Sonja Hensel

Rezension zu

Zheng, L., Niu, J., Zhong, L. & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments, 31*(9), 5650–5664.

Kommentierter Kurzbefund

Zheng, Niu, Zhong und Gyasi verfolgen mit ihrer Arbeit das Anliegen, eine umfassende Metaanalyse zu Effekten des Einsatzes künstlicher Intelligenz (KI) auf die (kognitive) Lernleistung und Einstellungen der Lernenden zum Lernen vorzulegen.

Dazu werteten sie in einer Metaanalyse zum Effekt von KI auf Lernleistungen 24 Studien mit insgesamt 2 908 Teilnehmerinnen und Teilnehmern aus den Jahren 2001 bis 2020 aus. Für die Auswahl der Studien führten sie zunächst eine umfassende Recherche in einschlägigen Datenbanken durch und grenzten das Gefundene dann anhand von Kriterien ein, etwa dem Vorhandensein von Experimental- und Kontrollgruppen oder von statistischen Daten, die die Berechnung von Effektstärken erlauben. Die Autorinnen und Autoren gehen in ihrer Metaanalyse 2 Forschungsfragen nach:

1. Wie wirksam ist der Einsatz von KI insgesamt in Bezug auf die Lernleistungen und die Wahrnehmung des Lernens? Wie groß ist der Effekt von KI auf Lernleistungen?
2. Wie beeinflussen verschiedene Moderatorvariablen die Auswirkungen von KI- Einsatz?

Als Moderatorvariablen definieren sie beispielsweise die Größe der Stichprobe, die Dauer einer Intervention oder die Klassenstufe, in der die Studie durchgeführt wurde.

Als Ergebnis finden sie zum einen eine große Effektstärke für den Einsatz von KI auf die Lernleistung und eine geringe auf die Wahrnehmung des Lernens über alle Studien hinweg. Als einflussreiche Moderatorvariablen erweisen sich die Größe der Stichprobe, die Lernstufe, die Fächergruppe (wie Natur- oder Sozialwissenschaften), die Sozialform, die Rolle der KI und die verwendete Hardware.

Die Ergebnisse sind für Lehrkräfte zum einen interessant, weil sie den Schluss nahelegen, dass der Einsatz von KI in sehr verschiedenen Kontexten einen positiven Einfluss auf die Lernleistung und in geringerem Maße auch auf die Wahrnehmung von Lernen haben kann. Zum anderen zeigt die Analyse Ansatzpunkte auf, um KI besonders gewinnbringend einzusetzen. Hier scheint die Arbeit in Gruppen beispielsweise effektiver zu sein als ein Einsatz bei individueller Arbeit.

Hintergrund

Die Möglichkeiten, die durch den Einsatz von KI in Bildungskontexten entstehen, sind seit einigen Jahren Gegenstand der Forschung (Chen, Xie, Zou & Hwang, 2020). Größere Metaanalysen in diesem Bereich

liegen aber nur vereinzelt vor und decken meist spezielle Fragestellungen ab. Diese Lücke füllt die vorliegende Studie, indem sie den Einfluss von KI auf die Lernleistung und die Wahrnehmung von Lernen für den Zeitraum 2001 bis 2020 untersucht.

Dabei wird Lernleistung nach Sung, Chang und Yang (2015) als der Zuwachs an Wissen gesehen, der durch standardisierte Tests gemessen wird. Unter Wahrnehmung von Lernen verstehen die Forschenden nach Chen (2015) die Meinungen der Lernenden zum Thema Lernen, wie die Einstellung zum Lernen, Lernmotivation oder Lernerfahrungen.

Insgesamt wird davon ausgegangen, dass KI Potenziale im Bereich der Individualisierung von Lernprozessen (Hwang, Xie, Wah & Gašević, 2020) und der Verringerung der Arbeitsbelastung von Lehrenden durch die Übernahme von Routinetätigkeiten hat (Chan & Zary, 2019; Chen et al., 2020).

Vor diesem Hintergrund formulieren Zheng et al. zwei Forschungsfragen:

1. Wie wirksam ist der Einsatz von KI insgesamt in Bezug auf die Lernleistungen und die Wahrnehmung des Lernens? Wie groß ist der Effekt von KI auf Lernleistungen?
2. Wie beeinflussen verschiedene Moderatorvariablen die Auswirkungen von KI- Einsatz?

Design

Die Forschenden nutzten für ihre Studie zum Effekt von KI auf Lernleistungen die Datenbanken *Web of Science*, *Scopus* und *ERIC* und durchsuchten diese nach einer Kombination aus Suchwörtern aus den Bereichen „KI“ und „Lernleistung“. Aus den so gefundenen 15 627 Artikeln in Fachzeitschriften wurden doppelte Texte, Studien, die sich nicht mit KI im Bildungsbereich, nicht-experimentelle oder nicht-quasi-experimentelle Studien, Studien ohne Kontrollgruppen und Studien, die nicht die nötigen statistischen Angaben enthielten, um Effektstärken zu berechnen, ausgeschlossen. Schließlich konnten 24 Studien mit 2 908 Teilnehmerinnen und Teilnehmern analysiert werden.

Zur Analyse der Studien bedienten sich die Forschenden eines Kodierschemas, das aus 14 Variablen sowie deren möglichen Ausprägungen bestand, wie z. B. zur Variable *Fächergruppe* die Ausprägungen Naturwissenschaften, Sozialwissenschaften, Ingenieurwissenschaften und Technik. Die 13 Variablen waren *Lernstufe*, *Stichprobengröße*, *Fächergruppe*, *Lernergebnisse*, *Lernmethoden*, *Forschungsdesign*, *Forschungsumfeld*, *Dauer der Intervention*, *Sozialform*, *Rolle der KI*, *Anwendungsbereiche des KI-Einsatzes*, *KI-Software*, *KI-Hardware* und *KI-Technologien*.

Für die Berechnung der Stärke des Effekts von KI auf Lernleistungen wurde diese zunächst für jede Studie kalkuliert. Anschließend wurde die gewichtete Gesamteffektstärke unter Verwendung von Hedges' g berechnet, wobei $g \geq 0.2$ einem kleinen, $g \geq 0.5$ einem mittleren und $g \geq 0.8$ einem großen Effekt entspricht. Zusätzlich wurde unter anderem der Kennwert Q_B berechnet, um zu prüfen, ob sich Effektstärken systematisch zwischen Gruppen bzw. verschiedenen Ausprägungen einer Variablen unterscheiden. Wird dieser Kennwert signifikant ($p \leq .05$), ist von einem systematischen Unterschied auszugehen. Zudem wurden mögliche Verzerrungen (*publication bias*) in Bezug auf die zugrundeliegenden Studien untersucht.

Ergebnisse

Rein deskriptiv zeigt sich, dass der überwiegende Teil der in die Metaanalyse inkludierten Studien aus dem Bereich der Hochschulbildung kommt, mit Stichprobengrößen überwiegend zwischen 101 und 300 Personen. Die meisten Studien fokussieren die Sozialwissenschaften sowie Ingenieurswissenschaften und Technik und nutzen die Lernmethode des problembasierten Lernens. Der überwiegende Teil der untersuchten Studien basiert auf quasi-experimentellen Forschungsdesigns und wurde im Unterrichtskontext durchgeführt. Die Interventionszeiträume sind dabei überwiegend kurz und lagen meist unter zwei Wochen. KI-Anwendungen werden zumeist in individueller Form eingesetzt und primär als intelligente Lernwerkzeuge genutzt. Inhaltlich liegt der Schwerpunkt der KI-Nutzung vor allem auf Leistungsdiagnostik und Evaluation sowie auf kombinierten Anwendungsbereichen. Auf Softwareebene dominieren intelligente tutorielle Systeme und adaptive Lernsysteme, während auf der Hardwareebene hauptsächlich konventionelle Computer verwendet werden. Als technologische Grundlage kommen überwiegend Expertensysteme oder agentenbasierte Systeme zum Einsatz, wobei regelbasierte KI-Algorithmen am häufigsten Anwendung finden.

In Bezug auf ihre erste Forschungsfrage (*Wie wirksam ist der Einsatz von KI insgesamt in Bezug auf die Lernleistungen und die Wahrnehmung des Lernens?* Wie groß ist der Effekt von KI auf Lernleistungen?*)* finden die Forscherinnen, dass der Einsatz von KI einen großen Effekt ($g = 0.812$) auf die Lernleistung und einen kleinen auf die Wahrnehmung des Lernens hat ($g = 0.208$). Beide Effektstärken sind heterogen ($Q = 261.054$, $p < .001$ bzw. $Q = 140.965$, $p < .001$), was bedeutet, dass die Effektstärken der in der Metaanalyse enthaltenen Studien stark variieren und sich systematisch voneinander unterscheiden.

Die zweite Forschungsfrage (*Wie beeinflussen verschiedene Moderatorvariablen die Auswirkungen von KI-Einsatz?*) kann nur im Hinblick auf die Lernleistung beantwortet werden. Da nur 6 der 24 Studien sich mit der Wahrnehmung von Lernen beschäftigen, konnten für diesen Bereich keine Effektstärken für Moderatorvariablen berechnet werden.

Die Ergebnisse zeigen, dass Studien zum KI-Einsatz in den Sekundarstufen (Junior und Senior High School) die größte Effektstärke aufweisen ($g = 1.022$), gefolgt von Studien an Hochschulen, mit Berufstätigen und in der Grundschule. In allen *Lernstufen* zeigen sich signifikante Effektstärken und der Unterschied der Effektstärken zwischen den Stufen ist ebenfalls signifikant ($Q_B = 13.069$, $p = .044$).

Bezüglich der *Stichprobengröße* hat die Kategorie > 300 die größte Effektstärke ($g = 1.380$), gefolgt von 1–50 ($g = 1.166$), 101–300 ($g = 0.718$) und 51–100 ($g = .707$). Auch hier sind die Unterschiede zwischen den Effektstärken signifikant ($Q_B = 14.801$, $p = .002$).

Bei den *Fächergruppen* finden die Forschenden in den Ingenieur- und Technikwissenschaften die größte Effektstärke ($g = 1.151$), gefolgt von Naturwissenschaften ($g = 0.642$) und Sozialwissenschaften ($g = 0.175$). Dabei unterscheiden sich die Fächergruppen in Bezug auf die Effektstärke signifikant voneinander ($Q_B = 9.513$, $p = 0.009$).

Die Effektstärke für kontextbezogenes Lernen erreicht in der Kategorie *Lernmethoden* die höchste Ausprägung ($g = 1.071$). Der Q_B-Kennwert ist hier nicht signifikant, d. h., es bestehen keine signifikanten Unterschiede in der Effektstärke im Vergleich zu den anderen untersuchten Lernmethoden ($Q_B = 9.291$, $p = 0.054$).

In Bezug auf das *Forschungsdesign* weisen experimentelle Designs die größte Effektstärke auf

(g = 1.173), gefolgt von quasi-experimentellen Designs (g = 0.743); beide zeigen signifikante Effektstärken. Der Unterschied zwischen den Designs ist jedoch nicht signifikant (Q_B = 1.748, p = 0.186).

Bezüglich des *Forschungsumfelds* ergibt sich die größte Effektstärke im Labor (g = 1.046), gefolgt von Klassenraum (g = 0.852), Blended-Learning (g = 0.631) und Fernunterricht (g = 0.619). Der Q_B-Kennwert ist nicht signifikant (Q_B = 2.431, p = 0.488).

Bei der *Dauer der Interventi*on ist eine Zeit von 2–4 Wochen am effektivsten (g = 1.089), gefolgt von 5–8 Wochen (g = 1.051), 9–24 Wochen (g = 0.618), < 2 Wochen (g = 0.611) und > 24 Wochen (g = 0.424). Diese Unterschiede sind aber nicht signifikant (Q_B = 7.077, p = 0.132).

Bei den *Sozialformen*, in denen KI zum Einsatz kommt, haben kooperative Lernformen die größte Effektstärke (g = 2.875), gefolgt vom Einsatz in der individuellen Arbeit (g = 0.753). Der Q_B-Kennwert ist signifikant (Q_B = 19.269, p < 0.001), d. h., es besteht ein signifikanter Unterschied zwischen den beiden Effektstärken.

Bei der Moderatorvariable *Rolle der KI* hat die Rolle eines Beraters in Politikprozessen die größte Effektstärke (g = 2.875). Alle KI-Rollen – also auch die als intelligenter Tutor (g = 0.639) und als intelligentes Lernwerkzeug (g = 0.809) – zeigen signifikante Effektstärken. Der Unterschied zwischen den Effektstärken ist signifikant (Q_B = 21.911, p < 0.001).

Bei den *Anwendungsbereichen der KI* hat der Einsatz zum Erstellen personalisierter Empfehlungen die größte Effektstärke (g = 1.084), gefolgt von Tutoring (g = .935). Die Unterschiede in den Effektstärken zwischen den Anwendungsbereichen sind nicht signifikant.

Bezüglich *KI-Software* weisen Agentensysteme die größte Effektstärke auf (g = 2.059), gefolgt von Test- und Diagnosesystemen (g = 1.661) – ohne signifikante Unterschiede in den Effektstärken (Q_B = 10.008, p = 0.124).

Bei *KI-Hardware* hat der Einsatz von mind. zwei unterschiedlichen Geräten die größte Effektstärke (g = 2.132), gefolgt von traditionellen Computern (g = 0.832), Smartphones (g = 0.588) und Tablets (g = 0.193). Der Unterschied in den Effektstärken ist signifikant (Q_B = 39.573, p < 0.001).

Bei *KI-Technologien* erzielt die Arbeit mit Sprachverarbeitungsprogrammen wie z. B. ChatGPT die höchste Effektstärke (g = 1.071), gefolgt vom gemischten Einsatz mehrerer Technologien (g = 0.919) und Experten-/Agentensystemen (z. B. ein automatischer Börsenhandels-Agent, g = 0.642). Die Unterschiede zwischen den verwendeten Technologien sind nicht signifikant (Q_B = 2.946, p = 0.229).

Zudem zeigt die statistische Auswertung, dass die Ergebnisse der vorliegenden Metaanalyse nicht vom *publication bias* betroffen sind.

Diskussion und Einschätzung

Zum Hintergrund

Die Metaanalyse bezieht sich auf die zum Datum ihrer Entstehung nur in geringem Maße vorhandenen Studien zum Einsatz von KI im Bildungsbereich. Aus ihr lassen sich Desiderate in Bezug auf den Effekt

von KI auf Lernleistungen ableiten.

Zum Design

Zwecks Durchführung der Studie findet eine intensive Recherche von vorhandenen Forschungsarbeiten statt. Die Auswahl von Arbeiten, die in die Metaanalyse zum Effekt von KI auf Lernleistungen einbezogen werden, erfolgt nach nachvollziehbaren Kriterien. Die statistischen Berechnungen zeigen, dass die Metaanalyse keiner Verzerrung durch *publication bias* unterliegt, sodass sie als aussagekräftig einzustufen ist. Die Kodierung anhand von Moderatorvariablen und deren Ausprägungen wird transparent gemacht. Zu beachten ist allerdings, dass nur Studien bis 2020 – also bevor der eigentliche KI-Boom durch die breite Verfügbarkeit von Large-Language-Modellen wie ChatGPT begann – Eingang fanden und nur solche, die auf Englisch erschienen sind, was die Frage nach der Übertragbarkeit der Ergebnisse auf deutsche Bildungseinrichtungen aufwirft.

Zu den Ergebnissen

In der vorliegenden Untersuchung zum Effekt von KI auf Lernleistungen zeigt sich, dass KI-Einsatz einen großen signifikanten Effekt auf die Lernleistung und einen kleinen signifikanten Effekt auf die Wahrnehmung des Lernens hat, d. h., dass Lernende, die KI nutzen, andere Lernende, die ohne KI lernen, in Bezug auf Leistung und die Wahrnehmung des Lernens, wie bspw. Motivation, übertreffen.

In Bezug auf die 13 untersuchten Moderatorvariablen haben die größten Effektstärken: ein Einsatz in den Sekundarstufen (*Lernstufe*), eine *Stichprobengröße* > 300, ein Einsatz in den Ingenieur-/Technikwissenschaften (*Fächergruppe*) und in einem Setting mit kontextbezogenem Lernen als *Lernmethode*, experimentelle *Forschungsdesigns*, eine Laborumgebung als *Forschungsumfeld*, eine *Interventionsdauer* von 2–4 Wochen, ein Einsatz in Gruppen (*Sozialform*), KI in der *Rolle* eines politischen Beraters, die Nutzung von KI als Tool für personalisierte Empfehlungen (*Anwendungsbereich*), die Arbeit mit mehreren *Hardware*geräten und die Nutzung von KI als Sprachverarbeitungsprogramm (*KI-Technologien*).

Signifikante Unterschiede zwischen Effektstärken eines Moderators gibt es bei der *Lernstufe*, der *Stichprobengröße*, der *Fächergruppe*, der *Sozialform*, der *Rolle der KI* und der genutzten *KI-Hardware*.

Keine signifikanten Unterschiede zwischen den Ausprägungen eines Moderators finden sich bei den Variablen *Lernmethode* (knapp nicht signifikant), *Forschungsdesign* (experimentell vs. quasi-experimentell), *Forschungsumfeld*, *Interventionsdauer*, *Anwendungsbereich der KI*, *KI-Software* und *KI-Technologie*.

Relevant ist die Studie zum einen deshalb, weil sie zeigt, dass ein Effekt von KI auf Lernleistungen vorhanden ist, d. h., dass der Einsatz von KI im Bildungsbereich große Chancen bietet. Ihre Nutzung sollte also grundsätzlich als Möglichkeit in vielen schulischen Kontexten mitgedacht werden. Zum anderen zeigt die Metaanalyse, dass die Rahmenbedingungen eines KI-Einsatzes einen großen Einfluss auf dessen Effektivität haben. Es ist also grundsätzlich nötig, diese Rahmenbedingungen genau zu betrachten und den KI-Einsatz immer wieder zu hinterfragen und anzupassen. Dazu bedarf es eines entsprechenden Knowhows auf Seiten der Lehrkräfte.

Reflexionsfragen für die Praxis

Nachfolgende Reflexionsfragen sind ein Angebot, die Befunde der rezensierten Studie auf das eigene Handeln als Lehrkraft oder Schulleitungsmitglied zu beziehen und zu überlegen, inwiefern sich Anregungen für die eigene Handlungspraxis ergeben. Die Befunde der rezensierten Studien sind nicht immer generalisierbar, was z. B. in einer begrenzten Stichprobe begründet ist. Aber auch in diesen Fällen können die Ergebnisse interessante Hinweise liefern, um über die eigene pädagogische und schulentwicklerische Praxis zu reflektieren.

Reflexionsfragen für Lehrkräfte

- Welchen Stellenwert hat die Arbeit mit KI in meinem Unterricht?
- In welchen Kontexten (Sozialformen, Vermittlungsformen etc.) setze ich KI ein?
- Inwieweit reflektiere ich die Rahmenbedingungen des KI-Einsatzes?
- Welche weiteren Möglichkeiten gibt es, die positiven Effekte von KI zur Förderung meiner Schülerinnen und Schüler und zur Erleichterung meiner eigenen Arbeit zu nutzen?
- Wie kann ich meine Kompetenzen in diesem Bereich ausbauen?

Reflexionsfragen für Schulleitungen

- Welchen Stellenwert hat die Arbeit mit KI an meiner Schule?
- Wo gibt es Potenziale, verstärkt die positiven Effekte von KI zur Förderung von Schülerinnen und Schülern und zur Erleichterung meiner eigenen Arbeit zu nutzen?
- Was kann ich tun, um die Kompetenzen meiner Lehrkräfte in diesem Bereich zu stärken?

Literatur

Chan, K. S. & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education, 5*(1), e13930. https://doi.org/10.2196/13930

Chen, K. T. C. (2015). Exploring college students' usage experiences, perceptions and acceptance of mobile English learning in Taiwan. *The International Technology Management Review, 5*(4), 162–171. https://doi.org/10.2991/itmr.2015.5.4.1

Chen, X., Xie, H., Zou, D. & Hwang, G. J. (2020). Application and theory gaps during the rise of Artificial Intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002, 1–20. https://doi.org/10.1016/j.caeai.2020.100002

Hwang, G. J., Xie, H., Wah, B. W. & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001, 1–5. https://doi.org/10.1016/j.caeai.2020.100001

Sung, Y. T., Chang, K. E. & Yang, J. M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review, 16*, 68–84. https://doi.org/10.1016/j.edurev.2015.09.001

Rezensent/-in

Dr. Sonja Hensel, Lehrerin am Berufskolleg in Siegburg sowie Lehrbeauftragte an der Universität Siegen. Arbeitsschwerpunkte: Rechtschreib-, Schreib- und Lesedidaktik, selbstreguliertes und kooperatives Lernen.

Zitiervorschlag

Hensel, S. (2026). Rezension zu Zheng, L., Niu, J., Zhong, L. & Gyasi, J. F. (2023). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. Interactive Learning Environments, 31(9), 5650–5664. *Forschungsmonitor Schule, 196*. Abgerufen von https://www-.forschungsmonitor-schule.de/print.php?id=191

Urheberrecht