

Die Integration generativer KI in die hochschulische Statistiklehre

Eine Untersuchung zur Förderung affektiver und kognitiver Zielgrößen¹

Kim Méliani und Kristina Kögler

Statistik-Veranstaltungen werden in erziehungswissenschaftlichen Studiengängen oft mit erheblichen Vorbehalten besucht. Dies ist größtenteils auf defizitäre Selbstkonzepte, aber auch Unsicherheiten und Wissensdefizite im Umgang mit den statistischen Methoden zurückzuführen. Generative KI bietet zwischenzeitlich große Lernpotenziale und es stellt sich die Frage, inwiefern sich diese in der Statistiklehre gezielt nutzen lassen. Der Beitrag zielt auf die Untersuchung der Wirksamkeit von KI-unterstützter Statistiklehre mittels ChatGPT und bezieht dabei sowohl affektive als auch kognitive Zielgrößen ein. Dabei werden quantitative Daten (N = 72) eines Statistikseminars zur Programmiersoftware R im Masterstudiengang Berufspädagogik der Universität Stuttgart herangezogen. Die Ergebnisse deuten auf signifikante positive Veränderungen des statistikbezogenen Selbstkonzeptes sowie der Statistikangst, aber auch der Wissensentwicklung in der Studierendengruppe hin, die mit generativer KI arbeitete.

The Integration of Generative AI in University-Level Statistics Education – Investigating the Effects on Affective and Cognitive Learning Outcomes

Statistical courses of educational science programs are often attended with considerable reservations, which are largely due to deficient self-concepts as well as uncertainties and knowledge gaps in dealing with statistical methods. In this context, generative AI offers substantial support and learning potential, raising the question to what extent these resources can be strategically utilized in statistics education. This study aims to investigate the effectiveness of AI-supported statistics education using ChatGPT, considering both affective and cognitive outcome measures. For this purpose, quantitative data (N =

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

72) from a statistics seminar on the programming software R in the Master's program in Vocational Education at the University of Stuttgart are used. The results indicate significant positive changes in statistics-related self-concepts and statistics anxiety, as well as knowledge development in the group of students that utilized generative AI.

Einleitung

Die Teilnahme an statistischen Lehrveranstaltungen wird von Studierenden erziehungswissenschaftlicher Studiengänge vielfach mit beträchtlichen Vorbehalten assoziiert. Diese Vorbehalte resultieren in erster Linie aus defizitären Selbstkonzepten sowie erheblichen motivationalen und emotionalen Beeinträchtigungen, welche das Lernverhalten negativ beeinflussen können (vgl. Faber/Drexler/Stappert 2018; Paulitsch 2013: 8; Bandalos/Finney/Geske 2003: 605). Die Ergebnisse empirischer Studien legen nahe, dass Studierende mit einem positiv ausgeprägten Fähigkeitsselbstkonzept über angemessenere Lernstrategien verfügen und dadurch bessere akademische Leistungen erzielen (vgl. Filiz et al. 2020; Eccles/Wigfield 2020; Fenning/May 2013; O'Mara et al. 2006). Zudem zeigen Befunde aus der Leistungsangstforschung, dass vor allem Besorgtheitskognitionen einen negativen Einfluss auf das Lernen und die Leistung ausüben (vgl. von der Embse, et al. 2018; Onwuegbuzie/Wilson 2003). Diese Kognitionen können die Anwendung geeigneter Lernstrategien und damit die akademische Leistung negativ beeinflussen, was den entscheidenden Unterschied zwischen Studierenden mit starkem und schwachem Fähigkeitsselbstkonzept weiter verdeutlicht. Die bereits vorhandene Skepsis der Studierenden gegenüber statistischen Konzepten und Methoden wird durch die zusätzliche Herausforderung, statistische Verfahren unter Zuhilfenahme von programmiernäher Statistik-Software wie R anzuwenden, weiter verstärkt. Mit der fortschreitenden Etablierung von künstlicher Intelligenz (KI) im Kontext von Lehren und Lernen eröffnen sich jedoch neue motivationale und didaktische Potenziale, insbesondere durch den Einsatz generativer KI als Lern- und Übungssystem, um in einer unterstützenden und interaktiven Lernumgebung Ängste abzubauen und Unsicherheiten im Umgang mit statistischen Konzepten zu überwinden (Ogunleye et al. 2024: 2). Der vorliegende Beitrag widmet sich der Frage, ob und inwiefern sich über den Einsatz generativer KI in der Statistiklehre positive Effekte auf die affektiven und kognitiven Outcomes auf Seiten der Studierenden erzielen lassen. Die Datengrundlage entstammt dem Statistik-Kurs »Einführung in die Pro-

grammiersprache R«, welcher an der Universität Stuttgart von Studierenden des Masterstudiengangs »Berufspädagogik und Personalentwicklung« als Pflichtmodul besucht wird. Im Rahmen einer quasi-experimentellen Interventionsstudie mit Experimental- und Kontrollgruppe wurden lehrbegleitend Daten gesammelt, die für die vorgestellten Analysen herangezogen werden.

Aufbau des Lehrkonzepts zur Förderung affektiver wie kognitiver Zielgrößen

Der Kurs »Einführung in die Programmiersprache R« ist ein unbenoteter Pflichtkurs, dessen Inhalt in drei Hauptmodule unterteilt ist: univariate, bivariate und multivariate Statistik. Im Rahmen der Lehrveranstaltung werden in Form von ILIAS-Tests und Skript-Abgaben formative Assessments durchgeführt, um die statistischen Inhalte sowie das Anwendungswissen im Umgang mit der Programmiersprache R formativ zu erfassen. Den Studierenden wird eine einwöchige Bearbeitungszeit eingeräumt, die kooperatives Arbeiten ermöglicht und sie gezielt auf die Anforderungen im Berufsleben vorbereiten soll. Seit dem Wintersemester 2023 wird zusätzlich der lehr- und lernbegleitende Einsatz generativer KI mittels ChatGPT durchgeführt. Hiermit kommt ein neuer Fokus hinzu, nämlich die implizite Förderung von AI Literacy. In mehreren Präsenzterminen wurden Themen wie die Potenziale und Grenzen des Sprachmodells behandelt, um ein kritisches Verständnis der Technologie zu fördern. Anhand konkreter Einsatzbeispiele wurden die Datenaufbereitung und -auswertung sowie das Nachvollziehen von Fehlermeldungen erläutert. Zudem wurde aufgezeigt, wie durch gezieltes Prompting komplexe Grafiken erstellt und durch präzise Input-Informationen die Qualität des Outputs sukzessive verbessert werden können. Im Verlauf des gesamten Semesters hatten Studierende außerdem die Möglichkeit, sich auch außerhalb der Sitzungstermine zu individuellen Anwendungsfällen beraten zu lassen.

Forschungsfragen und Instrumente

Ausgehend von dem skizzierten Lehrkonzept soll die folgende Fragestellung untersucht werden: Wie verändern sich die statistikbezogenen Selbstkonzepte und Wissensgrundlagen der Studierenden im Zeitverlauf mit und ohne Nutzung generativer KI in der Lehrveranstaltung? Zur Beantwortung dieser Fra-

gestellung kann auf Daten einer lehrbegleitenden Untersuchung im Rahmen des Kurses *Einführung in die Programmiersprache R* zurückgegriffen werden. Es stehen teilweise längsschnittliche Daten aus vier Semesterdurchgängen zur Verfügung, die zu Beginn (t₁) als auch am Ende des viermonatigen Semesters (t₂) erhoben wurden. Im Folgenden soll die Frage weiter ausdifferenziert werden, indem zwischen *affektiven* und *kognitiven Dimensionen* von Veränderungen unterschieden wird:

Affektive Ebene: Wie entwickelt sich das *statistische Fähigkeitsselbstkonzept* über das Semester hinweg? Reduziert sich die *Statistikangst* im Laufe des Semesters?

Kognitive Ebene: Wie verändert sich die Selbsteinschätzung *statistischen Wissens* über die Zeit hinweg? Wie schneiden die Studierenden beim *Wissenstest* ab, der am Ende des Semesters vor Ort durchgeführt wird?

Dabei soll genauer betrachtet werden, ob Unterschiede zwischen den Semestern, in denen KI in der Lehre eingesetzt wurde (Experimentalgruppe/EG) und den Semestern ohne KI in der Lehre (Kontrollgruppe/KG), bestehen. Mögliche Störfaktoren, wie parallele Veranstaltungen mit vergleichbaren Inhalten oder unterschiedliche Lehrende, können in diesem Setting ausgeschlossen werden, da die Datenerhebung ausschließlich in Lehrveranstaltungen derselben Dozierenden erfolgte. Zur Messung der Veränderung zwischen den Gruppen sowie über die Zeitpunkte hinweg wurden gemischte Varianzanalysen sowie t-Tests für abhängige und unabhängige Stichproben auf Basis von etablierten Fragebogen- und selbstentwickelten Testinstrumenten berechnet. Das *statistische Fähigkeitsselbstkonzept* wurde in Anlehnung an Marsh und Shavelson (1985) durch vier Items auf einer vierstufigen Likert-Skala erfasst. Die interne Konsistenz der Skala erreichte zu beiden Messzeitpunkten akzeptable Werte $\alpha_{t_1} = 0.76$ und $\alpha_{t_2} = 0.85$. Die *Statistikangst* (Faber/Drexler/Stappert 2018) wurde im Längsschnitt erhoben. Diese wurde mittels 17 Items, unterteilt in die Subdimensionen Besorgtheits-, Vermeidungs- sowie Aufgeregtheitskognitionen, auf einer vierstufigen Likert-Skala gemessen. Die interne Konsistenz der Skala erreichte zu beiden Messzeitpunkten exzellente Werte $\alpha_{t_1/t_2} = 0.95$. Infolge des späteren Einsatzes des Instruments können im Folgenden lediglich für die EG-Gruppe Entwicklungsdaten aufgezeigt werden. Zur Messung des selbst eingeschätzten Statistikwissens wurde eine Skala mit elf selbst konstruierten Items eingesetzt, die sich an den im Laufe des Semesters behandelten Statistikinhalten orientierte, wobei ebenfalls eine vierstufige

Likert-Skala zum Einsatz kam. Die interne Konsistenz der Skala erreichte zu beiden Messzeitpunkten sehr gute Werte $\alpha_{t1} = 0.90$ und $\alpha_{t2} = 0.87$.

Stichprobe und Ergebnisse

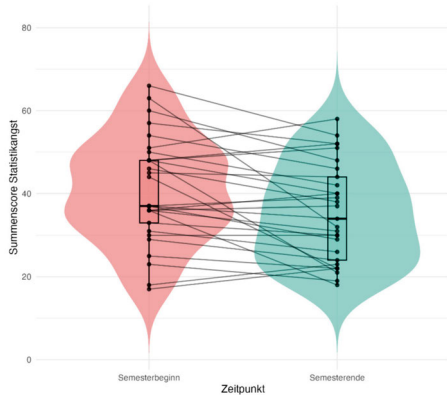
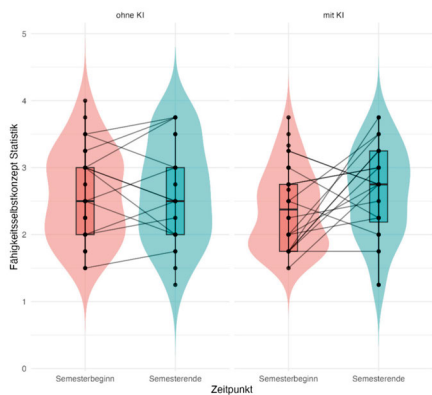
Die Gesamtstichprobe umfasst $N = 72$ Studierende, wobei pro Semester etwa 16 bis 20 Studierende vertreten sind. Davon sind $N = 60$ Studierende im Masterstudiengang Berufspädagogik und $N = 12$ Studierende im Masterstudiengang Digital Humanities. Die Auswertung der Stichprobe ergibt, dass 94 % ($N = 68$) der Befragten weiblich sind. Zudem zeigt sich, dass mit 72 % ($N = 50$) die Mehrheit der Studierenden über keine Vorerfahrung mit der R-Software verfügt. In der KG befinden sich das *Wintersemester 2022* und *Sommersemester 2023*, in der EG das *Wintersemester 2023* und das *Sommersemester 2024*. Zu Beginn des Semesters wurden die Studierenden gebeten eine Gesamteinschätzung (Note 1 bis Note 5) ihrer bisherigen Leistung in zuvor besuchten Statistik-Kursen abzugeben. Dabei fällt auf, dass die Studierenden der EG ihre eigene Leistung tendenziell schlechter beurteilen ($MW_{KG} = 3.36$ (1.09), $MW_{EG} = 3.71$ (1.01)), die Unterschiede sind jedoch nicht signifikant.

Veränderung des statistikbezogenen Selbstkonzepts

Die Studierenden aus der KG wiesen bereits zu Semesterbeginn einen relativ hohen, positiven Wert auf, wobei zwischen den beiden Messzeitpunkten lediglich eine minimale Veränderung zu verzeichnen ist ($MW_{t1} = 2.59$ (.68); $MW_{t2} = 2.63$ (.75)). In der EG-Gruppe zeigt sich hingegen eine tendenziell positivere Entwicklung ($MW_{t1} = 2.38$ (.59); $MW_{t2} = 2.63$ (.68)).

Dies wird insbesondere durch die Datenverläufe ersichtlich (vgl. Abb. 1), welche eine ausgeprägtere Verbesserung über die Zeit andeuten. Trotz dieser tendenziellen Steigerung blieb das Ergebnis der Varianzanalyse knapp über der Signifikanzgrenze ($p = .059$), was auf die geringe Stichprobengröße ($N \sim 30$) in den Subgruppen zurückgeführt werden könnte. Die Entwicklung der Statistikangst, hier dargestellt durch den Summenscore (vgl. Abb. 2), verdeutlicht, dass diese im Verlauf des Semesters hoch signifikant ($p < .001$) reduziert wurde ($MW_{t1} = 41$ (12.8); $MW_{t2} = 35.3$ (11.8)). Die berechnete Effektstärke nach Cohen ($d = 0.61$) deutet auf einen mittleren Effekt hin.

Abb. 1: Entwicklung des statistischen Fähigkeits- Abb. 2: Entwicklung der Statistikangst (EG)
selbstkonzepts



Veränderung der statistikbezogenen Kompetenzen

In beiden Gruppen kann eine signifikante Verbesserung der Selbsteinschätzung über das Semester hinweg verzeichnet werden (KG : $MW_{t1} = 2.25$ (.68); $MW_{t2} = 3.26$ (.49); EG : $MW_{t1} = 2.31$ (.74); $MW_{t2} = 3.29$ (.54). Die Ergebnisse der Varianzanalyse weisen auf einen hochsignifikanten Unterschied hin ($p < .001$), wobei die Effektstärke durch das partielle Eta-Quadrat ($\eta^2 = .73$) als sehr groß bezeichnet werden kann. An den Datenverläufen ist in beiden Gruppen eine mehrheitlich positive Entwicklung erkennbar (vgl. Abb. 3).

Als Gegenstück zu den formativen Studienleistungen, in denen die Studierenden über einen längeren Bearbeitungszeitraum die Möglichkeit haben, sich mit den Inhalten auseinanderzusetzen, findet zum Ende des Semesters in der Abschlusssitzung ein Wissenstest statt, der zentrale Inhalte des Semesters in acht Fragen erfasst. Der Test wird in ILIAS als E-Prüfung durchgeführt und umfasst fünf statistikbezogene sowie drei R-bezogene Fragen. Die Testergebnisse (Abb. 4) ermöglichen hierbei eine objektive und vor allem individuelle Leistungsbeurteilung, da keine Hilfsmittel vorgesehen sind. Die Ergebnisse des t-Tests weisen für die EG-Gruppe eine signifikant höhere Test-Lösungsquote auf als für die KG-Gruppe ($MW_{EG} = .76$ (.15); $MW_{KG} = .64$ (.18)). Die Effektstärke nach Cohen ($d = 0.73$) lässt auf einen Effekt mittlerer Größe schließen.

Abb. 3: Entwicklung der Selbsteinschätzung statistischer Kenntnisse

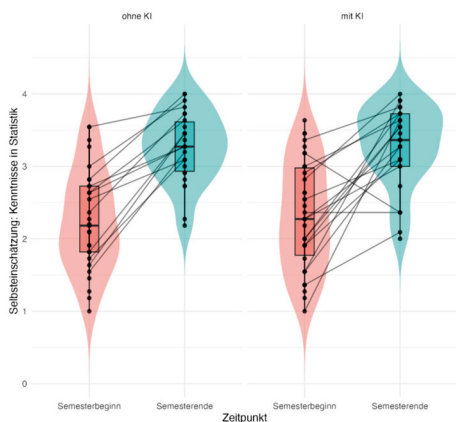
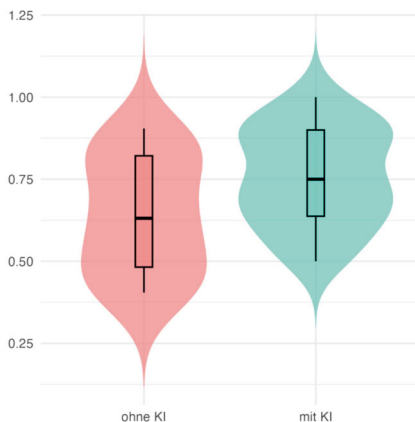


Abb. 4: Lösungsquote im Wissenstest zu Semesterende (t2)



Diskussion

In der Gesamtschau lässt sich eine positive Veränderung des statistikbezogenen Selbstkonzepts, der Statistikangst sowie des Statistikwissens feststellen. Diese Entwicklung ist insbesondere bei den Studierenden der EG zu beobachten, die in allen Aspekten eine tendenziell günstigere Entwicklung aufweisen. Der gezielte Einsatz generativer KI kann offenbar unterstützend für den Lehr- und Lernprozess wirken und positive Effekte auf den Lernerfolg zeitigen. Diese Entwicklung kann möglicherweise auf verschiedene Einflussfaktoren zurückgeführt werden: Der Chatbot ist 24/7 verfügbar, wodurch Studierende jederzeit auf Unterstützung zurückgreifen können. Die Fehlermeldungen des Statistikprogramms R wirkten sich häufig negativ auf die Studierenden aus: Die schnelle Bearbeitung und das sofortige Feedback, die durch den Einsatz von ChatGPT möglich sind, bieten einen deutlichen Zeitvorteil und tragen höchstwahrscheinlich zur Minderung dieser negativen Effekte bei. Ferner entfällt die soziale Bewertung im Lernprozess, was insbesondere im Kontext der 1:1-Interaktion einen geschützten Raum schafft, in dem Studierende Fragen stellen und experimentieren können. Schließlich führt die Möglichkeit, komplexere R-Grafiken mit Unterstützung der KI zu erstellen, zu positiven Selbstwirksamkeitserfahrungen. Die Studierenden gewinnen

dadurch Vertrauen in ihre eigenen Fähigkeiten, auch komplexe Aufgaben erfolgreich zu bewältigen. Es scheint zudem möglich, dass der Einsatz von KI einen Beitrag zur Reduktion von Besorgnis und Vermeidungstendenzen leisten könnte, die nach Faber et al. (2018) oft im Zusammenhang mit Statistikangst stehen. Studierende erleben weniger kognitive Überforderung, da sie durch die KI-gestützte Unterstützung gezielt und individuell mit den Inhalten interagieren können. Abschließend ist noch wichtig zu betonen, dass ein statistisches Grundlagenwissen für eine sinnvolle KI-Nutzung erforderlich ist. Andernfalls besteht die Gefahr, dass die ausschließliche Nutzung von KI zu einem *Deskilling* führt (siehe hierzu Reinmann 2023). Nur durch eine kritische und reflektierte Auseinandersetzung mit den Inhalten können langfristig Kompetenzen entwickelt werden, die sowohl das Wissen als auch das Fähigkeitsselbstkonzept stärken.

Literatur

- Bandalos, Deborah L./Finney, Sara J./Geske, Jenenne A. (2003): »A model of statistics performance based on achievement goal theory«, in: Journal of Educational Psychology, 95(3), S. 604–616.
- Eccles, Jacquelynne S./Wigfield, Allan (2020): »From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and socio-cultural perspective on motivation«, in: Contemporary Educational Psychology, 61.
- Faber, Günter/Drexler, Heike/Stappert, Alexander (2018): »BEVAST-EWL. Skala zu studentischen Besorgtheits-, Vermeidungs- und Aufregetheitskognitionen bezüglich statistischer Anforderungen in erziehungswissenschaftlichen Lehr-Lern-Kontexten«, in: Leibniz-Zentrum für Psychologische Information und Dokumentation ZPID, Hg. (Elektronisches Testarchiv).
- Fenning, Breanne E./May, Laura Negel (2013): »Where there is a will, there is an A: examining the roles of self-efficacy and self-concept in college students' current educational attainment and career planning«, in: Social Psychology of Education, 16(4), S. 635–650.
- Filiz, Mehmet/Early, Erin/Thurston, Allen/Miller, Sarah (2020): »Measuring and improving university students' statistics self-concept: A systematic review«, in: International Journal of Educational Research Open, 1, Article 100020. <https://doi.org/10.1016/j.ijedro.2020.100020>.

- Marsh, Herb/Shavelson, Richard (1985): »Self-concept: Its multifaceted, hierarchical structure«, in: *Educational Psychologist*, 20(3), S. 107–123.
- O'Mara, Alison/Marsh, Herb/Craven, Rhonda G./Debus, Raymond L. (2006): »Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis«, in: *Educational Psychologist*, 41(3), S. 181–206.
- Ogunleye, Bayode/Zakariyyah, Kudirat Ibilola/Ajao, Oluwaseuna/Olayinka, Olakunle/Sharma, Hemlata (2024): »A systematic review of generative AI for teaching and learning practice«, in: *Education Sciences*, 14(6), S. 636. <https://doi.org/10.3390/educsci14060636> vom 28.09.2024.
- Onwuegbuzie, Anthony J./Wilson, Vicki A. (2003): »Statistics anxiety: Nature, etiology, antecedents, effects and treatments – a comprehensive review of the literature«, in: *Teaching in Higher Education*, 8, S. 195–209.
- Paulitsch, Kerstin (2013): »Ich kann Statistik!« Zum Zusammenhang von Selbstwirksamkeit, Gedächtnisabruf und elektrokortikaler Aktivierung im Bereich der psychologischen Statistik. Diplomarbeit, Karl-Franzens-Universität Graz, Graz.
- R Core Team (2024): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reinmann, Gaby (2023): Deskillung durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik (= Diskussionspapier Nr. 25), Hochschulforum Digitalisierung. https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_25_Deskillung.pdf vom 28.09.2024.
- von der Embse, Nathaniel/Jester, Dane/Roy, Devlina/Post, James (2018): »Test anxiety effects, predictors, and correlates: A 30-year metaanalytic review«, in: *Journal of Affective Disorders*, 227, S. 483–493. <https://doi.org/10.1016/j.jad.2017.11.048> vom 28.09.2024.