

Chatbots im Unterricht: Welche Lernergebnisse werden unterstützt?

Die Integration von Chatbots in die Bildung ist ein wachsender Trend, der neue Möglichkeiten für individualisiertes Lernen und die Unterstützung von Lehrkräften bietet. Doch wie effektiv sind Chatbots wirklich, um unterschiedliche Lernerfolge¹ zu fördern und Wissen nachhaltig zu sichern? Welche Bedingungen beeinflussen ihre Wirksamkeit, und wo liegen ihre Grenzen? Diese Fragen untersuchen Deng und Yu (2023) in ihrer Studie, „A meta-analysis and systematic review of the effect of chatbot technology use in sustainable Education“². Darin analysieren die Autoren 32 Studien mit insgesamt 2.201 Teilnehmenden. Die Ergebnisse beleuchten die Potenziale und möglichen Einschränkungen von Chatbots im Bildungskontext.

EINLEITUNG. Die Integration von auf künstlicher Intelligenz basierenden Chatbots in den Unterricht gewinnt zunehmend an Bedeutung. Dies gilt insbesondere vor dem Hintergrund der digitalen Bildungsreform und dem Ziel, digitale Kompetenzen zu fördern. Für den MINT-Unterricht bieten Chatbots als Lehrassistenten und Tutoren neue Möglichkeiten, SchülerInnen individuell zu unterstützen und komplexe Inhalte zugänglicher zu machen. Dies steht im Einklang mit Bildungsstandards, die selbstgesteuertes Lernen und den Einsatz moderner Technologien betonen. In der Bildungsforschung wird jedoch kontrovers diskutiert, wie effektiv solche Technologien im Vergleich zu traditionellen Lehrmethoden tatsächlich sind, beispielsweise auch bezüglich der Motivation von Lernenden. Die Metaanalyse untersucht auf Basis von Theorien des technologiegestützten Lernens, wie Chatbots sieben unterschiedliche Lernergebnisse fördern können, wie Lernerfolg, Behaltensleistung und kritisches Denken. Außerdem analysiert sie

METAANALYSE IM ÜBERBLICK

Fokus der Studie	Wirksamkeit von Lernen mit Chatbots auf sieben verschiedene Lernergebnisse sowie Analyse von <u>Moderatorvariablen</u> , wie Chatbot-Rolle, Nutzungsdauer und Lerninhalt
Untersuchte Zielgruppe	2.201 SchülerInnen und Studierende verschiedener Bildungskontexte
Durchschnittliche Effektstärke	Mittlerer bis hoher Gesamteffekt (Cohen's $d = 0.79$)
Weitere Befunde	Starke Effekte bei expliziter Argumentation ($d = 1.19$) und Lernleistung ($d = 1.03$) keine signifikanten Effekte bei kritischem Denken, Engagement und Motivation

¹Die unterstrichenen Fachbegriffe sind auf unser [Online-Glossar](#) verlinkt.

²Nach Angaben der Zeitschrift wird ‚sustainable education‘ als das Finden nachhaltiger Lösungen für Umwelt-, soziale und wirtschaftliche Probleme durch Bildung definiert.

verschiedene Einsatzbedingungen, die die Effektivität von Chatbots beeinflussen, um praktische Handlungsempfehlungen für den Unterricht zu liefern.

WORUM GEHT ES IN DIESER STUDIE? In dieser Metaanalyse werden Chatbots von den AutorInnen als ein Dialogprogramm definiert, das Eingaben verarbeiten und dementsprechend Informationen durch verbale oder schriftliche Interaktion bereitstellen kann. Dabei werden experimentelle Designs mit Kontrollgruppen genutzt, um die Effektivität von Chatbots mit herkömmlichem Unterricht zu vergleichen. Hierfür beziehen sie 32 Studien mit insgesamt 2.201 Teilnehmenden ein. Die möglichen Einflüsse solcher Chatbots werden in drei zentralen Fragestellungen analysiert.

Erstens untersuchen sie, ob das Lernen mit Chatbots effektiver ist als das Lernen mit traditionellen Methoden, indem sie die Effekte von sieben verschiedenen Lernergebnissen zu einem Gesamteffekt kombinieren. Zweitens untersuchen die AutorInnen, die Auswirkungen von Chatbots auf jede der sieben verschiedenen Lernkomponenten wie z.B. Lernerfolg, Behaltensleistung, Motivation und Interesse (Definitionen und Effekte siehe Befunde zu Lernergebnisse). Drittens analysieren sie, ob Faktoren wie Dauer der Intervention, Rolle des Chatbots oder die Art der Lerninhalte die Effektivität der Chatbots beeinflussen.

Die Studien, die zwischen 2010 und 2022 veröffentlicht wurden, untersuchen Chatbots in verschiedenen Kontexten, darunter Sprachunterricht und technische Fächer. Dabei nutzen sie standardisierte Tests, Befragungen und Beobachtungen als Messmethoden. Zudem decken die Analysen ein breites Spektrum an Bildungskontexten ab, darunter Schulen und Hochschulen. Die Autoren differenzieren zudem zwischen drei Rollen, die Chatbots einnehmen können:

- Tutoren stellen und beantworten selbst Fragen und führen die SchülerInnen durch den Unterricht
- Lehrassistenten übernehmen eine unterstützende Funktion, indem sie professionelles Wissen bereitstellen, professionelles Feedback geben und den Lernprozess strukturieren
- Lernpartner kommunizieren und interagieren mit SchülerInnen – entweder über Text oder Sprache – weniger formal.

WAS FAND DIESE STUDIE HERAUS? Insgesamt zeigt diese Metaanalyse, dass der Einsatz von Chatbots im Unterricht einen mittleren bis großen positiven Einfluss auf den Lernergebnisse hat. Der durchschnittliche Gesamteffekt beträgt $d = 0.79$, was darauf hindeutet, dass Lernende, die mit Chatbots arbeiten, tendenziell bessere Ergebnisse erzielen als Lernende, die traditionelle Methoden nutzen. Darüber hinaus zeigen sich starke Effekte im Bereich der expliziten Argumentation ($d = 1.19$) und des Lernerfolgs ($d = 1.03$). Dies legt nahe, dass Chatbots besonders gut geeignet sind, argumentative Lernprozesse zu strukturieren und auch die Prüfungsleistung zu fördern.

Ebenso finden sich signifikant positive Effekt auf die Behaltensleistung ($d = 0.69$). Dies zeigt, dass Chatbots Lerninhalte nicht nur kurzfristig vermitteln, sondern auch langfristig im Ge-

dächtnis verankern können. Ein ähnlicher Effekt ist bei Interesse ($d = 0.84$) zu beobachten, was darauf hindeutet, dass Chatbots die Lernbereitschaft von SchülerInnen fördern können.

Bei den Auswirkungen von Chatbots gibt es nur sehr geringe Unterschiede zwischen verschiedenen Nutzungsarten, für verschiedene Fächer und für verschiedene Zeiträume. Die größten deskriptiven Effekte wurden jedoch bei der Verwendung von Chatbots als Lernassistent ($d = 1.63$), bei einer Dauer von weniger als fünf Wochen ($d = 1.06$) und in den technischen Fächern Informatik ($d = 0.70$) und Unterrichtstechnologie ($d = 0.93$) festgestellt. Dies deutet auf die Nützlichkeit von strukturiertem und klarem Feedback in diesen Bereichen hin. Die Befunde zu den einzelnen Lernergebnissen und weiteren Moderatoren sind der jeweiligen Gesamtübersicht zu entnehmen.

WIE BEWERTET DAS CLEARING HOUSE UNTERRICHT DIESE STUDIE? Die *Clearing House Unterricht Research Group* bewertet die Metaanalyse anhand der folgenden fünf Fragen und orientiert sich dabei an den Abelson-Kriterien (1995):

Wie substantiell sind die Effekte? Nach der Klassifizierung von Cohen (1988) war der Gesamteffekt von Chatbots im Unterricht positiv ($d = 0.79$). Das bedeutet, dass Lernende, die auf Chatbots basierendes Lernen nutzten, im Durchschnitt bessere Lernergebnisse (z. B. Leistung, Argumentation) erzielten als Lernende, die traditionelle Lernansätze verwendeten. Um diesen Effekt zu veranschaulichen: Wenn man aus jeder experimentellen Gruppe zufällig einen Lernenden auswählt, ist die Wahrscheinlichkeit, dass der mit Chatbots Lernende, bessere Ergebnisse erzielt, um etwa 70 % höher als bei dem Lernenden, der sie nicht nutzt.

Dieser Effekt war besonders stark bei den Lernleistungen ($d = 1.03$). Das bedeutet, dass etwa 85 % der SchülerInnen, die Chatbots verwenden, bessere Testergebnisse erzielen würden als der Durchschnitt der SchülerInnen, die traditionelle Lernansätze verwenden.

Alle Einzelstudien in der Metaanalyse zeigten positive Auswirkungen von Chatbots auf die Lernergebnisse. Die Metaanalyse zeigte jedoch keinen signifikanten Unterschied zwischen dem Lernen mit Chatbots und den traditionellen Ansätzen in Bezug auf die Lernergebnisse kritisches Denken, Lernmotivation und Lernengagement auf. Dennoch wurde festgestellt, dass Chatbots über verschiedene Moderatoren, wie Nutzungsdauer, verschiedene Lerninhalte und bei der Verwendung verschiedener Rollen, hinweg effektiv sind.

Erfahren Sie mehr über die Einschätzung von Effektstärken in unserem Handout.

Wie differenziert sind die Ergebnisse dargestellt? In dieser Studie wird die Wirkung des Lernens mit Chatbots auf verschiedene Weise differenziert. Die Effektgrößen werden nach Fachinhalten getrennt, einschließlich des MINT-Fachs Informatik. Darüber hinaus wurden Effektgrößen für sieben verschiedene Lernergebnisse angegeben: kritisches Denken, explizite Argumentation, Lernerfolg, Behaltensleistung, Lernengagement, Lernmotivation und Interesse. Die Effekte wurden auch in Abhängigkeit von der Rolle des Chatbots beim Lernen der SchülerInnen (d. h. Tutoren, Lernpartner oder Lehrassistenten) und der Dauer der Chatbot-Nutzung berichtet. Einschränkend ist festzuhalten, dass die Effektgrößen nicht nach

Altersgruppen oder Klassenstufen unterteilt werden. Da die Studien in dieser Metaanalyse Stichproben aus verschiedenen Schulformen der Primar- und Sekundarstufe sowie der Hochschulbildung umfassten, kann nicht dargestellt werden, ob sich die Auswirkungen von Chatbots zwischen diesen Schulstufen unterscheiden.

Wie verallgemeinerbar sind die Befunde? Laut der Metaanalyse haben Chatbots ihre Wirksamkeit für verschiedene Lernergebnisse und in mehreren Lernkontexten und Behandlungsbedingungen unter Beweis gestellt.

Im Lernkontext war die Wirkung von Chatbots über die Fächer hinweg größtenteils stabil. Ebenso waren die Lerneffekte durch Chatbots über unterschiedliche Zeiträume (d. h. von weniger als einer Woche bis zu mehr als zehn Wochen) und für verschiedene Rollen ähnlich. Diese Ergebnisse zeigen, dass die Effektivität von Chatbots für diese Lernkontexte tendenziell verallgemeinerbar ist.

In Bezug auf die Ergebnisse zeigte sich auch, dass das Lernen mit Chatbots gegenüber traditionellen Lernansätzen einen deutlichen Vorteil für den Lernerfolg bietet. Diese Effekte waren jedoch über die Vielzahl der untersuchten Lernergebnisse weniger verallgemeinerbar. Die Effekte waren robust für den Lernerfolg und basierten auf 31 Studien. Andere Ergebnisse, wie beispielsweise das Interesse basierten aber auf weniger Studien, was ihre Verallgemeinerbarkeit trotz möglicher positiver Trends einschränkte. Bei anderen Ergebnissen, wie z. B. kritisches Denken, unterschied sich der Einsatz von Chatbots nicht von traditionellen Methoden. Daher kann angenommen werden, dass die Wirkung von Chatbots vom Lernziel abhängt.

Einschränkend ist festzuhalten, dass andere Aspekte des Lernkontexts (z.B. Klassenstufe, geographische Lage) sowie unterschiedliche methodische Merkmale (z.B. Stichprobengröße, Studiendesign) nicht als Moderatoren mitberücksichtigt wurden, weswegen keine Aussage über die Generalisierbarkeit dieser Aspekte getroffen werden kann. Auch für eine differenzierte Aussage über die Wirksamkeit in anderen MINT-Fächern bedarf es weiterer experimenteller Studien.

Was macht die Metaanalyse wissenschaftlich relevant? Mit der wachsenden Verbreitung von KI und Chatbot-Technologie und ihrer Integration in den Unterricht nimmt die Forschung zur Lerneffektivität weiter zu. Da es sich um ein relativ neues Forschungsthema handelt, liefern erste Metaanalysen spezialisierte Informationen über aufkommende Trends, die Lehrkräften erste Einblicke in dieses dynamische Phänomen bieten. Da sich diese Studie auf den Einsatz von Chatbots in der Allgemeinbildung und nicht nur auf bestimmte Fachbereiche oder Chatbot-Rollen konzentriert, ist sie die erste, die Einblicke in die umfassendere Frage nach der Wirksamkeit von Chatbots für das Lernen im Bildungsbereich gibt.

Darüber hinaus sollten diese Ergebnisse im Bildungskontext zwar als vorläufig betrachtet werden, doch die strenge Auswahl der Studien, durch die AutorInnen (z.B. Äquivalenz von Kontroll- und Experimentalgruppe) und ihre Differenzierung zwischen den sieben Lernergebnissen, verleihen ihren Erkenntnissen zur Effektivität von Chatbots sowohl Stärke als auch Nuancen. Die Metaanalyse ist zudem wissenschaftlich bedeutsam, da sie ein besonders interessantes Ergebnis liefert: Im Gegensatz zur allgemeinen Meinung legen die bisherigen

Befunde nahe, dass Chatbots keinen deutlichen Einfluss auf Motivation, Engagement und kritisches Denken haben.

Wie methodisch verlässlich sind die Befunde? Die Offenlegung und Begründung des methodischen Vorgehens entspricht überwiegend den Kriterien gängiger Anforderungskataloge (z.B. APA Meta-Analysis Reporting Standards). Die Gesamtpunktzahl liegt im oberen Drittel. Somit werden das systematische Vorgehen der Metaanalyse bei der Suche nach relevanten Studien, die Auswahl der Studien, die Kodierung und die statistische Analyse weitgehend transparent und nachvollziehbar berichtet und beschrieben. Diese Studie zeigte beispielsweise besondere Stärken bei der Ermittlung relevanter Schlüsselwörter für die Datenbanksuche unter Verwendung einer neuartigen Clustering-Methode und der Durchführung mehrerer statistischer Prüfungen (z. B. Publikationsbias, Sensitivitätsanalyse) zur Sicherstellung der Validität. Die Datenbanksuche hätte jedoch durch zusätzliche Suchmethoden ergänzt werden können (z. B. Rückwärtssuche in Referenzlisten), um weitere relevante Studien zu finden, die einige Ergebniskategorien möglicherweise robuster gemacht hätten (z. B. kritisches Denken).

Weitere Informationen zur methodischen Beurteilung finden Sie in unserem Rating Sheet.

FAZIT FÜR DIE UNTERRICHTSPRAXIS. Die Ergebnisse dieser Metaanalyse bieten einen ersten Einblick in die Forschung zur Effektivität von Chatbots im Bildungsbereich. So erhalten Lehrkräfte einen ersten Überblick über bestimmte Lernergebnisse, Kontexte und Bedingungen, mit welchen die Integration von Chatbots erfolgreich gelingen kann. Was die Auswirkungen auf das Lernen betrifft, können Lehrkräfte positive Ergebnisse bei Leistungstests und möglicherweise auch einen positiven Einfluss auf explizite Argumentation und die Behaltensleistung erwarten.

Hinsichtlich des Lernengagements und der Lernmotivation der Lernenden scheinen Chatbots ihrem angenommenen Einfluss jedoch möglicherweise nicht gerecht zu werden. Weiterhin ist unklar, ob Chatbots kritisches Denken fördern (siehe dazu unser Kurzreview). Es ist auch erwähnenswert, dass ein potenzielles Risiko bei der Verwendung von Chatbots der Neuigkeitseffekt ist, bei dem sich die Ergebnisse anfangs verbessern, die Effekte aber nachlassen, sobald der Reiz des Neuen nachlässt (Clark, 1983; e.g., Fryer et al., 2017). Dies kann vermieden werden, indem sich SchülerInnen zunächst mit dem Chatbot vertraut machen, bevor sie ihn für Lernaktivitäten verwenden.

Die AutorInnen betonen, dass Chatbots in technischen Kursen, insbesondere als Tutoren oder Lernpartner die größten und robustesten Effekte erzielen, wenn sie über einen Zeitraum von weniger als fünf Wochen eingesetzt werden. In diesen Kontexten können sie gezielt eingesetzt werden, um explizite Argumentation oder die Behaltensleistung zu fördern.

STUDIENBEISPIEL

Eine Laborstudie von Ruan und Kollegen (2020) zeigte signifikante Auswirkungen eines Lerndesigns mit einem Chatbot-Tutor. Die Studie untersuchte eine interaktive mathematische Lernumgebung für SchülerInnen der 3. bis 5. Klasse und ihre Auswirkungen auf Engagement, Lernen und Wissenserhalt in Mathematik. Sie verglichen vier Varianten der Lernplattform:

- Gruppe A: Nur mathematische Textaufgabe (Kontrollgruppe)
- Gruppe B: Dieselbe mathematische Aufgabe eingebettet in eine märchenhafte Geschichte
- Gruppe C: Dieselbe mathematische Aufgabe in einer Geschichte mit Schritt-für-Schritt-Anleitung zur Lösung
- Gruppe D: Dieselbe Mathematische Aufgabe in einer Geschichte plus einen Chatbot-Tutor. Der Chatbot bot den SchülerInnen personalisierten Smalltalk, Ermutigung, wenn sie nicht weiterkamen, adaptive Hinweise im Konversationsstil und Fragen, um das Verständnis zu überprüfen.

Die Studie ergab, dass die erzählbasierte Aktivität plus Chatbot-Variante (D) im Vergleich zur reinen Aufgabenvariante (A) einen großen und signifikanten Effekt auf das Engagement ($ES = 0.90$), das Lernen ($ES = 0.88$) und den Wissenserhalt der Konzepte, eine Woche nach der Aktivität, hatte.

REFERENZEN.

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ahn, S., Ames, A. J. & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, 82 (4), 436-476. doi: 10.3102/0034654312458162
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53 (4), 445-459. doi: 10.3102/00346543053004445
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Aufl.). Hillsdale, NJ: Erlbaum.
- Deng, X. & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15 (4), 2940. doi: 10.3390/su15042940
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A. & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75, 461-468. doi: 10.1016/j.chb.2017.05.045
- Ruan, S., He, J., Ying, R., Burkle, J., Hakim, D., Wang, A., ... Landay, J. A. (2020). Supporting children's math learning with feedback-augmented narrative technology. *In Proceedings of the interaction design and children conference*, 567-580. doi: 10.1145/10.1145/3392063.3394400

LINKS.

Zur Metaanalyse von Deng & Yu, 2023

Zum Studienbeispiel von Ruan et al., 2020

ZITIEREN ALS.

Arvaneh, B., Farrell, M. & CHU Research Group (2024). Chatbots im Unterricht: Welche Lernergebnisse werden unterstützt? www.clearinghouse-unterricht.de, *Kurzreview* 34.



Dieses Kurzreview ist lizenziert unter einer [Creative Common Namensnennung - Keine Bearbeitungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-nc/4.0/). Verwendung und Verbreitung unter Namensnennung erlaubt, keine Veränderungen gestattet.