer Pendel-Experimente im Rahmen der Lehramtsausbildung untersucht.
Im Zentrum steht dabei der für diesen Zweck spezifisch entwickelte Chatbot ExperiMentor. Dieser unterstützt Lernende bei der Datenauswertung von Faden- und Federpendelversuchen durch strukturierende Hinweise, adaptive Rückfragen, grafische Visualisierungen und schrittweise Analysevorschläge, ohne fachliche Lösungen vorzugeben oder analytische Schritte autonom auszuführen. Die Interaktion ist bewusst so gestaltet, dass Lernende aktiv Entscheidungen treffen und Auswertungsschritte selbst anstoßen müssen. Ziel ist es, KI nicht als problemlösenden Ersatz, sondern als kognitives und metakognitives Scaffolding innerhalb eines instruktional angeleiteten Settings einzusetzen.

Das Ziel der Untersuchung ist herauszufinden, welchen Einfluss die verwendeten Werkzeuge sowohl auf den Lernerfolg bei der Datenauswertung als auch auf affektiv-motivationale Aspekte des Lernprozesses haben. Der Fokus liegt dabei explizit nicht auf einer Ersetzung didaktischer Prozesse, sondern auf der Analyse, inwiefern ein generativer KI-Chatbot Lernende bei komplexen Datenauswertungen unterstützen kann, ohne fachliche Lösungen vorzugeben.
Die Studie wurde als randomisiertes Pre-Post-Kontrollgruppendesign mit 50 Lehramtsstudierenden durchgeführt. Die Teilnehmenden analysierten Datensätze zu Faden- und Federpendel entweder mit dem KI Chatbot ExperiMentor oder mit Excel. Der Pre-Test erfolgte vor Kenntnis der Gruppenzugehörigkeit. Erhoben wurden fachliche Lernzuwächse über objektive Items sowie affektiv-motivationale Variablen mittels Likert-Skalen.
Die Ergebnisse zeigen, dass beide Gruppen signifikante Wissenszuwächse erzielten, wobei keine statistisch signifikanten Unterschiede in der kognitiven Performanz zwischen der KI- und der Excel-Gruppe festgestellt werden konnten. Damit verdeutlicht die Studie, dass KI-gestützte Datenauswertung traditionelle Werkzeuge hinsichtlich fachlicher Lernergebnisse nicht automatisch übertrifft. Gleichzeitig offenbaren sich jedoch deutliche Unterschiede auf der affektiv-motivationalen Ebene. Die KI-Gruppe zeigt signifikant höhere Werte in Motivation, wahrgenommener Methodenwirksamkeit und positiven emotionalen Lernerfahrungen bei gleichzeitig reduzierter Frustration und geringerer wahrgenommener kognitiver Belastung.
Diese Befunde lassen sich vor dem Hintergrund der Cognitive Load Theory sowie der Zone der proximalen Entwicklung interpretieren. ExperiMentor gewährleistet Unterstützung in Form von adaptivem Scaffolding, das extrinsische kognitive Belastung reduziert, indem operative und organisatorische Anforderungen der Datenauswertung strukturiert werden, ohne den fachlichen Denkprozess zu ersetzen. Dadurch verbleiben Lernende innerhalb ihres individuellen Kompetenzbereichs, während zugleich Ressourcen für konzeptuelles Verständnis und reflexive Auseinandersetzung freigegeben werden. Studie III zeigt, dass KI insbesondere dort wirksam wird, wo sie fachdidaktisch kontrolliert eingesetzt wird und professionelle Handlungsspielräume erweitert, statt sie zu ersetzen. Die Ergebnisse zeigen, dass im Rahmen der untersuchten Lernsettings affektiv-motivationale Entlastung auftritt und deuten darauf hin, dass diese Entlastung eine unterstützende Funktion für nachhaltige fachliche Lernprozesse haben kann.

6. Ergebnisse

6.1. Studie I

Der Autor der Dissertation war maßgeblich an der Konzeption und dem Design der Studie beteiligt. Ein zentraler inhaltlicher Beitrag des Autors bestand in der eigenständigen Entwicklung der problemorientierten Lernszenarien, sowie in der Auswahl und didaktischen Begründung der eingesetzten technologischen Plattformen. Darüber hinaus führte der Autor die Interviews mit Fachexpert:innen und die anschließende vollständige Analyse durch und war für die Organisation der Kooperation mit der Schule verantwortlich. Ebenso entwickelte der Autor die Fragebögen. Die Datenauswertung und -interpretation oblag ebenfalls dem Autor dieser Dissertation.

Im Hinblick auf die schriftliche Ausarbeitung verfasste der Autor der Dissertation den Erstentwurf des Manuskripts vollständig. Er verantwortete die inhaltliche und sprachliche Überarbeitung des gesamten Artikels bis zur Einreichung. Die abschließende Revision des Manuskripts lag ebenfalls bei dem Autor.

Die Beiträge der Koautor:innen umfassten insbesondere die Ko-Leitung der Entwicklung des pädagogischen Gesamtkonzepts durch Prof Dr. André Bresges, Lars Möhring und Carina Schatz sowie die Entwicklung und Evaluation von Fragebögen zur Untersuchung der Videos in der Engage-Phase durch Shalina Malik.

frontiers | Frontiers in Education

ORIGINAL RESEARCH
published: 21 June 2022
doi: 10.3389/feduc.2022.872637

How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training?

Jannik Henze*, Carina Schatz, Shalina Malik and André Bresges

Institute for Physics Education, University of Cologne, Cologne, Germany

OPEN ACCESS

Edited by:
Stamatios Papadakis,
University of Crete, Greece

Reviewed by:
Ángel Freddy Rodríguez Torres,
Central University of Ecuador, Ecuador
Sila Kaya-Capocci,
Ağrı İbrahim Çeçen University, Turkey

****Correspondence:***
Jannik Henze
jannik.henze@uni-koeln.de

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *09 February 2022*
Accepted: *20 April 2022*
Published: *21 June 2022*

Citation:
Henze J, Schatz C, Malik S and Bresges A (2022) How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training? Front. Educ. 7:872637. doi: 10.3389/feduc.2022.872637

Schools are searching for strategies to foster 4C competencies (Creativity, Cooperation, Communication and Critical Thinking) in children. Scientific Reasoning, Critical Thinking, and the ability to debunk myths are already important competencies that can be fostered with science education. How can we approach the majority of seventh grade students in a given school to create innovative approaches for the future, and leverage their skills in science, art and (digital) technology along the path? And are the teachers ready to guide them on this path? This article focuses on the questions: how did the teachers adopt both the STEAM approach, and the use of digital tools while being supervised by researchers and student teachers and how did this change their beliefs about technology in education. As a pathway, we aimed to connect Robotics, Coding, Artificial Intelligence (AI) with the Sustainable Development Goals (SDGs) of the United Nations. To end poverty, protect the environment, and ensure that all people enjoy peace and prosperity by 2030, the SDGs are incorporated into national policies and school curricula. With this, citizens, teachers, and governments alike struggle with strategies on how these goals can be reached by 2030, facing the growing challenges in an ever increasingly complex and insecure world. It is clear that technology will take a dominant role in this development. Based on the STEAM paradigm and the 5E approach of the Biological Sciences Curriculum Study (BSCS), we have developed a pedagogical concept that encompasses both the technological aspects, AI and the SDGs. We tested this concept as part of an on-the-job teacher training project with 60 education science student teachers and 8 teachers in their classrooms, together with their 116 7th grade students and found out that STEAM-based projects with a sixth phase in addition to the 5E approach can be carried out promisingly with the help of digital creativity tools. We found that the 5E model with an additional sixth phase is well suited for bringing STEAM into the classroom.

Keywords: STEAM, robotics, computational thinking (CT), Sustainable Development Goals (SDGs), teacher—education, Piaget, Vygotski

INTRODUCTION

Scientific reasoning, critical thinking, and the ability to debunk myths are important competencies that can be fostered with science education. But how can a majority of students in a given school leverage their skills in science, art and technology to create innovative paths that will lead them to a positive future, and how can teachers guide them on their journey?

The 5E model, which dates to the 1980s (Bybee et al., 2006), serves as the basis for this study. Since then, many digital innovations have found their way into the lives of students. Likewise, their everyday living has changed. Due to becoming an internal part of the modern school system, it became necessary to investigate whether sustainability and digitalization are compatible with a 40 year old teaching model. Furthermore, modern, and digital education is lacking in the German school system.

In a large STEAM (Science, **T**echnology, **E**ngineering, **A**rts, and **M**athematics)-based project we aimed to connect Robotics, Coding, Artificial Intelligence (AI) with the Sustainable Development Goals (SDGs) of the United Nations. These Goals are implemented worldwide into curricula "as a universal call to action to end poverty, protect the planet, and ensure that by 2030 all people enjoy peace and prosperity" (United Nations Development Programme, 2021). This means that governments and education must strive to develop and implement strategies on how these goals can be communicated within their classrooms and how it is even possible to reach them before 2030. Can advancing the "smart" use of technology be a possible solution to achieve these goals?

In the following, the theoretical framework of the research will first be outlined. This includes the presentation of developmental psychological aspects, the 5E model, based on the works of Bybee et al. (2006) and Bybee (2009), the explanation of what digital creativity tools is as well as the connection between STEAM education and the SDGs. We then describe the research questions, our approach, and the materials and methods we used. Finally, we present and discuss the results, draw a conclusion and give a brief outlook on possible future research.

THEORETICAL FRAMEWORK

When we look at digital creativity tools, at the first glance they remind us of the toys that students are playing games with. This is our motivation to start by briefly examining the developmental psychological perspective on the process of playing as a concept of learning. Next, the 5E-model, on which the field study is based, the aspect of STEAM digital creativity tools and STEAM education will be presented. Lastly, a brief description of the school where the test was conducted is given.

Developmental Psychological Aspects of Playing as a Concept of Learning

Playing can mean several different things that children can engage in. According to psychologist Lev Vygotski, *playing*, be it with toys or a game, is triggered by situations that might be relevant for children's lives and engages them to transform these certain situations into a game. For example: when a child observes a stagecoach driving by, it might react by playing "stagecoach driver." Within this game-situation, the child prepares himself to engage in a situation where it might become, eventually, an actual stagecoach driver.

Further on, according to psychologist Jean Piaget, playing—as a concept of how Vygotsky described it—can be divided into two different developmental stages that describe how, and to what extent, a child benefits from playing. The first of these stages is practice. Here, the physical development with respect to play takes place by imitating known basic principles and understanding the uses of objects, thus satisfying the intrinsic urge to explore (Leong and Bodrova, 1996), which can be applied to this study by letting the children explore the given tools and partaking in construction games. The next stage, according to Piaget, is symbolic play, in which mental models are created, where every object can be a placeholder for something else, which are then applied in play (Leong and Bodrova, 1996). The advantage of play is that it gives the learners a sense of self-control, which serves as a base to take on new challenges more self-efficiently (Leong and Bodrova, 1996). Both Lev Vygotsky and Jean Piaget assume an interiorization process in their theories, in which learners develop their conceptions, ideas and models with the help of concrete actions (Aebli, 1985). This means that, through playful situations, complex interrelationships can be modeled in an understandable way (Kircher et al., 2014). One of the big ideas of STEAM Education and the Maker Movement is linking basic knowledge in physics with everyday technology using construction games. Within these games, students can explore complex socio-technical issues in a playful situation. This enables learners to be creative during the construction process and thus to realize many ideas (Kircher et al., 2014).

The 5E-Model

To facilitate the learning of physical concepts, learners must be engaged in appropriate learning activities. These activities should be designed in three parts to be as effective as possible. In the first part, goals should be identified. Following this, the current learning status should be discussed. The last part should determine the means by which the learners can reach the identified goal from their current position (Etkina et al., 2006). This tripartition can be expanded into more parts to allow learners to delve deeper into the given subject matter. The 5E-model was developed based on constructivist learning theory and cognition psychology as well as proven methods in science education (Duran and Duran, 2004) to create lessons in a student instead of teacher centered way (Turan, 2021). The model can be used within single or few hours as well as for longer units. Teachers that participated in studies in which the 5E-Model was applied, said to have more confidence and are more comfortable in teaching sciences (Duran and Duran, 2004). Nevertheless, studies also showed that it is hard for teachers to find suitable activities and materials for different phases of the 5E-Model (Namdar and Kucuk, 2018). Furthermore, several studies have shown different barriers, like classroom management and time issues, that hinder teachers from implementing the 5E into their own lesson planning (Turan, 2021).

The 5E-Model consists of the following five phases: Engagement, Exploration, Explanation, Elaboration, and Evaluation. In the first phase, learners are confronted with the learning content, which activates their existing knowledge and their curiosity (Bybee et al., 2006). It is also possible to determine what students might already know about the topic or what (mis)conceptions they have (Duran and Duran, 2004). Accordingly, the learners are confronted with a problem to solve. The phase is successful when the pupils are engaged in the problem and are intrinsically motivated to solve it (Bybee, 2009). In Vygotski's approach, the motivation and the need for action are to be located here.

In the Exploration phase, learners are given the opportunity to freely explore and become familiar with the essential skills and concepts that are made necessary by the problem posed in the engagement phase (Duran and Duran, 2004). This phase should be designed so that all learners have the same experience to build knowledge and skills. The role of the teacher in this phase can be seen as merely accompanying to allow students to explore as freely as possible (Bybee, 2009) and explicitly not giving away any kind of explanation, which is reserved for the following phase (Duran and Duran, 2004).

The Explanation phase allows learners to demonstrate their understanding of the concepts by explaining certain aspects or the entire concept itself (Bybee et al., 2006). In this way, the Explanation phase helps to ensure that learners develop a consistent vocabulary related to the problem, and present the concepts, information, and skills they have grasped in an understandable way (Bybee, 2009). Furthermore, a teacher should only fill in with explanations, if the Student's way of explaining is not sufficient or contains misconceptions (Duran and Duran, 2004; Namdar and Kucuk, 2018).

In the Elaboration phase, learners can consolidate their abilities and understanding regarding the topic, thereby leading them to a deeper understanding and adapted skills (Bybee et al., 2006). In this phase, learners can build on the concepts and skills they have already understood by, for example, applying them to new concepts within the problem. For this purpose, the interaction between learners in groups can be seen as a major factor for the success of the phase. The group discussions and collaborations provide opportunities to receive feedback from other learners on the one hand and to enter an exchange about their knowledge on the other hand. The goal of the elaboration phase is the transfer of knowledge from previous phases to new problems (Bybee, 2009). Here, as in the Exploration phase, the playful approach emphasized by Vygotsky is followed.

In the Evaluation phase, the learners are given the opportunity to reflect on their learning journey (Bybee et al., 2006). In this final phase they also receive feedback on their learning progress, skills, and insights (Bybee, 2009). This should give the teachers proof of the Student's learning success and can be conducted in a formal or informal way (Duran and Duran, 2004).

In the Exchange phase, a sixth phase we added to the 5E in the last week of the field study, we provided an opportunity for all participants to reflect and exchange on what and how they learned. This phase was added to emphasize the exchange between learners as well as between learners and teachers. We found this to be a very profitable addition to the 5E-Model, to get insight into the students as well as the teachers' experience of the whole project to enrich the Evaluation phase. Accordingly, this phase focuses more on meta-cognitive skills than the other phases. In Vygotskian thinking, the Engage phase would stimulate the children to open their Zone of Proximal Development, while the Explore and Elaborate phase provide the necessary playground for the learners to simulate the situation they engage in, test and improve their competencies, and simulate possible outcomes. The Explain and Exchange phase with their focus on inter-group communication provide the students with the necessary opportunity to negotiate the rules of their game in Vygotskian theory. From a social-constructivist perspective, these phases provide the opportunity to exchange insights, models and world-views and assess the relevance for life in the view of their peers. While the "Explore" phase is generally open and playful, the "Elaborate" Phase targets the development of a testable prototype that might be evaluated in the subsequent "Evaluate" Phase. This connects to the learning theory of Piaget, where children test their hypotheses by play.

STEAM Education

The core idea behind STEM (Science, Technology, Engineering and Math) Education is to connect the sciences, rather than teaching them in isolation (Krakower, 2018). But even though the relationship between different disciplines was recognized, the creative aspects of them were missing. Due to becoming more influential and significant in this digital and global world, such aspects were incorporated into the STEM framework (Yakman, 2008), resulting in the existence of the STEAM approach. The natural science disciplines are not only complemented by the arts, but also by methods to encourage creativity and innovation. These methods, like visual thinking, were derived from artistic fields (Thomas and Huffman, 2020). If Art would be used in a narrow sense, e.g., just in the form of painting, learners would not see where this is connected to and relevant for STEAM problems.

Art can only be integrated into the learning process if it is used in a broader sense. Here learners progress by integrating the arts in the area of problem solving (Quigley et al., 2020). By integrating the Arts aspect, more individuals can be reached, who have little interest in traditional STEM contexts (Thomas and Huffman, 2020). In the context of STEAM education, collaboration, and mutual feedback among learners worked very well, as has been observed by Cassie Quigley from the University of Pittsburgh. This was due in part to the use of technology and assignments that encourage collaborative work. Each learner in a group was assigned a task according to their abilities to solve a problem cooperatively as a group (Quigley et al., 2020). This cooperative and problem-solving approach of learning is at the forefront of STEAM education (Jackson et al., 2020).

Sustainable Development Goals

The United Nations formulated 17 goals to improve human life on earth in the near future. They are known as the Sustainable Development Goals or SDGs. Each of these goals aims for different aspects of life and contains different targets and possible actions to reach it. Some of the SDGs are already covered by the STEAM education definition. For example, SDG 09 promotes to "Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation" (United Nations Department of Economic and Social Affairs, 2021). SDG 07 is to "Ensure access to affordable, reliable, sustainable, and modern energy for all" (United Nations Department of Economic and Social Affairs, 2021). The Art aspect includes considerations of societal developments, and the aspect of Engineering relates to the SDGs via the creative as well as logical use of technical tools to solve global problems (Yakman, 2008).

STEAM Digital Creativity Tools

STEAM tools aim to support the training of creative thinking as well as the competence of problem solving and critical thinking (Makeblock, 2019). The intrinsic drive for discovery postulated by Piaget can also be nurtured with the help of STEAM tools by encouraging learners to find creative solutions to specific problems. In this regard, digital creativity tools, such as those used in the field study we conducted, are suitable for this purpose, as they are child-friendly and contain many technical resources which are relevant for the teaching of physics. Digital creativity tools refer to various devices that, among other things, can be used to stimulate the learners' creativity in order to find solutions to problems. A digital creativity tool can be used to integrate STEAM education in schools and develop problem solving skills, creativity, and boost the students motivation (Kalogiannakis et al., 2021). A study commissioned by LEGO Education, conducted in 2019 by Harris Insights and Analytics, examined Students' confidence in the context of STEAM education and digital creativity tools. In this study, only 14% of German students reported being very confident in learning STEAM content (Harris Insights and Analytics, 2019). Furthermore, a study on the physical area of light and optics showed a significant increase in both learning success and creativity among students who learnt these topics using STEAM methods (Wandari et al., 2018). Accordingly, compared to traditional instruction, there is a significant positive difference in the use of digital technologies in STEAM-based instruction (Tamim, 2011). A study that examined the use of another digital creativity tool (BBC micro: bit) found teachers being more open about using such a tool if this has a connection to the everyday life of a student and is generally useful (Kalogiannakis et al., 2021). They also show to be more positive about using digital tools if they have multiple uses in school contexts and allows students and teachers to learn from it (Papadakis, 2022).

Recent past works like the ones from Kalogiannakis et al. (2021) and Papadakis (2022) mainly focused on the usage of digital tools like apps and programming languages in a school context and lacked physical tools. Therefore we wanted to gain more insight on several educational tools presented in the following.

In our study, the results of which are presented in this paper, we used four different tools (Makeblock mTiny, Makeblock Cody Rocky, Makeblock Neuron, and DJI Tello Edu Drone) each of which has different characteristics.

The Makeblock mTiny was chosen for the project because it offers screen-free programming and thus reduces screen time, on the other hand it enables inexperienced students to experience and understand complex programming in a playful and uncomplicated way, while at the same time teaches the basics of computational thinking. A meta study regarding ScratchJr which is similar to the way the mTiny is programmed, shows it to be useful in introducing young students to STEAM education (Papadakis, 2022).

The Makeblock Codey Rocky was selected because this tool is a further development of the mTiny. It contains many sensors with which learners can program various commands and then see if Codey Rocky reacts to them. It can also be controlled directly using an app or be programmed using a block-based programming language. It was chosen as an addition to the mTiny, because this robot cannot be controlled via Joystick and has to be programmed or controlled with the help of a tablet device.

The Makeblock Neuron set was chosen for the project because its properties allow it to be easily used as a versatile construction kit. This is based on a number of sensors that allow various measurements and, on the other hand, a large number of actuators that can be attached and controlled, even remotely to simulate an Internet of Things (IoT) environment. In addition, it is possible to connect the Neuron set with the Codey Rocky and thus exploit a potential for mobile or robotics applications. The set can be programmed without a screen by connecting individual blocks in a certain order to build simple measuring devices. In addition, it can be programmed via app in a block-based coding environment, which, according to Kalogiannakis et al. (2021) seems to help students understand the general concept of a programming language.

We selected to use a DJI Tello Edu drone for the project because drones can enable students to study and control an object that can freely move in a 3-dimensional, Cartesian space. In mathematics and physics education, this option was only accessible in simulations or thought experiments before the introduction of drones. In addition, drones are becoming increasingly present in today's world and students should therefore learn how to handle them in a safe manner that obeys rules of privacy. We chose the Tello Edu drone as its small size and weight allows it to be used in the classroom, making it very suitable for this project. An app makes it possible to program this drone using a block-based language. For classroom use, it is important that students can test their code within the app in a simulated flight environment before they are provided with the actual drone. This enhances safety, reduces the need to load batteries and reduces the overall cost for the school.

MATERIALS AND METHODS

Digital STEAM Creativity Tools were used for teaching and learning Robotics, Sensors, Artificial Intelligence and Computational Thinking together with a Vygotskian teaching approach in a large scale, school-spanning field study. For our research, a mixed method design was conducted with different focus areas.

The field study project was structured using a modified 5E model (Bybee et al., 2006; see **Figure 1**), which was used in the context of the university with the English terms but translated for the students with appropriate German terms. The individual phases and activities are briefly described below. Across all the phases of the project, each unit was transparently accompanied by appropriate presentations by the staff. At the beginning of each lesson, the learners were thus offered a classification of the respective day in the overall project as well as an overview of the daily schedule.

The *Engagement* phase took place in the first week of the project. The thematic introduction was done by means of two videos on different SDGs, of which each learning landscape watched one video. The first video focused on SDGs 2 *Zero Hunger* and 6 *Clean Water and Sanitation* (see **Supplementary Figure 1**), whereas the second video focused on SDGs 3 *Good Health and Wellbeing* and 11 *Sustainable Cities and Communities* (see **Supplementary Figure 2**).

Following these videos, the students reflected on what they had seen with the help of in-depth questions on a padlet and conducted their own research on the SDGs. In this phase, the staff of the University of Cologne continued to explain the entire project process to the 7th grade students in detail. In the following second project week, the *Exploration* phase, the students, guided by university students, got to know the digital creativity tools. The teachers only played a minor role in supporting the 7th graders as well as the university students in terms of classroom management. Again, a distinction was made between the learning landscapes, so that in learning landscape A the devices mTiny, Codey Rocky and Neuron were used, and in learning landscape B the Codey Rocky, Neuron and the Tello (Edu) drone were used to test how the provided tools influence the designed solutions. During the *Explanation* phase, which was carried out in the third week, learners had to explain the possibilities of one of the devices they tested on a digital worksheet in Google Classroom. This was then evaluated as part of the study. In this way, the positive and negative characteristics, functions, and programming possibilities of the digital creativity tools in the perspective of the student could be studied. In addition, core groups had to display and explain the STEAM tools they had researched to other core groups so that an exchange could take place about all the devices and each student saw a short presentation about each tool. The *Elaboration* phase was extended to the fourth to sixth week. The phase started with the learners working in small groups to choose a topic related to the SDGs and one of the creativity tools, and then working on either their own or pre-determined research questions. Their aim was to find a solution to a problem that could be modeled using the tools.

This led to the fifth phase, the *Evaluation* phase. In this phase the students mainly prepared the presentation of their projects to a public audience. Since the learners were free to work on their project, prepare a presentation or do both at the same time, this phase blurs with the preceding Elaboration until the day of presentation: The Barcamp. This event was designed to resemble a design pitch to raise venture capital, or to raise public awareness for a project. The learners presented their solutions to the public in the form of a video conference. Access to the video conference was possible for everyone after prior registration. After each presentation, the audience had the opportunity to ask the presenters questions and give feedback on their prototypes. In the *Exchange* phase, the sixth phase we added to the 5E, in the last week of the field study, there was an opportunity for both learners and teachers to share and reflect on the project. An intervention on Artificial Intelligence was also conducted during this week to give the students a perspective on what modern technology could enable their own projects to do. Furthermore, the students filled out various surveys regarding final university student theses. Because of the burden of the surveys on the students we chose not to collect measurable data.

Due to the COVID-19 pandemic, the project had to be launched exclusively digitally. Accordingly, the students took part in a video conference led by University of Cologne staff. In the following week, due to the pandemic situation, it was possible to switch to a hybrid state that lasted for another 2 weeks. In this state, the groups were separated into subgroups, which alternated in daily visits to the school. While one group was able to go to school, the second group was connected to the lessons with the help of tablet PCs. From the fourth week of the project onward, the restrictions were eased, and the core groups were then present until the end of the project. Students who were not employed at the University of Cologne and who conducted some parts of the project were connected via tablet PC video conferencing during the project in order to minimize the risk of contagion for all involved.

The project was carried out at the Helios School—Inclusive University School of the City of Cologne. This school is designed by University of Cologne education scientist Kersten Reich in the tradition of John Dewey's laboratory school at Chicago University, but under today's conditions (Reich, 2018). Dewey anticipated already 100 years ago the needs of education that we consider crucial today, namely the multiperspectivity and broad access to learning. His vision of a school included the participation of students in social processes where they would

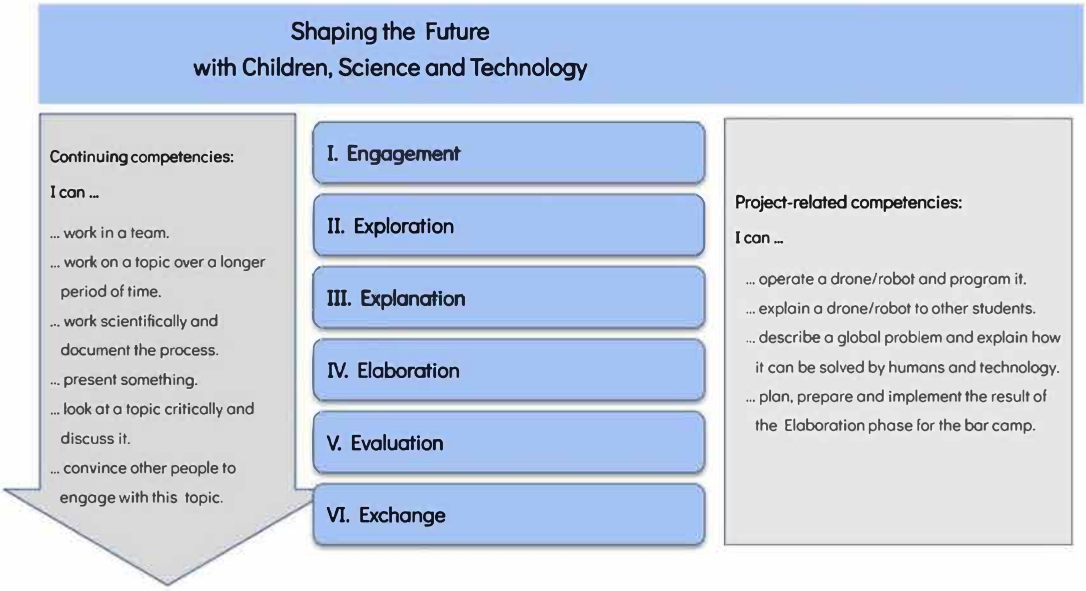

FIGURE 1 | Advanced organizer used in the project to show the students the competencies they can acquire.

build on their skills in communication as well as problem solving. One of the schools main foundations is, according to Deweys as well as Reichs research, the principle of learning and teaching through *learning by doing* (Reich, 2018). The Helios School was founded under a constructivist perspective toward education but had to face two major problems. The first being the heavy focus of the German educational system on the attainment of a degree rather than social equity. The second problem lies in the German teacher training system, which is split into theoretical and practical units (Reich, 2018).

The participants comprised about 116 7th grade students of the Helios Inclusive University School of Cologne, Germany together with their 14 teachers. The age range of the teachers was between 28 and 46 years. The teachers had been in teaching for between less than one and more than 16 years at the time of the study. The teachers' subjects ranged from social studies over languages to STEM and physical education, also one of the teachers was a special education teacher who did not specify further subjects. Furthermore over 200 students of the Bachelor and Master programs of University of Colognes STEM Teacher Training Department took part in this study, of which 40 were actively involved in the implementation of this field study, while the rest supported them with templates and feedback. All 40 actively involved university students and 116 7th graders took part in a 7 week on-the-job training program that was part of the regular 7th grader classes. The pupils were divided into the two learning landscapes A and B with each three different Stem Groups, which is the equivalent to a school class at the IUS.

To conclude the evaluation of the specific tools used, we used a pre-post-test on the partaking teachers, as well as a pre-post-test on the 7th grade students to evaluate the usage of videos. Furthermore, university students and teachers were interviewed regarding their view on the whole project at different times of the field study (see **Figure 2**).

The pre-post-tests regarding the evaluation of the digital creativity tools were formulated according to the rules for formulating questions for qualitative surveys (Döring et al., 2016). The questionnaires of the pre- and post-test differ in a few questions, which are only useful in each case in the pre- or post-test, for example, when first thoughts about a respective tool or experiences from the field study are asked. For each of the devices it was asked what thoughts the teachers had in each case when they saw the device for the first time. Teachers were also asked what they liked and disliked about each tool. This was intended to identify certain advantages or criticisms of the tools. Regarding possible points of criticism, the teachers could also suggest possible improvements. Also, a possible place of use away from the field study in combination with the willingness to use a digital creativity tool in the classroom was asked. This question gives first indications whether the field study has changed the willingness of the teachers to use digital creativity tools in the classroom. General desires for a digital creativity tool were also considered. This should highlight certain characteristics that digital creativity tools should have in order to be considered by teachers for use in the classroom. Teachers' responses were anonymized but coded so that pre-tests and post-tests could be matched without revealing teachers' identities.

The pre-post-tests regarding the use of videos were modeled after the *IPN Interessensstudie* (Measuring Students' interest in physics) from Häussler (1987). The original test assumed that student interest is not one-dimensional and not constant, but a complex situative variable that must be modeled along the three dimensions *topic*, *context* and *activity*. Sample item questions were "Do you want to learn more (*activity*) about how colors occur (*topic*) in the sky (*context*)" or "Do you want to

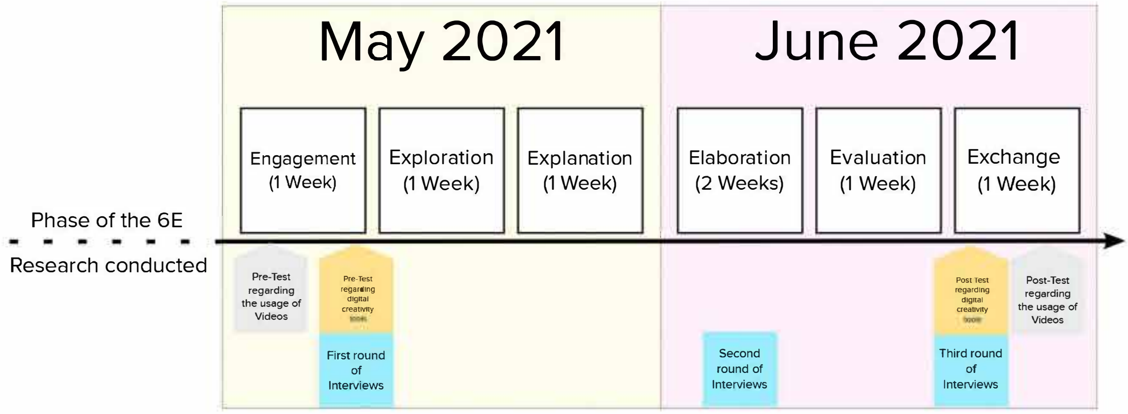

FIGURE 2 | Research process.

discuss and evaluate (*activity*) the positive and negative effect of micro-electronics (*topic*) on our lives (*context*)" (Häussler, 1987). To introduce students into the subject and to compensate for different prior knowledge of the participants, a short one-page introductory text was provided for each topic. After reading these texts, participants should indicate their interest to follow the topic in different contexts and with different activities. Interest was indicated using a five-point scale ranging from "My interest in this (item) is very high" to "My interest in this (item) is very low" (Häussler, 1987).

A test of the mathematical model conducted by Häussler on 4,034 students between 11 and 16 years revealed that the 3 dimensions are largely independent of each other, with interaction terms between the dimensions explaining only 2% of the variance (Häussler, 1987). Hence, it is reasonable to assure that the topics and contexts can be modified or exchanged independently of each other. We chose contexts that were derived from the SDGs for the pre-post-testing in this study. The one-page introductory texts of the original study were replaced by the introductory videos in the engagement phase. The proposed activities (to learn more, to construct, to discuss) of the 6E process were similar to the original study, and students could indicate on a 5-point scale if they are interested to take part in it. In addition, there were open-ended questions in which the students could independently write down activities they would take part in. These answers were clustered into *suggested* activities that are connected to the topic and context of the video, and *independent* activities that might still be connected to the context of the video (e.g., climate change) but did not have any connection to the lessons and the topic of the video (e.g., using public transportation to reduce CO_2 emission).

The questionnaire was tested with students to ensure that their understanding of the questions was comparable to the original study. Pre-testing took place immediately after the videos were shown. 91 Students took part in the pre-test (83%). The post-test that was conducted about 6 weeks later, after all activities took place. 83 students took part in the post-test (75%). The anonymous surveys ensured the privacy of the students. Since no code was generated and no socio-demographic data was collected, no conclusions can be drawn about individual students.

For the interviews, five participating teachers from the cooperating school were interviewed in three rounds each at the beginning, between the exploration and explanatory phase, in the elaboration phase, and after the end of the field study. The teachers were two women and three men. In the interviews, many open-ended questions were asked, which encouraged the persons to tell their stories freely and to follow up where, for example, dissatisfaction could be suspected. Of interest in the interviews were negative as well as positive personal experiences and aspects, learning situations, attitudes toward technology and cooperation with the school. The aim was to capture as many views as possible and to record the learning process of the individuals.

After the transcription, the qualitative data were analyzed according to Kuckartz (2018) using MAXQDA. Example main categories are praise, positive experiences, growth, learning process, attitude toward technology, criticism, negative experiences, wishes. Subcategories were then inductively derived from the data.

Since the research was conducted in German, the data is also mainly in German. Furthermore, conclusions about individuals could be drawn from the interviews despite greatest efforts to anonymize them. Therefore, the appendix of the interviews (**Appendix A**) is not distributed publicly but can be viewed on request.

RESULTS

How Can STEAM Education Based on the 5E Model Be Introduced in Schools?

One of two different videos was shown in the Engagement Phase in each of the different learning groups A or B. As a result of the survey regarding the effectiveness of the used videos, 23.8% of the students in learning group A and 26.1% in learning group B formulated ideas after watching the video on how to improve the life of people around the world. Another 21.4% of learning landscape A and 8.7% of the learning landscape B described ideas suggested by the respective video. Only 4.8% (A) and 10.9% (B) said they would have no ideas. In each case 50% (A) and 54.3% (B) made no statement (**Appendix B**, Chapter 4, Diagram 1). Through the video analysis survey, it was found that there was a tendency for increased interest in physics among the learners at the Inclusive University School prior to the project implementation. Even though the initial interest was measured immediately after watching the respective video, the results indicated a decrease in the interest (**Appendix B**, Chapter 6, Diagram 21 and 27).

Overall, through the pre-post survey regarding the use of video, it was found that engaging videos were instrumental in generating students interest in the subject matter. A video 4–5 min in length was sufficient for the interviewed students (**Appendix B**, Chapter 8, Item 15) if all essential problems and solution ideas were presented.

Due to pandemic teaching modes, the Exploration phase could not be conducted with all students at the same time. The participating teachers did not feel that the involvement of students via distance learning was adequate, causing frustration. Students found it difficult to participate in class via video conferencing (**Appendix A**, L4, Interview 2, pos. 5). As an alternative, for example, a more targeted use of university students in online teaching could be identified by having them help develop programs, with those who are not in school, that could then be tested on site (**Appendix A**, L1, Interview 1, pos. 21).

Many of those involved in the project commented positively in connection with the playful and practical opportunities offered by the devices. It was emphasized several times that not only the students had fun with the tools, but also the adult members of the project (**Appendix A**, L4, Interview 1, pos. 25). Several teachers as well as students wished for an extension of the Exploration Phase (**Appendix A**, L1, L3, L4, B1).

From the Students' presentations and completed worksheets, it can already be concluded that through the Explore phase, they learned about many of the positive and negative features of each device and understood how to achieve possible goals with these devices (**Appendix B**, Chapter 12, Summarized Evaluation). This highlighted the simplicity and intuitiveness of the devices, as the learners only had 90 min to get to know each one, but most importantly, it reinforced the success of the previous exploration phase.

Concluding this phase, the playful introduction of the devices aroused the interest of the learners encouraging them to expand the capabilities with the device. This has been shown that they were able to recognize the advantages and disadvantages as well as potential, with help, in the short time available. Furthermore, it seems to make sense to extend this phase to give all students the opportunity to get to know each device intensively, instead of only being able to try out three of the four devices for about 90 min each, as was the case in this project.

Sharing learning outcomes across learning landscapes in the Explanation phase was seen by teachers as critical for students because learning landscapes had little contact and additional connectivity issues would have limited already difficult communication. However, the fact that the students had to explain the devices to each other was seen positively (**Appendix A**, L2, Interview 2, pos. 7–9).

It can be concluded that while mutual exchange is important, it should be limited to the known peer group and, ideally under non-pandemic conditions, should take place in person. This means, for example, that there is no inhibition of communication that could arise from speaking in front of other children. This is an aspect that could be investigated in further studies in the future.

Regarding the Elaboration phase, a teacher reported that the students did not understand why, despite being fully present in class again, they should still interact with university students in videoconferences (**Appendix A**, L2, Interview 2, pos. 13). Therefore, the help that the university students were supposed to represent was not accepted by the pupils. Which is why, from the moment when all pupils were back in class, the university students perceived the negative reaction to support via videoconferencing (**Appendix A**, B1, Interview 3, pos. 7, 26).

This phase was described as particularly stressful by both teachers and members of the university team. They were forced to deliver the intervention, manage the classroom, and provide technical support to multiple groups simultaneously (**Appendix A**, L2, Interview 2, pos. 21; M3, Interview 1, pos. 21–41). The projects the students worked on were deepened and revised by them to solve a selected problem connected to the SDGs. They organized themselves into groups and worked on their projects without further instruction from the teachers or university students. No further motivation was needed than handing out the digital creativity tools and giving them a short overview of the schedule.

Extending this phase was mentioned afterward as a possible improvement (**Appendix A**, L1, Interview 3, pos. 9), since the students only had about 12 h over a 2 week period to work on their projects.

Summing up this phase, it can be said that the students had a good opportunity to work on their own projects. In order to create a more relaxed environment for all involved, including the teachers and in our case students of the university, this phase could be extended to allow more time on the one hand and on the other hand to give the teachers more possibilities to interact.

In the Evaluation phase the presentation of the Students' projects to a public audience took place. Since the learners were free to work on their project, prepare a presentation or do both at the same time, this phase blurs with the preceding Elaboration until the day of presentation.

The participation of the learners in the oral feedback in the Exchange phase was excellent and helped us to understand their perception of the project as well as providing insight on what could be improved going forward.

As mentioned with the Evaluate Phase, this phase has been added to the 5E model to allow for sharing of the learning journey. This exchange should only refer to the learning process and explicitly not to the learning outcome, so that the students can give unevaluated feedback, whereupon the learning process can be better adapted for them in the future.

The Post-Survey regarding the usage of videos in the learning landscapes A and B showed that 38.6% (A) and 35.9% (B) of the 7th grade students had their own ideas on how they could improve the life of people, which is an increase of + 14.8% (A) and + 9.8% (B) in contrast to the pre-Survey. Further 22.7% (A) and 23.1% (B) gave ideas suggested by the respective Videos they had watched. Nevertheless 27.3 (A) and 25.6% (B) of the students said they would have no ideas, which is a drastic increase of + 22.5% (A) and + 14.7% (B). Additionally, 11.4% (A) and 15.4% (B) did not answer this question in the Post-test (**Appendix B**, Chapter 5, Diagram 14). A possible Explanation for the increase of students saying to have no ideas is the decrease in students not answering this question. They might have just answered with no intention of giving an idea but unwilling to not-answer to this question.

The results show that interest in physics decreased after the 7-week project period, which could be associated with a kind of routine and saturation that occurred among the students (**Appendix B**, Chapter 6, Diagram 21, Diagram 27).

The project was well received by the teachers involved, especially regarding the cooperation between the university and the school, the motivation that the pupils experienced through the project, the equipment used as well as the learning paths taken by the learners (see **Figure 3**). In terms of the learning process, communication with students, the use of technology in the classroom, and programming were emphasized, and the interdisciplinary teaching was, among other things, also praised (see **Figure 4**). The main points of criticism relate to the usage of video conferences and didactic decisions and content. Also, the wish to strengthen teamwork was often mentioned, as well as more transparency in terms of organization (see **Figure 5**). Due to the COVID-19 pandemic, some phases of the project had to be designed either through distance learning or hybrid learning. This is also the biggest point of criticism from those involved. Because this will (hopefully) no longer be a problem in the future, this point of criticism should not be overestimated. The individual phases of the project also suffered from distance learning and hybrid learning, especially around the Explore and Elaborate phases. In a renewed implementation or consistent further development of the project, more time for these important practical parts should be considered. Furthermore, in a renewed implementation special incentives and insights could be created through possible links with experts on the respective topic.

What Is the Attitude of Teachers Toward the Adoption of STEAM Tools in the Context of STEAM Teaching and How Does It Change in the Course of a 7-Week, On-the-Job Training Program?

In the pre-post-test, the teachers expressed confidence in their Students' ability to work with the devices prior to the project, since they had great trust in their Students' abilities (**Appendix B**, Chapter 11, pos. 6, K10). Their belief that the students had already grown up with technology and thus had a high affinity for technology served as an important factor, which is why intuitive handling was to be expected (**Appendix B**, Chapter 11, pos. 6, K101). The teachers also reported little fear of contact on the part of the students and a high degree of curiosity. (**Appendix B**, Chapter 11, pos. 6, K102). All teachers at the project school indicated in a survey that they had not previously used any of the devices used in this project, nor had they used similar devices, in the classroom (**Appendix B**, Chapter 11, pos. 5, G50). Most of the persons interviewed showed a positive attitude toward technology in school lessons and emphasized on the advantages of it; but not without mentioning the importance of critical thinking while using technology (see **Figure 6**).

When asked in the post-test whether they would use other devices in the future, other than the ones used in the project, two teachers indicated *yes*, whereas one indicated *no* (**Appendix B**, Chapter 11, pos. 5, G50). Due to the wording of the question, it remains unclear at this point whether teachers would use the devices used in the project for teaching in the future. The teachers who stated *yes* to the above question named other projects and workshops as possible reasons along with other devices (**Appendix B**, Chapter 11, pos. 5, G51). When asked why the teachers have not yet used any digital creativity tools in their lessons, a lack of experience or the lack of the necessary equipment were the main reasons (**Appendix B**, Chapter 11, pos. 5, G50). Nevertheless, in many

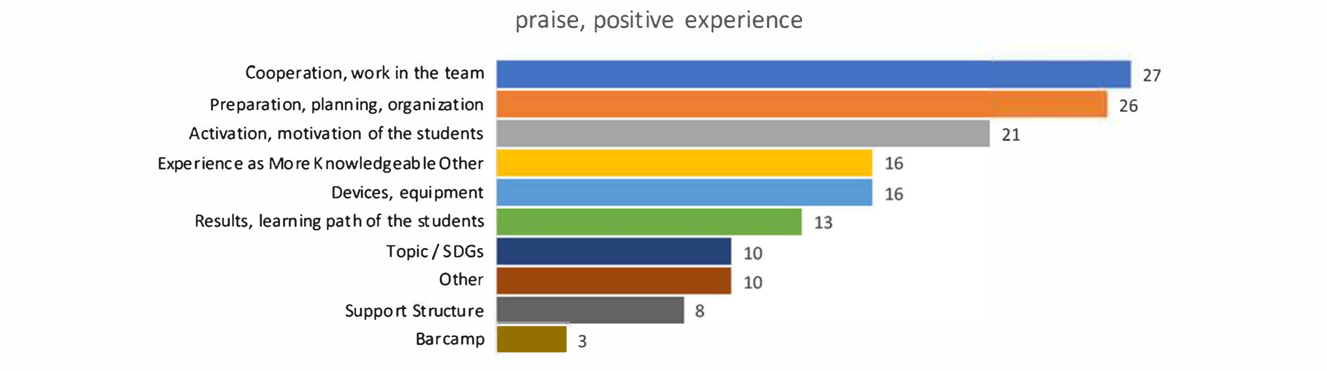

FIGURE 3 | Frequencies of interviews with mentions of the listed categories regarding praise and positive experiences in the project.

FIGURE 4 | Frequencies of interviews with mentions of the listed categories regarding growth and learning in the project.

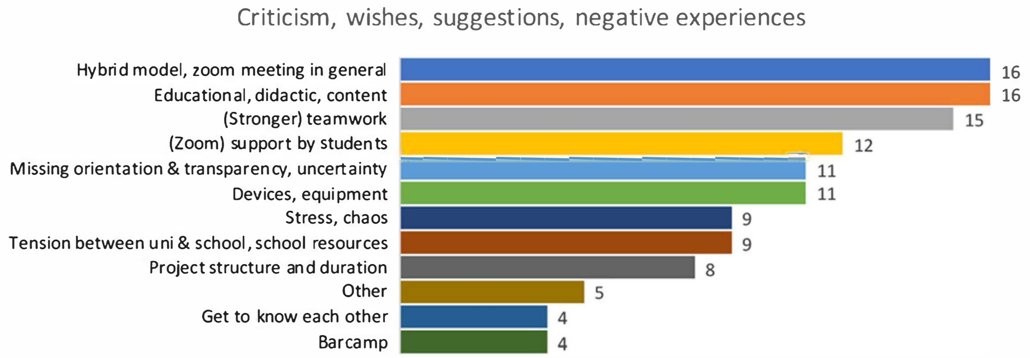

FIGURE 5 | Frequencies of interviews with mentions of the listed categories regarding criticism, wishes, suggestions and negative experiences in the project.

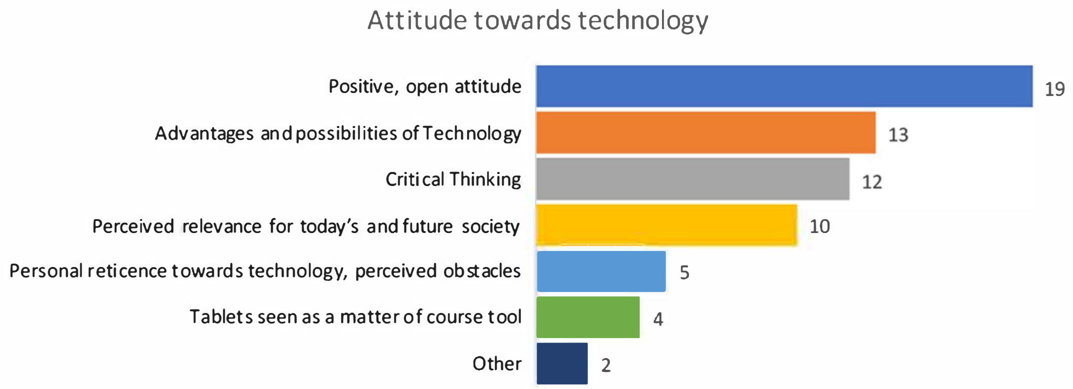

FIGURE 6 | Frequencies of interviews with mentions of the listed categories regarding the attitude toward technology.

interviews it was made clear that all participants were open toward using technology in their lessons and highlighted the advantages of it.

The digital creativity tools used in the project were generally well received by the teachers and students involved in the project. Nevertheless, from the point of view of the study participants, there are also possibilities for improving the tools, which draw attention to the disadvantages of the devices.

The easiest accessible device, the mTiny, was also rated the least popular by the participating teachers of all the devices. This can be explained with the target group (age 3 and older) that is usually addressed by this device. The mTiny can therefore only be recommended to a very limited extent for use in the seventh grade or higher, as it offers too few options for this age group, which is why students who already have experience with digital products quickly reach the limits of the device (**Appendix B**, Chapter 11, pos. 1, G10.4). It would therefore be an option to improve the mTiny by creating the possibility of programming using a tablet PC. This would enable more complex tasks for higher grades as well as technical enhancements.

The Codey Rocky, on the other hand, is much more suitable for the project's target group according to the data available. This can be concluded from the fact that the complexity is appropriate and variable, i.e., it is very easy to get started with the tool, but at the same time very complex problems can be processed. The given robustness against falls is also a factor that can play a central role in everyday school life (**Appendix B**, Chapter 11, pos. 2, G20.4).

Some teachers suspect that it looks too childish for seventh-grade students, which could create a barrier to learning. In contrast, however, the appearance was also viewed positively by other teachers as well as students. However, it was also suggested that a neutral version be developed for adolescent learners (**Appendix B**, Chapter 11, pos. 2, G20.1). Of the teachers involved in the project, four out of five stated in the pre-test that they would not use Codey Rocky in their lessons outside of the project. The main reasons for this were uncertainty in dealing with digital creativity tools and a lack of ideas for integrating the tool in a project in a meaningful way. One teacher stated that she would use the device in grades four to seven to reduce fear of contact with technical devices. In the post-test, on the other hand, one of the teachers who could not yet imagine using the Codey Rocky in the classroom in the pre-test, stated that she would want to use this device in the sixth or seventh grade in the context of programming. Another teacher, who stated in the pre-test that she had no ideas for the usage of it, answered in the post-test that she still had no ideas, but that she would build on the Students' results from the project to see how they could be transferred into reality or what possibilities already existed. In total, four teachers responded in the post-test that they could imagine using it in school (**Appendix B**, Chapter 11, pos. 2, G20.4).

The Neuron set is perceived as very positive both individually and as an extension of the Codey Rocky. The color scheme of the individual building blocks signalizing the purpose of the blocks was also positively emphasized. According to the teachers, this reinforces the inclusive character of the set and thus makes it easier to work with. The variable complexity, as with the Codey Rocky, also ensures a wide range of applications (**Appendix B**, Chapter 11, pos. 4, G40.2).

Overall, this digital creativity tool was also well received by the subjects of this study, as already in the pre-test three of the four teachers who answered this question stated that they would use the device in their own lessons outside of the project as a toy on the one hand and as an experimental kit for learners on the other. In addition, the set is intuitive and can be used from grade six in creative contexts without prescribing concrete tasks, since the urge to discover can be acted out here. One of the teachers also stated that she did not want to limit the use of the set to one grade level but wanted to use it in all grades. She confirmed this in the post-test and added that the complexity showed great variability. In the post-test, four of the five respondents said they would use the device outside of the project. It should be added here that one person would use it in grades five to seven, and another person noted that the Neuron Set was useful in science projects on the one hand, and as a pastime during breaks on the other. One teacher seemed to be particularly enthusiastic about the Neuron Set, stating that she would choose the Neuron Set if she were allowed to choose only one device for school, as it could be used in a variety of ways in science, arts, and social studies subjects. However, this teacher emphasized that she would never buy such a set because she was convinced that technology is always developing and therefore such a set could quickly become obsolete. Only one teacher stated in the pre- or post-test that they did not know whether they would use the Neuron set. However, these are two different teachers who did not fill out the corresponding test, so that no change can be determined here. The variability of the Neuron Set was described by many teachers as a positive aspect. It was also frequently mentioned that the set promotes the urge to discover and to be creative. The possibility of combining the set with the Codey Rocky was also emphasized by the teachers as a positive aspect. The haptics of the individual blocks, complexity and yet simplicity and highlighting the individual functions of the building blocks were also mentioned. Also, the Neuron Set promotes inclusive learning opportunities and ties into learners' interests (**Appendix B**, Chapter 11, pos. 4, G40.2).

The DJI Tello drone is suitable as a means of addressing several aspects of math and science education in the classroom. However, the math and science aspects should be central to reach this purpose, as it could otherwise distract too much from the actual subject matter. (**Appendix A**, Interview Transcription: René Foellmer, pos. 30).

One example was the discussion of possible flight paths for a load of water after being dropped from the drone on a plant. This discussion resembled an item of the Force Concept Inventory (FCI) which is regularly used in physics education. Nevertheless, this discussion was observed in the *elaborate* phase of a group concerned with the SDG 2: *promote sustainable agriculture* and was initiated by the problem solving process, without intervention by teachers. The possibility to program the drone, instead of controlling it, is positively emphasized by teachers (**Appendix B**, Chapter 11, pos. 3, G30.2).

Of the teachers involved in the project, two stated in the pre-test that they would also use the drones presented outside of the project, for example to take aerial photos. The teachers considered the drones equally suitable for higher grades, since responsible handling of the drones is important, and many questions can be raised. The other person who would use the drones outside of the project would use them in a foreign language and humanities class in grades seven and eight. Another teacher stated in the pre-test that she would not use the drone in her classes outside of the project because she did not have the confidence to develop a didactic concept and also did not have the subject expertise. This opinion changed in the post-test, with this teacher now being confident enough to use the drone in her own lessons after the project. Overall, of the five teachers who answered this question in the post-test, three said they would use the drone outside of the project. One person would continue to use it in projects and only one person answered that they did not know what they would use the drone for. Again, for the drone, grades seven and eight were indicated as possible settings (**Appendix B**, Chapter 11, pos. 3, G30.4).

CONCLUSION

This research gives a brief answer to the first research question *How can STEAM education based on the 5E model be introduced in schools?*

The 5E-Model with an additional sixth phase has proven to be a good foundation on how to implement STEAM into school lessons with the help of digital creativity tools. Adding the Exchange phase as a sixth phase to the already established 5E Model seems to be a profitable expansion. On the one hand, it allows exchange between students and students, and students and teachers. On the other hand, it allows both teachers and educational researchers to collect more insights into the Students' way of learning by examining Student's presentations and prototypes. Finally, teachers get to know their students better and can prepare their future teaching in a more adjusted way. The effectiveness of this must be proven in further studies but this and another study conducted by the university of cologne emphasizing on six instead of five phases indicates the possible impact of this addition.

The use of videos to introduce the 7th grade students into the topic proved to be extremely beneficial and it became clear through the interviews and student results that a differentiated examination of the videos can be sufficient to motivate the learners. The devices used were quickly and persistently understood by the students through the introductions designed by university students. This is supported by the observation made within the Exchange phase, seeing the students having designed intelligent examples to explain how the devices work. The fact that most of the supporting students were not on site in the hybrid situation and the pupils therefore had to learn how to use the tools on their own supports the idea that STEAM tools are easy and intuitive to handle (at least compared to typical equipment in a traditional science lab). It also supports promising ideas of STEAM tools as tools to foster creativity, and reduce the workload on teachers, since all activities were guided only by work instructions. The potential of the devices is visible in various projects the students developed. After asking the teachers what a digital creativity tool suitable for STEAM should be able to do, it became clear that the previously mentioned aspect of intuition was the most important. Also, further features like sturdiness just as a prerequisite to promote creativity were mentioned by teachers as something that should be characteristic for a digital creativity tool. Those features can all be found in the tools used in this project as well as many other tools on the market. As we have furthermore seen, the project itself as well as the used digital tools were able to expand and deepen the 4C competencies (Creativity, Cooperation, Communication, and Critical Thinking) and further competencies according to the teachers' assessments.

The research question *What is the attitude of teachers toward the adoption of STEAM tools in the context of STEM Teaching and how does it change in the course of a 7-Week, on-the-Job training program?* is difficult to answer due to the data situation.

The teachers participating in the project mentioned many different features a digital creativity tool should offer. What seems to be important to many of them is that the tool should be intuitive to use. In other words, it should be obvious at first glance how the device can be used, so that with the help of such tools, basic computer literacy can be taught in a playful manner at an early age. In contrast, it was also mentioned that a digital creativity tool should have a certain complexity so that students remain motivated not only to learn on the device but also to explore its different facets. Other frequently mentioned characteristics are that such a tool should, above all, promote creativity and explorative learning. Furthermore, it should be versatile and combinable so that it is able to implement most of the ideas and conceptions of the students. Features that are important for everyday school life, such as robustness and safety, were also mentioned by the teachers.

What furthermore seems to be important, is the possibility to individualize the devices so that the students can build up a personal relationship with them. From a special education perspective, it was also important to the school's teachers that a digital creativity tool could be used by all students in one way, that being ideally for those unable to read or write.

Overall, only a small change is observable, since only a few teachers who completed the pre-test also completed the post-test. However, a tendency toward more readiness can be observed when the tools are considered individually. In the post-test, more people indicated that they wanted to use these or similar tools in the classroom. The research question cannot be answered in general terms, but at least the described tendency can be derived from the available data, since in the pre-test, none of the teachers stated that they had previously used a digital creativity tool in the classroom, whereas in the post-test several teachers stated that they would consider doing so in the future.

In order to answer the title question *How might we raise interest in Robotics, Coding, AI, STEAM and Sustainable Development in university and on-the-job teacher training?* conclusively it requires more research in the future and could be focusing on different areas of the basis laid with this paper. One example could be more in-depth research regarding individual tools or certain activities for the respective phases of the expanded 5E-Model. It is also possible to adapt the field study for other schools and try to get more teachers to answer the research forms, to gain more insight on this concept and also avoid having to use video conferences and involve the university students in a better way.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://uni-koeln.sciebo.de/s/aWF3sU4PomZVQk4.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AB, CS, and JH contributed to the conception and design of the study, and wrote sections of the manuscript. AB led the design of the pedagogical concept. JH led the design of the problem based learning scenarios for computational thinking and selected the appropriate technological platforms, performed the analysis of the interview with domain experts, and wrote the first draft of the manuscript, acted as the main author and revised the manuscript. JH and CS organized the cooperation with the school and conducted jointly the interviews. CS performed the analysis of the interview with teachers. SM developed and evaluated questionnaires to assess the influence of the videos in the engage phase on the students. All authors approved the submitted version.

FUNDING

This project was part of the "Qualitätsoffensive Lehrerbildung," a joint initiative of the Federal Government and the Länder which aims to improve the quality of teacher training. The programme was funded by the Federal Ministry of Education and Research (No. 01JA1815).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2022.872637/full#supplementary-material

Supplementary Figure 1 | Screenshot from the video regarding SDGs 2 zero hunger and 6 clean water and sanitation.

Supplementary Figure 2 | Screenshot from the video regarding SDGs 3 good health and wellbeing and 11 sustainable cities and communities.

Supplementary Figure 3 | 7th grade students use the modular neuron-set to build their own prototype of a remote-controlled robot that carries a water pump. In later versions they included a water tank as well as a measuring device for soil moisture to determine when the pump should be activated to water certain spots.

Supplementary Figure 4 | Students elaborate on how to apply significant payload to a small drone while maintaining stable flight.

REFERENCES

Aebli, H. (1985). Das operative prinzip. *Math. Lehren* 11, 4–6.

Bybee, R. (2009). "The BSCS 5E instructional model and 21st century skills," in *Paper Prepared for the Workshop on Exploring the Intersection of Science Education and the Development of 21st Century Skills.* (Ottawa: National Research Council).

Bybee, R., Taylor, J., Gardner, A., van Scotter, P., Carlson Powell, J., Westbrook, A., et al. (2006). *The BSCS 5E Instructional Model: Origins, Effectiveness and Applications.* Available Online at: http://www.bscs.org/bscs-5e-instructional-model (accessed April 12, 2021).

Döring, N., Bortz, J., Pöschl, S., Werner, C. S., Schermelleh-Engel, K., Gerhard, C., et al. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften.* Berlin: Springer.

Duran, L. B., and Duran, E. (2004). The 5E instructional model: a learning cycle approach for inquiry-based science teaching. *Sci. Educ. Rev.* 3, 49–58.

Etkina, E., van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S., et al. (2006). Scientific abilities and their assessment. *Phys. Rev. Spec. Top. Phys. Educ. Res.* 2:020103. doi: 10.1103/PhysRevSTPER.2.020103

Harris Insights & Analytics (2019). *Confidence in Learning Poll.* Available Online at: https://mojebambino.pl/lego-education/wp-content/uploads/2022/01/ankieta-harris-poll.pdf (accessed November 15, 2021).

Häussler, P. (1987). Measuring students' interest in physics - design and results of a cross-sectional study in the Federal Republic of Germany. *Int. J. Sci. Educ.* 9, 79–92. doi: 10.1080/0950069870090109

Jackson, H., Basham, J., Thomas, K., and Hunt, C. (2020). "Designing STEAM learning environments," in *Challenges and Opportunities for Transforming from STEM to STEAM Education*, eds K. Thomas and D. Huffman (Hershey, PA: IGI Global), 1–23. doi: 10.5642/steam.20200402.04

Kalogiannakis, M., Tzagaraki, E., and Papadakis, S. (2021). "A systematic review of the use of BBC micro: bit in primary school," in *Proceedings of the 10th Virtual Edition of the International Conference New Perspectives in Science Education*, Crete, 379–384.

Kircher, E., Girwidz, R., and Häußler, P. (eds). (2014). *Physikdidaktik: Theorie und Praxis.* Berlin: Springer.

Krakower, B. (2018). *Getting Started with STEAM: Practical Strategies for the K-8 Classroom.* Milton: Routledge.

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung.* Weinheim: Beltz Juventa.

Leong, D., and Bodrova, E. (1996). *Play: A Vygotskian Approach.* Available Online at: https://video.alexanderstreet.com/watch/play-a-vygotskian-approach (accessed March 2, 2021).

Makeblock. (2019). *How to Teach STEAM Education in the Classroom.* Available Online at: https://education.makeblock.com/blogs/blog-primary-education/48109/ (accessed March 4, 2021).

Namdar, B., and Kucuk, M. (2018). Preservice science teachers' practices of critiquing and revising 5E lesson plans. *J. Sci. Teach. Educ.* 29, 468–484. doi: 10.1080/1046560X.2018.1469188

Papadakis, S. (2022). Can preschoolers learn computational thinking and coding skills with scratchJr? A systematic literature review. *Int. J. Educ. Reform* 1–34. doi: 10.1177/10567879221076077

Quigley, C., Herro, D., Shekell, C., Cian, H., and Jacques, L. (2020). Connected learning in STEAM classrooms: opportunities für engaging youth in science and math classrooms. *Int. J. Sci. Math. Educ.* 18, 1441–1463. doi: 10.1007/s10763-019-10034-z

Reich, K. (2018). Chapter 1. The Helios school – inclusive university school in the city of cologne. *Int. Res. High. Educ.* 3, 4–22. doi: 10.5430/irhe.v3n1p4

Tamim, R. (2011). What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Rev. Educ. Res.* 81, 4–28. doi: 10.3102/0034654310393361

Thomas, K., and Huffman, D. (2020). "Becoming STEAM: perspectives from school leaders," in *Challenges and Opportunities for Transforming From STEM to STEAM Education*, eds K. Thomas and D. Huffman (Hershey, PA: Business Science Reference), 104–129. doi: 10.4018/978-1-7998-2517-3.ch005

Turan, S. (2021). Pre-service teacher experiences of the 5E instructional model: a systematic review of qualitative studies. *Eurasia J. Math. Sci. Tech. Ed.* 17:em1994. doi: 10.29333/ejmste/11102

United Nations Department of Economic and Social Affairs (2021). *Sustainable Development Goals Report 2021*. New York, NY: United Nations Department of Economic and Social Affairs.

United Nations Development Programme (2021). *The SDGs in action: What are the Sustainable Development Goals?*. Available Online at: https://www.undp.org/sustainable-development-goals#:~:text=The%20Sustainable%20Development%20Goals%20(SDGs)%2C%20also%20known%20as%20the,people%20enjoy%20price%20and%20prosperity (accessed January 11, 2022).

Wandari, G. A., Wijaya, A. F. C., and Agustin, R. R. (2018). The effect of STEAM-based learning on students' concept mastery and creativity in learning light and optics. *J. Sci. Learn.* 2, 26–32. doi: 10.17509/jsl.v2i1.12878

Yakman, G. (2008). "STΣ@M education: an overview of creating a model of integrative education," in *Proceedings of the Pupils Attitudes Towards Technology*, Netherlands.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

APPENDIX

Appendix A | Interview data with teachers (L) and students (B/M) participating in the project and manufacturer Yu Hu as well as special education teacher René Foellmer.
Appendix B | Data regarding video analysis and survey data from participating teachers regarding digital creativity tools.

6.2. Studie II

Der Autor der Dissertation war verantwortlich für die inhaltliche Konzeption und die methodische Ausgestaltung der Studie. Der Eigenanteil des Autors umfasste insbesondere die Planung und Durchführung der empirischen Untersuchung, die systematische Aufbereitung und Kuratierung der erhobenen Daten, sowie deren vollständige formale Analyse und Auswertung. Die Interpretation der Daten und deren didaktisch fundierte Einordnung erfolgten ebenfalls durch den Autor der Dissertation. Die Erstellung der Visualisierungen zur Darstellung der Ergebnisse lag vollständig in der Verantwortung des Autors der Dissertation.

Im Hinblick auf die schriftliche Ausarbeitung verfasste der Autor den vollständigen Erstentwurf des Manuskripts. Die Überarbeitung, inhaltliche Präzisierung und redaktionelle Finalisierung des Artikels erfolgten ebenfalls durch den Autor unter Einbezug von Rückmeldungen von Julia Lademann, Prof. Dr. Sebastian Becker-Genschow und Prof Dr. André Bresges. Die weiteren Beiträge der Koautor:innen umfassten Beratung, Bereitstellung von Ressourcen und wissenschaftliche Betreuung durch Prof. Dr. Sebastian Becker-Genschow und Prof Dr. André Bresges. Die Gesamtverantwortung für Konzeption, Durchführung, Auswertung und Verschriftlichung der Studie lag bei dem Autor der Dissertation.

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 18 December 2025
DOI 10.3389/feduc.2025.1707534

Check for updates

OPEN ACCESS

EDITED BY
Ilya Levin,
Holon Institute of Technology, HIT, Israel

REVIEWED BY
Kleopatra Nikolopoulou,
National and Kapodistrian University of Athens, Greece
Li Li,
The University of Ontario Institute of Technology, Canada

*CORRESPONDENCE
Jannik Henze
jannik.henze@uni-koeln.de

RECEIVED 17 September 2025
REVISED 21 October 2025
ACCEPTED 21 November 2025
PUBLISHED 18 December 2025

CITATION
Henze J, Lademann J, Bresges A and Becker-Genschow S (2025) Iterative development of an AI intervention for pre-service physics teachers from a Vygotskian perspective.
Front. Educ. 10:1707534.
doi: 10.3389/feduc.2025.1707534

Iterative development of an AI intervention for pre-service physics teachers from a Vygotskian perspective

Jannik Henze[1]*, Julia Lademann[1,2], André Bresges[1] and Sebastian Becker-Genschow[2]

[1]Institute for Physics Education, University of Cologne, Cologne, Germany, [2]Digital Education Research, University of Cologne, Cologne, Germany

While artificial intelligence (AI) is transforming many sectors, its integration into pre-service teacher education in higher education remains limited. This study investigates the iterative development and effects of a concise, two-session educational intervention designed to foster AI literacy among pre-service physics teachers. Following a design-based research approach, the intervention was implemented in two iterations at the University of Cologne ($n = 31$ across two cohorts). Structured according to the 5E instructional model, the intervention required students to use generative AI tools as didactic instruments to create lesson plans and reflect on their usage. AI literacy was measured using a validated 30-item test, while attitudes toward AI were assessed via a 4-point Likert survey. Results indicate only small, non-significant increases in overall AI literacy, with selective gains observed in competencies explicitly supported by hands-on activities and targeted scaffolding. However, attitudinal measures demonstrated that even brief interventions can strengthen participants' openness toward AI and their perceived preparedness to use AI tools in teaching. Additionally, the iterative comparison highlighted format-sensitive effects. These findings suggest that while short design-based interventions can selectively activate elements of AI literacy and foster professional confidence, they are insufficient for broader skill Acquisition. Consequently, more sustained, context-rich engagements are likely required to achieve comprehensive and durable AI literacy development in pre-service teacher education.

KEYWORDS

design based research, AI literacy, pre-service teacher education, Vygotsky, attitude, knowledge gain

1 Introduction

The rapid evolution of artificial intelligence (AI) technologies has created both opportunities and challenges across various sectors, including education (Zawacki-Richter et al., 2019). In recent years, research on AI in education has seen significant growth, with Chen et al. (2020) demonstrating a substantial increase in related publications from 2015 onward. However, despite this growing interest, the practical integration of AI into educational contexts remains in its early stages and requires further empirical investigation (Farrokhnia et al., 2024), especially the impact of AI on higher education (Kuleto et al., 2021).

Educational systems are increasingly focused on addressing diverse learning needs and preparing students for future challenges, moving beyond the transmission of factual knowledge toward the development of comprehensive skill sets (Roll and Wylie, 2016). Within this evolving landscape, the concept of AI literacy has emerged as a critical component of modern education. Long and Magerko (2020) define AI literacy as 17 different competencies enabling effective engagement with AI technologies. It involves a thorough understanding of AI concepts, reflective judgment, and ethical sensitivity, empowering individuals to engage thoughtfully with AI and make responsible choices across education, work and everyday life (Chiu, 2025).

For educators, developing AI literacy has become essential as they navigate the integration of these technologies into their teaching practices. As Zawacki-Richter et al. (2019) note, teachers who possess adequate AI literacy are better positioned to make informed pedagogical choices and guide their students in an increasingly technology-driven world. However, research by Bewersdorff et al. (2023) indicates that learners often display limited technical understanding of AI, evidenced by confusion over basic AI definitions and concepts.

This knowledge gap is particularly concerning in teacher education programs, where future educators are being prepared to teach in classrooms that will increasingly incorporate AI technologies. Research by Zhao et al. (2022) suggests that structured training programs could be an effective way to promote AI literacy among educators, enabling them to better implement technological advancements in educational frameworks. Lee and Perret (2022) found that professional development training on technology use increased teachers' confidence in teaching AI.

Despite potential benefits of AI integration in education, including personalized feedback, tailored instructional materials, and enhanced data analysis capabilities (Chounta et al., 2022), concerns remain regarding accessibility, equity, and the preservation of critical human elements in the educational process (Padma and Rama, 2022; UNESCO, 2021).

Given these challenges, there is a clear need for targeted educational interventions to support pre-service teachers in developing AI literacy. This study investigates the design and iterative refinement of a concise, two-session AI intervention for pre-service physics teachers, developed within a Design-Based Research framework to address the practical challenge of integrating AI literacy into teacher education. Within the educational intervention, students engage in tasks that involve generative AI tools as part of activities to critically analyze and apply AI in teaching contexts.

The study is guided by two research questions:

How does the iterative design of a two-session AI intervention support the development of AI literacy among pre-service physics teachers? (RQ1).

What are pre-service teachers' attitudes towards AI in education and how does the intervention influence them? (RQ2).

2 Theoretical background

2.1 Design based research (DBR)

DBR serves as a coherent methodology that links theoretical inquiry with educational practice by examining intervention designs (The Design-Based Research Collective, 2003). This approach aims to develop effective, context-sensitive educational interventions through an iterative process that relies on continuous feedback to improve each successive iteration (van Zyl and Karsten, 2022). The methodology unfolds in systematic phases, progressing from initial problem analysis to prototype development, testing, and evaluation, with ongoing refinement of designs based on empirical evidence (van Zyl and Karsten, 2022). DBR involves designing educational interventions and generating insights about their effectiveness while at the same time promoting positive change in learning environments (Minichiello and Caldwell, 2021).

DBR's applicability is particularly evident in technological interventions (Anderson and Shattuck, 2012), where it generates innovative practices and principles that can be adapted to diverse contexts (The Design-Based Research Collective, 2003). DBR provides a framework for both reporting the problem and its background as well as presenting and evaluating a tested solution (van Zyl and Karsten, 2022). On the other hand, using DBR to develop a concise intervention can take a long time and participants often differ from each iteration to the next (van Zyl and Karsten, 2022).

2.2 The BSCS instructional model

The intervention was designed after the 5E instructional model, which is grounded in constructivist learning theory and consists of the five phases Engage, Explore, Explain, Elaborate, and Evaluate (Bybee, 2009; Bybee et al., 2006; Duran and Duran, 2004). This model provides a structured yet flexible framework (Duran and Duran, 2004) for consolidating knowledge and fostering scientific interest (Bybee et al., 2006). The 5E-framework has been shown to increase teacher confidence (Duran and Duran, 2004) as well but also presents challenges regarding the availability of suitable activities (Namdar and Kucuk, 2018) and time for implementation (Turan, 2021).

Building on this foundation, we adopted an extended version of the model, the 6E-framework, which adds a final Exchange phase to emphasize collaborative reflection. This extension was originally developed and discussed in our earlier work on teacher training in robotics, coding, and artificial intelligence (Henze et al., 2022; Figure 1).

2.3 Vygotskian principles in educational design

The psychologist Lev Vygotsky conceptualized development as unfolding on two levels. The first represents what an individual learner can accomplish independently, while the second encompasses what the learner can achieve with guidance or support (Leong and Bodrova, 1996). The gap between these two levels is referred to as the *Zone of Proximal Development (ZPD)*, which indicates the range within which learning and development can occur (Pedapati, 2022). This gap bridged through scaffolding provided by teachers or peers (Rigopouli et al., 2025).

In Vygotsky's view, tools are most effective for learning when they serve multiple purposes rather than being narrowly designed for a single concept (Leong and Bodrova, 1996), therefore AI with its manifold of possibilities could prove to be an effective tool for education. Within Vygotsky's ZPD, technology-assisted designs, using tools such as an AI chatbot, support learners and teachers in moving beyond their current possibilities (Rigopouli et al., 2025).

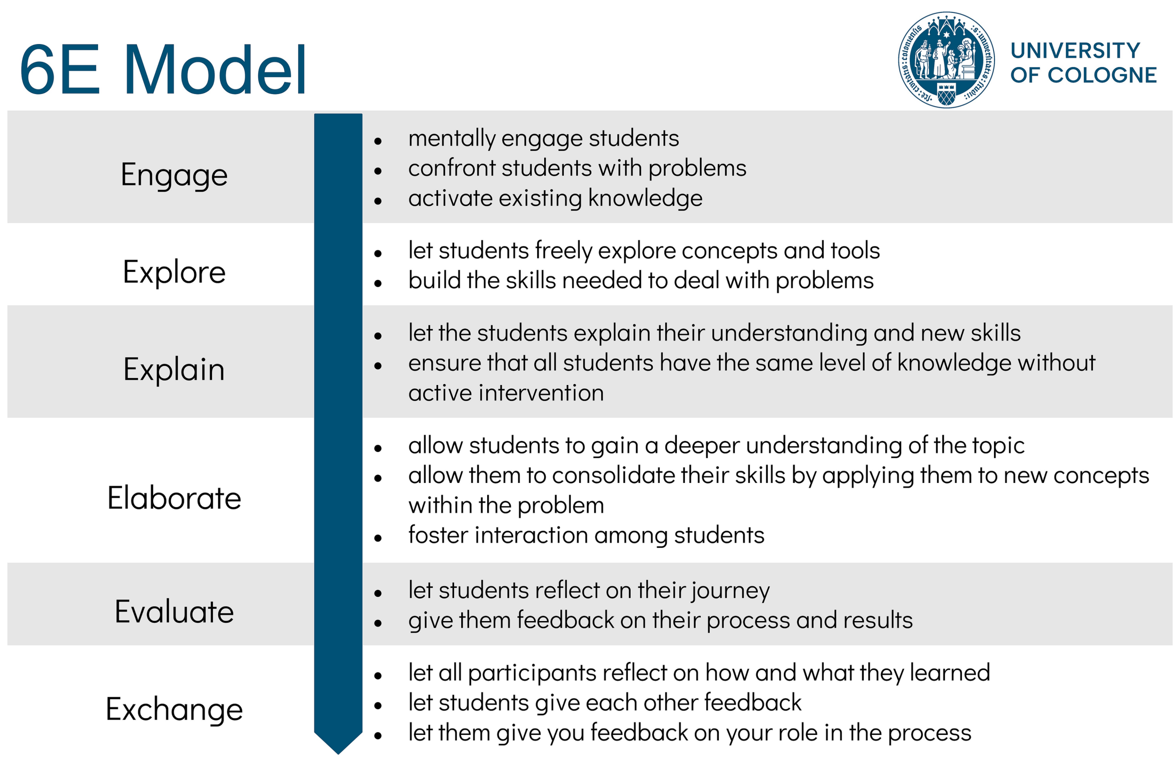

FIGURE 1
6E model phases and description. Based on the 5E Model from Bybee et al. (2006), extended by Henze et al. (2022).

2.4 State of AI in education

Recent literature points a steady rise in application of AI in educational context (Mahligawati et al., 2023). Chen et al. (2020) demonstrate a significant increase in research papers on AI in education from 2015 onward, while Roll and Wylie (2016) summarize a growing number of AI developments, refinements and research initiatives. Nevertheless, it is important to emphasize that the use of AI tools like ChatGPT in education is still at an early stage, highlighting the need for further empirical investigation (Farrokhnia et al., 2024).

While most research on AI in education originates from STEM fields and relies mostly on quantitative methods (Zawacki-Richter et al., 2019), recent studies have begun focusing on university contexts regarding ChatGPT-based interventions (Lo, 2023). AI adoption in schools currently happens through individual teachers rather than institutional policies (Vincent-Lancrin and van der Vlies, 2020).

2.5 AI literacy

According to Long and Magerko (2020), AI literacy encompasses the competencies enabling effective engagement with AI technologies both professionally and privately. For educators, developing AI literacy has become essential as they navigate the integration of these technologies into their teaching practices (Zawacki-Richter et al., 2019). Self-report surveys regarding AI literacy show positive relations between understanding AI, evaluating applications and ethical awareness (Zhao et al., 2022). Hornberger et al. (2023) developed a questionnaire based on Long and Magerko's (2020) competencies and objectively verifiable knowledge, finding a strong positive relationship between AI literacy and factors such as self-efficacy, interest, and positive attitude toward AI. Research also indicates that AI literacy positively correlates with educational background, suggesting higher levels of education predict greater AI literacy (Zhao et al., 2022). Despite these studies, research shows that learners often display limited technical understanding of AI, evident from confusion over basic AI definitions and concepts (Bewersdorff et al., 2023).

2.6 Teacher education for AI integration

Structured training programs represent a promising approach to promote AI literacy among educators, enabling them to use and implement technological advancements in educational frameworks (Zhao et al., 2022). Such trainings can also help teachers to better estimate their knowledge of AI, as educators who overestimate their knowledge might face difficulties due to misconceptions, while those who underestimate their capabilities may avoid using AI systems despite having adequate skills (Chounta et al., 2022).

Lee and Perret (2022) found that teachers across different backgrounds reported increased confidence in teaching AI after receiving professional training focused on technology utilization. Professional development in this area can enhance educators' abilities and strengthen students' learning experiences (Salas-Pilco et al., 2022).

Studies suggest that pre-service teachers can benefit from using ChatGPT in their professional development, showing improved understanding and higher academic achievement (Lo, 2023). However, students lacking knowledge of AI functionality tend to uncritically accept generated responses (Ding et al., 2023), highlighting the need to develop specific AI competencies within teacher education programs (Hornberger et al., 2023).

The integration of AI in educational contexts is viewed positively by teachers, who value it's potential for providing personalized feedback, creating tailored instructional materials, and analysing student data (Chounta et al., 2022). Generative AI chatbots demonstrate potential for education by supporting the explanation of complex concepts and problem-solving processes (Santos, 2023).

Students similarly find AI-driven tools appealing due to their high interactivity, motivating elements and opportunities for experimentation and simulation (Kuleto et al., 2021). In the context of special needs education, AI additionally offers opportunities for individualized learning (Padma and Rama, 2022).

2.7 Challenges and concerns

Despite these advantages, the incorporation of AI raises important concerns regarding accessibility and equity. Students with diverse barriers, be it disabilities or lack of access, may face significant challenges when using AI-supported tools (Varsik and Vosberg, 2024). Similarly, economically disadvantaged students may face challenges due to restricted access to essential technologies and support (UNESCO, 2021).

Critical reflection remains an important challenge, as research showed that students relying on ChatGPT performed worse in problem-solving tasks compared to peers who relied on search engines (Krupp et al., 2023). ChatGPT users predominantly copied questions and answers directly, indicating limited reflection and critical thinking (Krupp et al., 2023).

A representative study by the Vodafone Foundation (Vodafone Stiftung Deutschland gGmbH, 2023) in Germany found that while respondents generally anticipate significant educational changes through AI, a majority also viewed AI in schools as more of a risk that an opportunity, particularly regarding negative impacts on learning skills and creativity.

Concerns persist about excessive usage of technology, especially among younger learners who may develop dependency and addiction, negatively impacting their physical, emotional and psychological well-being (Padma and Rama, 2022). Additionally, AI currently cannot replace the critical human elements of personal interaction and meaningful teacher-student relationships, which remain integral to effective education (Padma and Rama, 2022). Furthermore, initial enthusiasm associated with AI integration might partly be explained by the novelty effect (Long et al., 2024).

3 Methodology

3.1 Study design

The study followed an experimental pre-post-design to analyze the changes between the pre- and post-test in two consecutive semesters with two different iterations of an intervention to improve AI literacy. In each iteration students completed a two-session intervention, each session lasting 90 min and separated by 1 week (see *Description of the intervention* for details). Pre- and post-testing allowed to determine whether this short treatment increased objectively measured AI literacy.

Because the attitudes survey was handled differently across the semesters, the design combines cross-sectional and longitudinal elements. In winter 2023/24 only a single survey (n = 14) was taken to establish baseline perceptions of AI after the intervention. In summer 2024 a survey was administered both before and after the intervention (n = 17), allowing attitude change scores parallel to the AI literacy measure. Subsequent sections describe the participant characteristics, instruments and statistical analyses in detail.

3.2 Participants

The study involved pre-service physics teachers enrolled in either the bachelor's or master's program of the University of Cologne. Participants represented typical stages of German secondary-school teacher education, which prepares educators to teach students aged approximately ten to eighteen. The pre-service physics teachers who took part in the intervention in winter 2023/2024 were either in the bachelor's program and in their third to fifth semester, or in the master's program in their first semester, the participants of the summer 2024 iteration were solely in the bachelor's program and in their third to fifth semester. Both cohorts were comparable in age distribution and gender composition, reflecting the typical demographic structure of the program. Detailed demographic characteristics of both cohorts are presented in Table 1.

3.3 Description of the intervention

The AI intervention was embedded in a mandatory module of the physics teacher education curriculum at the University of Cologne and carried out by members of the research team. This two-session intervention is designed to familiarize pre-service physics teachers with AI tools for classroom use. Following DBR, the intervention was iteratively implemented in two consecutive semesters (winter 2023/24 and summer 2024), with systematic variation in the second iteration due to external constraints. This shift was treated not as a limitation but as a design variation, providing insights into how different instructional formats (in-person vs. self-study) affect learning processes and outcomes.

The structure of both sessions was structured after the 6E instructional framework (Bybee, 2009; Henze et al., 2022), which extends the established 5E model by adding a final Exchange phase for collaborative reflection. The winter semester followed the full 6E cycle, while the summer iteration tested a shortened version without the Exchange phase, thereby allowing DBR-driven comparison of the effects of individual reflection versus collective negotiation.

Table 2 shows the phases and corresponding activities in detail. In the Engage phase, students discussed real-world AI examples and analyzed different AI-generated definitions all written by different generative AI tools—ChatGPT0F[1], Perplexity1F[2] and Raina2F[3],

1 https://chat.openai.com/ (registration required).
2 https://www.perplexity.ai/
3 https://app.magicschool.ai/raina (registration required).

TABLE 1 Sample description.

Characteristic		Winter 2023/2024		Summer 2024		Total	
n		14 (13 matched)		17		31 (30 matched)	
Gender (m/f)		7/7		10/7		17/14	
Age (years)	Mean (SD)	24.6 (7.6)		23.4 (4.6)		24 (6.1)	
	Median	22.5		22		22	
	Range	20–50		20–40		20–50	
		Pre	Post	Pre	Post	Pre	Post
Self-assessed AI knowledge	Mean (SD)	33.4 (23.2)	48.1 (26.3)	46.4 (18.6)	54.4 (24.5)	41.1 (21.7)	50.4 (24.7)
	Median	20	52	50	57	48	56.5
	Range	10–74	6–86	11–75	1–85	10–75	1–85

TABLE 2 Oversight of the interventions.

Session	Phase	Contents	
		Winter 23/24	Summer 24
1	Engage	• Discussion of real-world AI examples (10 min) • Comparison of AI definition texts (10 min)	
	Explore	• Hands-on image-classifier built with Teachable Machine (50 min)	
	Explain	• Interactive lecture on technological concepts (20 min)	
2	Elaborate	• Teams generate lesson plans, tests, and multiple-choice items on Ohm's Law with ChatGPT & Perplexity (30 min)	• Design of a 90-min Ohm's Law lesson (lesson plan, experiment, assessment) with AI; compare models • Usage of AI to generate literature lists on "Experiments in physics teaching"; verify entries online;
	Evaluate	• Critique of AI outputs (30 min)	• Short written reflection on accuracy of AI-generated literature lists • Reflection on applicability of designed lesson plans
	Exchange	• Whole-class "Chance vs. Risk" debate (30 min)	• not included due to self-study format

activating prior knowledge and surfacing misconceptions. In the Explore phase, students trained a simple classifier using Google Teachable Machine[4], a simple browser-based tool that can be trained in a very short time using simple image or pose classifiers, used as a multipurpose learning tool aligned with Vygotsky's notion that such tools foster deeper learning (Leong and Bodrova, 1996). The Explain phase was addressed through interactive input on technological foundations, clarifying key concepts and vocabulary.

The Elaborate and Evaluate phases provided opportunities for applying and critically assessing AI tools in authentic teacher tasks. Students generated lesson materials on Ohm's Law with ChatGPT and Perplexity, annotated factual and pedagogical shortcomings, and reflected on AI's reliability by comparing generated literature lists with real sources. These activities aimed to test and push the limits of each individuals ZPD, as students worked at the boundary of their independent ability, supported by an AI chatbot as well as peer collaboration and scaffolding from the instructors (Rigopouli et al., 2025). The generated lesson plans and working materials were discussed by the students in the winter semester in plenary to illustrate the capabilities and limitations of AI. The purpose was to practice critical evaluation, not the creation of ready-to-teach materials. In the summer semester the evaluation and reflection were made by each student individually based on theoretical knowledge.

Finally, in the Exchange phase of the winter semester, students engaged in a structured debate, deliberately swapping positions to broaden perspectives. This activity embodied Vygotsky's emphasis on social negotiation (Leong and Bodrova, 1996). In contrast, the summer iteration omitted this phase, relying instead on written reflections. From a DBR perspective, this allowed to examine whether individual evaluation could substitute for social discourse, thereby producing design insights about the non-substitutable role of collaborative Exchange in fostering critical engagement.

3.4 Instruments

To verify the effectiveness of the respective interventions in terms of increasing AI literacy and changing the attitudes of pre-service teachers toward AI, the research instruments described below were used.

3.4.1 AI literacy

The AI literacy test from Hornberger et al. (2023) serves as a multifaceted evaluation tool designed to measure an individual's understanding and knowledge of AI reliably across various dimensions with a total of 30 items in 14 areas of competence after Long and Magerko (2020). It assesses participants' abilities to recognize AI

4 https://teachablemachine.withgoogle.com/

10.3389/feduc.2025.1707534

applications, distinguish between human and AI interactions, and comprehend the utilization of AI systems in daily life and specialized fields. The test based on objectively verifiable knowledge and measures best at an average level of AI literacy rather than on a very high or low level (Hornberger et al., 2023).

3.4.2 Emotional-motivational variables

The emotional-motivational dimension of the study was assessed using self-designed self-report questionnaires focused on attitudes toward AI in education. The structure and administration of the survey differed between the two semesters. The attitude questionnaires were specifically designed for this study. Its items reflect key themes, such as teachers competencies, perceived opportunities and risks, and ethical aspects.

In the winter semester 2023/2024, an eight-item cross-sectional survey was administered once to capture a baseline snapshot of students' attitudes toward AI in educational contexts. The statements addressed key themes such as the importance of AI-related teacher competencies, perceived risks and opportunities, and openness to integrating AI in classroom practice. Responses were given on a four-point Likert scale ranging from 0 (Do not agree at all) to 3 (Fully agree). To test whether item scores deviated significantly from a neutral position, values were compared against a midpoint reference value of 1.5. As this was a one-time measurement, the results serve as a contextual reference only and do not allow for measuring change over time.

Although the attitudinal questionnaire used in this study was self-designed and not based on an existing standardized instrument, its focus on pre-service teachers' perceptions of AI aligns conceptually with recent work by Ishmuradova et al. (2025). Their study developed and validated a four-factor scale capturing attitudes toward the use of generative AI in science education.

The structural change in the test methodology in the summer semester aimed to evaluate the effectiveness of the intervention more precisely and to gain deeper insights into potential changes in students' attitudes. Therefore, the questionnaire was expanded, including all items from the winter semester, and administered in a pre–post design to assess changes in attitudes over the course of the intervention. The revised version included 20 statements and the same four-point Likert scale (0–3) was used.

3.5 Analysis procedure

All data were collected via a digital survey-tool. Completion time was approximately 14 min per survey. The data analysis was carried out in three steps. First, descriptive statistics were calculated separately for each group at pre- and post-test to obtain an overall performance profile. Statistical analysis was performed at the overall test level and the competency level after Hornberger et al. (2023). The pre–post difference scores for each participant and the overall-level were subjected to a Shapiro–Wilk normality test. The resulting *p*-value is reported as p_{sw}. If the normality assumption held, the change was evaluated with a paired-samples t-test, whose p-value is denoted p_t. If normality was violated, the Wilcoxon signed-rank test was applied, reported as p_w.

4 Results

4.1 Results RQ1

4.1.1 Iteration-wise descriptive summary

Across both iterations, overall AI literacy scores showed small upward shifts but no substantial changes. In the winter semester, scores ranged from 8 to 24 at pre-test and 8 to 22 at post-test, with the mean increasing slightly from 13.92 (SD 4.68) to 14.54 (SD 3.93) and the median rising from 13 to 14.

In the summer semester, the range broadened from 5 to 21 at pre-test and 7 to 25 at post-test with the mean showing a modest

TABLE 3 Overview of the descriptive data for both semesters on competency level.

Competency summarizing items	Item Code	Winter		Summer	
		Pre	Post	Pre	Post
		The data is presented in the following form: Mean (SD), Median			
C01 Recognizing AI	Q1, Q2	0.62 (0.65), 1	0.54 (0.66), 0	0.71 (0.59), 1	0.53 (0.51), 1
C02 Understanding Intelligence	Q5, Q6	1.54 (0.66), 2	1.85 (0.38), 2	1.41 (0.71), 2	1.18 (0.88), 1
C03 Interdisciplinarity	Q3, Q4	0.54 (0.78), 0	0.62 (0.77), 0	0.59 (0.80), 0	1.00 (0.94), 1
C04 General vs. Narrow	Q8, Q9	0.85 (0.8), 1	1.23 (0.44), 1	0.71 (0.69), 1	0.88 (0.70), 1
C05 AI's Strengths & Weaknesses	Q10, Q11	0.77 (0.83), 1	1.08 (0.76), 1	0.76 (0.66), 1	0.76 (0.75), 1
C07 Representations	Q12, Q13	0.46 (0.52), 0	0.54 (0.66), 0	0.47 (0.51), 0	0.41 (0.62), 0
C08 Decision-Making	Q14, Q15, Q16	0.92 (0.76), 1	0.92 (0.86), 1	0.65 (0.61), 1	0.65 (0.79), 0
C09 ML-Steps	Q17, Q18, Q19	0.92 (1.12), 1	1.00 (0.91), 1	0.71 (0.92), 0	1.12 (0.93), 1
C10 Human Role in AI	Q20, Q21	0.69 (0.63), 1	0.54 (0.52), 1	0.47 (0.72), 0	0.53 (0.72), 0
C11 Data Literacy	Q23	0.54 (0.52), 1	0.62 (0.51), 1	0.59 (0.51), 1	0.65 (0.49), 1
C12 Learning from Data	Q24, Q25	1.38 (0.77), 2	1.08 (0.76), 1	0.71 (0.69), 1	1.24 (0.66), 1
C13 Critically Interpreting Data	Q26	1.00 (0.00), 1	1.00 (0.00), 1	0.76 (0.44), 1	0.76 (0.44), 1
C16 Ethics	Q27–Q31	2.69 (1.03), 3	2.54 (1.05), 2	2.35 (1.32), 2	2.12 (1.22), 2
C17 Programmability	Q22	1.00 (0.00), 1	1.00 (0.00), 1	0.76 (0.44), 1	0.88 (0.33), 1

increase from 11.65 (SD 4.46) to 12.71 (SD 5.52) but the median remaining unchanged at 12. Variability in the data remained high in both semesters.

Table 3 summarizes for each of the 16 AI literacy competencies (C01–C17), the pre- and post-test means, medians, and standard deviations in both iterations. The table also shows which items belong to which competency. In the winter semester, competency means ranged between 0.54 and 2.69 (SD 0.00–1.03) at pre-test and 0.54 to 2.54 (SD 0.00–1.05) at post-test. In the summer semester, the corresponding ranges were 0.47 to 2.35 (SD 0.00–1.32) and 0.41 to 2.12 (SD 0.33–1.22).

Some competencies demonstrated stability across both semesters, such as *C13 Critically Interpreting Data* (constant at a mean of 1.0 in winter and 0.76 in summer) and *C17 Programmability* with stable medians and only little variance. Others showed mixed patterns: *C02 Understanding Intelligence* increased in winter from mean 1.54 to 1.85 but declined in summer from 1.41 to 1.18, while *C12 Learning from Data* decreased in winter from mean 1.38 to 1.08 but improved in summer from 0.71 to 1.24. Ethical considerations (*C16*) remained comparatively high but showed slight decreases from pre- to post-test in both semesters.

Table 4 presents the item-level overview (Q1–Q31). In the winter semester the pre-test means ranged from 0.08 to 1.00 (SD 0.00–0.52) and the post-test means also ranged from 0.08 to 1.00 (SD 0.00–0.5). In the summer semester the pre-test means ranged from 0.12 to 0.82 (SD 0.33–0.51) while the post-test means ranged from 0.06 to 0.88 (SD 0.24–0.51).

TABLE 4 Overview of the descriptive data for both semesters on question level.

Question	Winter		Summer	
	Pre	Post	Pre	Post
	The data is presented in the following form: Mean (SD), Median			
Q1	0.23 (0.44), 0	0.15 (0.38), 0	0.35 (0.49), 0	0.12 (0.33), 0
Q2	0.38 (0.51), 0	0.38 (0.51), 0	0.35 (0.49), 0	0.41 (0.51), 0
Q3	0.23 (0.44), 0	0.38 (0.51), 0	0.35 (0.49), 0	0.53 (0.51), 1
Q4	0.31 (0.48), 0	0.23 (0.44), 0	0.23 (0.44), 0	0.47 (0.51), 0
Q5	0.69 (0.48), 1	1.00 (0.00), 1	0.82 (0.39), 1	0.65 (0.49), 1
Q6	0.85 (0.38), 1	0.85 (0.38), 1	0.59 (0.51), 1	0.53 (0.51), 1
Q8	0.54 (0.52), 1	0.77 (0.44), 1	0.35 (0.49), 0	0.41 (0.51), 0
Q9	0.31 (0.48), 0	0.46 (0.52), 0	0.35 (0.49), 0	0.47 (0.51), 0
Q10	0.31 (0.48), 0	0.54 (0.52), 1	0.41 (0.51), 0	0.47 (0.51), 0
Q11	0.46 (0.52), 0	0.54 (0.52), 1	0.35 (0.49), 0	0.29 (0.47), 0
Q12	0.08 (0.28), 0	0.23 (0.44), 0	0.12 (0.33), 0	0.12 (0.33), 0
Q13	0.38 (0.51), 0	0.31 (0.48), 0	0.35 (0.49), 0	0.29 (0.47), 0
Q14	0.38 (0.51), 0	0.23 (0.44), 0	0.23 (0.44), 0	0.18 (0.39), 0
Q15	0.15 (0.38), 0	0.31 (0.48), 0	0.29 (0.47), 0	0.12 (0.33), 0
Q16	0.38 (0.51), 0	0.38 (0.51), 0	0.12 (0.33), 0	0.35 (0.49), 0
Q17	0.38 (0.48), 0	0.08 (0.28), 0	0.12 (0.33), 0	0.12 (0.33), 0
Q18	0.38 (0.51), 0	0.54 (0.52), 1	0.35 (0.49), 0	0.53 (0.51), 1
Q19	0.23 (0.44), 0	0.38 (0.51), 0	0.23 (0.44), 0	0.47 (0.51), 0
Q20	0.46 (0.52), 0	0.31 (0.48), 0	0.18 (0.39), 0	0.35 (0.49), 0
Q21	0.23 (0.44), 0	0.23 (0.44), 0	0.29 (0.49), 0	0.18 (0.39), 0
Q22	1.00 (0.00), 1	1.00 (0.00), 1	0.76 (0.44), 1	0.88 (0.33), 1
Q23	0.54 (0.52), 1	0.62 (0.51), 1	0.59 (0.51), 1	0.65 (0.49), 1
Q24	0.85 (0.38), 1	0.62 (0.51), 1	0.47 (0.51), 0	0.71 (0.47), 1
Q25	0.54 (0.52), 1	0.46 (0.52), 1	0.23 (0.44), 0	0.53 (0.51), 0
Q26	1.00 (0.00), 1	1.00 (0.00), 1	0.76 (0.44), 1	0.76 (0.44), 1
Q27	0.38 (0.51), 0	0.15 (0.38), 0	0.12 (0.33), 0	0.06 (0.24), 0
Q28	0.46 (0.52), 0	0.31 (0.48), 0	0.65 (0.49), 1	0.65 (0.49), 1
Q29	0.92 (0.28), 1	0.92 (0.28), 1	0.82 (0.39), 1	0.71 (0.47), 1
Q30	0.23 (0.44), 0	0.62 (0.51), 1	0.47 (0.51), 0	0.47 (0.51), 0
Q31	0.69 (0.48), 1	0.54 (0.52), 1	0.29 (0.47), 0	0.24 (0.44), 0

Robust performance was observed for items such as Q5 (Intelligence of AI), which remained relatively high across both semesters (winter: mean from 0.69 to 1.00; summer mean from 0.82 to 0.65, median stable). Similarly, Q22 (programmability) showed consistent high values in both cohorts. By contrast, items tapping into *data literacy and decision-making* posed greater challenges: Q12 (knowledge representation) stayed very low in both semesters (mean below 0.23), and Q27 (ethical principles) even declined in summer from a mean of 0.12 to 0.06. Isolated improvements occurred, for example in Q3 (AI systems), which rose slightly in both semesters, and Q30 (risks of AI), which showed a larger pre–post gain in the winter cohort.

4.1.2 Statistical analysis

Inferential tests were conducted to examine whether the observed descriptive tendencies for each of the iterations reached statistical significance. The Shapiro–Wilk tests on the pre–post difference scores confirmed an approximately normal distribution of overall change in both cohorts ($p = 0.491$ in the winter and $p = 0.592$ in the summer). Accordingly, paired t-tests were used to evaluate change in total AI literacy score. In both semesters, the mean increase was not statistically significant.

The Shapiro–Wilk tests on competency-level difference scores revealed that only C08 Decision-Making ($p = 0.195$) met the normality assumption in the summer cohort, all other competencies violated normality. Therefore, a paired t-test was applied for C08 and Wilcoxon signed-rank tests for the remaining competencies in each semester.

The statistical analyses shown in Table 5 demonstrate that neither iteration produced significant pre–post gains in total AI literacy scores. In the winter cycle, the overall mean increased slightly but failed to reach statistical significance and a similar pattern emerged in the summer cycle. On the competency level, no significant changes were observed, although individual competencies such as understanding intelligence (C02) and general vs. narrow AI (C04) in the winter semester as well as interdisciplinarity (C03) and learning from data (C12) in the summer semester approached significance in one of the two cycles. At the item level, only Q30 in the winter iteration showed a significant change ($p = 0.037$).

The item-level analysis shows no normality and no significant changes between pre- and post-test responses (see Table 5). All comparisons, except for Q30 in the winter semester ($p_w = 0.037$), fail to reach statistical significance, indicating that the intervention did not produce notable shifts in participants' answers.

4.2 Attitudes & motivation (RQ2)

4.2.1 Winter semester

Table 6 presents the item-level descriptive statistics for the eight attitudinal statements. Means ranged from 0.94 to 2.35, medians from 1.0 to 2.0, and standard deviations from 0.44 to 0.90 on the 0–3 agreement scale. The observed values spanned the full range on most items.

Figure 2 shows the distribution of responses to the attitude statements administered in the winter semester. Most participants agree that teachers should have AI skills (over 90% agree or fully agree), whereas opinions on automated grading (MW03) are divided, with about half disagreeing. Views on AI replacing teachers (MW07)

TABLE 5 Results of the Wilcoxon test p_w and *t*-test p_t on the total level, competency wide and for each individual question.

Level	Winter p_w	Summer p_w
Total (*t*-test)	$p_t = 0.27461$	$p_t = 0.29926$
C01	0.85011	0.39295
C02	0.07186	0.12943
C03	0.77283	0.06498
C04	0.0726	0.23304
C05	0.12943	1
C07	0.77283	0.77681
C08	1	$p_t = 1$
C09	0.78972	0.06498
C10	0.42371	0.77681
C11	0.77283	0.77283
C12	0.24017	0.07758
C13	-	1
C16	0.59407	0.34049
C17	-	0.48402
Q1	0.76559	0.12943
Q2	1.00000	0.76559
Q3	0.34578	0.23304
Q4	1.00000	0.07186
Q5	0.07186	0.23304
Q6	No Variation	1.00000
Q8	0.14891	1.00000
Q9	0.42371	0.42371
Q10	0.23304	0.77283
Q11	1.00000	0.76559
Q12	0.34578	1.00000
Q13	0.76559	0.76559
Q14	0.34578	0.76559
Q15	0.42371	0.14891
Q16	1.00000	0.07186
Q17	0.14891	No Variation
Q18	0.34578	0.23304
Q19	0.42371	0.12943
Q20	0.34578	0.14891
Q21	1.00000	0.42371
Q22	No Variation	0.48402
Q23	0.77283	0.77283
Q24	0.23304	0.12943
Q25	0.77283	0.07260
Q26	No Variation	1.00000
Q27	0.23304	0.77283
Q28	0.34578	1.00000
Q29	1.00000	0.42371
Q30	0.03689+	No Variation
Q31	0.42371	0.76559

Significant results are marked with +.

tend strongly toward disagreement, and most reject the notion that AI in schools is more threat than opportunity (MW08).

All eight items violated normality, so the non-parametric Wilcoxon test was used throughout. Table 6 shows the results of the Wilcoxon signed-rank tests against the neutral midpoint. The three noticeable statements mentioned above, MW03, MW07 and MW08 displayed significant deviations from neutrality.

4.2.2 Summer semester

The analysis examined changes in the attitudes and knowledge of the participants regarding AI in the educational field. Participants' responses show slight shifts in their perceptions and usage of AI from pre- to post-test (see Figure 3). Some items, like regular use of AI in studies (MS01), frequency of AI use in daily life (MS02), and voluntary engagement with AI (MS03) showed small increases in mean scores and either stable or slightly improved medians and standard deviations (see Table 7).

TABLE 6 Descriptive statistics for winter-semester attitudes and results of the Wilcoxon test p_w.

Item	Mean	Median	SD	p_w
MW01	2.35	2.00	0.47	0.239
MW02	2.29	2.00	0.66	0.287
MW03	1.59	1.00	0.71	0.0014+
MW04	1.06	1.00	0.44	0.0624
MW05	1.00	1.00	0.71	0.0782
MW06	2.41	2.00	0.51	0.117
MW07	2.24	2.00	0.66	0.0012+
MW08	1.06	1.00	0.90	0.000827+

Significant results are marked with +.

Shapiro–Wilk tests on each item's pre-post change score showed that none of the motivational change scores met the normality assumption. Accordingly, all item-level pre–post comparisons were conducted with Wilcoxon signed-rank tests. Table 7 shows certain areas, like the belief that AI will significantly change everyday teaching (MS04) or the notion that AI in schools is more of a threat than an opportunity (MS05), displayed minimal fluctuations in means and varying adjustments in standard deviations, without major directional changes. A few items, such as the importance of teacher training on AI (MS11, MS16) or integrating AI into everyday teaching (MS17), showed small but not significant positive shifts in mean and median values (see Table 7). Table 7 further displays significant changes only occurred regarding the feeling of being prepared to integrate AI into their own teaching (MS19) and using AI regularly in their everyday life (MS02).

5 Discussion

5.1 AI literacy (RQ1)

The winter and summer iterations represent two successive cycles, each providing insights into the stability and adaptability of the intervention design under varying conditions. These conditions are namely a full 6E circle after Henze et al. (2022) in the first iteration in winter 2023/2024 or a 5E circle with self-study elements after Bybee et al. (2006).

At the competency level, covering all AI literacy competencies by Hornberger et al. (2023) after Long and Magerko (2020), the pre-post comparisons yielded no statistically significant changes in either iteration. Nevertheless, the overall trend and consistency of these changes provide insights into how effective the intervention was and how the instructional design contributed to its impact.

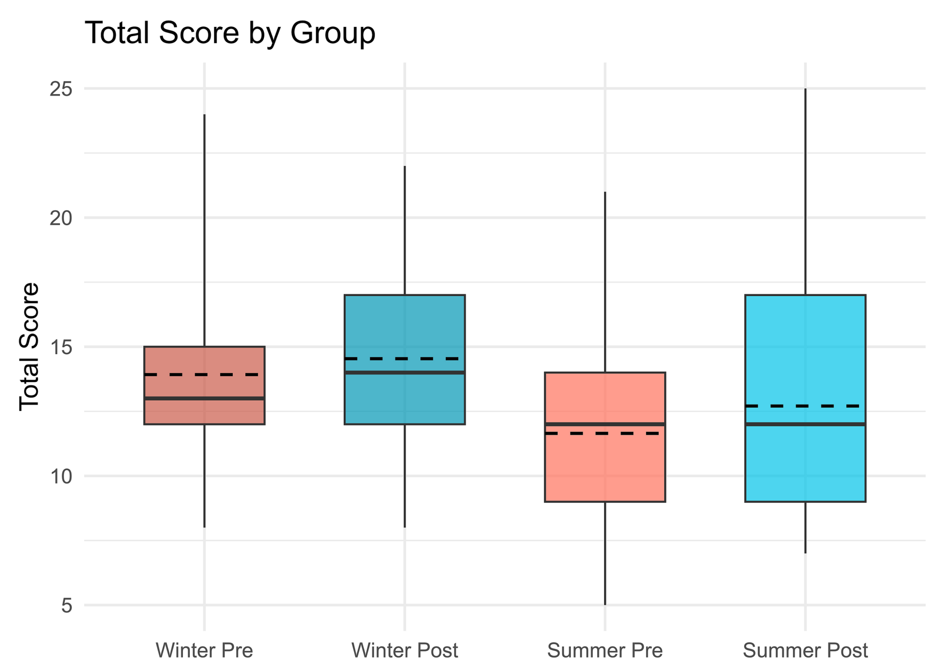

FIGURE 2
Comparison of total scores in the winter and summer group, the dashed line shows the mean and the solid line the median.

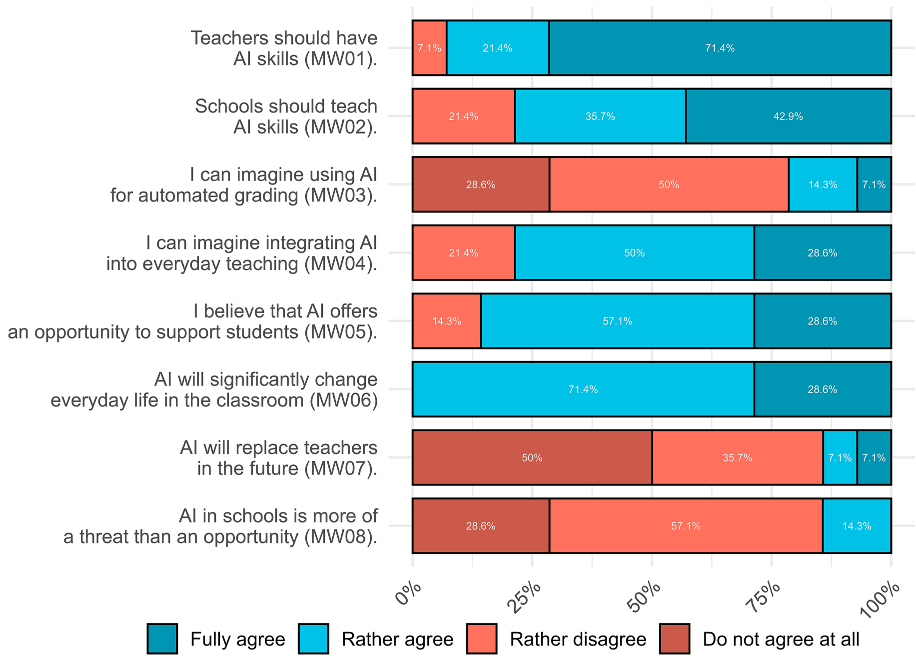

FIGURE 3
Participants perspectives on AI in education, indicating levels of agreement across various statements.

TABLE 7 Descriptive statistics and results of the Normality-test p^{sw} and the Wilcoxon test p_w.

Statement	Mean		SD		Median		Wilcoxon p_w
	Pre	Post	Pre	Post	Pre	Post	
MS01	1.88	2.06	1.05	0.83	2	2	0.233
MS02	1.24	1.59	0.90	0.71	1	2	0.048+
MS03	0.71	1.00	0.77	0.71	1	1	0.110
MS04	2.35	2.25	0.49	0.68	2	2	0.484
MS05	1.24	0.88	0.56	0.49	1	1	0.095
MS06	2.29	2.29	0.47	0.59	2	2	1.000
MS07	2.47	2.47	0.51	0.62	2	3	1.000
MS08	2.82	2.76	0.39	0.44	3	3	0.766
MS09	0.29	0.29	0.47	0.47	0	0	1.000
MS10	2.12	2.24	0.70	0.75	2	2	0.572
MS11	1.94	2.00	0.66	0.71	2	2	0.821
MS12	2.18	2.35	0.53	0.49	2	2	0.345
MS13	1.59	1.81	0.71	0.40	2	2	0.129
MS14	1.06	0.71	0.44	0.59	1	1	0.089
MS15	1.00	0.94	0.71	0.66	1	1	0.850
MS16	2.41	2.47	0.51	0.62	2	3	0.766
MS17	2.24	2.53	0.66	0.51	2	3	0.089
MS18	1.06	1.06	0.90	0.90	1	1	1.000
MS19	0.76	1.65	0.75	0.61	1	2	0.004+
MS20	1.06	1.29	0.77	0.85	1	1	0.182

Significant results are marked with +.

The mean scores for the competencies Recognizing AI (C01) and Ethics (C16) declined slightly in both iterations, likely reflecting cognitive recalibration as participants reassessed assumptions after comparing AI-generated definitions and encountering complexities

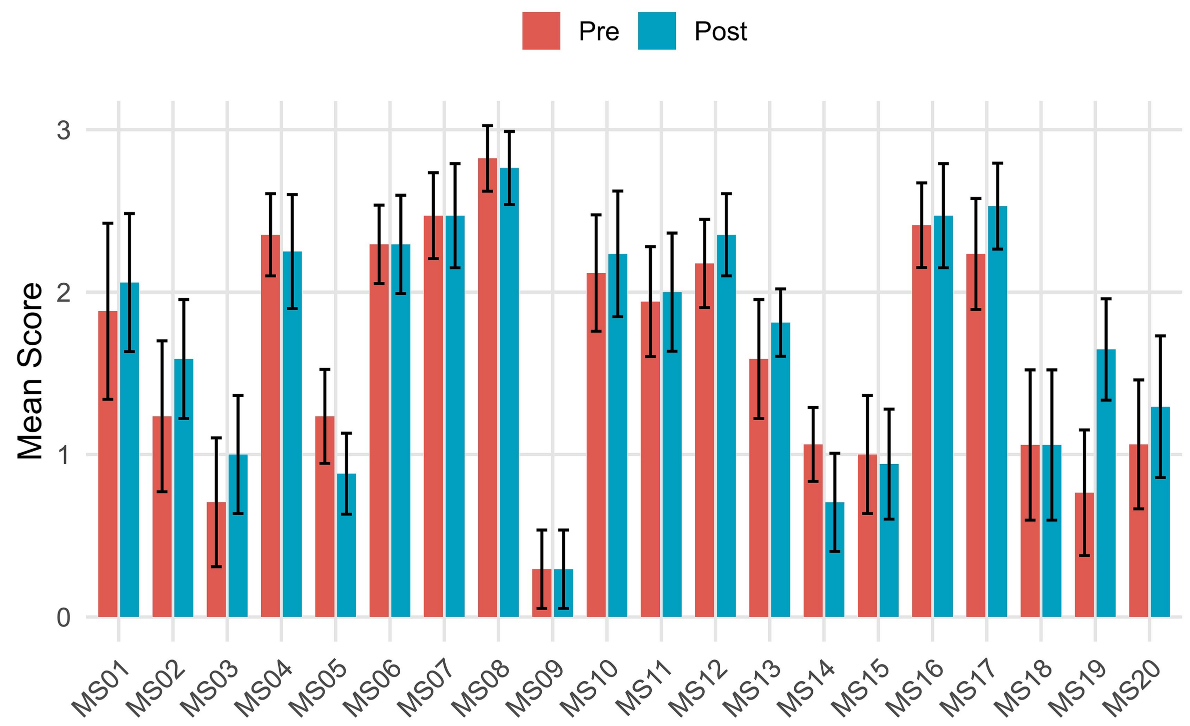

FIGURE 4
Pre-post comparison of means with 95%-confidence interval as error bars.

like hallucinated outputs. This aligns with findings by Bewersdorff et al. (2023), who observed confusion among learners over basic AI concepts, and Zhao et al. (2022), who argue that understanding AI's ethical implications requires more than surface-level familiarity. These topics require abstraction and comparison, which benefit from social interaction. Without sufficient scaffolding, learners may (falsely) question prior knowledge, leading to declining values. In winter, debate and collective reflection in the Exchange-Phase appeared to slightly stabilize understanding, whereas in summer, the absence of exchange limited reflection to the individual level may have fostered uncertainty.

The competencies Decision-Making (C08) and Critically Interpreting Data (C13) remained unchanged, indicating competencies without explicit instructional focus are unlikely to develop. Despite exercises with Teachable Machine, the transfer from simple, self-trained models to larger AI systems appeared unsuccessful or partial. This means that simply using basic tools like Teachable Machine is apparently not enough for the transfer to more complex AI systems. None of the phases of the 5E or 6E model addressed these skills directly and in depth in any of the iterations. The lack of specific tasks may have led to stagnation. Since skills do not develop automatically without being explicitly promoted, this highlights a specific area for improvement in the next iteration.

Other competencies showed divergent trends: Understanding Intelligence (C02) increased in winter but declined in summer, likely due to instructor-led scaffolding versus self-study format, even though the basic definitions and content regarding AI was transmitted in the same way. AI's Strengths & Weaknesses (C05) improved slightly in winter but remained unchanged in summer, again underscoring the value of peer dialogue. Representations (C07) increased slightly in winter but decreased in summer, suggesting that collaborative exploration in person supported deeper engagement than self-study. From a Vygotskian perspective, the absence of peer interaction and teacher scaffolding in summer limited movement beyond the zone of proximal development (Leong and Bodrova, 1996). These skills appear particularly dependent on the Exchange phase, and their decline in summer highlights the sensitivity of such competencies to format variations.

The Human Role in AI (C10) decreased in winter but increased in summer, with independent work possibly prompting deeper reflection on educational uses of AI. Learning from Data (C12) dropped in winter but rose in summer, suggesting that self-guided critique enhanced understanding despite limited instruction. Programmability (C17) remained unchanged in winter, likely due to a ceiling effect while increasing slightly in summer. Overall, these trends indicate that while the intervention activated selective knowledge, its effectiveness varied by content and format. The self-study setting appeared to support individual reflection on teachers' roles (C10) and data critique (C12), while the winter format emphasized collective error analysis. Thus, certain competencies (C10, C12) may benefit more from individual engagement.

Four competencies showed positive development across both iterations: Interdisciplinarity (C03) increased, especially in summer, as participants connected AI to educational contexts. The competencies General vs. Narrow (C04) and Machine-Learning-Steps (C09) improved in both groups likely due to specific instructional focus using concrete examples. Data Literacy (C11) showed slight gains in both groups, possibly from critically assessing AI-generated outputs. Participants were apparently able to build on prior knowledge and exceed their ZPD through scaffolding. Here, Explore (practical tools such as Teachable Machine), Explain (technical concepts) and Elaborate (lesson plan with AI) worked synergistically. Even in the summer, in the self-study format of the Elaborate phase, the tasks were sufficient to achieve progress. This stability indicates robust design elements such as contextually embedded learning effectively support foundational AI literacy regardless of format.

The only statistically significant item-level change occurred in winter for Q30 (ethical implications of predictive policing), suggesting themes intersecting with ethical, societal, or real-world controversies may resonate more strongly with learners, prompting them to reassess their stance more readily than with purely technical or conceptual content. Without Exchange phase in the summer, this effect was missing, so there was no change. The comparison of the iterations shows that ethical content requires collective discussion.

Several knowledge items exhibited either complete stability or no measurable variation across the two iterations (e.g., Q6, Q22, Q26 in winter and Q17, Q30 in summer; see Table 5). This can either be explained with ceiling effects, items such as Q22 and Q26 in the winter were already answered correctly by all participants in the pre-test. When prior knowledge is already high, even effective instruction produces limited statistical movement. Secondly, conceptual and ethical complexity may explain the results of items such as Q30 (predictive policing). In the winter iteration, this question showed significant improvement, while in the summer's self-study format, it showed no variation. This contrast underscores that socially mediated reasoning and collective discourse seem to have been critical for addressing morally or societally charged content. Without opportunities for peer negotiation and instructor scaffolding, participants may have lacked the dialogic space needed to compare ethical standpoints and reconstruct their understanding. From a design-based research perspective, such non-changes are informative because they expose where the intervention's reach was limited. Rather than representing a null result, they indicate the boundaries of current design effectiveness.

In winter, questions initially answered correctly by all participants (Q22, Q26) remained stable, while some low-baseline items (Q12, Q30) showed increased means, suggesting that topics previously less understood benefitted to a degree from the intervention. In summer, slight mean increases emerged in items that started lower (Q19, Q20, Q24), though improvements were not statistically significant. The summer's self-study format may have limited engagement compared to winter's face-to-face sessions. The fact that participants still displayed considerable variability in their responses indicates that learners internalized the material unevenly. While some individuals may have deepened their comprehension, others remained uncertain or unconvinced. Together, these findings illustrate that while certain competencies and items can be reinforced through individual reflection, more complex or abstract elements require structured scaffolding and collective engagement within the 6E framework to foster consistent development.

Although all differences between the two iterations are interpreted towards scaffolding and social negotiation, this remains theoretical. No qualitative process data were collected that could directly show how scaffolding or peer interaction affected the participants. However, the quantitative pattern of competency change developed only when peer discussion was available and is consistent with Vygotskian accounts of scaffolded learning. Therefore, the link between scaffolding and observed effects should be understood as a theory-based inference grounded in established learning principles rather than a directly measured causal mechanism.

These trends show short interventions can activate selective elements of AI literacy when grounded in hands-on experience and critical evaluation. This resonates with Long and Magerko (2020), who emphasize the importance of hands-on, reflective learning in developing AI literacy, while Zhao et al. (2022) identified a positive correlation between hands-on AI applications and an improved understanding of AI ethics.

However, as Viberg et al. (2023) noted, professional development should focus on enhancing fundamental AI knowledge rather than comprehensive technical methodologies. Therefore, short-term interventions may be insufficient for fostering comprehensive AI literacy, underscoring the need for repeated, context-rich engagements to address misconceptions, as Lin et al. (2022) suggested.

5.2 Attitudes and motivation (RQ2)

5.2.1 Winter semester

For the emotional-motivational measures, the winter semester survey showed statistically significant deviations from neutrality in three items, specifically regarding the use of AI for automated grading (MW03), concerns about AI replacing teachers (MW07), and perceptions of AI as more of a threat than opportunity (MW08).

Hesitancy toward applying AI directly for automated grading (MW03) indicates deeper concerns about trusting AI with important evaluative tasks, reflecting broader apprehensions about delegating core pedagogical decisions to AI and highlighting the need for ethical AI training that fosters trust while retaining teacher agency. The concern that AI might eventually replace teachers (MW07) highlights a slight apprehension that automation could undermine the role of human educators, but most of the students (85.7%) think AI will not replace teachers. Similarly, the view that AI in schools poses more of a threat than an opportunity (MW08) reveals the students do not fully agree with that. Together, these two results highlight that while participants accept the necessity for AI competencies and foresee positive changes, they draw a clear line at ceding core pedagogical functions to machines. The participants' perspectives underscore the importance of integrating ethical considerations and human-centered approaches into AI literacy programs, like Zhao et al. (2022) as well as Padma and Rama (2022) also proposed.

Even though not significantly different from the neutral point, the participants showed pronounced agreement that teachers should possess AI skills (MW01) and that schools should explicitly integrate AI into their curricula (MW02), although the intensity of that support and the exact nature of integrating AI into curricula may still be subject to individual interpretation, highlighting the importance of developing contextually relevant AI educational resources. This aligns with Zhao et al. (2022), who emphasize the critical role teacher preparedness plays in the effective adoption of AI technologies. The support for teacher training reflects a general awareness among future educators of AI's relevance in modern educational settings.

A moderate, however not statistically significant level of agreement that AI could be integrated into everyday teaching (MW04) complements the earlier findings. Although students generally view AI usage as favourable, the variability in the answers highlights that comfort with large-scale AI utilization may still be developing.

Similarly, the idea that AI offers an opportunity to support students (MW05) was met with agreement, again reflecting a positive perspective. The belief that AI will significantly change everyday classroom life (MW06) is widely accepted, but the absence of statistical significance suggests that, while change is anticipated, the participants are not consistently convinced about the scale of this transformation. This indicates that, although they acknowledge AI's potential impact, they remain cautious about predicting the range and speed at which these changes will happen.

From a Vygotskian perspective, attitudes toward automated grading (MW03) may lay outside participants' ZPD, as the lack of trust in AI required intensive scaffolding and collective negotiation. Within the 5E/6E framework the absence of the Exchange phase in summer therefore likely limited opportunities to address such fears, whereas in winter, the available discussion space may have supported collective processing of concerns. Within a DBR perspective, this iteration underscores that sensitive issues such as automated assessment must be deliberately addressed in the design of interventions.

Concerns about AI replacing teachers (MW07) and the perception of AI as a threat rather than an opportunity (MW08) similarly point to the value of peer interaction. Vygotskian theory suggests that social contexts helped participants learn through dialogue (Leong and Bodrova, 1996). The Exchange phase in winter appears to have provided conditions for perspective-taking, whereas in summer, the lack of such peer dialogue may have reduced opportunities to resolve cognitive conflict. From a DBR standpoint, these results argue for the inclusion of structured debate elements in future iterations to ensure controversial topics are productively explored.

By contrast, the high baseline agreement that teachers should acquire AI skills (MW01) and that AI should be integrated into curricula (MW02) indicates that these items already fell within the learners' ZPD. Here, the Engage and Explain phases were sufficient to reinforce existing beliefs, without requiring additional elaboration or exchange. Therefore, such stable convictions can be treated as anchors, while future interventions should focus on more fragile or contested areas of belief and practice.

5.2.2 Summer semester

The results from the summer semester group show that even a relatively brief intervention can influence participants' perceptions, knowledge, and confidence in engaging with AI within educational settings, albeit to a small extent. The pre-post comparison revealed statistically significant shifts in two items. Students reported feeling significantly more prepared to integrate AI into their teaching practices (MS19) and demonstrated an increased use of AI in daily life (MS02) (see Table 7). These findings indicate the intervention's effect on perceived pedagogical readiness and practical AI engagement. The MS02 increase suggests learners began viewing AI as a tool integrated into everyday routines or recognizing AI more often, while MS19 captures increased professional confidence. These shifts demonstrate that task-based exposure to AI systems can bridge the gap between conceptual understanding and applied confidence, resonating with the emphasis of Zhao et al. (2022) on direct interaction to foster self-efficacy. These improvements occurred despite the self-study format, suggesting autonomous exploration with authentic tasks can yield self-perceived competence gains. The self-reflective task structure encouraged participants to critically appraise AI for teaching needs, enhancing both usage behaviour and pedagogical preparedness, indicating even short, professionally relevant interventions can stimulate meaningful change in personal and instructional domains of AI readiness.

Other attitudinal items showed minor, non-significant variations (see Figure 4; Table 7). Slight gains in voluntarily educating oneself about AI (MS03) suggest rising intrinsic motivation as participants discover AI's relevance, with Zhao et al. (2022) noting motivation as key to improving AI literacy. The data shows agreement that schools should teach AI competencies (MS07), reflecting recognition of AI literacy's growing importance in curricula, as supported by Hornberger et al. (2023).

Beliefs about AI's impact on everyday teaching (MS04) slightly receded, indicating stronger anticipation of fundamental change, while items regarding teacher involvement in AI tool decisions (MS12) and AI fostering independent learning (MS13) showed small improvements. Chounta et al. (2022) emphasize educators' role as mediators in AI adoption. The stable mean for AI supporting students (MS06) and small increase in AI promoting classroom diversity and inclusion (MS10) highlight participants' recognition of AI's potential to address equity challenges, aligning with Lin et al. (2022). This stability reflects participants' sustained confidence in AI's supportive classroom role, with recent studies highlighting personalized learning pathways (Chounta et al., 2022) enhancing inclusivity (Zhao et al., 2022). While initial enthusiasm may decrease as participants gain more realistic understanding, dimensions like teacher agency and student autonomy gain traction as participants consider practical aspects of AI integration.

Results show a small decline in perceived risk of AI in schools (MS05), suggesting a gradual reduction of concerns rather than excitement. The stable but very low perception of AI replacing teachers (MS09) contrasts with participants' readiness to accept teacher training (MS11, MS16) and teacher involvement in AI decisions (MS12). The strong agreement that teachers should possess AI competencies (MS08) underscores the necessity for educators to navigate AI integration, with Zhao et al. (2022) highlighting that equipping teachers with AI skills is critical.

Items showing small increases in mean values, such as prioritizing teacher training (MS16) and integrating AI into everyday teaching (MS17), also saw improved medians and reduced variability, suggesting participants are becoming more consistent in these positions. Participants favor a future where educators remain central figures who leverage AI to enhance instruction rather than give up control (Chounta et al., 2022; Lin et al., 2022).

While certain perceptions advanced, attitudes toward automated grading (MS18) and institutional support for AI use (MS20) showed minimal change at relatively low levels, suggesting resistance to short-term interventions. MS18 showing no changes suggests that ethical or high-stakes educational decisions are not easily reconsidered through brief, individual reflection, but seem to require collective negotiation and contextualized experience. Even when participants recognise AI's potential benefits, they tend to maintain clear boundaries regarding professional autonomy. Also, topics of items linked to institutional or policy-level constructs were not explicitly addressed in the learning activities and depend largely on perceived external structures. Chounta et al. (2022) similarly report teachers would prefer using AI to help with grading rather than fully automated grading. Views on transparency of data protection policies (MS14) remained low, while belief that parents should be involved in AI classroom decisions

(MS15) maintained high agreement levels with slight increases post-intervention.

From a Vygotskian perspective, significant gains in AI use in daily life (MS02) and preparedness for teaching (MS19) indicate that the intervention may have directly engaged the learners' ZPD: independent, authentic tasks seem to be able to foster a sense of competence. Within the 5E/6E model, the self-reflective Evaluate phase appears to have supported this development, demonstrating that even without peer exchange, individual reflection can drive perceived readiness. Smaller gains in voluntary engagement (MS03), teacher training (MS11, MS16), and integration into everyday teaching (MS17) underline that intrinsic motivation and professional awareness can be strengthened when learners generate their own teaching materials and apply concepts practically. Declines in expectations about AI fundamentally transforming teaching (MS04) suggest cognitive recalibration, as initial enthusiasm seems to be tempered by realistic engagement. By contrast, persistent low means for abstract or sensitive issues such as automated grading (MS18), data protection (MS14), and institutional support (MS20) indicate that without scaffolding and structured peer dialogue, these items remain difficult to shift. The absence of an Exchange phase in the summer group likely limited collective negotiation around such contested topics.

5.3 Broader implications

The iterative design of the two-session intervention provided selective but limited support for the development of AI literacy among pre-service physics teachers. Across both iterations, total literacy scores showed small, non-significant increases, indicating that short interventions can activate specific competencies but are insufficient for comprehensive progress. Competencies with direct instructional focus and hands-on activities displayed consistent gains in both iterations. By contrast, competencies without targeted scaffolding remained stable, while ethically or conceptually demanding areas sometimes declined, suggesting cognitive recalibration in the absence of sustained guidance.

Given the sample size of $n = 31$ from a single program and the design-based, two-iteration context, the results should be interpreted primarily as context-specific insights rather than universally generalizable causal effects. Effect estimates are exploratory and sensitive to sampling error and the value of the present study lies in the design principles it surfaces for this setting.

Nevertheless, the iterative DBR design highlighted the role of instructional format in this context. The winter semester, which included collective Exchange and in-person scaffolding, supported competencies that rely on social negotiation and teaching. In contrast, the summer self-study version promoted more individual reflection, with small gains in competencies such as learning from data or the human role in AI but failed to sustain progress in socially embedded competencies.

The 6E instructional structure provided an effective scaffold, particularly through Explore, Elaborate, and Evaluate, which enabled hands-on practice and critical engagement with AI tools. However, the absence of the Exchange phase in the summer iteration limited opportunities for collaborative reflection, diminishing development in ethical and conceptual dimensions. The DBR framework proved valuable for uncovering design-sensitive effects, such as gains in some areas regardless of format, but also format-dependent differences that emphasize the importance of social interaction, scaffolding, and iterative refinement.

The stable results observed across both iterations offer valuable insights for interpreting the boundaries of the current design. Rather than signaling ineffectiveness, the non-change outcomes highlight where learning may require greater depth, scaffolding and time. Ethical and policy-related constructs such as data protection, automated grading, or predictive policing appear particularly dependent on dialogic engagement and contextualized examples, suggesting that future iterations should extend opportunities for collaborative reasoning and adapt item difficulty to capture incremental growth.

Beyond the physics context, these findings also hold relevance for other subject areas in teacher education. The 6E framework, particularly the Exchange phase, can be adapted to foster AI literacy in disciplines such as mathematics, languages, or the social sciences. Collaborative reflection and peer negotiation could support subject-specific objectives, help evaluate AI-generated problem solutions, essays, or data visualizations, and address ethical and pedagogical challenges.

Similar patterns have been reported in other teacher-education contexts. For instance, early-childhood education students valued ChatGPT for lesson planning but highlighted the need to critically evaluate its outputs critically (Nikolopoulou, 2024). These parallels suggest that the 6E framework may serve as a transferable scaffold for supporting AI literacy development across disciplines beyond physics.

Ishmuradova et al. (2025) identified clusters ranging from enthusiastic to skeptical among pre-service science teachers when analyzing perceptions of generative AI. The present study's finding of simultaneous optimism and critical caution among physics pre-service teachers mirrors this broader tendency, suggesting that the ambivalence toward generative AI is not domain-specific but characteristic of teacher education.

Taken together, the findings indicate that iterative, design-based interventions grounded in the 6E model and Vygotskian principles can selectively foster AI literacy. However, to achieve broader and more sustained development, interventions require repeated engagements, explicit scaffolding for complex competencies, and dedicated space for collaborative reflection.

Regarding RQ2, pre-service teachers across both semesters demonstrated a consistent perspective that embraces AI as a pedagogical tool while establishing boundaries against delegating core teaching functions to automated systems. This manifested in resistance to automated grading despite general openness to AI integration,

The significant increase in participants' perceived readiness to integrate AI into teaching (MS19) alongside increased daily AI use (MS02) demonstrates that even brief interventions can bridge the gap between abstract knowledge and practical application confidence. This transformation represents a crucial step in preparing future teachers for AI-enhanced educational environments.

6 Implications and limitations

This study offers insights into the challenges and opportunities of integrating AI literacy into teacher training programs, while also highlighting several methodological and contextual limitations. First, the sample is small and drawn from a single teacher-education

10.3389/feduc.2025.1707534

program, which limits statistical power and external validity. Accordingly, findings should be treated as context-bound design principles rather than generalizable causal effects. Future studies should aim for larger, more diverse participant pools across multiple teacher education programs to enhance generalizability and statistical power.

The findings emphasize the importance of embedding AI literacy as a component of teacher education, with a focus on sustained and iterative learning. Although the lack of significant changes in most survey items suggests that the interventions had limited measurable impact, the single significant result in the winter semester indicates that targeted areas of AI literacy can be positively influenced through focused instructional efforts.

The short-term nature of the intervention points to the need for more sustained efforts. Future research should implement longitudinal programs spanning multiple sessions with structured follow-up activities. Researchers might consider establishing communities of practice where participants can continue developing their AI skills beyond formal interventions.

The study's limitations reveal areas for improvement in future research and practice. The lack of sensitivity in certain survey items, evidenced by no variability in several knowledge questions, suggests a need for revised and more targeted assessments. Furthermore, the reliance on quantitative methods alone may have restricted the ability to capture shifts in understanding or attitudes. Incorporating qualitative approaches could provide a more comprehensive picture of participants' experiences and the intervention's impact. Pre-existing knowledge and attitudes toward AI may have influenced the results, underscoring the importance of using pre-assessments to tailor interventions to initial competency levels. Future interventions could implement adaptive learning environments that respond to participants' knowledge and provide personalized learning experiences. Researchers should also consider controlling for technology self-efficacy and prior AI exposure as potential confounding variables in their analyses. Moreover, the proposed link between scaffolding and performance differences is inferred rather than directly observed, therefore subsequent DBR iterations should combine quantitative results with qualitative process data to substantiate this mechanism.

7 Conclusion

The study was conducted in two iterations, one in the winter semester 2023/2024 and one in the summer semester 2024. The iterative design of the two-session AI intervention provided selective but limited support for developing AI literacy among pre-service physics teachers. Across both iterations, total literacy scores showed only small, non-significant increases, suggesting that short interventions can activate specific competencies but are insufficient for comprehensive progress. Competencies with explicit instructional focus and hands-on activities displayed consistent gains, while those lacking targeted scaffolding remained stable, and ethically or conceptually demanding areas sometimes declined, indicating cognitive recalibration in the absence of sustained guidance.

Future designs should therefore strengthen scaffolding for complex competencies through structured peer discussions and systematically embed ethics. The 6E model proved effective through all original 5E phases, but the Exchange phase seems to be essential. Its absence in summer reduced socially negotiated learning, underscoring the need to secure space for collaborative reflection even in self-study contexts. Balancing peer dialogue with opportunities for individual reflection may best support diverse learning processes.

The DBR approach revealed format-sensitive effects, such as in-person scaffolding and exchange supported socially embedded competencies, while self-study fostered individual reflection on data use and the human role in AI. Large individual differences further suggest that differentiated tasks of varying complexity are needed to meet learners at different points in their ZPD. Iterative refinement should now test hybrid formats, extend hands-on engagement beyond simple models, and evaluate longitudinal effects across multiple sessions.

Taken together, these findings highlight that iterative, design-based interventions grounded in Vygotskian principles can selectively foster AI literacy. To achieve broader and more sustainable development, however, repeated engagements, explicit scaffolding, and collaborative reflection remain indispensable.

Future research should explore more extensive interventions that provide repeated, contextualized engagements with AI in educational settings. By enhancing AI literacy in pre-service teacher education through sustained, professionally relevant interventions, teacher education programs can better prepare future educators to navigate AI-enhanced educational environments with both critical awareness and practical competence.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

Ethics approval was not required for this study, as the data collected were anonymized, and the survey did not involve sensitive or health-related topics. Participants were not exposed to any risk or burden, and the information gathered was and is not personally identifiable. Additionally, the survey was conducted solely for research purposes related to general study conditions, without any impact on participants' academic assessment, rights, or well-being. The study was conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. JL: Conceptualization, Data curation, Investigation, Methodology, Writing – review & editing. SB-G: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. AB: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The ComeMINT project is funded by the Federal Ministry of Education and Research as part of the BMBF funding line 'Competence Centers for digital and digitally supported teaching in schools and continuing education in the STEM Field', funding code 01JA23M06G.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. This manuscript was linguistically developed and revised for better readability using OpenAI ChatGPT 4o, Anthropic Claude 3.7 Sonnet and the translation tool DeepL.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2025.1707534/full#supplementary-material

References

Anderson, T., and Shattuck, J. (2012). Design-based research. *Educ. Res.* 41, 16–25. doi: 10.3102/0013189X11428813

Bewersdorff, A., Zhai, X., Roberts, J., and Nerdel, C. (2023). Myths, mis- and preconceptions of artificial intelligence: a review of the literature. *Comp. Educ. Artif. Intellig.* 4:100143. doi: 10.1016/j.caeai.2023.100143

Bybee, R. (2009). The BSCS 5E instructional model and 21st century skills: A commissioned paper prepared for a workshop on exploring the intersection of science education and the development of 21st century skills

Bybee, R., Taylor, J., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., et al. (2006). The BSCS 5E instructional model: origins and effectiveness. Available online at: https://media.bscs.org/bscsmw/5es/bscs_5e_full_report.pdf (Accessed November 10, 2021).

Chen, L., Chen, P., and Lin, Z. (2020). Artificial intelligence in education: a review. *IEEE Access* 8, 75264–75278. doi: 10.1109/ACCESS.2020.2988510

Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interact. Learn. Environ.* 33, 3225–3229. doi: 10.1080/10494820.2025.2514372

Chounta, I.-A., Bardone, E., Raudsep, A., and Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in Estonian K-12 education. *Int. J. Artif. Intell. Educ.* 32, 725–755. doi: 10.1007/s40593-021-00243-5

Ding, L., Li, T., Jiang, S., and Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *Int. J. Educ. Technol. High. Educ.* 20. doi: 10.1186/s41239-023-00434-1

Duran, L. B., and Duran, E. (2004). The 5E instructional model: a learning cycle approach for inquiry-based science teaching. *Sci. Educ. Rev.* 5, 49–58. Available online at: https://files.eric.ed.gov/fulltext/EJ1058007.pdf

Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2024). A SWOT analysis of ChatGPT: implications for educational practice and research. *Innov. Educ. Teach. Int.* 61, 460–474. doi: 10.1080/14703297.2023.2195846

Henze, J., Schatz, C., Malik, S., and Bresges, A. (2022). How might we raise interest in robotics, coding, artificial intelligence, STEAM and sustainable development in university and on-the-job teacher training? *Front. Educ.* 7:872637. doi: 10.3389/feduc.2022.872637

Hornberger, M., Bewersdorff, A., and Nerdel, C. (2023). What do university students know about artificial intelligence? Development and validation of an AI literacy test. *Comp. Educ. Artif. Intellig.* 5:100165. doi: 10.1016/j.caeai.2023.100165

Ishmuradova, I. I., Zhdanov, S. P., Kondrashev, S. V., Erokhova, N. S., Grishnova, E. E., and Volosova, N. Y. (2025). Pre-service science teachers' perception on using generative artificial intelligence in science education. *Contemp. Educ. Technol.* 17:ep579. doi: 10.30935/cedtech/16207

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., et al. (2023). Unreflected acceptance: investigating the negative consequences of ChatGPT-assisted problem solving in physics education. Available online at: http://arxiv.org/pdf/2309.03087v1 (Accessed November 28, 2024).

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M. D., Păun, D., et al. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability* 13:10424. doi: 10.3390/su131810424

Lee, I., and Perret, B. (2022). Preparing high school teachers to integrate AI methods into STEM classrooms. *Proc. AAAI Conf. Artif. Intell.* 36, 12783–12791. doi: 10.1609/aaai.v36i11.21557

Leong, D., and Bodrova, E. (1996). Play: a Vygotskian approach. Available online at: https://video.alexanderstreet.com/watch/play-a-vygotskian-approach (Accessed November 12, 2021).

Lin, X.-F., Chen, L., Chan, K. K., Peng, S., Chen, X., Xie, S., et al. (2022). Teachers' perceptions of teaching sustainable artificial intelligence: a design frame perspective. *Sustainability* 14:7811. doi: 10.3390/su14137811

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Educ. Sci.* 13:410. doi: 10.3390/educsci13040410

Long, T., Gero, K. I., and Chilton, L. B. (2024). Not just novelty: a longitudinal study on utility and customization of an AI workflow. Available online at: http://arxiv.org/pdf/2402.09894v2 (Accessefd November 28, 2024).

Long, D., and Magerko, B. (2020). "What is AI literacy? Competencies and design considerations" in ACM digital library, proceedings of the 2020 CHI conference on human factors in computing systems. eds. R. Bernhaupt, F. Mueller, D. Verweij and J. Andres (New York, NY, USA: Association for Computing Machinery), 1–16.

Mahligawati, F., Allanas, E., Butarbutar, M. H., and Nordin, N. A. N. (2023). Artificial intelligence in physics education: a comprehensive literature review. *J. Phys. Conf. Ser.* 2596:012080. doi: 10.1088/1742-6596/2596/1/012080

Minichiello, A., and Caldwell, L. (2021). A narrative review of design-based research in engineering education: opportunities and challenges. *Stud. Eng. Educ.* 1:31. doi: 10.21061/see.15

Namdar, B., and Kucuk, M. (2018). Preservice science teachers' practices of critiquing and revising 5E lesson plans. *J. Sci. Teach. Educ.* 29, 468–484. doi: 10.1080/1046560X.2018.1469188

Nikolopoulou, K. (2024). ChatGPT for educational-academic activities: preschool student teachers' experiences. *Intellig. Technol. Educ.* doi: 10.53761/ITED/1.6

Padma, C., and Rama, C. (2022). A study of artificial intelligence in education system & role of AI in Indian education sector. *International Journal of Scientific Research in Engineering and Management* 6:1854. doi: 10.55041/IJSREM11854

Pedapati, K. (2022). Piagetian and Vygotskian concepts of cognitive development: a review. *Indian J. Ment. Health* 9:239. doi: 10.30877/IJMH.9.3.2022.227-239

Rigopouli, K., Kotsifakos, D., and Psaromiligkos, Y. (2025). Vygotsky's creativity options and ideas in 21st-century technology-enhanced learning design. *Educ. Sci.* 15:257. doi: 10.3390/educsci15020257

Roll, I., and Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *Int. J. Artif. Intell. Educ.* 26, 582–599. doi: 10.1007/s40593-016-0110-3

Salas-Pilco, S., Xiao, K., and Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: a systematic review. *Educ. Sci.* 12:569. doi: 10.3390/educsci12080569

Santos, R. P. d. (2023). Enhancing physics learning with ChatGPT, Bing chat, and bard as agents-to-think-with: a comparative case study. Available online at: http://arxiv.org/pdf/2306.00724 (Accessed November 26, 2024).

The Design-Based Research Collective (2003). Design-based research: an emerging paradigm for educational inquiry. *Educ. Res.* 32, 5–8. doi: 10.3102/0013189X032001005

Turan, S. (2021). Pre-service teacher experiences of the 5e instructional model: a systematic review of qualitative studies. *Eurasia J. Math. Sci. Technol. Educ.* 17:em1994. doi: 10.29333/ejmste/11102

UNESCO (2021). AI and education: Guidance for policy-makers. Paris, France: UNESCO.

van Zyl, A., and Karsten, I. (2022). Design-based research (DBR) as an effective tool to create context-sensitive and data-informed student success initiatives. *J. Stud. Affairs Afr.* doi: 10.24085/jsaa.v10i1.3706

Varsik, S., and Vosberg, L. (2024). The potential impact of artificial intelligence on equity and inclusion in education: OECD artificial intelligence papers. OECD EDUCATION WORKING PAPERS SERIES: Vol. 23.

Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., et al. (2023). What explains teachers' trust of AI in education across six countries? Available online at: http://arxiv.org/pdf/2312.01627v2 (Accessed November 26, 2024).

Vincent-Lancrin, S., and van der Vlies, R. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. OECD education working papers no. 218, 218

Vodafone Stiftung Deutschland gGmbH (2023). Aufbruch ins Unbekannte: Schule in Zeiten von künstlicher Intelligenz und ChatGPT [Into the Unknown. Schools in Times of Artificial Intelligence and ChatGPT]

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int. J. Educ. Technol. High. Educ.* 16. doi: 10.1186/s41239-019-0171-0

Zhao, L., Wu, X., and Luo, H. (2022). Developing AI literacy for primary and middle school teachers in China: based on a structural equation modeling analysis. *Sustainability* 14:14549. doi: 10.3390/su142114549

6.3. Studie III

Der Autor der Dissertation war für die Konzeption und die inhaltliche Ausrichtung der Studie verantwortlich. Er plante und führte die empirische Untersuchung, die systematische Kuratierung und Aufbereitung der erhobenen Daten, sowie die vollständige formale Analyse und Auswertung der Ergebnisse durch. Die methodische Gestaltung der Studie sowie die Interpretation der Befunde erfolgten ebenfalls durch den Autor der Dissertation. Darüber hinaus verantwortete er die Erstellung sämtlicher Visualisierungen zur Darstellung der Ergebnisse.

Im Hinblick auf die schriftliche Ausarbeitung verfasste der Autor der Dissertation den vollständigen Erstentwurf des Manuskripts. Die inhaltliche Überarbeitung, redaktionelle Präzisierung sowie die finale Zusammenführung der Texte bis zur Einreichung lagen ebenfalls in seiner Verantwortung.

Die Beiträge der Koautor:innen umfassten unterstützende Tätigkeiten in einzelnen Phasen der Studie. Julia Lademann wirkte an der kritischen Durchsicht des Manuskripts mit und unterstützte die empirische Untersuchung sowie Datenaufbereitung. Prof. Dr. André Bresges übernahm Aufgaben im Bereich der wissenschaftlichen Betreuung, der Bereitstellung von Ressourcen und Software, sowie der Drittmittelakquise und unterstützte die Überarbeitung des Manuskripts. Prof. Dr. Sebastian Becker-Genschow brachte sich in die konzeptionelle und methodische Beratung der Studie ein, unterstützte die formale Analyse und wirkte an der kritischen Revision des Manuskripts mit. Die Gesamtverantwortung für Konzeption, Durchführung, Auswertung und Verschriftlichung der Studie lag bei dem Autor der Dissertation.

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 18 February 2026
DOI 10.3389/feduc.2026.1719670

Check for updates

OPEN ACCESS

EDITED BY
Rogers Kaliisa,
University of Oslo, Norway

REVIEWED BY
Cecep Kustandi,
Jakarta State University, Indonesia
Víctor Robledo-Rella,
Tecnologico de Monterrey, Mexico

*CORRESPONDENCE
Jannik Henze
jannik.henze@uni-koeln.de

RECEIVED 06 October 2025
REVISED 14 December 2025
ACCEPTED 12 January 2026
PUBLISHED 18 February 2026

CITATION
Henze J, Lademann J, Becker-Genschow S and Bresges A (2026) AI-supported data analysis boosts student motivation and reduces stress in physics education. *Front. Educ.* 11:1719670.
doi: 10.3389/feduc.2026.1719670

AI-supported data analysis boosts student motivation and reduces stress in physics education

Jannik Henze[1]*, Julia Lademann[2], Sebastian Becker-Genschow[2] and André Bresges[1]

[1]Institute for Physics Education, University of Cologne, Cologne, Germany, [2]Digital Education Research, Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, Germany

The integration of artificial intelligence (AI) into education presents new opportunities for supporting learning processes. This study investigates the impact of AI-assisted versus traditional *Excel*-based data analysis on both learning outcomes and emotional-motivational responses in a physics education context. A custom *GPT*-based chatbot, *ExperiMentor*, was developed to support student teachers in analyzing experimental data from thread and spring pendulum experiments. Fifty student teachers were randomly assigned to either the AI or *Excel* group, with both groups completing identical tasks in a guided setting. Learning progress was measured using pre- and post-tests, while emotional and motivational variables were assessed through structured surveys. Both groups demonstrated significant learning gains, with no statistically significant differences found between them in terms of cognitive performance. However, the AI group reported substantially higher levels of engagement, enjoyment, and perceived method effectiveness compared to the *Excel* group. These findings suggest that interactive AI tools may enhance the affective dimensions of learning, even when cognitive outcomes remain comparable to traditional methods. The results underscore the importance of integrating AI not as a replacement for instructional design, but as a supportive element within pedagogical frameworks. Future research should explore long-term retention effects, the role of learner diversity, and comparisons with other forms of pedagogical support.

KEYWORDS

artificial intelligence, data analysis, excel, learning, motivation, physics education

1 Introduction

The role of digital technologies in education is expanding, with artificial intelligence (AI) attracting particular interest as a potential support for complex learning processes. Across disciplines, educators and researchers are exploring how AI tools might assist students in developing not only subject-specific knowledge, but also broader skills for navigating increasingly data-rich and cognitively demanding learning environments (Roll and Wylie, 2016; Winkelmann et al., 2021). As educational strategies evolve toward adaptability, the integration of AI emerges as a pivotal factor at both school and university levels that is redefining the boundaries of teaching and learning (Kuleto et al., 2021; Bacia et al., 2024).

One of the most widely discussed developments in this context is the emergence of Large Language Models (LLMs), such as *ChatGPT*. In the context of this study, the term AI primarily refers to LLMs such as *ChatGPT*, which represent a specific subset of AI focused on natural language processing.

In physics education, LLMs can be used to support tasks like problem solving or interpreting experimental data. Studies suggest these models can reduce barriers to engagement by making learning more flexible and responsive (Küchemann et al., 2024).

Furthermore *ChatGPT* can simplify the process of tackling complex physics problems by offering step-by-step explanations, providing a conversational partner and adapting to different learning styles, making physics more accessible (Tong et al., 2024). The growing capabilities of AI in processing and interpreting data have opened up new possibilities for integrating these tools into physics instruction (Farrokhnia et al., 2024).

Despite the availability of AI tools such as *ChatGPT*, students often struggle to apply them productively, especially when lacking guidance or a solid understanding of underlying physics concepts (Kechel and Wodzinski, 2015; Low and Kalender, 2023). Therefore, explicit instructional strategies are necessary to help students bridge conceptual gaps and improve their ability to leverage AI effectively.

These technological shifts necessitate a critical evaluation of educators' roles and pedagogical methodologies (Popenici and Kerr, 2017). As AI-powered applications gain growing importance in educational strategies (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024) at school and university levels institutions must develop robust frameworks to effectively incorporate this technology, addressing specific challenges such as technological complexity, data privacy concerns, ethical considerations and other potential barriers to equal learning opportunities.

Given the rapid pace of development and ongoing debate, it is crucial to examine not only what AI tools can offer, but also where their limits lie. While AI-based tools show potential, concerns remain regarding their accuracy, potential bias, and long-term pedagogical value. This study conducts a comprehensive analysis through a comparison of two learning tools used in a physics context. A custom GPT-based chatbot called *ExperiMentor* was developed to support students in analyzing data from physics experiments. In a controlled setting, one group of student teachers, university students studying to become teachers after their master's degree, used *ExperiMentor* to evaluate pendulum experiments, while a second group used *Microsoft Excel*. Both groups received the same tasks, allowing for a comparison of their learning outcomes as well as their emotional and motivational experiences.

The study focuses on two key questions: first, how do students' learning outcomes change within each group, and how do the groups compare? Second, how do students experience each tool emotionally and motivationally, in terms of engagement, frustration, and perceived effectiveness? These questions, derived from a self-developed questionnaire, explore the conditions under which AI tools might support student learning, and where challenges or limitations may arise. The paper is structured as follows: Section 2 reviews current research on AI in education. Section 3 outlines the study design and methods. Section 4 presents the findings, followed by a discussion in Section 5. Section 6 concludes with implications and directions for future research.

2 State of research

2.1 Educational potential and technical foundations

Recent developments in AI research have drawn attention to the educational potential of LLMs. To contextualize the present study, this section first outlines how LLMs such as *ChatGPT* function and why they are gaining relevance in educational discourse. From a technical perspective, LLMs like *ChatGPT* are trained to predict the next token (e.g., word or character) in a sequence, using massive datasets to model patterns in language (Bender et al., 2021; Ouyang et al., 2022). *ChatGPT*-4, the model used in this study, is multimodal and can process both text and image inputs. It was trained to improve its factual accuracy and responsiveness (OpenAI et al., 2023) compared to predecessor models. The educational potential of such technology has led to growing research interest (Roll and Wylie, 2016; Chen et al., 2020; Mahligawati et al., 2023) along with calls for educators to understand and engage AI in order to prepare for future teaching and learning contexts (Zawacki-Richter et al., 2019). Current findings remain often preliminary, diverse and context dependent. There is no clear consensus yet on the long-term impact or optimal use of AI tools in education. Nevertheless studies show that ChatGPT can increase learning gains (Alarbi et al., 2024; Kestin et al., 2025) and enhance engagement and motivate learners (Wang and Fan, 2025). AI integration in education has mostly taken place at the individual level, not institutionally (Vincent-Lancrin and van der Vlies, 2020). As a result, teacher training is considered crucial for meaningful implementation (Salas-Pilco et al., 2022).

2.2 Applications and opportunities in physics education

Beyond their technical foundations, LLMs have increasingly been applied in subject-specific domains. Studies suggest that AI tools like *ChatGPT* support interactive, student-centered learning by offering real-time explanations and adapting to individual input, particularly in explaining complex physics concepts and problem-solving strategies (Popenici and Kerr, 2017; Lin et al., 2022; Liang et al., 2023; Santos, 2023; Küchemann et al., 2024). However, it remains unclear how effective this feedback is compared to traditional forms (Farrokhnia et al., 2024).

Previous studies suggest that *ChatGPT* may be a useful assistant tool, surpassing traditional tools such as *Microsoft Excel* in certain tasks (Halaweh, 2023) while also providing students with the advantage of actively engaging with educational materials, allowing a flexible learning environment (Padma and Rama, 2022; La Ossa et al., 2024; Ng et al., 2024). These capabilities have been studied in relation to their potential to support structured reasoning and dataset interpretation (Mustofa et al., 2024). In one study, *ChatGPT-4* outperformed engineering students on the Force Concept Inventory (FCI), suggesting its potential for physics instruction (Kieser et al., 2023).

2.3 Limitations, accuracy and concerns

Despite these applications, recent research also highlights important limitations of LLMs in educational contexts, particularly in terms of accuracy, reliability, and their influence on students' critical thinking. For instance, the performance of *ChatGPT* across physics tasks is highly variable, ranging from 15% to 80% accuracy, which may partly reflect its English-language training base (Revalde et al., 2025). *ChatGPT* performs well on conceptual physics questions but struggles with open-ended or calculation-heavy tasks, reflecting its strength in language rather than quantitative reasoning (Bender et al., 2021; Gregorcic and Pendrill, 2023; Revalde et al., 2025).

Students often lack the experience to critically assess such outputs, and may accept them without questioning them (Ding et al., 2023). In this context, encouraging students to examine responses carefully can become an important part of *ChatGPT's* educational use. When used thoughtfully, the tool may help develop critical thinking by prompting reflection on varying outputs. It may also serve as a starting point for discussions about the strengths and limitations of AI tools and their role in assessing the reliability of information (Bitzenbauer, 2023).

2.4 Learning theory integration

The Self-Determination Theory examines people's basic psychological needs that drive motivation as well as learning (Ryan and Deci, 2000). This theory distinguishes between two types of motivation: Intrinsic motivation is defined as the natural drive toward learning, skill development, curiosity and discovery, while extrinsic motivation means performing an activity to achieve a specific external goal (Ryan and Deci, 2000). Both types of motivation do not combine cumulatively (Deci, 1975).

Self-Determination Theory therefore highlights the significance of autonomy and competence in fostering intrinsic motivation for learning (Deci and Ryan, 2008). To effectively cultivate these conditions, it is essential to design educational experiences that not only support student motivation but also account for cognitive load. This refers to the demand placed on working memory by incoming information.

Cognitive load theory is an instructional framework grounded in understanding human cognitive architecture, specifically designed to address working memory constraints (Sweller, 2012). It seeks to enhance productive cognitive processing to facilitate knowledge acquisition. There are three different types of cognitive load, that can be aggregated (Sweller, 2012). Intrinsic cognitive load refers to the mental effort required due to the difficulty of the material being learned (Sweller, 2012). Extraneous cognitive load results from instructions that disregard working memory capacity and diverts mental resources away from building knowledge structures and germane cognitive load represents the cognitive demand generated by focused learning that leads to building knowledge (Sweller, 2012). Further, favorable emotional states that strengthen mental functions and reduce inherent complexity can generate additional working memory capacity that can be redirected to amplify learning experiences (Young et al., 2021).

Cognitive load theory emphasizes the importance of managing mental effort during learning, which aligns closely with Vygotsky's concept of the zone of proximal development (ZPD). This zone describes the gap between what learners can accomplish on their own and what they can achieve with guidance by a More Knowledge Other (MKO) (Rigopouli et al., 2025). Technology-enhanced approaches, including AI, function as scaffolding within the ZPD to help students reach beyond their independent capabilities (Rigopouli et al., 2025).

Generative AI like ChatGPT can act like an MKO and support learning processes by enabling learners to interact more thoroughly with the content (Wang and Fan, 2025). By guiding and supporting learners, extraneous cognitive load can be lowered and working memory resources released, potentially allowing for increased germane cognitive load (Sweller, 2012) and therefore better learning outcomes.

2.5 Emotional and motivational impacts

In addition to cognitive performance, several studies have explored how AI-based tools affect students' emotional and motivational experiences during learning. A study from Stanford found that while *ChatGPT* performs good on well-defined physics problems, it struggles with open-ended questions that require assumptions or contextual understanding (Wang et al., 2024). Fluent AI responses can create an illusion of understanding (Dahlkemper et al., 2023), while students' tendency to copy answers without reflection further limits critical engagement (Krupp et al., 2023a). This tendency can make them more likely to accept false or misleading content without questioning it (Kasneci et al., 2023; Kortemeyer, 2023), which in turn may reinforce misconceptions and negatively affect future learning (Ding et al., 2023). At the same time, students have rated AI as moderately important in physics education but viewed *ChatGPT* as a valuable tool for various applications (Bitzenbauer, 2023).

On the other hand, AI also can be beneficial in encouraging critical thinking and engagement. A recent study comparing moderated and unmoderated LLM use as well as simple internet searches suggests that the use of moderated LLMs can potentially promote critical thinking and enhance engagement strategies (Krupp et al., 2023b).

Possible positive motivational effects of an AI tool (Hanum Siregar et al., 2023) appear linked to the interactive and novel nature of the tools (Kuleto et al., 2021; Kestin et al., 2025; Wang and Fan, 2025). New technologies often attract attention to the tool itself rather than their educational function, which may result in superficial or unfocused use (Miguel-Alonso et al., 2024).

Turning to the impact on learning processes, studies indicate that students were able to learn effectively through *ChatGPT*, leading to improved understanding and higher achievement (Lo, 2023; Alarbi et al., 2024). AI learning tools have been reported to affect university students' learning experiences (Wen et al., 2024), increasing engagement and promoting positive attitudes toward learning (Gada and Chudasana, 2024). AI chatbots have been examined for their potential to support the development of practical skills in learning contexts. Research indicates that their use enhances cognitive abilities and supports skill development (Essel et al., 2024).

When used in educational settings, AI chatbots have been found to affect several aspects of student learning, including performance, motivation, interest, self-efficacy, and anxiety. However, factors like educational level and intervention duration moderate these effects (Wu and Yu, 2024). In higher education, where most studies on

ChatGPT have been conducted (Lo, 2023), students generally view AI tools like *ChatGPT* as useful and engaging, though they also expect high levels of reliability, user-friendliness, and personalized support (Wen et al., 2024).

A comprehensive analysis of various studies on the impact of *ChatGPT* on education reveals that students' emotional engagement with a chatbot is primarily influenced by the perceived benefits, performance expectations and the quality of information output (Lo, 2023). Nevertheless, students utilizing generative AI demonstrate elevated cognitive as well as emotional involvement (Guo et al., 2025). Additional factors, such as enjoyment, self-directed regulation, ease of use and the alignment of outcomes with prior expectations, further contribute to engagement with the tool (Lo, 2023). Learners who view generative AI as beneficial for their studies display stronger positive emotional involvement, indicating that attitudes toward the technology significantly influence its impact (Guo et al., 2025). However, this enthusiasm with technology coexists with significant public skepticism (Vodafone Stiftung Deutschland gGmbH., 2023).

In summary, while AI tools show potential for enhancing engagement and offering personalized support, their impact on learning outcomes remains inconsistent, particularly in tasks requiring deeper conceptual understanding. This study takes up that challenge by comparing AI-supported learning with the more traditional tool *Microsoft Excel* in the context of physics experiments.

3 Methodology

To compare the educational impact of AI-assisted vs. *Excel*-based data analysis in physics, this study implemented a controlled intervention with student teachers analyzing pendulum experiment data. The intervention was implemented as a structured task in which participants evaluated two physics experiments, one involving a thread pendulum and the other a spring pendulum, using either the AI chatbot *ExperiMentor* or *Excel*.

The study followed a two-stage testing procedure to evaluate participants' experiences and skills. One group worked with *ExperiMentor*, a specialized conversational agent developed specifically for this study based on the GPT-4o model. *ExperiMentor* was implemented using the OpenAI GPT Creator and utilized between June and August 2024. It is not a publicly hosted tool apart from the OpenAI *ChatGPT GPT*-Store[1], nor integrated into an external platform. *ExperiMentor* was developed solely for research purposes without any commercial interest. To ensure transparency and reproducibility, the full system instructions and a screenshot of a sample interaction with *ExperiMentor* are provided in the Supplementary material.

ExperiMentor is a custom *GPT*-based chatbot that supports data analysis by automatically generating and executing *Python* code. Students received only the outputs (e.g., results, visualizations), with optional code previews, but no direct interaction with the code environment. *ExperiMentor's* behavior was governed by a system prompt designed around three core pillars:

First, contextual integrity and anonymity: the prompt enforces a strict authentication protocol. Before any pedagogical interaction begins, *ExperiMentor* is programmed to persistently request a generated identification code (based on demographic data) to link chat logs with pre-post survey data. The instructions explicitly forbid from proceeding without this code, ensuring data validity.

Second, pedagogical scaffolding and role definition: the system prompt defines *ExperiMentor's* role as a mentor or guide rather than a solver. It explicitly suppresses direct answer-giving behavior. Instead, the prompt instructs the model to guide students through data collection, analysis and interpretation. It promotes critical thinking by providing explanations, formulas, and hints only upon request or when errors are detected.

Third, interaction flow control: to manage cognitive load, the prompt enforces a step-by-step interaction model. After each analytical step, *ExperiMentor* is instructed to pause, display intermediate results, and ask the student *"What should we do next?"*. This design prevents the AI from hallucinating a full solution path and forces the student to maintain agency over the scientific process.

The prompts also included negative constraints to mitigate common LLM issues like long answers or over-helpfulness. These measures increased internal validity by keeping the pedagogical behavior and analytical guidance of *ExperiMentor* consistent throughout the intervention, regardless of individual student phrasing.

To evaluate the impact of the intervention, a randomized pre-post control group design was employed (Döring et al., 2016). Paired observations were used to assess change, comparing pre- and post-intervention responses within the same participants (Hedderich and Sachs, 2016). Both groups, one using *Excel*, the other *ExperiMentor*, completed identical tasks, allowing for direct comparison of outcomes.

3.1 Research questions

Despite the growing interest in AI tools, there is still no conclusive empirical evidence on their impact on learning processes. This emphasizes the urgent need for clear guidance on their role in educational practices, including their potential benefits, limitations, and challenges (Ständige Wissenschaftliche Kommission der Kultusministerkonferenz, 2024). As tools like *ChatGPT* are still at an early stage of adoption, ongoing research is crucial to understanding their effectiveness and ensuring their responsible integration into education (Farrokhnia et al., 2024). To help address this research gap, the present study investigates the following questions.

Research Question 1 (RQ1): how do students' results change from pre- to post-test within each group, and how do the learning gains compare between the *Excel* and AI groups? This research question combines both intragroup improvements and intergroup comparisons. It aims to analyze whether participants show measurable gains in their understanding of physical concepts and in their ability to apply data analysis methods after the

1 ExperiMentor can be accessed via: https://chatgpt.com/g/g-WZ4GsKZbZ-experimentor-studie (to interact with it, make up an identification code it will ask for).

intervention. It also explores whether the learning gains vary depending on the tool used.

Research Question 2 (RQ2): how do students' emotional responses differ when analyzing data using *Excel* vs. *ExperiMentor*? This question moves beyond cognitive learning outcomes and considers the emotional dimension of the learning experience. It explores whether the type of tool used influences students' emotional reactions, such as uncertainty, enjoyment, or frustration, while working on physics data analysis tasks. The goal is to develop a more holistic understanding of how digital tools affect engagement and motivation.

3.2 Procedure

The intervention, i.e., the instructional activity, was carried out by two different groups consisting of randomly assigned participants. Both groups evaluated one experiment on the thread pendulum and one on the spring pendulum using given data sets. Both groups were guided through the evaluation step by step with the help of identical tasks given on paper (see Supplementary material for details). The aim of the thread pendulum experiment was to determine the magnitude of the gravitational acceleration g. The spring pendulum experiment aimed to calculate the spring constant k.

The participating students selected seats at random from pre-prepared workstations, each pre-loaded with either *Excel* or *ExperiMentor*. The software configuration was hidden during selection to ensure unbiased assignment. The pre-test was carried out by the participants before they knew which group they belonged to. The task sheets were only made available after completion of the pre-test and removed again before the start of the post-test.

While *Excel* is not a pedagogical method, it was chosen as a control condition due to its widespread use in university-level data analysis. The goal was to provide a familiar, static environment that allowed for a contrast with the dynamic and responsive features of the AI-based chatbot. The *Excel* group worked without external resources to simulate a closed-book setting. To minimize potential cross-group influence, particularly given that both groups worked in the same physical space, participants were explicitly instructed not to discuss the tasks or digital tools during the session. The room layout was arranged to separate the groups and the intervention was monitored to ensure compliance with these instructions. Although minimal cross-group communication was possible, the room layout ensured working independence.

3.3 Material

The pre- and post-tests focused on evaluating participants' knowledge and skills in analyzing physical experiments. Emotional and motivational variables were assessed through structured survey instruments (see Supplementary material for details).

As Table 1 shows, the study followed a pre-post design consisting of five sections (Parts A to E), with a total of 31 items in the pre-test and 39 items in the post-test. Part A collected demographic information in the pre-test and group affiliation in the post-test.

Part B focused on emotional and motivational variables, using Likert-type items. In the pre-test, technology commitment (B1) was assessed using 7 validated items from the short scale for measuring technology commitment (Neyer et al., 2016), rated on a 5-point Likert scale. Participants also rated their interest and perceived ease of experimental evaluation (B2) using two 4-point Likert scales.

In the post-test, emotional responses were addressed in Section B3. Here, positive emotional learning experiences (items B3.1 and B3.2) captured participants' feelings of enjoyment and success, while negative emotional experiences (items B3.3 to B3.5) reflected emotional states such as uncertainty, stress, or frustration. Sections B4 and B5 examined participants' perception of the method and the perceived difficulty of concepts and topics, respectively. Section B6 focused on method effectiveness, which was subdivided into four constructs: Effectiveness of the method for understanding physics and supporting data analysis (items B6.1, B6.3), perceived learning gains regarding comprehension and progress (B6.2, B6.6), Motivation to engage with the experiment and evaluating data (B6.4, B6.5), and a comparison between the AI method and *Excel* (B6.7, B6.8).

Parts C to E assessed participants' conceptual understanding of thread and spring pendulum physics, as well as their data analysis competencies. The tasks in these sections consisted of objective, single-choice items designed to assess factual knowledge, conceptual reasoning, and the ability to interpret and analyze experimental data. These measures were used to evaluate actual learning gains between the pre- and post-tests.

The group working with *Excel* received an *Excel* spreadsheet in which the respective measured values of the experiment were already entered. They could edit the table freely and were not subject to any restrictions, except that they were not allowed to search for help outside of *Excel* and the printed task sheets (see Supplementary material). The group using the AI tool *ExperiMentor* completed the tasks within the chat environment of the custom GPT *ExperiMentor*. They also received an *Excel* spreadsheet containing the measurement results. They had the option of transferring the data manually, copying it or uploading the entire table to the chat.

3.4 Data collection and analysis

Data were recorded using a digital survey tool and analyzed in *R Studio* (version 4.3.3) (Hanum Siregar et al., 2023). Descriptive parameters included mean (M), median (MDN), and standard deviation (SD (Miguel-Alonso et al., 2024). The Shapiro-Wilk test was used to check distribution assumptions (Miguel-Alonso et al., 2024) because of its sensitivity to deviations from normality (Kortemeyer, 2023).

3.4.1 Performance data

To address the first research question, learning gains within groups between pre and post were examined. Parts C, D, and E of the questionnaire were analyzed first, followed by the three

TABLE 1 Structure of the pre- and post-test instruments, including item count and response format by section and focus area.

Section	Focus	# Items	Scale
Part A (Pre/Post)	Demographics	4 Pre 1 Post	
Part B (Pre/Post)	Emotional & motivational Variables	9 Pre 20 Post	Likert Scales
B1 (Pre)	Technology commitment	7	5-point Likert
B2 (Pre)	Interest & perceived ease of experimental evaluation	2	4-point Likert
B3 (Post)	Emotional learning experience	5	4-point Likert
B4 (Post)	Perception of method	3	4-point Likert
B5 (Post)	Difficulty of concepts & topic	4	4-point Likert
B6 (Post)	Method Effectiveness	8	4-point Likert
Part C (Pre/Post)	Thread pendulum physics	6	Nominal (Single Choice)
Part D (Pre/Post)	Spring pendulum physics	6	Nominal (Single Choice)
Part E (Pre/Post)	Data analysis competencies	6	Nominal (Single Choice)

subject areas thread pendulum, spring pendulum as well as evaluation methods and finally individual questions. The *Shapiro-Wilk test* indicated non-normal distribution throughout. A two-sided *Wilcoxon signed-rank test* ($\alpha \leq 0.05$) was used to check for significant directional changes between the groups and the pre- and post-test (Wollschläger, 2014). To assess effect sizes the rank-biserial correlation was used (Kerby, 2014).

An *ANCOVA* was used examining the intervention's impact while controlling for pre-test scores (Bortz and Schuster, 2010; Wollschläger, 2014). Despite non-normal distribution, ANCOVA is robust against such violations (Glass et al., 1972; Bortz and Schuster, 2010; Schmider et al., 2010), especially with equal group sizes ($n = 25$ each) (Bortz and Schuster, 2010). To evaluate the *ANCOVA* assumption of homogeneous regression slopes, it was tested whether the covariate interacted with the experimental factors. The interaction between Group and Pre-Test was not significant in any model, indicating that the basic assumption of parallel slopes across groups was met. However, the higher-order interactions involving the covariate, particularly Group × Measurement × Pre were significant in all models (see Table 6). This indicates that the relationship between the covariate and the dependent variable varied across measurement points and groups, meaning that full homogeneity of regression slopes was not strictly satisfied. Consequently, the ANCOVA results should be interpreted with caution.

The analysis began with the total sum of items to compare general learning gains, with pre-test scores as covariates. The three main categories were then analyzed separately, followed by individual items within each category. The *Hake Index g* was calculated for combined and separate group data. This metric measures average normalized learning growth (Hake, 1998) as the proportion of actual to maximum possible improvement (McKagan et al., 2022). The gain of averages method was used to evaluate the effectiveness of the learning interventions (McKagan et al., 2022). The index shows minimal correlation with pre-test scores, indicating independence from prior knowledge. In contrast, the average post-test score and the average gain are less suitable for comparing the course success rate across different groups (Hake, 2002; Coletta and Steinert, 2020). Values for g range from 0 (no gain) to 1 (maximum gain; Hake, 2002).

3.4.2 Emotional-motivational data

Descriptive statistics and Shapiro-Wilk tests were conducted for all questions. Internal reliability was evaluated using *Cronbach's* α, which represents the average correlation among items (Döring et al., 2016) and the proportion of test variance that can be attributed to shared factors across the items (Cronbach, 1951). The *Wilcoxon rank-sum test* measured differences between groups, as it does not require normal distribution (Wollschläger, 2014), which was not given after analyzing the results of the *Shapiro-Wilk test*.

3.5 Sample

The $n = 50$ participants were student teachers from the University of Cologne: 33 male, 17 female, aged 19–37 years ($M = 24.04$). At the time of the study, 35 were in a Bachelor of Arts program and 15 in a Master of Education program. The Excel group ($n = 25$) had an average age of $M = 24.36$ years, with 17 Bachelor and 8 Master students. The AI group ($n = 25$) had an average age of $M = 23.72$ years, with 18 Bachelor and 7 Master students. All participants were training to become teachers for various school types.

4 Results

This section presents the results of the experimental study in accordance with the two research questions. Section A reports learning-related outcomes based on participants' pre- and post-test performance, while Section B addresses emotional and motivational outcomes assessed through survey responses.

4.1 Learning outcomes (RQ1)

4.1.1 Descriptive statistics

Descriptive statistics were calculated to provide an overview of participants' performance before and after the intervention (see Tables 2, 3 and Figure 1). For both groups, *mean scores (M)*, *medians (MDN)*, and *standard deviations (SD)* were determined for the total score and the three domains: thread pendulum (TP), spring pendulum (SP), and data evaluation (DE).

Items within the thread pendulum domain focused on conceptual understanding and application of the physical principles of a simple thread pendulum, including the influence of length, gravitational acceleration and probing mass independence as well as scenarios involving altered gravity or acceleration. The spring pendulum domain included items regarding the factors determining the period of a spring pendulum, such as dependence on mass, gravitational acceleration and the spring constant. Items within the data evaluation (DE) domain required knowledge on the interpretation of graphical data, application of linear regression and understanding of statistical metrics. Overall, both groups tended to perform better after the intervention, with increases in scores seen across the total results and in most of the tested areas. Score variability remained stable overall, with a slight narrowing in the AI group's post-test data.

TABLE 2 Descriptive data for the excel group.

Item	M		SD		MDN	
	Pre	Post	Pre	Post	Pre	Post
Total	7.72	8.96	3.27	3.51	7	9
TP	3.32	3.16	1.80	1.77	3	3
SP	1.56	1.96	1.39	1.54	1	2
DE	2.84	3.84	1.31	1.28	3	4

TABLE 3 Descriptive data for the AI group.

Item	M		SD		MDN	
	Pre	Post	Pre	Post	Pre	Post
Total	7.64	9.92	3.43	2.96	7	9
TP	3.24	3.72	1.45	1.40	3	4
SP	1.40	2.20	1.29	1.29	1	2
DE	3.00	4.00	1.61	1.22	3	4

4.1.2 Intragroup learning gains

To assess the learning gains within each group, the results of the pre- and post-tests were analyzed separately for the *Excel* and AI groups using the *Wilcoxon signed-rank test*. Effect sizes (*rank biserial correlation*) indicate the magnitude of learning gains across domains. The effect sizes for the rank-biserial correlation are categorized as very small (< |0.10|), small (|0.10| to |0.29|), moderate (|0.30| to |0.49|), or large (≥ |0.50|; López-Martín and Ardura, 2023).

The results of each group are presented in Tables 4, 5. The subsequent analysis applies the same statistical approach to the

TABLE 4 Wilcoxon test and effect sizes excel group.

Item (excel)	*p*-value	Rank biserial correlation
Total	0.04	−0.54
TP	0.60	
SP	0.32	
DE	<0.01	−0.78

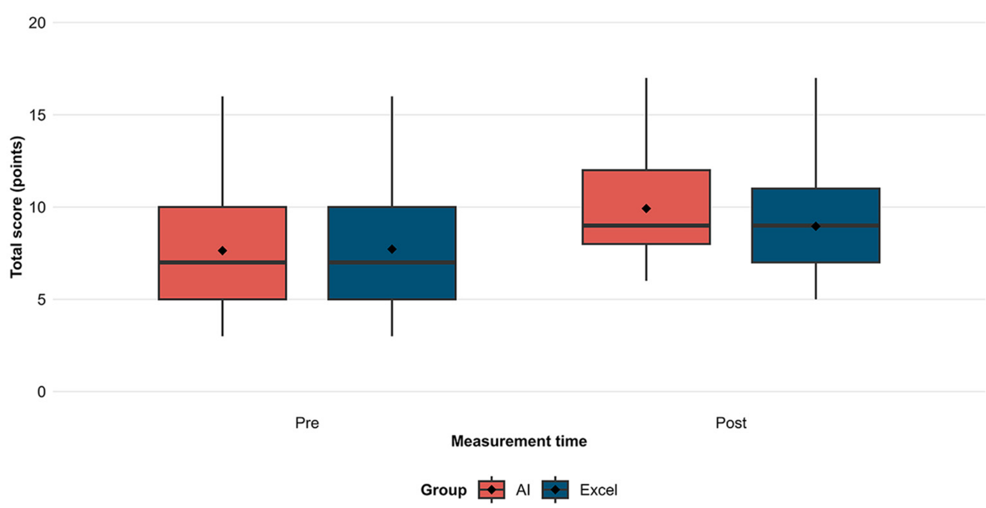

FIGURE 1
Boxplots of total scores before and after the intervention for the AI and Excel groups with the horizontal line inside the box indicating the group median and the dot representing the group mean.

second group, enabling a consistent assessment of learning progress across domains.

Both groups showed significant learning gains from pre- to post-test, though the magnitude and distribution of these gains differed by domain. In the *Excel* group, only the data evaluation domain showed a statistically significant improvement, with a large effect size. The other two domains showed no significant change. By contrast, the AI group exhibited significant gains in both the spring pendulum and data evaluation domains, with very large effect sizes. The thread pendulum domain reached a borderline significance threshold, suggesting a possible effect that warrants further investigation. Overall, the total scores improved significantly in both groups, with the AI group displaying a notably larger effect.

4.1.3 Intergroup learning gains

To investigate differences in learning gains between the two groups, an *ANCOVA* was conducted (see Table 6). The dependent variable was the post-test performance, the independent variable was the group assignment, and the pre-test scores served as the covariate. This design allows for a more accurate comparison of group outcomes by adjusting for individual differences in performance. *ANCOVA* was used despite non-normality (Glass et al., 1972; Bortz and Schuster, 2010). Partial η^2 values showed that the group factor accounted for a negligible amount of the variance in all models (see Table 6). In contrast, the measurement factor showed small to moderate effects, particularly for the total score and the DE items.

In addition to that, the *Hake Index (g)* was calculated (see Table 6) to quantify normalized learning gains. This measure provides a standardized way to compare relative learning progress regardless of initial performance levels, where values below 0,3 indicate low gains and values near 0 or negative suggest minimal or no improvement (Neyer et al., 2016).

No statistically significant differences were found between the AI and *Excel* groups at the overall level or in any of the specific domains. The total *Hake Index* value of $g = 0{,}02$ also indicates only marginal learning progress when combining both groups.

While *ANCOVA* offers adjusted comparisons between groups, the *Hake Index* complements this by summarizing the magnitude of within-group progress relative to potential gains. To better understand group-specific improvements, the Hake Index was also calculated separately for each condition and is presented in Table 7.

Overall, the results from both the *ANCOVA* and *Hake Index* calculations indicate that no statistically significant differences were observed between the groups across the measured domains. The *Hake Index* values indicated minimal learning gains in both groups.

TABLE 5 Wilcoxon test and effect sizes AI group.

Item (AI)	*p*-value	Rank biserial correlation
Total	<0.01	−1
TP	0.05[a]	−0.53
SP	< 0.01	−0.8
DE	< 0.01	−0.86

[a]Borderline significant.

TABLE 7 Hake indices per group.

Item	Hake *g* AI	Hake *g* excel
Total	0.02	0.01
TP	0.01	−0.004
SP	0.02	0.01
DE	0.02	0.02

TABLE 6 ANCOVA results and Hake Index summary.

Item	*p*-value group × measurement	Slope homogenity pre × group × measurement	Partial η^2		Hake *g* (total)
Total	0.39	<0.01	Group	<0.01	0.02
			Measurement	0.08	
			Group×Measurement	<0.01	
TP	0.30	< 0.01	Group	<0.01	< 0.01
			Measurement	<0.01	
			Group × Measurement	0.01	
SP	0.45	<0.01	Group	<0.01	0.01
			Measurement	0.05	
			Group × Measurement	<0.01	
DE	1	<0.01	Group	<0.01	0.02
			Measurement	0.13	
			Group × Measurement	0	

4.2 Emotional and motivational outcomes (RQ2)

4.2.1 Pre-intervention differences in emotional-motivational attitudes

Before the intervention, participants in both groups completed survey items designed to capture their emotional-motivational attitudes toward technology and experimental data analysis. Section B1 included seven previously validated items on technology commitment (Neyer et al., 2016), while Section B2 contained two items assessing interest in and perceived ease of evaluating experiments. Given the small number of items in B2 and the prior validation of B1, no reliability analysis was performed for the pre-test.

Due to non-normality, group comparisons were conducted using *Wilcoxon rank-sum tests* (see Table 8). Across all items, no statistically significant differences were found between the *Excel* and AI groups. This suggests that the groups started the intervention with comparable attitudes toward both technology and the task of evaluating experimental data.

4.2.2 Post-intervention differences in emotional-motivational attitudes

After completing the experimental tasks, participants responded to a broader set of questions grouped into eight constructs reflecting their emotional and motivational experience (see II. B. in the Supplementary material). These included categories such as enjoyment, frustration, perceived effectiveness, and motivation. Internal consistency of the constructs was evaluated using *Cronbach's* α. Most constructs reached acceptable to good reliability, with only two falling slightly below the conventional threshold. The calculated α as well as mean scores, significant values p and effect sizes are presented in Table 9.

Violin plots revealed clear differences in responses (see Figure 2). Responses in the AI group clustered toward the higher end of the scale, particularly in constructs such as *Positive Emotional Learning Experience*, *Method Effectiveness*, and *Motivation*. In contrast, the *Excel* group exhibited broader distributions with more lower-end responses, especially in Negative *Emotional Learning Experience* and *Method Comparison*.

TABLE 8 Wilcoxon rank-sum test results for pre-test items.

Item	Formulated	*p*-value
B1.1	I am very curious about technical innovations.	0.41
B1.2	I am often afraid of failing when dealing with modern technology.	0.36
B1.3	If I had the opportunity, I would use technical products much more frequently than I currently do.	0.85
B1.4	I quickly develop a liking for technical innovations.	0.39
B1.5	I find dealing with new technology difficult – I usually can't manage it.	0.71
B1.6	Dealing with technical innovations mostly overwhelms me.	0.99
B1.7	Whether I succeed in using modern technology depends mainly on me.	0.17
B2.1	I find it easy to evaluate experiments.	0.57
B2.2	I find evaluating experiments interesting.	0.87

TABLE 9 Group differences in emotional-motivational constructs: mean scores, significant values (*p*), effect sizes (rank biserial correlation) and internal consistency (Cronbach's α).

Construct	Items	α	Mean		*p*-value	Effect size
			AI	Excel		
Positive emotional Learning experience	B3.1 B3.2	0.77	2.26	1.74	0.01*	−0.392
Negative emotional Learning experience	B3.3 B3.4 B3.5	0.82	0.64	1.56	<0.01*	0.717
Perception of methods	B4.1 B4.2 B4.3	0.73	2.44	1.68	<0.01*	−0.65
Challenges	B5.1 B5.2 B5.3 B5.4	0.80	0.93	1.13	0.33°	0.294°°
Method effectiveness	B6.1 B6.3	0.85	2.52	1.88	<0.01*	−0.459
Perceived learning gains	B6.2 B6.6	0.68	1.38	1.26	0.66	−0.07
Motivation	B6.4 B6.5	0.62	2.28	1.56	< 0.01*	−0.0584
Method comparison	B6.7 B6.8	0.81	1.04	1.89	< 0.01*	0.646

°t-test.
°°Cohen's d.
*Significant.

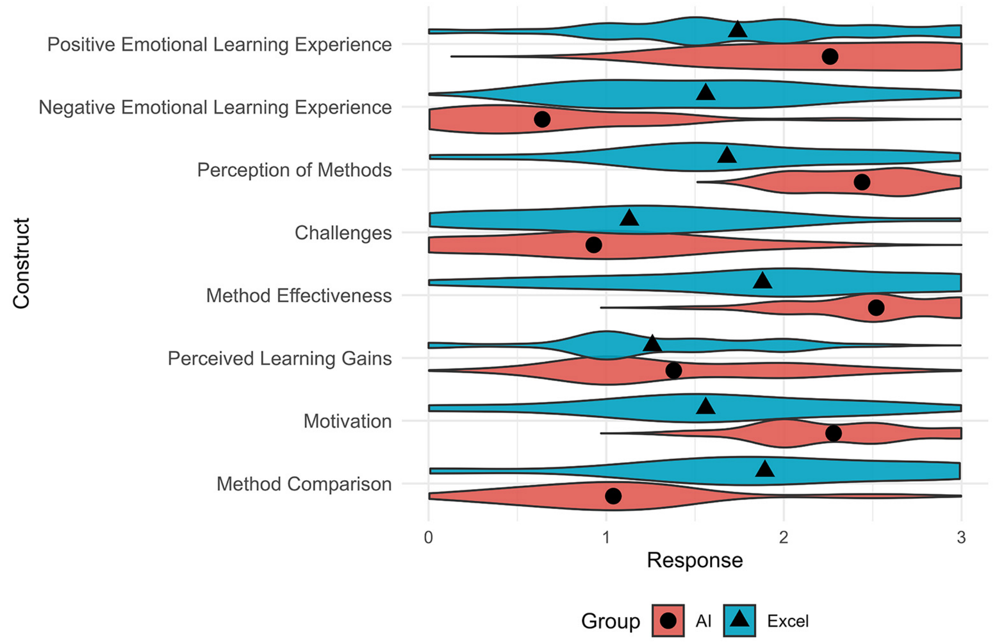

FIGURE 2
Violin plot of emotional and motivational responses in the post-test. Each plot illustrates the density of responses; black shapes represent group means. AI-Group = red; Excel group = blue.

Overall, the data indicates clear differences in emotional and motivational outcomes between the groups. The constructs *Positive Emotional Learning Experience, Perception of Methods, Method Effectiveness, Motivation*, and *Method Comparison* showed statistically significant differences between the AI and *Excel* groups, with medium to large effect sizes. In contrast, *Negative Emotional Learning Experience* was significantly lower in the AI group. Other constructs such as *Challenges* and *Perceived Learning Gains* did not show significant differences. These results reflect distinct patterns in how students experienced and evaluated the respective learning tools.

5 Discussion

This section interprets the study's findings considering the research questions and the existing literature. It considers both the learning-related outcomes and the emotional-motivational dimensions of the intervention, before reflecting on broader methodological implications and practical considerations.

5.1 Interpretation of learning outcomes (RQ1)

The first research question focused on the extent to which students benefited from the intervention in terms of content learning. This section discusses both within-group improvements and between-group comparisons.

5.1.1 Intragroup learning gains

The Wilcoxon analyses indicate that both tools supported some degree of learning, although the size and distribution of these gains varied by domain. The *Excel* group showed a moderate improvement overall, with the most notable gain occurring in the domain of data evaluation. Improvements in the *Excel* group likely reflect familiarity with the tool and the structured task design, rather than the tool's pedagogical value. No statistically significant gains emerged for the thread or spring pendulum tasks, which suggests that the structured Excel workflow may reinforce procedural aspects of data handling but offers limited support for more conceptually demanding domains. From a Cognitive Load Theory perspective *Excel's* technical interface with many interactive elements may have imposed excessive extraneous cognitive load through its manual data manipulation requirements, making the content appear complex and challenging (Sweller, 2012). Lacking adaptive feedback, students likely expended their limited working memory on extraneous load rather than germane load. Consequently, cognitive resources that could have been devoted to germane processing and conceptual understanding were instead consumed by managing operational demands. This is consistent with the relatively low Hake values in this group. Unlike *ExperiMentor*, *Excel* provided no responsive feedback, which may have contributed to lower motivation and higher frustration.

In contrast, the AI group showed significant gains across most domains, particularly in data evaluation and spring pendulum tasks. The effect sizes observed suggest that *ExperiMentor* was effective in supporting students' learning progress. Its capability to generate explanations, visualizations, and real-time feedback likely contributed to this outcome. This finding can be interpreted through Vygotsky's ZPD, where *ExperiMentor* functioned as a

MKO by providing scaffolded support tailored to students' current knowledge levels. The stronger gains in the AI group can therefore be understood as the result of learning activities that remained within students' ZPD while still extending their independent performance. The AI's capacity to offer immediate, contextualized guidance may have enabled learners to navigate tasks beyond their independent capabilities, effectively working within their ZPD. By managing cognitive load and providing adaptive scaffolding, the AI tool could have reduced extraneous demands while supporting germane cognitive processes necessary for conceptual development (Ratniyom et al., 2025). This aligns with findings from other studies that emphasize the potential of AI tools to enhance engagement, guide reasoning, and provide tailored guidance in complex tasks (Gada and Chudasana, 2024; Küchemann et al., 2024; Wen et al., 2024). The Hake indices for this group were also small in absolute terms.

Limited gains in the thread pendulum task suggest that neither tool provided sufficient conceptual scaffolding for more abstract problem domains. As prior research has shown, *ChatGPT* tends to perform better on well-structured, theory-based prompts than on tasks requiring mathematical modeling or nuanced physical interpretation (Wang et al., 2024; Revalde et al., 2025). This domain-specific limitation might explain the absence of a stronger effect in this area, despite the learning gains in the AI group. The data evaluation tasks involve more procedural elements, where progress may appear limited when initial performance is already moderate. Conversely, the thread pendulum and spring pendulum require abstract conceptual reasoning, and low normalized gains likely reflect the inherent difficulty of this domain rather than the ineffectiveness of either tool.

The findings reflect the differentiated potential of each tool. While *Excel* remains a reliable tool for basic data processing, it appears to offer limited support for learning gains in conceptual domains without additional instructional input. In contrast, the AI-assisted approach demonstrates potential for scaffolding more complex reasoning and analytical thinking. Other studies have shown the practicality and success of AI instructors in academic environments (Kestin et al., 2025). This further resonates with previous findings on the motivational and cognitive benefits of intelligent tutoring systems (Liang et al., 2023; Lo, 2023). However, these benefits were not uniform across all content areas, indicating that AI tools should not be viewed as a universal solution but rather as one component within a broader instructional strategy (Chounta et al., 2022).

5.1.2 Intergroup learning gains

While both groups improved from pre- to post-test, the *ANCOVA* results indicated no statistically significant differences in learning gains between the AI-assisted and *Excel*-based approaches. Across all learning domains, the group by measurement interaction was non-significant and the associated partial η^2 values were negligible, showing that the choice of tool did not meaningfully alter learning progress when initial performance was controlled. The covariate explained a substantial proportion of the variance in post-test scores, consistent with the well-established link between prior knowledge and later achievement. Assumption checks indicated that the relationship between pre- and post-test performance varied slightly across measurement points, suggesting that regression slopes were not fully homogeneous and that adjusted means should be interpreted with some caution. Nevertheless, the overall pattern remained clear: after controlling for baseline differences, both tools produced comparable learning outcomes.

Similarly, the *Hake Index* values were generally low across both groups. This does not necessarily indicate that students made negligible progress. Rather, they reflect the domain-specific nature of the tasks. Within this context, the AI tool's broader intragroup improvements and higher effect sizes reflect enhanced support for reasoning processes without producing measurable group-level advantages in the statistical analysis. This is consistent with the *ANCOVA* findings, where the tool variable explained only a minimal fraction of the variance and further with prior research noting that AI tools like *ChatGPT* tend to perform more reliably on conceptual or theory-based problems than on open-ended or quantitative tasks (Wang et al., 2024; Revalde et al., 2025).

Despite higher emotional engagement, the AI tool did not lead to significantly better performance outcomes compared to *Excel* under structured conditions. This discrepancy suggests a form of affective-cognitive dissonance, where learners feel more supported and confident without necessarily achieving superior objective learning gains. Because AI smooths the learning path by providing immediate answers and structural support, students may perceive the content as easier and themselves as more competent than their independent performance warrants. The tool effectively bridges the gap between the learner's ability and the task requirements, potentially masking the cognitive effort required to truly master the material. This aligns with earlier research that points to the context-dependence of AI tool effectiveness. While some studies report that tools like *ChatGPT* can outperform traditional platforms like *Excel* on specific tasks (Halaweh, 2023), other research emphasizes that these advantages often depend on how AI is embedded into the learning environment, the nature of the task, and the learner's ability to engage critically with the tool's output (Bitzenbauer, 2023; Gada and Chudasana, 2024).

The structured instructional design may have compensated for differences in tool affordances by providing explicit scaffolding that managed cognitive load for both groups. While *ExperiMentor* acted as an MKO within students' ZPD, the *Excel* group received comparable support through carefully designed task structures. In both cases, learners were provided with scaffolding that helped them operate near the upper boundary of their ZPD, which may explain the absence of pronounced between group differences in performance. This suggests that the presence of scaffolding, whether technology-based or instructionally embedded, can be more influential than the specific tool used.

Overall, these results show that AI tools do not automatically lead to superior learning outcomes compared to traditional methods. Instead, the effectiveness of any tool depends on how it is implemented. Similar to the findings of Halaweh (2023), this study found a not significant but noticeable performance advantage in the AI group. Both tools were embedded within a guided setting, which may explain the absence of major differences in learning outcomes.

5.2 Interpretation of emotional and motivational outcomes (RQ2)

Understanding students' attitudes regarding the used tools is key to evaluating its broader educational impact. This section reflects on participants' emotional and motivational responses, both before and after the tasks.

5.2.1 Pre-intervention attitudes

Before the intervention began, students in both groups held similar attitudes toward working with digital tools and evaluating experimental data. This was reflected in nearly identical mean scores across all items and negligible effect sizes. These findings indicate that neither group had an emotional or motivational advantage at the outset.

This comparability suggests that the emotional and motivational differences observed later are less likely to stem from pre-existing preferences or dispositions. Rather, they appear to be linked to the experience of working with the respective tools during the task itself.

5.2.2 Post-intervention attitudes

After the intervention, clear distinctions, as shown in Figure 2 emerged in how participants experienced the two tools. Learners in the AI group reported significantly more positive emotional learning experiences, including greater enjoyment and a stronger sense of success. These differences point to a learning environment that better supports basic psychological needs. The emotional advantages of the AI tool suggest that its interactivity and responsiveness fostered engagement and reduced negative emotional learning experiences like frustration. *ExperiMentor* appears to have better satisfied students' basic psychological needs for autonomy, relatedness and competence (Deci and Ryan, 2008), allowing more positive emotional learning experiences the students through personalized, responsive interactions (Chiu, 2025), adaptive feedback and conversational cooperative-seeming engagement (Ratniyom et al., 2025). Environments that support these needs promote better learning compared to environments that blocks these needs (Ryan and Deci, 2000). The AI tool's ability to provide individualized support likely enhanced students' sense of competence while its interactive nature fostered a perception of cooperation, even in a digital environment. As visualized in Figure 2, the distributional differences between groups further highlight these effects, particularly in constructs such as motivation and emotional experience.

These findings align with earlier studies suggesting that AI tools, especially those that simulate human-like interactions such as chatbots, can foster emotional engagement and reduce anxiety through scaffolding and adaptive feedback (Popenici and Kerr, 2017; Liang et al., 2023; Farrokhnia et al., 2024). At the same time, these outcomes should be interpreted with care. The emotional and motivational advantages may partly reflect novelty effects or the perception of receiving personalized assistance rather than a demonstrable educational benefit (Long et al., 2024).

However, attributing these differences solely to novelty risks oversimplification. It is critical to reflect on the role of feedback quality as a distinct design variable. The experimental setup compared a tool offering responsive scaffolding through *ExperiMentor* against *Excel*, offering no adaptive feedback. Consequently, the observed motivational gap may stem less from the AI itself and more from the discrepancy in interactivity. The static nature of the Excel condition represents a lack of feedback, whereas the AI condition provided continuous, conversational support. The positive outcomes in the AI group likely reflect the benefits of adaptive feedback mechanisms, which could theoretically be implemented via non-AI means, rather than the unique properties of AI alone.

The *Excel* group reported higher levels of negative emotional responses such as stress and frustration, emotional reactions can be seen as a consequence of elevated extraneous load. This contrast was associated with a very large effect size, reinforcing the interpretation that the AI-supported environment may have offered greater emotional comfort. Again, the static nature of *Excel* may have failed to manage extraneous cognitive load effectively, forcing students to simultaneously handle technical operations and conceptual understanding, contributing to increased frustration and diminished emotional engagement. The AI tool's ability to respond flexibly to user input and offer contextual hints may have provided users with a sense of support, thereby reducing uncertainty during problem-solving, a benefit noted in recent research on AI-supported learning (Liang et al., 2023; Gada and Chudasana, 2024; Mustofa et al., 2024; Tong et al., 2024).

Students' perception of the method also differed significantly between the groups. Participants rated the AI-based environment as more interesting, intuitive, and useful for future applications, with a very large effect size, suggesting a strong difference in tool preference. In SDT terms, this preference can be interpreted as a sign that students experienced the AI environment as more autonomy supportive and competence enhancing than the *Excel* setting. This reflects findings in the literature that associate AI tools with a higher degree of user satisfaction and perceived usefulness (Lo, 2023). Similarly, participants perceived the AI-supported method as more effective, particularly in helping them navigate complex or unfamiliar tasks, an interpretation supported by research suggesting that AI tools can strengthen learners' self-perceived performance (Wu and Yu, 2024).

Despite differences in engagement, both groups perceived the tasks as equally challenging, indicating that tool preference did not stem from task difficulty. This suggests that the more positive experience in the AI group was not due to an easier task, but perhaps due to the type and quality of feedback received.

While the two groups reported comparable perceived learning gains, motivation levels were notably higher in the AI group. This difference was also reflected in a large effect size highlighting a meaningful motivational advantage associated with the AI tool. This finding is consistent with research indicating that AI tools can increase learner motivation through interactivity and responsiveness (Hanum Siregar et al., 2023; Liang et al., 2023; Wu and Yu, 2024) and is consistent with Self Determination Theory, which attributes higher quality motivation to environments that satisfy learners' needs for autonomy, competence and relatedness

(Deci and Ryan, 2008). The motivational advantage observed aligns with Self-Determination Theory's emphasis on intrinsic motivation emerging from satisfaction of basic psychological needs (Deci and Ryan, 2008). The AI tool's personalized responses likely enhanced perceived autonomy by allowing students to explore at their own pace, while its supportive feedback fostered competence perceptions. Furthermore, as an MKO, the AI provided scaffolding within students' ZPD. Similarly, through creating a climate of affective encouragement and acknowledgment, another study showed learners perceived themselves as capable of achieving their highest capabilities (La Ossa et al., 2024). Nevertheless, the disconnect between this high motivation and the comparable test scores reinforces the presence of an affective-cognitive dissonance. Students experienced a greater sense of competence and autonomy facilitated by *ExperiMentor*, even if this did not translate into superior quantitative performance.

Finally, responses to the Method Comparison construct showed a strong preference for the AI tool even among *Excel* users. The students across both groups saw clear advantages in the AI environment and believed it could have helped them complete the task more effectively, this could be by having reduced manual calculation effort or possible time savings by not having to manually navigate in the spreadsheet of Excel. Further, the additional feedback support and guidance embedded in *ExperiMentor*, which the *Excel* group lacked could be a reason for this tool preference. Lastly, since the AI handled the calculations, preparations and presentation, the extraneous cognitive load of the students might be reduced, which may lead to focus on the conceptual understanding and therefore preferring the AI tool. This redistribution of cognitive resources is exactly what Cognitive Load Theory describes as a shift from extraneous toward germane load. This cognitive redistribution, combined with the AI's role as an adaptive MKO operating within learners' ZPD, may explain the consistent preference across both groups.

The patterns displayed in Figure 2 confirm the differences identified in the quantitative analysis. The more concentrated distributions in the AI group suggest a consistently more positive experience, particularly regarding emotional support and perceived tool effectiveness. The *Excel* group showed greater variation, with several constructs demonstrating broader spreads toward lower ratings. This reinforces the interpretation that the AI-assisted environment offered a more uniformly positive experience.

5.3 Methodological reflections

This study explored the effects of AI-assisted vs. traditional data analysis using a randomized pre-post design in a real educational setting. It has limitations in duration, design, and sample characteristics. The intervention was relatively short, allowing only limited time for participants to fully engage with the tools. Consequently, the study may have captured only early-stage effects, rather than sustained learning trajectories. The pre- and post-tests focused on factual knowledge and practical application. However, closed-ended tasks may have missed complex learning forms such as conceptual change or skill transfer. The *Hake Index* and effect sizes provided useful information about average learning gains, but averages can mask individual differences. This study compared two different learning environments, not just two tools. The *Excel* group worked with a spreadsheet without feedback beyond written instructions. The AI group interacted with *ExperiMentor*, which responded to input, gave hints, offered feedback, and visualized results. This difference in interactivity shaped both performance and the learning experience. Although groups were seated separately and instructed not to interact, the shared physical space represents a potential source of cross-group influence. While incidental communication is unlikely to have substantially affected outcomes, it cannot be ruled out entirely.

5.4 Implications and limitations

The findings offer insights into the use of AI-assisted tools like *ExperiMentor* in physics education. However, several limitations must be considered when interpreting the results and assessing their broader relevance.

The relatively small, homogeneous sample of 50 student teachers from a single university restricts generalizability. The participants shared institutional context, prior knowledge, and cultural background limits the ability to draw conclusions about how *ExperiMentor* might function across diverse learning environments, educational systems, or disciplinary contexts beyond physics.

The limited impact observed in this study can be attributed to its single-session design. Research demonstrated that ChatGPT's influence on academic achievement becomes minimal with short-term applications under 1 week (Wang and Fan, 2025). Therefore, the short intervention duration limited deep tool exploration and made it difficult to evaluate lasting learning gains or long-term motivational effects. The single-session design offers limited insight into longitudinal outcomes or transfer effects to real classroom settings, where teachers would need to apply learned skills repeatedly and adapt them to varying student needs. Furthermore, the intervention relied on pre- and post-performance comparisons and self-reported emotional-motivational variables. Self-report instruments are prone to response biases (Kreitchmann et al., 2019), so emotional-motivational outcomes should be viewed as subjective indicators rather than objective evidence. Also, the study did not differentiate between intrinsic and extrinsic motivation. Future studies would benefit from more differentiated motivational instruments and qualitative data. Also aggregated metrics like mean scores are useful for identifying trends but may mask learner differences. Future research should include distributional analyses and subgroup comparisons to better understand variation across learners.

The tools differed substantially in their nature. The AI group used a dynamic, responsive chatbot with immediate feedback and contextual assistance, while the *Excel* group worked with a static spreadsheet without external help. This disparity likely influenced both performance and perceptions. *ExperiMentor* acted as a backend interpreter for Python-based queries, enabling interactive assistance that *Excel* could not match. Additionally, *ExperiMentor's* outputs were not systematically validated for scientific correctness during the intervention. While researcher oversight indicated plausible and coherent responses, it cannot be ruled out that misleading or incorrect information influenced some learners' interpretations. Further, observed motivational advantages may reflect novelty effects. Students may have rated the AI tool more favorably because of its newness. Recent studies suggest, that the sense of novelty regarding AI remains intact even after longer exposure (Long et al., 2024).

To overcome these limitations and build a more robust understanding of AI-mediated learning in physics education, future work could explore several directions:

Longitudinal studies could examine whether AI tool advantages are sustained across multiple exposures and whether skills developed with *ExperiMentor* transfer to real classroom teaching contexts. Multi-session interventions spanning several weeks or months would clarify whether motivational benefits persist and whether performance gains consolidate over time. Also, mixed-methods approaches combining quantitative performance measures with qualitative interviews and interaction log analyses could illuminate the source for motivational differences and reveal how learners engage with AI assistance. Diverse samples and contexts including in-service teachers and participants from multiple institutions and countries would strengthen generalizability claims. Comparative studies across different physics topics and educational levels would clarify boundary conditions for AI tool effectiveness.

Refined motivational measurement using instruments that differentiate intrinsic and extrinsic motivation, along with measures of self-efficacy, autonomy and competence, would provide more nuanced understanding of affective outcomes. Subgroup analyses examining how learners with different prior knowledge, attitudes toward technology or learning preferences respond to AI assistance would reveal for whom these tools work best. Further, design variations comparing *ExperiMentor* to other support strategies like human tutoring, enhanced static materials or hybrid approaches and testing different AI interaction styles would help isolate which features drive observed benefits. *ExperiMentor's* interactive feedback could be leveraged through gamification elements like achievement systems or progress visualization.

6 Conclusion

While both AI-supported (*ExperiMentor*) and traditional tools (*Excel*) facilitated measurable learning improvements, *ExperiMentor* led to significantly higher emotional and motivational engagement. However, these benefits did not translate into superior conceptual gains, suggesting that AI functions best as a complementary support rather than a standalone solution.

The observed engagement likely stems from the AI functioning as a More Knowledgeable Other, offering scaffolding that keeps learners within their Zone of Proximal Development. Furthermore, from a Cognitive Load Theory perspective, the tool's automation seems to have managed extraneous load (Sweller, 2012), while its interactivity satisfied Self-Determination Theory needs for autonomy and competence (Ryan and Deci, 2000). The AI-supported approach resulted in greater enjoyment, motivation, and perceived effectiveness. These results align with previous research indicating that AI tools promote engagement through interactive feedback and user-centered design through interactivity, adaptability and timely feedback (Popenici and Kerr, 2017; Liang et al., 2023; Farrokhnia et al., 2024). The observed motivation boost may partly reflect novelty effects or perceived usability advantages. This aligns with similar criticisms raised in educational research, such as by Buchner and Kerres (2023), who argue that media comparison studies often mask the conditions under which specific tools are effective. While the AI-based tool *ExperiMentor* showed motivational advantages, performance outcomes did not differ significantly from the *Excel* condition. This underscores the need to move beyond the question of whether AI "works better" than traditional tools, and toward understanding under what conditions, for which learners, and in what instructional designs AI tools provide added value.

These insights reinforce the idea that digital tools are not pedagogical strategies in and of themselves. Their effectiveness depends on how they are embedded within instructional settings, how much support they offer, and how well they align with educational goals. In this study, the AI and *Excel* groups not only differed in interface but also in the type and degree of feedback and guidance provided, which makes it difficult to isolate the effect of AI from the broader learning environment.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for this study, as the data collected were anonymized, and the survey did not involve sensitive or health-related topics. Participants were not exposed to any risk or burden, and the information gathered was and is not personally identifiable. Additionally, the survey was conducted solely for research purposes related to general study conditions, without any impact on participants' academic assessment, rights, or well-being. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JH: Data curation, Investigation, Conceptualization, Methodology, Writing – review & editing, Visualization, Writing – original draft, Formal analysis. JL: Investigation, Writing – review & editing, Data curation. SB-G: Supervision, Formal analysis, Writing – review & editing, Methodology, Investigation, Conceptualization. AB: Supervision, Software, Funding acquisition, Resources, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The ComeMINT project was funded by the Federal Ministry of Education and Research as part of the BMBF funding line "Competence Centers for digital and digitally supported teaching in schools and continuing education in the STEM Field," funding code 01JA23M06G.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors AB, SB-G declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. This manuscript was linguistically developed and revised for better readability using OpenAI ChatGPT 4o, Anthropic Claude 3.7 Sonnet and the translation tool DeepL.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feduc.2026.1719670/full#supplementary-material

References

Alarbi, K., Halaweh, M., Tairab, H., Alsalhi, N. R., Annamalai, N., Aldarmaki, F., et al. (2024). Making a revolution in physics learning in high schools with ChatGPT: a case study in UAE. *Eurasia J. Math. Sci. Tech. Ed.* 20:em2499. doi: 10.29333/ejmste/14983

Bacia, E., Belafi, C., Dohmen, D., Klingemann, J., Kummer, B., and Müller, F. (2024). *Innovatives Lernen mit Intelligenten Tutoriellen Systemen: Eine Analyse der bildungspolitischen Gelingensbedingungen [Innovative Learning with Intelligent Tutoring Systems. An Analysis of the Conditions for Educational Policy Success].* In German. Available online at: https://www.vodafone-stiftung.de/wp-content/uploads/2024/05/vsd_leitfaden_innovative_lernumgebungen.pdf

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). "On the dangers of stochastic parrots," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA: ACM), 610–623. doi: 10.1145/3442188.3445922

Bitzenbauer, P. (2023). ChatGPT in physics education: a pilot study on easy-to-implement activities. *Cont. Ed. Technol.* 15:ep430. doi: 10.30935/cedtech/13176

Bortz, J., and Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler [Statistics for Human and Social Scientists]*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-12770-0

Buchner, J., and Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Comp. Educ.* 195:104711. doi: 10.1016/j.compedu.2022.104711

Chen, L., Chen, P., and Lin, Z. (2020). Artificial intelligence in education: a review. *IEEE Access* 8, 75264–75278. doi: 10.1109/ACCESS.2020.2988510

Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interact. Learn. Environ.* 33, 3225–3229. doi: 10.1080/10494820.2025.2514372

Chounta, I-. A., Bardone, E., Raudsep, A., and Pedaste, M. (2022). Exploring teachers' perceptions of artificial intelligence as a tool to support their practice in estonian K-12 education. *Int. J. Artif. Intell. Educ.* 32, 725–755. doi: 10.1007/s40593-021-00243-5

Coletta, V. P., and Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Phys. Rev. Phys. Educ. Res.* 16:010108. doi: 10.1103/PhysRevPhysEducRes.16.010108

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Dahlkemper, M. N., Lahme, S. Z., and Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality. *Phys. Rev. Phys. Educ. Res.* 19:010142. doi: 10.1103/PhysRevPhysEducRes.19.010142

Deci, E. L. (1975). *Intrinsic Motivation*. New York, NY: Plenum Press. doi: 10.1007/978-1-4613-4446-9

Deci, E. L., and Ryan, R. M. (2008). Self-determination theory: a macrotheory of human motivation, development, and health. *Can. Psycholo.* 49, 182–185. doi: 10.1037/a0012801

Ding, L., Li, T., Jiang, S., and Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *Int. J. Educ. Technol. High. Educ.* 20:63. doi: 10.1186/s41239-023-00434-1

Döring, N., Bortz, J., Pöschl, S., Werner, C. S., Schermelleh-Engel, K., Gerhard, C., et al. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research Methods and Evaluation in the Social and Human Sciences]*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-41089-5

Essel, H. B., Vlachopoulos, D., Essuman, A. B., and Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: receiving instant responses from AI-based conversational large language models (LLMs). *Comp. Educ. Artif. Intell.* 6:100198. doi: 10.1016/j.caeai.2023.100198

Farrokhnia, M., Banihashem, S. K., Noroozi, O., and Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations Educ. Teach. Int.* 61, 460–474. doi: 10.1080/14703297.2023.2195846

Gada, T., and Chudasana, S. (2024). Impact of artificial intelligence on student attitudes, engagement, and learning. *IRJMETS* 6, 2695–2698. doi: 10.56726/IRJMETS56247

Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev. Educ. Res.* 42:237. doi: 10.3102/00346543042003237

Gregorcic, B., and Pendrill, A-. M. (2023). ChatGPT and the frustrated Socrates. *Phys. Educ.* 58:35021. doi: 10.1088/1361-6552/acc299

Guo, F., Zhang, L., Shi, T., and Coates, H. (2025). Whether and when could generative ai improve college student learning engagement? *Behav. Sci.* 15:1011. doi: 10.3390/bs15081011

Hake, R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* 66, 64–74. doi: 10.1119/1.18809

Hake, R. (2002). Lessons from the physics education reform effort. *Conserv. Ecol.* 5:28. doi: 10.5751/ES-00286-050228

Halaweh, M. (2023). ChatGPT in education: strategies for responsible implementation. *Contemp. Educa. Technol.* 15:ep421. doi: 10.30935/cedtech/13036

Hanum Siregar, F., Hasmayni, B., and Lubis, A. H. (2023). The analysis of chat GPT usage impact on learning motivation among scout students. *Int. J. Res. Rev.* 10, 632–638. doi: 10.52403/ijrr.20230774

Hedderich, J., and Sachs, L. (2016). *Angewandte Statistik: Methodensammlung mit R [Applied Statistics. Collection of Methods with R]*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individual Differ.* 103:102274. doi: 10.1016/j.lindif.2023.102274

Kechel, J., and Wodzinski, R. (2015). "Methoden zur Erfassung von Schwierigkeiten bei Schülerexperimenten [Variety of Prerequisites in Science Education]," *in Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht: Tagungsband Jahrestagung in Bremen 2014 [Heterogeneity and Diversity - Conference Proceedings, Annual Meeting in Bremen 2014]*, ed. S. Bernholt (Kiel). Available online at: https://gdcp-ev.de/wp-content/tb2015/TB2015_304_Kechel.pdf

Kerby, D. S. (2014). The simple difference formula: an approach to teaching nonparametric correlation. *Compr. Psychol.* 3, 2165–2228. doi: 10.2466/11.IT.3.1

Kestin, G., Miller, K., Klales, A., Milbourne, T., and Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Sci. Rep.* 15:17458. doi: 10.1038/s41598-025-97652-6

Kieser, F., Wulff, P., Kuhn, J., and Küchemann, S. (2023). Educational data augmentation in physics education research using ChatGPT. *Phys. Rev. Phys. Educ. Res.* 19:020150 doi: 10.1103/PhysRevPhysEducRes.19.020150

Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Phys. Rev. Phys. Educ. Res.* 19:15. doi: 10.1103/PhysRevPhysEducRes.19.010132

Kreitchmann, R. S., Abad, F. J., Ponsoda, V., Nieto, M. D., and Morillo, D. (2019). Controlling for response biases in self-report scales: forced-choice vs. psychometric modeling of likert items. *Front. Psychol.* 10:2309. doi: 10.3389/fpsyg.2019.02309

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., et al. (2023a). Unreflected acceptance: investigating the negative consequences of ChatGPT-assisted problem solving in physics education. *arXiv*. doi: 10.3233/FAIA240195

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., et al. (2023b). Challenges and opportunities of moderating usage of large language models in education. *arXiv*. doi: 10.1145/3701571.3701590

Küchemann, S., Steinert, S., Kuhn, J., Avila, K., and Ruzika, S. (2024). Large language models-valuable tools that require a sensitive integration into teaching and learning physics. *Phys. Teach.* 62, 400–402. doi: 10.1119/5.0212374

Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M. D., Păun, D., et al. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability* 13:10424. doi: 10.3390/su131810424

La Ossa, L. A. M. D., Garzón, N. N., Pulido, C. P., Herrán, B. E. G., and Hoyos, J. H. (2024). Evaluation of artificial intelligence as a source of motivation in the teaching of chemistry: a study with tenth grade students at the adventist school of ibagué. *Esiculture* 8, 2193–2208. doi: 10.70082/esiculture.vi.1721

Liang, Y., Zou, D. i., Xie, H., and Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learn. Environ.* 10:52. doi: 10.1186/s40561-023-00273-7

Lin, X-. F., Chen, L., Chan, K. K., Peng, S., Chen, X., Xie, S., et al. (2022). Teachers' perceptions of teaching sustainable artificial intelligence: a design frame perspective. *Sustainability* 14:7811. doi: 10.3390/su14137811

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Educ. Sci.* 13:410. doi: 10.3390/educsci13040410

Long, T., Gero, K. I., and Chilton, L. B. (2024). Not just novelty: a longitudinal study on utility and customization of an AI workflow. *arXiv*. doi: 10.1145/3643834.3661587

López-Martín, E., and Ardura, D. (2023). The effect size in scientific publication. *EducXX1* 26, 9–17. doi: 10.5944/educxx1.36276

Low, A., and Kalender, Z. Y. (2023). Data dialogue with ChatGPT: using code interpreter to simulate and analyse experimental data. *arXiv*. doi: 10.48550/arXiv.2311.12415

Mahligawati, F., Allanas, E., Butarbutar, M. H., and Nordin, N. A. N. (2023). Artificial intelligence in physics education: a comprehensive literature review. *J. Phys. Conf. Ser.* 2596:12080. doi: 10.1088/1742-6596/2596/1/012080

McKagan, S., Sayre, E., and Madsen, A. (2022). Normalized gain: What is it and when and how should I use it? Available online at: https://www.physport.org/recommendations/Entry.cfm?ID=93334 (Accessed September 14, 2025).

Miguel-Alonso, I., Checa, D., Guillen-Sanz, H., and Bustillo, A. (2024). Evaluation of the novelty effect in immersive virtual reality learning experiences. *Virtual Reality* 28:27. doi: 10.1007/s10055-023-00926-5

Mustofa, H. A., Bilad, M. R., and Grendis, N. W. B. (2024). Utilizing AI for physics problem solving: a literature review and ChatGPT experience. *Jurnal. Kependidikan. Fisika* 12:78. doi: 10.33394/j-lkf.v12i1.11748

Neyer, F. J., Felber, J., and Gebhardt, C. (2016). *Kurzskala Technikbereitschaft [Short Scale for Technology Commitment]*. Mannheim, Germany: ZIS - GESIS Leibniz Institute for the Social Sciences.

Ng, D. T. K., Tan, C. W., and Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: a pioneering pilot study. *Brit. J. Educ. Tech.* 55, 1328–1353. doi: 10.1111/bjet.13454

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., et al. (2023). GPT-4 technical report. *arXiv*. doi: 10.48550/arXiv.2303.08774

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., et al. (2022). Training language models to follow instructions with human feedback. *arXiv*. doi: 10.48550/arXiv.2203.02155

Padma, C., and Rama, C. (2022). A study of artificial intelligence in education system & role of AI in Indian education sector. *Int. J. Sci. Res. Eng. Manage.* 6. doi: 10.55041/IJSREM11848

Popenici, S. A. D., and Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res. Pract. Technol. Enhanc. Learn.* 12:22. doi: 10.1186/s41039-017-0062-8

Ratniyom, J., Panmas, W., Rattanakorn, P., and Tientongdee, S. (2025). Generative AI-assisted phenomenon-based learning: exploring factors influencing competency in constructing scientific explanations. *Eur. J. Ed. Res.* 14, 1087–1103. doi: 10.12973/eu-jer.14.4.1087

Revalde, G., Zholdakhmet, M., Abola, A., and Murzagaliyeva, A. (2025). Can ChatGPT pass a physics test? *Tech. Know. Learn.* 30, 2459–2478. doi: 10.1007/s10758-025-09814-0

Rigopouli, K., Kotsifakos, D., and Psaromiligkos, Y. (2025). Vygotsky's creativity options and ideas in 21st-century technology-enhanced learning design. *Educ. Sci.* 15:257. doi: 10.3390/educsci15020257

Roll, I., and Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *Int. J. Artif. Intell. Educ.* 26, 582–599. doi: 10.1007/s40593-016-0110-3

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68

Salas-Pilco, S., Xiao, K., and Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: a systematic review. *Educ. Sci.* 12:569. doi: 10.3390/educsci12080569

Santos, R. P. D. (2023). Enhancing physics learning with ChatGPT, bing chat, and bard as agents-to-think-with: a comparative case study. *arXiv*. doi: 10.2139/ssrn.4478305

Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M. (2010). Is it really robust? *Methodology* 6, 147–151. doi: 10.1027/1614-2241/a000016

Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (2024). *Large Language Models und ihre Potenziale im Bildungssystem: Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz [Large Language Models and their potential in the education system. Impulse paper of the Standing Scientific Commission (SWK) of the Standing Conference of the Ministers of Education and Cultural Affairs]*. Bonn: Ständige Wissenschaftliche Kommission der Kultusministerkonferenz.

Sweller, J. (2012). "Implications of cognitive load theory for multimedia learning," in *The Cambridge Handbook of Multimedia Learning*, ed. R. Mayer (Cambridge: Cambridge University Press), 19–30. doi: 10.1017/CBO9780511816819.003

Tong, D., Tao, Y., Zhang, K., Dong, X., Hu, Y., Pan, S., et al. (2024). Investigating ChatGPT4's performance in solving physics problems and its potential implications for education. *Asia Pac. Educ. Rev.* 25, 1379–1389. doi: 10.1007/s12564-023-09913-6

Vincent-Lancrin, S., and van der Vlies, R. (2020). *Trustworthy Artificial Intelligence (AI) in Education: Promises and Challenges. OECD Education Working Papers No. 218.* Paris: OECD.

Vodafone Stiftung Deutschland gGmbH. (2023). *Aufbruch ins Unbekannte: Schule in Zeiten von künstlicher Intelligenz und ChatGPT [Into the Unknown. Schools in Times of Artificial Intelligence and ChatGPT]*. Deutschland gGmbH: Vodafone Stiftung.

Wang, K., Burkholder, E., Wieman, C., Salehi, S., and Haber, N. (2024). Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Front. Educ*. 8:1330486. doi: 10.3389/feduc.2023.1330486

Wang, J., and Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanit. Soc. Sci. Commun.* 12. doi: 10.1057/s41599-025-04787-y

Wen, Z., Bai, E., and Li, M. (2024). An evaluation of the impact of artificial intelligence on university students' learning. *JID* 6, 22–25. doi: 10.54097/f3vvmp82

Winkelmann, J., Freese, M., and Strömmer, T. (2021). Schwierigkeitserzeugende Merkmale im Physikunterricht [Difficulty-Inducing Features in Physics Education]. *Prog. Sci. Edu.* 5, 6–23. doi: 10.25321/prise.2022.1168

Wollschläger, D. (2014). *Grundlagen der Datenanalyse mit R [Basics of Data Analysis with R]*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-45507-4

Wu, R., and Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *Brit. J. Educ. Tech*. 55, 10–33. doi: 10.1111/bjet.13334

Young, J. Q., Thakker, K., John, M., Friedman, K., Sugarman, R., van Merriënboer, J. J. G., et al. (2021). Exploring the relationship between emotion and cognitive load types during patient handovers. *Adv. Health Sci. Educ. Theory Pract.* 26, 1463–1489. doi: 10.1007/s10459-021-10053-y

Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education - where are the educators? *Int. J. Educ. Technol. High Educ.* 16:39. doi: 10.1186/s41239-019-0171-0

7. Diskussion & Synthese

7.1. Technologische Scaffolding-Strukturen zur Förderung digitaler Souveränität

Die Entwicklung digitaler Souveränität lässt sich mit Vygotskys Konzept der Zone der proximalen Entwicklung beschreiben, in der Lernende durch einen More Knowledgeable Other so unterstützt werden, dass sie Aufgaben bewältigen, die ihre aktuelle Kompetenz überschreiten (Leong & Bodrova, 1996; Pedapati, 2022; Rigopouli et al., 2025). In den vorliegenden Studien übernimmt KI die Rolle eines digitalen MKO, indem sie komplexe Aufgaben strukturiert sowie individuelle Erklärungen und Feedback bereitstellt (Henze et al., 2026; Henze et al., 2025). Dadurch sinkt die sachfremde kognitive Belastung und kognitive Kapazitäten für konzeptionelles Verständnis können freigesetzt werden (Sweller, 2012). Die Teilnehmenden berichteten ein deutlich gesteigertes Gefühl der eigenen Kompetenzen und mehr Motivation, auch wenn der objektive Wissenszuwachs im Vergleich zu traditionellen Vorgehensweisen nicht signifikant höher ausfiel (Henze et al., 2026; Henze et al., 2025).
Diese Schlussfolgerungen machen potenzielle Grenzen des KI-Scaffoldings sichtbar. Die empirischen Ergebnisse zeigen, dass Studierende mit KI-Unterstützung zwar komplexe Aufgaben lösen und der Lernzuwachs innerhalb der Gruppe gegeben ist, sich dieser aber von dem Lernzuwachs der Kontrollgruppe nicht signifikant unterscheidet (Henze et al., 2026). Es ist möglich, dass die KI den Lösungsweg so stark vorbereitet, dass kognitive Herausforderungen teilweise wegfallen.
Die Studienergebnisse machen zudem deutlich, dass technisches Scaffolding ohne sozialen Austausch nur selektive Kompetenzen fördert. Erst wenn Lernende und Lehrende gemeinsam über Funktionsweisen, Risiken und Fehleinschätzungen von KI reflektieren, stabilisieren sich komplexe und insbesondere ethische Dimensionen von AI Literacy (Henze et al., 2025). Digitale Souveränität entsteht folglich nicht aus der bloßen Nutzung von KI, sondern aus der Fähigkeit, automatisiertes Scaffolding mit menschlichem, dialogischem Scaffolding zu kombinieren (Krupka, 2020; Stubbe, 2017; Thimm, 2023; Thyssen et al., 2023). Eine mögliche Lösung stellt eine duale MKO-Struktur dar, in der KI Arbeitsprozesse strukturiert und entlastet, während menschliche Akteur:innen Kontext, Bewertung und kritischen Austausch bereitstellen. Diese Kombination verlangt einen didaktischen Rahmen, der technisches und soziales Scaffolding vereint.

7.2. Das 6E-Modell als strukturierendes Ordnungsprinzip

Für die Lehrer:innenbildung stellt sich die Frage, wie ein didaktischer Rahmen gestaltet wird, der die zuvor diskutierte, duale MKO-Struktur systematisch nutzt. Das 6E-Modell (Henze et al., 2022) bietet hierfür eine geeignete Struktur, weil es phasenweise organisiert, wann KI primär zur Entlastung und Strukturierung eingesetzt werden kann und wann der Fokus auf Reflexion und Bewertungsprozessen liegt. Es basiert auf dem 5E-Modell nach Bybee (2009; 2006) und überträgt dessen konstruktivistische Ansätze auf technisches und soziales Scaffolding.

In der **Engage-Phase** wird die motivationale Grundlage durch Probleme oder gesellschaftliche Fragestellungen geschaffen, an die spätere Erfahrungen mit KI-Werkzeugen anschließen können. In dieser Phase ist vor allem die professionelle Handlungsfähigkeit der Lehrkraft zentral, während KI hier nur eine untergeordnete Funktion, beispielsweise als Impulsgeberin, erfüllt.

In der **Explore-Phase** steht das niedrigschwellige Erproben von zur Verfügung stehenden Werkzeugen im Vordergrund, sodass Berührungsängste abgebaut und positive Nutzungserfahrungen gesammelt werden. Dies stellt eine Voraussetzung für die Bereitschaft zur vertieften Auseinandersetzung dar (Henze et al., 2022).

In der **Explain-Phase** werden die in der Explore-Phase gemachten Erfahrungen mit KI fachlich und didaktisch eingeordnet. Hier ist gezieltes Scaffolding durch die Lehrperson entscheidend, um die Funktionsweisen und Grenzen von KI transparent nachvollziehbar zu machen. Die Explain-Phase leistet damit einen wesentlichen Beitrag zur Entwicklung digitaler Souveränität, da sie Orientierung bietet sowie Fehlannahmen korrigiert.

In der **Elaborate-Phase** zeigt sich, wie KI-gestützte Ausbildungsformate die Handlungsfähigkeit angehender Lehrkräfte stärken können. In den vorliegenden Studien gestaltet sich diese Phase so, dass Studierende mit Hilfe von Chatbots Unterrichtsentwürfe entwickeln (Henze et al., 2025) oder experimentelle Daten auswerten und dabei unmittelbares, adaptives Feedback erhalten (Henze et al., 2026). Damit wird deutlich, dass generative KI Lernprozesse vertiefen und fachliche Auseinandersetzungen unterstützen kann, was sich sowohl in verbesserten Lernergebnissen als auch in positiven Einschätzungen ihrer Nützlichkeit durch Studierende und Lehrkräfte widerspiegelt (Almasri, 2024; Wang & Fan, 2025). Fortbildungskonzepte, die KI in dieser Phase einsetzen, können die sachfremde kognitive Belastung durch Routinetätigkeiten reduzieren und setzen damit kognitive Ressourcen für didaktische Entscheidungen frei (Stadler et al., 2024). Die Teilnehmenden erleben dadurch gesteigerte Methodenkompetenz und Motivation, was ihre Offenheit gegenüber digitalen Werkzeugen erhöht (Henze et al., 2025; Henze et al., 2022) und damit eine wesentliche Voraussetzung für den Einsatz im eigenen Unterricht schafft (Huwer et al., 2024). Entscheidend ist, dass Technologie nicht isoliert geschult wird, sondern von vornherein mit fachlichen und gesellschaftlich relevanten Themen verschränkt ist (Krupka, 2020; Stubbe, 2017) und die Technologie selbst kritisch thematisiert wird (Loroff et al., 2017).

In der **Evaluate-Phase** wird die erbrachte Lernleistung formal bewertet, beispielsweise über benotete Ergebnisse wie Unterrichtsentwürfe. Dabei ist entscheidend, dass die Bewertungsmaßstäbe transparent bleiben und die Verantwortung für die endgültige Leistungsbeurteilung bei der Lehrperson liegt (Henze et al., 2025). KI kann unterstützend eingesetzt werden, beispielsweise zur sprachlichen Überarbeitung, sie ersetzt jedoch weder fachliche Kriterien noch didaktische Qualitätsurteile.

Die Erweiterung um die **Exchange-Phase** macht das 6E-Modell für die Lehrkräftebildung besonders wirksam, weil sie einen strukturellen Raum für diskursive Auseinandersetzung schafft, um digitale Souveränität aufbauen zu können. In dieser sechsten Phase reflektieren Lehrende und Lernende gemeinsam ihre Lernwege, Schwierigkeiten und beispielsweise die Rolle der KI im eigenen Arbeitsprozess. Kritische Reflexion, ethische Bewertung und der Transfer von Wissen können durch kollektiven Austausch und Diskussion besser vertieft werden als im reinen Selbststudium (Henze et al., 2025). Die Exchange-Phase adressiert damit gezielt die affektiv-kognitive Dissonanz (siehe Kapitel 7.3), die durch motivierende KI-Unterstützung bei gleichzeitig begrenztem Wissenszuwachs entsteht, indem sie die Diskrepanz zwischen gefühlter und tatsächlicher Kompetenz thematisiert und möglichst korrigiert. Ausbildungskonzepte, die diese Phase auslassen oder in individuelles Selbststudium verlagern, verfehlen die Entwicklung einer kritisch-reflexiven Haltung. Die Ergebnisse zeigen, dass tiefgehende Reflexion über Risiken und didaktische Integration von KI dort gelingen kann, wo Prozesse diskutiert werden. Erst in diesem Diskurs stabilisieren sich ethische Dimensionen von AI Literacy und die Fähigkeit, die Grenzen von KI-Systemen einzuschätzen (Henze et al., 2025). KI-bezogene Professionalisierung sollte daher ganzheitlich angelegt sein und nicht auf technische Wissensvermittlung reduziert werden. Erst erfahrungsbasierte und dialogische Lernprozesse unterstützen die Ausbildung grundlegender KI-Kompetenzen und ethischer Orientierungen nachhaltig (Lin et al., 2022; Viberg et al., 2023). Fortbildungsformate, die das 6E-Modell nutzen, können Lehrkräfte damit von einer vorwiegend technischen Nutzungskompetenz zu einer umfassenden digitalen Souveränität führen, die technisches Verständnis, didaktische Integration und ethische Bewertung verbindet (Blossfeld et al., 2018; Loroff et al., 2017; Stubbe, 2017).

Die empirischen Befunde zeigen allerdings auch, dass eine kurze 6E-Intervention die Diskrepanz zwischen emotionaler Aktivierung und objektivem Kompetenzerwerb nicht vollständig auflösen kann (Henze et al., 2025). Die Exchange-Phase markiert daher weniger den Endpunkt, sondern den Beginn eines längerfristigen Professionalisierungsprozesses. Aus- und Fortbildungskonzepte sollten folglich nicht auf Einzelinterventionen setzen, sondern iterative Formate etablieren, in denen Lehrkräfte wiederholt zwischen KI-gestützter Praxis und sozialem Austausch wechseln.

7.3. Affektiv-kognitive Dissonanz als Ausgangspunkt digitaler Souveränität

Die vorliegenden Studien zeigen, dass Lehrende und Lernende bereit sind, digitale Technologien einzusetzen, wenn sie diese nicht als Bedrohung wahrnehmen, sondern als ein Werkzeug zur Verbesserung und Unterstützung der eigenen Lehr- und Lernerfahrung (Henze et al., 2025). Aus- und Fortbildungskonzepte, die digitale Souveränität im Umgang mit KI fördern sollen, sollten demnach zunächst Berührungsängste abbauen und eine positive Einstellung gegenüber neuen Technologien fördern und demnach Werkzeuge verwenden, die kreative Problemlösung sowie exploratives Lernen ermöglichen. Vielseitigkeit, Individualisierbarkeit und Adaptivität sind ebenfalls wichtige Faktoren, die eine positive Einstellung zur Nutzung digitaler Werkzeuge unterstützen (Henze et al., 2022). Besonders das Motivationspotenzial digitaler Werkzeuge entspricht Vygotskys Beobachtung, wonach Werkzeuge Lernprozesse dann fördern, wenn sie sich zu diversen Zwecken einsetzen lassen und damit diversifiziertes Lernen ermöglichen (Leong & Bodrova, 1996). Erfolgserlebnisse, beispielsweise mit KI-gestützten Werkzeugen, erzeugen positive emotionale Lernerfahrungen, die Offenheit für neue Technologien schaffen (Henze et al., 2026).

In den aufgeführten Studien zeigte sich das Phänomen einer affektiv-kognitiven Dissonanz. Die Teilnehmenden der Studien fühlten sich durch KI-Tools kompetenter und motivierter, obwohl der fachliche Lernzuwachs nicht stärker war als bei Verwendung traditioneller Methoden (Henze et al., 2026). Ähnliches wurde in vergleichbaren Studien gefunden (Lademann et al., 2025). Ebenso führte die Nutzung von KI zu einem höheren Engagement, wie auch andere Studien zeigen (Gada & Chudasana, 2024; Liang et al., 2023), und ging mit einer reduzierten sachfremden kognitiven Belastung einher. Dennoch waren keine signifikanten Leistungsunterschiede nachweisbar (Henze et al., 2026). Abbildung 2 macht deutlich, dass positive Lernzustände wie Flow nicht automatisch mit einem hohen Kompetenzniveau gleichzusetzen sind. KI kann Lernende emotional entlasten und motivieren, ohne dass damit zwingend ein entsprechend hoher fachlicher Lernzuwachs verbunden ist (Henze et al., 2026).
Ebenso führten beim Aufbau von AI Literacy kurze Interventionen zu einer gesteigerten Kompetenzwahrnehmung, gleichzeitig jedoch nur zu geringen Wissenszuwächsen (Henze et al., 2025). Die Dauer der Einheit wurde auch in anderen Studien bereits als maßgeblicher Faktor für die Entwicklung inhaltlicher Kompetenzen identifiziert (Wang & Fan, 2025). Die Auswirkung auf die Motivation blieb in jeweils beiden Studien (Henze et al., 2026; Henze et al., 2025) stabil, möglicherweise durch einen Neuheitseffekt, der auch über längere Kontaktzeiten nachgewiesen werden konnte (T. Long et al., 2024).

Die Ergebnisse machen deutlich, dass wahrgenommene Kompetenz nicht zwangsläufig mit tatsächlichem Wissenszuwachs einhergeht. Lehrkräfte, die ihre eigenen Kompetenzen überschätzen, könnten aufgrund von Fehlvorstellungen auf Schwierigkeiten stoßen, während eine Unterschätzung der eigenen Fähigkeiten dazu führen kann, dass KI-Anwendungen trotz vorhandener Kompetenzen vermieden werden (Chounta et al., 2022). Ergänzend zeigen Bewersdorff et al. (2023), dass insbesondere ein begrenztes technisches Verständnis verbreitet ist, welches sich in Unsicherheiten bei zentralen Konzepten wie maschinellem Lernen, neuronalen Netzen oder Deep Learning äußert.

Daraus ergibt sich, dass Aus- und Weiterbildungskonzepte zur Steigerung digitaler Souveränität diese motivationalen Aspekte aufgreifen sollten, um durch sie die Selbstwirksamkeit und die Autonomie der Teilnehmenden zu stärken (Deci & Ryan, 2008). Andererseits sollte das gesteigerte Kompetenzerleben durch Diskurse aktiv einbezogen werden, um Fehleinschätzungen zu vermeiden. Der Grund dafür ist, dass die überzeugende Art, mit der KI Antworten, unter Umständen auch falsche, formuliert (Gregorcic & Pendrill, 2023; Kasneci et al., 2023), dazu führen kann, das eigene Verständnis zu überschätzen und kritische Reflexionsprozesse zu überspringen. Dies zeigte sich auch in anderen Studien (Krupp et al., 2023), besonders bei Lernenden, die nicht wussten, wie KI zu benutzen ist (Ding et al., 2023).

Die Überwindung dieser affektiv-kognitiven Dissonanz erfordert einen Professionalisierungsprozess, der über die anfängliche Begeisterung oder Skepsis hinausgeht und strukturiert in Aus- und Fortbildungskonzepten enthalten sein sollte. Sie sollten ermöglichen, dass eine realistische Einschätzung von Technologie erlernt wird, ohne diese jedoch abzulehnen. Dazu gehört die Erkenntnis, dass KI fehlerhaft sein kann und zu falschen Antworten neigt (Gregorcic & Pendrill, 2023; Kasneci et al., 2023). Idealerweise führt eine solche strukturierte Fortbildung damit zu einer souveränen Haltung, die Chancen und Risiken der Technologie differenziert abwägt.

Digitale Souveränität lässt sich insgesamt nur fördern, wenn über die technische Bedienbarkeit hinaus ein umfassendes Verständnis für Funktionsweisen und Grenzen sowie handlungsorientierte und praxisnahe Anwendungen der betrachteten Technologie thematisiert werden. Daher ist es notwendig, sowohl Lehrende als auch Lernende dazu weiter- und auszubilden, dass sie digitale Werkzeuge nicht nur passiv rezipieren, sondern aktiv nutzen und gestalten.

7.4. Implikationen für die physikalische Lehrer:innenbildung

Die zuvor beschriebenen Erkenntnisse über das Spannungsfeld zwischen Motivation, Kognition und Reflexion haben direkte Konsequenzen für die Gestaltung von Aus- und Fortbildungskonzepten in der physikalischen Lehrer:innenbildung.

Die Studien belegen, dass handlungsorientierter Umgang mit Technologie notwendig ist, um Berührungsängste abzubauen und Interesse zu wecken. Intuitive Werkzeuge wie Lernroboter oder KI-gestützte Analysewerkzeuge können auch ohne tiefe Vorkenntnisse genutzt werden und ermöglichen sofortige Erfolgserlebnisse (Henze et al., 2022). Kompetenzzuwächse wurden vor allem dort verzeichnet, wo Studierende praktisch mit Tools arbeiteten oder eigenständig KI-gestützte Unterrichtsentwürfe erstellten (Henze et al., 2025; Karataş & Ataç, 2024). Daher sollte zuerst das Werkzeug praktisch erlebt, bevor die theoretischen Grundlagen vertieft werden. Ein zentrales Ziel dieser Praxisorientierung ist das Erleben von Kompetenz und Autonomie im Sinne der Self-Determination Theory (Ryan & Deci, 2017). KI-Tools können das Gefühl der Methodenkompetenz und Motivation signifikant steigern, selbst wenn der objektive Lerngewinn im Vergleich zu standardisierten Lernmethoden, zunächst gleich bleibt (Henze et al., 2026). Die Interventionen zeigten, dass sich Lernende und Lehrende nicht nur nach professioneller Kontaktzeit signifikant besser vorbereitet fühlten, KI im Unterricht einzusetzen (Henze et al., 2025), sondern sich auch Vorbehalte gegenüber dem Einsatz von KI reduzierten (Lademann et al., 2026). Selbstwirksamkeitserfahrungen bilden demnach die Basis, auf der fachliche Tiefe aufgebaut werden kann. Lehrkräfte müssen das Gefühl haben, die Kontrolle über die KI zu behalten, statt von ihr ersetzt zu werden.

Technische Hürden führten zu Frustration, wenn die didaktische Einbettung fehlte (Henze et al., 2022). Technisches Scheitern, beispielsweise durch fehlerhafte KI-Antworten, darf jedoch nicht vermieden, sondern muss didaktisch konstruktiv verwendet werden. Die Lehrer:innenbildung sollte daher geschützte Lernräume in Form von Weiterbildungskonzepten bieten, in denen technisches Scheitern als Anlass für Problemlösung und kritische Reflexion genutzt wird. Ein wesentlicher Teil digitaler Souveränität ist der Umgang mit der Technik und ihrer Fehler. Die Befunde zeigen außerdem, dass ethische Kompetenzen und das Erkennen von Risiken nur dann nachhaltig gelernt werden, wenn sie im sozialen Austausch explizit thematisiert werden (Henze et al., 2025).

Die Ergebnisse verdeutlichen, dass KI zwar als technisches Scaffolding-System fungieren kann, aber für tiefgreifende Lernprozesse nicht ausreicht. Zwar ermöglicht KI individuelle Unterstützung und kann Motivation fördern (Henze et al., 2026; Wang & Fan, 2025; Wu & Yu, 2024), die soziale Aushandlung von Werten, ethischen Fragen und komplexen didaktischen Entscheidungen bleibt jedoch an menschliche Bezugspersonen gebunden (Henze et al., 2025). KI sollte demnach als Assistent für standardisierbare Prozesse wie Coding oder das

Erstellen von Entwürfen eingeführt werden, während durch Diskussionsformate die Bewertung und ethische Einordnung erfolgt.

Ein fundiertes Verständnis und digitale Souveränität lassen sich nicht durch einzelne Workshops erreichen. Während kurze Interventionen die Einstellung und gefühlte Bereitschaft verbessern können, bleiben tiefgehende Kompetenzzuwächse oft aus (Henze et al., 2026; Henze et al., 2025; Henze et al., 2022). Der Erwerb digitaler Souveränität sollte daher als iterativer und langfristiger Prozess verstanden werden, der wiederholte Auseinandersetzung und vertiefte Praxisphasen erfordert. Um digitale Souveränität dauerhaft zu verankern, braucht es strukturierte, wiederkehrende Formate in Form von Längsschnitt-Curricula, die über reine Tool-Schulungen hinausgehen. Frameworks wie das DPACK-Modell (Thyssen et al., 2023) oder DiKoLAN (Becker et al., 2020) bzw. DiKoLAN^{KI} (Huwer et al., 2024) bieten einen Orientierungsrahmen, um den Aufbau von technologischem, pädagogischem und fachlichem Wissen gemäß einem systematisiertem Kompetenzraster einzuordnen.

8. Limitationen & Ausblick

Studienübergreifend stellte sich die Interventionsdauer als eine zentrale Herausforderung dar. Kurzzeitige Interventionen konnten zwar in einzelnen Punkten einen Kompetenzerwerb hervorrufen, reichten jedoch nicht aus, um eine umfassende AI Literacy im Sinne digitaler Souveränität zu etablieren. Die zeitliche Einschränkung, die sich vor allem aus den Rahmenbedingungen der universitären Lehrer:innenbildung ergab, limitierte die Tiefe der Auseinandersetzung mit digitalen Werkzeugen. Andere Studien zeigen, dass längerfristig angelegte Fortbildungskonzepte diese Schwierigkeit überwinden können (Lademann et al., 2026).
Die Studien basieren auf homogenen Stichproben aus dem universitären oder diesem nahestehenden schulischen Kontext. Die Beschränkung auf Lehrer:innen der *Heliosschule – Inklusive Universitätsschule der Stadt Köln* und Lehramtsstudierende der Universität zu Köln mit vergleichbarem Vorwissen und kulturellem Hintergrund schränkt die Generalisierbarkeit der Ergebnisse ein. Daher sind diese vor allem als kontextspezifische Designprinzipien für Aus- und Fortbildungskonzepte zu digitaler Souveränität mit dem Schwerpunkt KI zu verstehen. Eine Übertragung auf heterogenere Lerngruppen könnte Aufschluss darüber geben, inwiefern die gezeigten Ergebnisse verallgemeinerbar sind.
Alle Studien wurden mithilfe von Selbstauskunftsinstrumenten durchgeführt. Emotional-motivationale Ergebnisse repräsentieren damit vor allem subjektive Wahrnehmungen und keine objektive Evidenz. Zusätzlich sind durch aggregierte Werte aus den Messdaten individuelle Unterschiede zwischen einzelnen Teilnehmenden nicht mehr identifizierbar und die fehlende Differenzierung zwischen intrinsischer und extrinsischer Motivation stellt eine konzeptionelle Lücke dar. Auch die vorrangig quantitative Herangehensweise maskiert individuelle Veränderungen, sodass prozessuale Daten, die auf konkrete Erlebnisse eingehen, nicht aufgedeckt wurden. Mixed-Methods-Designs, die quantitative Erhebungen mit qualitativen Datenerhebungen verbinden, könnten die Quellen motivationaler Unterschiede aufdecken und prozessbezogene Dynamiken sichtbar machen.
Die wissenschaftliche Korrektheit der KI-generierten Ausgaben wurde nicht systematisch validiert. Obwohl plausible und kohärente Antworten beobachtet wurden, kann nicht ausgeschlossen werden, dass irreführende oder fehlerhafte Informationen einzelne Interpretationen beeinflussten.

Aus den identifizierten Limitationen ergeben sich mehrere zentrale Ansatzpunkte für nachfolgende Forschungsarbeiten. Zukünftige Studien könnten Interventionen über mehrere Wochen oder Monate implementieren, um zu klären, ob motivationale Vorteile bestehen bleiben, ob Leistungsgewinne auftreten oder sich verfestigen und inwieweit entwickelte Kompetenzen in reale Unterrichtskontexte übertragen werden können.

KI-Tools sollten systematisch mit alternativen Unterstützungsstrategien wie menschlicher Lernbegleitung, erweiterten statischen Materialien oder hybriden Ansätzen verglichen werden. Außerdem sollten unterschiedliche KI-Interaktionsstile getestet werden, um einen möglichen Einfluss dessen auf die beobachteten Ergebnisse zu ermitteln. Adaptive Lernumgebungen, die auf initiale Kompetenzniveaus reagieren, könnten personalisierte Lernerfahrungen ermöglichen.

Die systematische Validierung von KI-Ausgaben sollte integraler Bestandteil zukünftiger Studien werden, um sicherzustellen, dass Lernende mit korrekten Informationen arbeiten und gleichzeitig kritische Evaluationskompetenz entwickeln. Vergleichsstudien über verschiedene Fachinhalte, Bildungsstufen und Domänen hinweg könnten prüfen, inwieweit sich die entwickelten Konzepte auf andere Fachbereiche übertragen lassen.

9. Fazit

Die vorliegende Dissertation entwickelt ein theoretisch fundiertes und empirisch erprobtes Konzept zur Förderung digitaler Souveränität von Lehrkräften. Die Untersuchung zeigt, wie Aus- und Fortbildungskonzepte gestaltet werden können, die über reine Technologievermittlung hinausgehen.

Die empirischen Befunde der drei Teilstudien belegen, dass digitale Werkzeuge und insbesondere KI-Systeme zwar als technisches Scaffolding kognitive Entlastung bieten und Motivation fördern können (Henze et al., 2026; Henze et al., 2025; Henze et al., 2022), jedoch ohne menschliche Bezugspersonen keine umfassende Erweiterung digitaler Souveränität ermöglichen. KI übernimmt dabei die Rolle eines digitalen More Knowledgeable Other, stellt strukturierende Hilfestellungen bereit und reduziert die sachfremde kognitive Belastung (Stadler et al., 2024). Gleichzeitig zeigen die Studien, dass ethische Reflexion, kritische Bewertung von Technologie und die Entwicklung einer professionellen Haltung an soziale Aushandlungsprozesse gebunden bleiben. Weder die vollständige Delegation von Lernprozessen an KI noch der Verzicht auf technologische Unterstützung erweisen sich als gewinnbringend zur Steigerung digitaler Souveränität. Diese entsteht erst durch die systematische Kombination beider Scaffolding-Formen.

Die Erweiterung des etablierten 5E-Modells (Bybee, 2009; Bybee et al., 2006) um die Exchange-Phase ermöglicht die strukturierte Integration von technischem und sozialem Scaffolding in Aus- und Fortbildungskonzepten. Studie I identifizierte die Notwendigkeit eines strukturierten Austauschs zwischen Lernenden und Lehrenden, um Lernerfahrungen zu reflektieren und technologische Werkzeuge professionell einordnen zu können. Studie II konkretisierte diese Beobachtung durch den direkten Vergleich zweier Iterationen einer Intervention und wies nach, dass Kompetenzzuwächse in ethisch anspruchsvollen Bereichen nur dann auftraten, wenn die Exchange-Phase systematisch implementiert wurde. Studie III ergänzte diese Befunde um die affektiv-motivationale Dimension und zeigte, dass technisches Scaffolding zwar emotionale Entlastung, jedoch ohne signifikante Vorteile im Lernzuwachs, schafft.

Das 6E-Modell bietet einen praktizierbaren didaktischen Rahmen, der festlegt, wann KI primär zur Strukturierung eingesetzt werden kann und wann der Fokus auf sozialer Aushandlung liegen sollte. Es vereint Vygotskys Konzept der Zone der proximalen Entwicklung (McLeod, 2025; Pedapati, 2022; Rigopouli et al., 2025) mit konstruktivistischen Lernprinzipien und macht diese für die Gestaltung von Lehrkräftefortbildungen im Kontext digitaler Technologien nutzbar. Die hinzugefügte Exchange-Phase bietet Raum für die Diskussion von Grenzen und Fehlern von KI-Systemen. Teilnehmende lernen, dass motivierende Nutzungserfahrungen nicht automatisch mit fachlicher Korrektheit einhergehen und entwickeln Strategien zur kritischen Bewertung KI-generierter Inhalte (Henze et al., 2025).

Die Befunde haben direkte Konsequenzen für die Gestaltung von Aus- und Fortbildungskonzepten in der Physikdidaktik und darüber hinaus. Erfolgreiche Professionalisierung erfordert einen handlungsorientierten Zugang zu Technologie, um Berührungsängste abzubauen und positive Nutzungserfahrungen zu ermöglichen (Döbeli Honegger, 2021; Henze et al., 2022; Loroff et al., 2017; Stubbe, 2017). Weiterhin müssen Formate über Einzelinterventionen hinausgehen und iterative Strukturen etablieren, in denen Lehrkräfte wiederholt zwischen praktischer Anwendung und reflexivem Austausch wechseln (Henze et al., 2025; Wang & Fan, 2025).
Frameworks wie DiKoLAN, DiKoLAN^{KI} oder DPACK bieten Orientierung für die systematische Verknüpfung von technologischem, pädagogischem und fachlichem Wissen (Döbeli Honegger, 2021; Huwer et al., 2024; Thyssen et al., 2023). Das 6E-Modell ergänzt diese Kompetenzraster um eine prozessorientierte Perspektive, die konkrete Handlungsmöglichkeiten für die Sequenzierung von Lernphasen liefert.
Digitale Souveränität entsteht demnach weder durch technische Schulungen allein, noch durch theoretische Reflexion ohne Praxisbezug (Krupka, 2020; Stubbe, 2017; Thimm, 2023). Sie entwickelt sich in einem iterativen Prozess, der technisches und soziales Scaffolding systematisch verbindet. Das 6E-Modell mit der Exchange-Phase bietet einen didaktischen Rahmen, der diese Integration strukturiert und für die Lehrkräftebildung operationalisierbar macht.

Die Fähigkeit von Lehrkräften, Technologie kritisch zu bewerten und fachdidaktisch sinnvoll einzusetzen, gewinnt immer weiter an Bedeutung. Die Erkenntnisse dieser Dissertation tragen dazu bei, Professionalisierungskonzepte zu entwickeln, die über Technikbegeisterung und Technikskepsis hinausgehen und eine reflektierte, souveräne Haltung im Umgang mit digitalen Werkzeugen fördern.

Literaturverzeichnis

Almasri, F. (2024). Exploring the Impact of Artificial Intelligence in Teaching and Learning of Science: A Systematic Review of Empirical Research. *Research in Science Education*, *54*(5), 977–997. https://doi.org/10.1007/s11165-024-10176-3

Anderson, T. & Shattuck, J. (2012). Design-Based Research. *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Backfisch, I., Lachner, A., Stürmer, K. & Scheiter, K. (2021). *Gelingensbedingungen beim Einsatz digitaler Medien im Unterricht – Kognitive und motivationale Voraussetzungen von Lehrpersonen.* https://doi.org/10.15496/publikation-52635

Ballone Duran, L. & Duran, E. (2004). The 5E Instructional Model: A Learning Cycle Approach for Inquiry-Based Science Teaching. *The Science Education Review*, *5*(2), 49–58.

Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K. & Sailer, M. (2025). Looking Beyond the Hype: Understanding the Effects of AI on Learning. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10020-8

Becker, S., Meßinger-Koppelt, J. & Thyssen, C. (Hrsg.) (2020). *Digitale Basiskompetenzen: Orientierungshilfe und Praxisbeispiele für die universitäre Lehramtsausbildung in den Naturwissenschaften.*

Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (S. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., Kasneci, G., Zhai, X. & Nerdel, C. (2025). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences*, *118*, 102601. https://doi.org/10.1016/j.lindif.2024.102601

Bewersdorff, A., Zhai, X., Roberts, J. & Nerdel, C. (2023). Myths, mis- and preconceptions of artificial intelligence: A review of the literature. *Computers and Education: Artificial Intelligence*, *4*, 100143. https://doi.org/10.1016/j.caeai.2023.100143

Bliss, J. (1996). Piaget und Vygotsky. Ihre Bedeutung für das Lehren und Lernen der Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften : ZfDN*, *2.* https://doi.org/10.25656/01:31493 (Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 2 (1996) 3, S. 3-16).

Blossfeld, H.-P., Bos, W., Daniel, H.-D., Hannover, B., Köller, O., Lenzen, D., McElvany, N., Roßbach, H.-G., Seidel, T., Tippelt, R. & Wößmann, L. (2018). *Digitale Souveränität und Bildung: Gutachten* (1. Auflage). *Aktionsrat Bildung*. Waxmann.

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7., vollständig überarbeitete und erweiterte Auflage). *Springer-Lehrbuch*. Springer-Verlag Berlin Heidelberg. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10448295 https://doi.org/10.1007/978-3-642-12770-0

Bybee, R. (2009). *The BSCS 5E Instructional Model and 21st century skills: A commissioned paper prepared for a workshop on exploring the intersection of science education and the development of 21st century skills.*

Bybee, R., Taylor, J., Gardner, A., van Scotter, P., Carlson Powell, J., Westbrook, A. & Landes, N. (2006). *The BSCS 5E Instructional Model: Origins and Effectiveness.*

Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, *32*(10), 6187–6203. https://doi.org/10.1080/10494820.2023.2253861

Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments*, *33*(5), 3225–3229. https://doi.org/10.1080/10494820.2025.2514372

Chounta, I.-A., Bardone, E., Raudsep, A. & Pedaste, M. (2022). Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. *International Journal of Artificial Intelligence in Education*, *32*(3), 725–755. https://doi.org/10.1007/s40593-021-00243-5

Deci, E. L. (1975). *Intrinsic motivation*. *Perspectives in social psychology*. Plenum Press. https://permalink.obvsg.at/AC02080079

Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, *39.* https://doi.org/10.25656/01:11173 (Zeitschrift für Pädagogik 39 (1993) 2, S. 223-238).

Deci, E. L. & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology / Psychologie canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

Dedehayir, O. & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, *108*, 28–41. https://doi.org/10.1016/j.techfore.2016.04.005

Ding, L., Li, T., Jiang, S. & Gapud, A. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00434-1

Döbeli Honegger, B. (2017). *Mehr als 0 und 1: Schule in einer digitalisierten Welt* (2., durchgesehene Auflage). hep der Bildungsverlag.

Döbeli Honegger, B. (2021). Covid-19 und die digitale Transformation in der Schweizer Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, *39*(3), 412–422. https://doi.org/10.25656/01:23693

Esbenshade, L., Sarkar, S., Nucci, D., Edwards, A., Nielsen, S., Rosenberg, J. M., Liu, A., Tian, Z., Sun, M., Zhang, Z., Han, T., Lapicus, Y. & He, K. (2025). *Emerging Patterns of GenAI Use in K-12 Science and Mathematics Education.* https://doi.org/10.48550/arXiv.2509.10747

Essel, H. B., Vlachopoulos, D., Essuman, A. B. & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, *6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR) (2024 & i.d.F.v. 12.07.2024). https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32024R1689

European Broadcasting Union (Hrsg.) (2025). *News Integrity in AI Assistants: An international PSM study*. https://www.ebu.ch/files/live/sites/ebu/files/Publications/MIS/open/EBU-MIS-BBC_News_Integrity_in_AI_Assistants_Toolkit_2025.pdf

Ferdinand, J.-P., Richter, S. & Engelhardt, S. von (2017). Digitale Partizipation in Wissenschaft und Wirtschaft. In V. Wittpahl (Hrsg.), *Digitale Souveränität* (S. 27–42). Springer Berlin Heidelberg.

Gada, T. & Chudasana, S. (2024). Impact of Artificial Intelligence on student attitudes, engagement, and learning. *International Research Journal of Modernization in Engineering Technology and Science*, *06*(05), 2695–2698. https://doi.org/10.56726/IRJMETS56247

Gartner, Inc. (Hrsg.) (2025). *Gartner Hype Cycle Identifies Top AI Innovations in 2025*. https://www.gartner.com/en/newsroom/press-releases/2025-08-05-gartner-hype-cycle-identifies-top-ai-innovations-in-2025

Gerjets, P. H. & Hesse, F. W. (2004). When are powerful learning environments effective? The role of learner activities and of students' conceptions of educational technology. *International Journal of Educational Research*, *41*(6), 445–465. https://doi.org/10.1016/j.ijer.2005.08.011

Gesellschaft für deutsche Sprache e. V. (Hrsg.) (2025). *Wörter des Jahres 2025*. https://gfds.de/worter-des-jahres-2025/

Gregorcic, B. & Pendrill, A.-M. (2023). ChatGPT and the frustrated Socrates. *Physics Education*, *58*(3), 35021. https://doi.org/10.1088/1361-6552/acc299

Guo, F., Zhang, L., Shi, T. & Coates, H. (2025). Whether and When Could Generative AI Improve College Student Learning Engagement? *Behavioral sciences (Basel, Switzerland)*, *15*(8). https://doi.org/10.3390/bs15081011

Henze, J., Lademann, J., Becker-Genschow, S. & Bresges, A. (2026). AI-Supported Data Analysis Boosts Student Motivation and Reduces Stress in Physics Education. *Frontiers in Education*, *11*, Artikel 1719670. https://doi.org/10.3389/feduc.2026.1719670

Henze, J., Lademann, J., Bresges, A. & Becker-Genschow, S. (2025). Iterative development of an AI intervention for pre-service physics teachers from a Vygotskian perspective. *Frontiers in Education*, *10*, Artikel 1707534. https://doi.org/10.3389/feduc.2025.1707534

Henze, J., Schatz, C., Malik, S. & Bresges, A. (2022). How Might We Raise Interest in Robotics, Coding, Artificial Intelligence, STEAM and Sustainable Development in University and On-the-Job Teacher Training? *Frontiers in Education*, *7*, Artikel 872637. https://doi.org/10.3389/feduc.2022.872637

Hornberger, M., Bewersdorff, A. & Nerdel, C. (2023). What do university students know about Artificial Intelligence? Development and validation of an AI literacy test. *Computers and Education: Artificial Intelligence*, *5*, 100165. https://doi.org/10.1016/j.caeai.2023.100165

Huwer, J., Becker-Genschow, S., Thyssen, C., Thoms, L.-J., Kotzebue, L. von, Finger, A., Kremser, E., Berber, S., Brückner, M., Maurer, N., Bruckermann, T. & Meier, M. (2024). Kompetenzen für den Unterricht mit und über Künstliche Intelligenz in den Naturwissenschaften: DiKoLAN^{KI}. In J. Huwer, S. Becker-Genschow, C. Thyssen, L.-J. Thoms, A. Finger, L. von Kotzebue, E. Kremser, M. Meier & T. Bruckermann (Hrsg.), *Kompetenzen für den Unterricht mit und über Künstliche Intelligenz.: Perspektiven, Orientierungshilfen und Praxisbeispiele für die Lehramtsausbildung in den Naturwissenschaften* (S. 4–59). Waxmann Verlag GmbH.

Huwer, J., Irion, T., Kuntze, S., Schaal, S. & Thyssen, C. (2019). Von TPaCK zu DPaCK: Digitalisierung im Unterricht erfordert mehr als technisches Wissen. *MNU Journal*(5), 358–364. https://www.researchgate.net/publication/335748844_Von_TPaCK_zu_DPaCK_-_Digitalisierung_im_Unterricht_erfordert_mehr_als_technisches_Wissen

Initiative D21 e.V. (Hrsg.) (2025). *D21- D21-Digital-Index 2024/25: Jährliches Lagebild zur Digitalen Gesellschaft*. https://initiatived21.de/uploads/03_Studien-Publikationen/D21-Digital-Index/2024-25/D21_Digital_Index_2024_2025_final.pdf

Ishmuradova, I. I., Zhdanov, S. P., Kondrashev, S. V., Erokhova, N. S., Grishnova, E. E. & Volosova, N. Y. (2025). Pre-service science teachers' perception on using generative artificial intelligence in science education. *Contemporary Educational Technology*, *17*(3), ep579. https://doi.org/10.30935/cedtech/16207

Jackson, H., Basham, J., Thomas, K. & Hunt, C. (2020). Designing STEAM Learning Environments. In K. Thomas & D. Huffman (Hrsg.), *Challenges and Opportunities for Transforming From STEM to STEAM Education* (S. 1–23).

Jude, N., Klusmann, U., Richter, D., Selçik, F. & Sichma, A. (2025). *Deutsches Schulbarometer 2025: Befragung Lehrkräfte. Ergebnisse zur aktuellen Lage an allgemein- und berufsbildenden Schulen*. https://www.bosch-stiftung.de/de/publikation/deutsches-schulbarometer-lehrkraefte-2025

Karataş, F. & Ataç, B. A. (2024). When TPACK meets artificial intelligence: Analyzing TPACK and AI-TPACK components through structural equation modelling. *Education and Information Technologies.* Vorab-Onlinepublikation. https://doi.org/10.1007/s10639-024-13164-2

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103.* https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T. & Ponti, G. (2025). AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific reports*, *15*(1). https://doi.org/10.1038/s41598-025-97652-6

Kortemeyer, G. (2023). Could an Artificial-Intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, *19*(1), Artikel 010132, 15. https://doi.org/10.1103/PhysRevPhysEducRes.19.010132

Krupka, D. (2020). Dimensionen digitaler Souveränität: Ein Überblick. In Gesellschaft für Informatik e.V. (Hrsg.), *Arbeitspapier: Schlüsselaspekte digitaler Souveränität* (S. 4–7).

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., Küchemann, S. & Karolus, J. (2023). *Unreflected Acceptance: Investigating the Negative Consequences of ChatGPT-Assisted Problem Solving in Physics Education*. http://arxiv.org/pdf/2309.03087v1

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (4. Auflage). *Grundlagentexte Methoden*. Beltz Juventa. http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-7799-3682-4

La Ossa, L. A. M. D., Garzón, N. N., Pulido, C. P., Herrán, B. E. G. & Hoyos, J. H. (2024). Evaluation of Artificial Intelligence as a Source of Motivation in the Teaching of Chemistry: A Study with Tenth Grade Students at the Adventist School of Ibagué. *EVOLUTIONARY STUDIES IN IMAGINATIVE CULTURE*, 2193–2208. https://doi.org/10.70082/esiculture.vi.1721

Lademann, J., Henze, J. & Becker-Genschow, S. (2025). Augmenting learning environments using AI custom chatbots: Effects on learning performance, cognitive load, and affective variables. *Physical Review Physics Education Research*, *21*(1). https://doi.org/10.1103/PhysRevPhysEducRes.21.010147

Lademann, J., Henze, J., Honke, N., Wollny, C. & Becker-Genschow, S. (2026). Teacher training in the age of AI: impact on AI literacy and teachers' attitudes. *Frontiers in Education*, *10*, Artikel 1671306. https://doi.org/10.3389/feduc.2025.1671306

Lee, G., Yun, M., Zhai, X. & Crippen, K. (2025). Artificial Intelligence in Science Education Research: Current States and Challenges. *Journal of Science Education and Technology.* Vorab-Onlinepublikation. https://doi.org/10.1007/s10956-025-10239-8

Leong, D. & Bodrova, E. (1996). *Play: A Vygotskian Approach*. https://video.alexanderstreet.com/watch/play-a-vygotskian-approach.

Liang, Y., Di Zou, Xie, H. & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments*, *10*(1). https://doi.org/10.1186/s40561-023-00273-7

Lin, X.-F., Chen, L., Chan, K. K., Peng, S., Chen, X., Xie, S., Liu, J. & Hu, Q. (2022). Teachers' Perceptions of Teaching Sustainable Artificial Intelligence: A Design Frame Perspective. *Sustainability*, *14*(13), 7811. https://doi.org/10.3390/su14137811

Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, *13*(4), 410. https://doi.org/10.3390/educsci13040410

Long, D. & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In R. Bernhaupt, F. Mueller, D. Verweij & J. Andres (Hrsg.), *ACM Digital Library, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (S. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Long, T., Gero, K. I. & Chilton, L. B. (2024). *Not Just Novelty: A Longitudinal Study on Utility and Customization of an AI Workflow*. http://arxiv.org/pdf/2402.09894v2

Loroff, C., Lindow, I. & Schubert, M. (2017). Bildung als Voraussetzung digitaler Souveränität. In V. Wittpahl (Hrsg.), *Digitale Souveränität.* Springer Berlin Heidelberg.

McLeod, S. (2025). *Vygotsky's Theory of Cognitive Development.* Simply Psychology. https://www.simplypsychology.org/vygotsky.html

Minichiello, A. & Caldwell, L. (2021). A Narrative Review of Design-Based Research in Engineering Education: Opportunities and Challenges. *Studies in Engineering Education*, *1*(2), 31. https://doi.org/10.21061/see.15

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (Hrsg.) (2025). *Kernlehrplan für die Sekundarstufe II Gymnasium / Gesamtschule in Nordrhein-Westfalen: Physik*. https://lehrplannavigator.nrw.de/system/files/media/document/file/klp-entwurf_vb_sii_gost_physik.pdf

Nikolopoulou, K. (2024). ChatGPT for educational-academic activities: Preschool student teachers' experiences. *Intelligent Technologies in Education.* Vorab-Onlinepublikation. https://doi.org/10.53761/ITED/1.6

Pedapati, K. (2022). Piagetian and Vygotskian Concepts of Cognitive Development: A Review. *Indian Journal of Mental Health*, *9*(3). https://doi.org/10.30877/IJMH.9.3.2022.227-239

Pfiffner, M., Sterel, S. & Hassler, D. (2021). *4K und digitale Kompetenzen: Chancen und Herausforderungen*. *4K kompakt: Bd. 1*. hep Verlag AG.

Repenning, A. (2012). Programming goes back to school. *Communications of the ACM*, *55*(5), 38–40. https://doi.org/10.1145/2160718.2160729

Rigopouli, K., Kotsifakos, D. & Psaromiligkos, Y. (2025). Vygotsky's Creativity Options and Ideas in 21st-Century Technology-Enhanced Learning Design. *Education Sciences*, *15*(2), 257. https://doi.org/10.3390/educsci15020257

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M. & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Ryan, R. M. & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Schröter, L., Brase, B. & Struck, Y. (2021). *Positionspapier der Deutschen Physikalischen Gesellschaft e.V.: Digitale Bildung im Fach Physik*. https://www.dpg-physik.de/veroeffentlichungen/publikationen/stellungnahmen-der-dpg/bildung-wissenschaftlicher-nachwuchs/positionspapier_digitale_bildung_physikunterricht

Seidel, T. & Krapp, A. (Hrsg.) (2014). *Pädagogische Psychologie: Mit Online-Materialien zum Download* (6., vollständig überarbeitete Aufl.). Beltz. http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1142309

Sekretariat der Kultusministerkonferenz (Hrsg.) (2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen: Themenspezifische Handlungsempfehlung* [Beschluss der Bildungsministerkonferenz vom 10.10.2024]. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pdf

Stadler, M., Bannert, M. & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior*, *160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (Hrsg.) (2024). *Large Language Models und ihre Potenziale im Bildungssystem: Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz*. http://dx.doi.org/10.25656/01:28303

Stubbe, J. (2017). Von digitaler zu soziodigitaler Souveränität. In V. Wittpahl (Hrsg.), *Digitale Souveränität* (S. 43–59). Springer Berlin Heidelberg.

Sweller, J. (2012). Implications of Cognitive Load Theory for Multimedia Learning. In R. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 19–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.003

Thimm, C. (2023). Digitale Souveränität und personale Autonomie: Herausforderungen für eine Digitalpädagogik. *mediendiskurs*, *2*(104), 84–87. https://mediendiskurs.online/data/hefte/ausgabe/104/thimm_digitale_souveraenitaet_md104.pdf

Thyssen, C., Huwer, J., Irion, T. & Schaal, S. (2023). From TPACK to DPACK: The "Digitality-Related Pedagogical and Content Knowledge"-Model in STEM-Education. *Education Sciences*, *13*(8), 769. https://doi.org/10.3390/educsci13080769

Tiemann, J., Gumz, J. D., Kupi, M., Opiela, N., Wachsmann, D. & Weber, M. (2024). *Mythen der Digitalisierung.* Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS. https://www.oeffentliche-it.de/publikationen/mythen-der-digitalisierung/Mythen%20der%20Digitalisierung.pdf

Traga Philippakos, Z. A. & Rocconi, L. (2025). AI Literacy: Elementary and Secondary Teachers' Use of AI-Tools, Reported Confidence, and Professional Development Needs. *Education Sciences*, *15*(9), 1186. https://doi.org/10.3390/educsci15091186

United Nations Department of Economic and Social Affairs (2021). *The Sustainable Development Goals Report 2021*. United Nations. https://www.un-ilibrary.org/content/books/9789210056083

van Zyl, A. & Karsten, I. (2022). Design-Based Research (DBR) as an Effective Tool to Create Context-Sensitive and Data-Informed Student Success Initiatives. *Journal for Students Affairs in Africa.* Vorab-Onlinepublikation. https://doi.org/10.24085/jsaa.v10i1.3706

Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., Wasson, B., Tømte, C., Spikol, D., Milrad, M., Coelho, R. & Kizilcec, R. F. (2023). *What Explains Teachers' Trust of AI in Education across Six Countries?* http://arxiv.org/pdf/2312.01627v2

Wang, J. & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-04787-y

Wu, R. & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Yakman, G. (2008). *STΣ@M Education: An overview of creating a model of integrative education*. https://www.researchgate.net/publication/327351326_STEAM_Education_an_overview_of_creating_a_model_of_integrative_education
Young, J. Q., Thakker, K., John, M., Friedman, K., Sugarman, R., van Merriënboer, J. J. G., Sewell, J. L. & O'Sullivan, P. S. (2021). Exploring the relationship between emotion and cognitive load types during patient handovers. *Advances in health sciences education : theory and practice*, *26*(5), 1463–1489. https://doi.org/10.1007/s10459-021-10053-y
Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0171-0
Zhao, L., Wu, X. & Luo, H. (2022). Developing AI Literacy for Primary and Middle School Teachers in China: Based on a Structural Equation Modeling Analysis. *Sustainability*, *14*(21), 14549. https://doi.org/10.3390/su142114549

Abbildungsverzeichnis

Benutzte Hilfsmittel

Für die empirische Forschung, die Datenanalyse sowie sprachliche Revision der wissenschaftlichen Texte wurden unterschiedliche Hilfsmittel herangezogen.

Datenerhebung

- LimeSurvey zur Durchführung und Verwaltung der empirischen Datenerhebung.

Datenanalyse

- MAXQDA zur qualitativen Datenanalyse.
- RStudio zur statistischen Auswertung der Daten und zur Erstellung von Grafiken.
- Microsoft Excel zur Organisation, Auswertung und Visualisierung von Daten sowie für grundlegende statistische Berechnungen.

Literaturverwaltung und Texterstellung

- Citavi 6 zur Unterstützung der Literaturarbeit und Literaturverwaltung.
- Microsoft Word zur Erstellung und Überarbeitung der Manuskripte sowie der Dissertation.

KI-gestützte Werkzeuge zur sprachlichen Unterstützung

- Modelle Künstlicher Intelligenz von OpenAI, darunter ChatGPT-3.5, ChatGPT-4, ChatGPT-4o, ChatGPT-o1, ChatGPT-5 und ChatGPT-5.2, zur sprachlichen Überarbeitung der einzelnen Manuskripte sowie der Dissertation und der Hilfe bei der Erstellung von Programmcodes zur Auswertung der Studiendaten mit R.
- Gemini 3 Pro von Google zur Unterstützung bei sprachlichen Überarbeitungen.
- Claude 3.5 Sonnet, Claude 4.5 Sonnet von Anthropic zur Unterstützung bei sprachlichen Überarbeitungen.
- DeepL zur Übersetzung von Texten.