

# LERNSPIEL VS. KI-AVATAR: DIGITALE TRAININGSFORMATE ZUR FÖRDERUNG DIAGNOSTISCHER KOMPETENZ IN DER INKLUSIVEN LEHRKRÄFTEBILDUNG

## THEORIE & FRAGESTELLUNG

Die Förderung diagnostischer Kompetenz ist eine zentrale Aufgabe der Lehrkräftebildung im inklusiven Schulsystem, da Lehrkräfte Lernstände erfassen und gezielte Unterstützungsmaßnahmen ableiten müssen (Gebhardt & Jungjohann, 2023; Kultusministerkonferenz & Hochschulkonferenz, 2015). Simulationsbasierte Lernumgebungen bieten hierfür ein hohes Potenzial, weil sie die wiederholte und reflektierte Anwendung diagnostischer Prozesse in realitätsnahen Situationen ermöglichen (Chernikova et al., 2020; Zellner et al., 2024). KI-basierte Avatare erweitern diese Möglichkeiten, indem sie dialogische und adaptive Interaktionen schaffen, die insbesondere kommunikative und reflexive Aspekte diagnostischer Kompetenz fördern können (Robinson, 2023). Gleichzeitig beeinflussen technologische Kompetenz sowie die wahrgenommene kognitive Belastung das Lernerleben und den Wissenserwerb entscheidend (Fink et al., 2024; Krieglstein et al., 2023). **Vor diesem Hintergrund interessiert uns, wie sich ein KI-basierter Avatar im Vergleich zu einem klickbasierten Lernspiel auf die Entwicklung diagnostischer Kompetenz auswirkt, in welchem Zusammenhang technologische Kompetenz und kognitive Belastung mit den Lernprozessen stehen und welche Gestaltungsmerkmale simulationsbasierter Lernumgebungen insbesondere für unterschiedliche Expertisegrade geeignet sind.**

## METHODE

- 1. Fragebogen:** Selbsteinschätzung diagnostischer Kompetenz mit dem DaKI-Instrument (Jungjohann & Gebhardt, 2023) & Erfassung technologischer Kompetenz mit TPACK (Schmidt et al.; 2009).
- 2. 1. Simulation (randomisiert):** Die Teilnehmenden werden per Zufallszuweisung einem von zwei Formaten zugeteilt und bearbeiten entweder (A) das strukturierte klickbasierte Diagnosespiel (Zellner & Gebhardt, 2025) oder (B) den KI-generierten Lehrkraft-Avatar (GPTAvatar; Robinson, 2023).
- 3. 2. Simulation (konterbalanciert):** Bearbeitung des jeweils alternativen Simulationsformats (A) bzw. (B) mit anschließenden Reflexionsfragen und Förderplanung.
- 4. Fragebogen nach Simulation 1 und 2:** Subjektive kognitive Belastung (Intrinsic, Extraneous, Germane Load) nach Krieglstein et al.; 2023. Nach der Avatar-Bedingung zusätzlich: Wahrnehmung der Agenten-Persona (API; Baylor & Ryu, 2003).

## KONTAKT

Jakob Koch\*  
Jakob.Koch@  
edu.lmu.de



**Project cooperation partners:**  
Judith Zellner\*  
Priska Hagmann-von Arx\*\*  
Liliana Tönnissen\*\*  
Pierre-Carl Link\*\*  
Markus Gebhardt\*  
\*Ludwig-Maximilians-Universität München  
\*\*Interkantonale Hochschule für Heilpädagogik Zürich

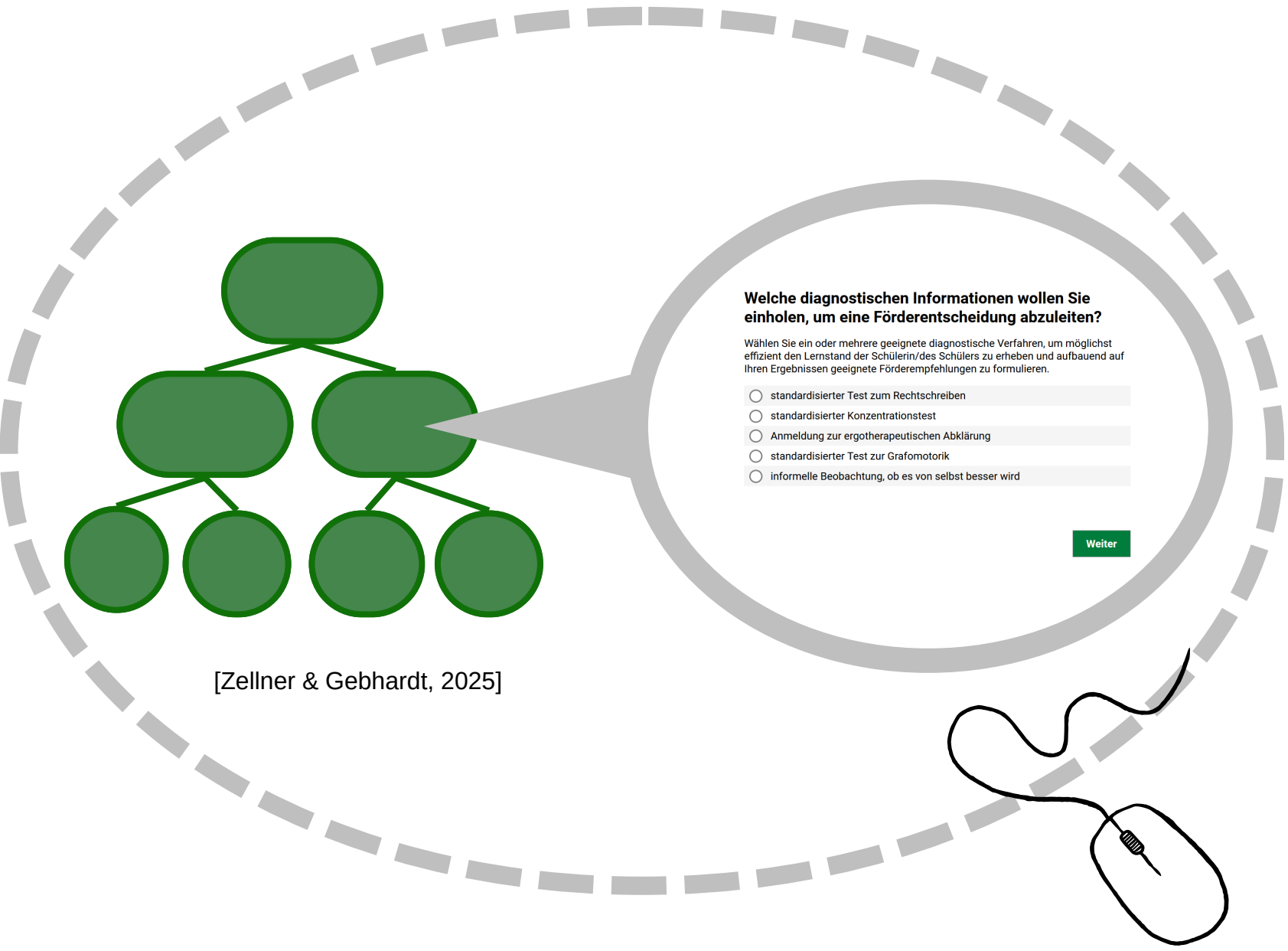
## SIMULATION

Experiment: A-B-Crossover-Design mit wiederholter Messung

### A: Strukturiertes Klickspiel im Entscheidungsbaumformat

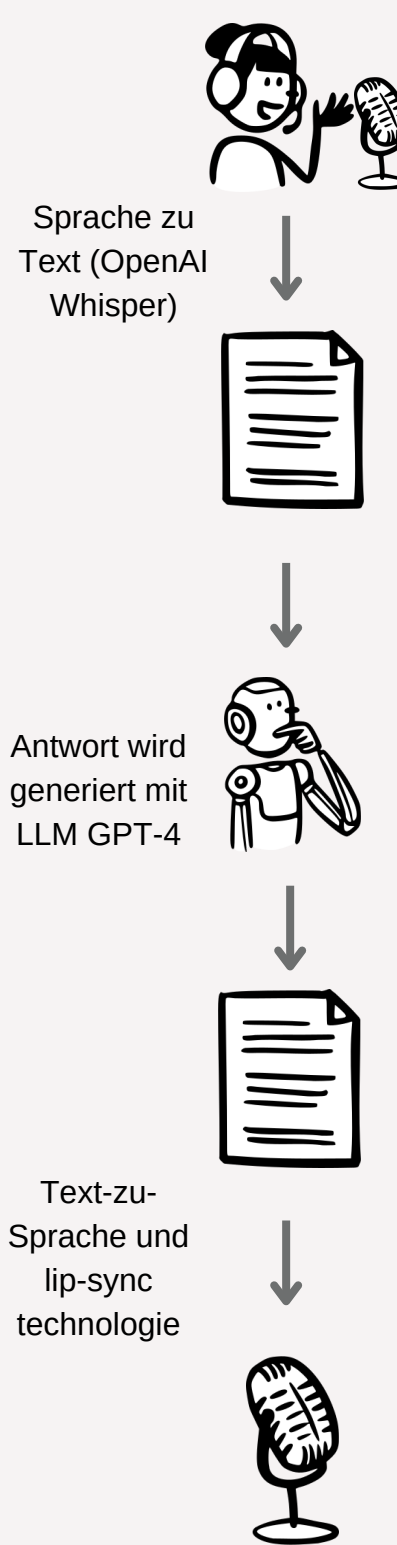
#### Fallvignette Janina

**Zusammenfassung:** Janina schreibt Wörter noch nicht phonetisch und hält sich nicht an die Rechtschreibregeln. Sie verwendet die Verlängerung und Verschärfung uneinheitlich. Ein standardisierter Rechtschreibungstest zeigt deutlich unterdurchschnittliche Ergebnisse, wobei häufig Fehler bei der korrekten Verwendung der Verlängerung „h“ auftreten.

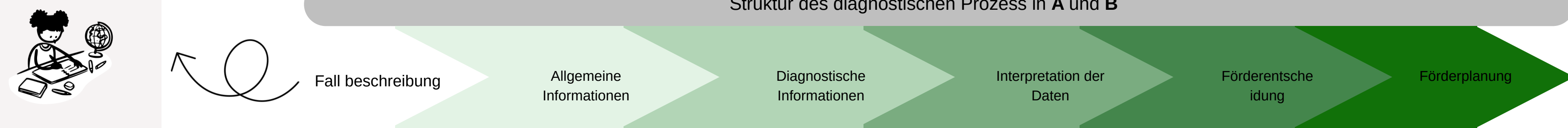


[Zellner & Gebhardt, 2025]

### B: Diagnostisches Gespräch mit KI-Lehrer-Avatar



[Fink et al., 2024]



## ERGEBNISSE

### Stichprobe

**Teilnehmende:**  
N = 171 Studierende  
N = 58 Dozierende

**Randomized Case Order:**

- Studierende: 42,86 % A, 57,14 % B
- Dozierende: 50 % A, 50 % B

– Die Abbrecherquote der Studierenden lag bei 72,25 %, wenn der KI-generierte Avatar zuerst durchgeführt wurde.

### Selbsteinschätzung der Diagnostischen Kompetenz

Dozierende wiesen signifikant höhere Werte in der selbst eingeschätzten diagnostischen Kompetenz auf ( $M = 3,85, SD = 0,59$ ) als Studierende ( $M = 3,13, SD = 0,67$ ); ( $t(17,26) = 3,74, p = .002, d = 1,07$ ). Große Effektstärken zeigten sich in den Subdimensionen Status-/Leistungsdiagnostik ( $d = 1,04$ ), sonderpädagogischer Förderdiagnostik ( $d = 1,01$ ) und Lernverlaufsdiagnostik ( $d = 1,16$ ). Nur die fachdidaktische Diagnostikkompetenz wies einen kleineren Unterschied auf ( $d = 0,38$ ). Über die Gesamtschichtprobe hinweg lagen die Werte im mittleren bis guten Bereich ( $M = 3,25, SD = 0,71$ ).

### Selbsteinschätzung der technischen Kompetenz

Die technische Kompetenz lag insgesamt auf einem mittleren Niveau ( $M = 3,39, SD = 0,86$ ). Studierende ( $M = 3,35, SD = 0,88$ ) und Dozierende ( $M = 3,61, SD = 0,75$ ) unterschieden sich nicht signifikant, ( $t(71) = 1,02, p = .312$ ). Inhaltlich deutet die Kombination aus mittlerem TPACK-Niveau und der ANOVA-Interaktion darauf hin, dass technologische Kompetenz vor allem bei Studierenden eine puffernde Funktion gegenüber extrinsischer Belastung erfüllt.

### Wahrnehmung des Avatars gemessen durch API

Die mittlere Wahrnehmung des Avatars war insgesamt positiv ( $M = 3,33, SD = 0,87$ ). Zwischen Studierenden ( $M = 3,34, SD = 0,91$ ) und Dozierenden ( $M = 3,26, SD = 0,71$ ) zeigten sich keine signifikanten Unterschiede, ( $t(19) = 0,37, p = .718$ ). Allerdings zeigten sich bei Studierenden deutliche Zusammenhänge zwischen der Bewertung der Avatarpersona und der wahrgenommenen kognitiven Belastung. Je positiver der Avatar eingeschätzt wurde, desto geringer waren intrinsische und extrinsische Belastung und desto höher war die lernförderliche Germane Load.

### Diagnostische Qualität nach Simulation

Gruppe	Avatar (AV)	Lernspiel (LS)	Gesamt
Studierende	72,1–83,6 %	95,1–100 %	67%
Dozierende	75–91,7 %	91,7–100 %	83%
Gesamt	83,6–94,5 %	95,9–98,6 %	70%

Die diagnostische Qualität wurde als Anteil richtiger Förderentscheidungen operationalisiert. Insgesamt trafen 69,9 % der Teilnehmenden die korrekte Entscheidung. Dozierende erzielten höhere Raten (83,3 %) als Studierende (67,2 %). Differenziert nach Bedingung zeigte sich ein deutlicher Vorteil für das Lernspiel ( $M = 91,8 \%$ ) im Vergleich zur Avatarbedingung ( $M = 83,6–94,5 \%$ ). Der Unterschied war bei Studierenden stärker ausgeprägt (+14 Prozentpunkte) als bei Dozierenden.

## DISKUSSION

Das klickbasierte Lernspiel unterstützt eine besonders strukturierte und präzise diagnostische Entscheidungsfindung, während der KI-Avatar insbesondere bei positiv erlebter Interaktion ebenfalls lernförderlich wirkt und reflexive Prozesse anregt. Technologische Vorerfahrungen und subjektive Wahrnehmungen beeinflussen kognitive Belastung und Diagnostikqualität wesentlich. Simulationsbasierte Lernumgebungen wirken differenziell und sollten an den Expertisegrad der Lernenden angepasst werden. Für die Lehrkräftebildung bedeutet dies, dass Klickspiele den Einstieg in diagnostisches Denken erleichtern, während KI-Avatare fortgeschrittene Reflexions- und Kommunikationskompetenzen fördern. Insgesamt bieten KI-Simulationen eine skalierbare und evidenzbasierte Möglichkeit, diagnostische Kompetenz im inklusiven Bildungssystem zu professionalisieren (Koch et al., 2025; Zellner et al., 2025). **Implikation für die Lehrkräftebildung:** Strukturierte Simulationen eignen sich besonders für frühe Ausbildungsphasen, während KI-basierte Avatarformate fortgeschrittene diagnostische Kompetenzen fördern können.

## LITERATUR

Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: Benefits and challenges. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1446397> / Heilmann, N., Seidel, T., Opat, A., Heilmann, A., Wecker, C., Fischer, M., Ufer, S., Schmidmaier, R., Neuhaus, B., Siebeck, M., Stürmer, K., Obersteiner, A., Reiss, K., Ginziv, R., & Fischer, F. (2019). Facilitating Diagnostic Competencies in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontiers in Learning Research*, 7(4), 1–24. / Jungjohann, J., & Gebhardt, M. (2023). Fragebogen zur Erfassung der diagnostischen Kompetenz von Lehrkräften in der inklusiven Schule (DaKI, Version 0.2). Regensburg: Universitätsbibliothek. <https://doi.org/10.52020/sub-34248/> / Koch, J., Hagmann-von Arx, P., Tönnissen, L., Zellner, J., Link, P.-C., & Gebhardt, M. (2025). Schulung diagnostischer Kompetenzen in der Heil- und Sonderpädagogik: Eine Pilotstudie zum Einsatz eines GPT-gestützten Avatars. SGBF-SGL-Jahreskongress 2025, Pädagogische Hochschule Luzern. / Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stoekhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review*, 35. <https://doi.org/10.1007/s12664-023-09782-2> / Kultusministerkonferenz & Hochschulkonferenz. (2015). Leitbild für eine Schule der Vielfalt. Gemeinsame Erklärung von Hochschulkonferenz und Kultusministerkonferenz. Beschluss der Kultusministerkonferenz vom 12.03.2015/Beschluss der Hochschulkonferenz vom 18.03.2015. Verfügbar unter: [https://www.kmk.org/Dateien/Sonderpädagogische%20Schulung\\_beschluss20152015\\_03\\_12\\_Schule-der-Vielfalt.pdf](https://www.kmk.org/Dateien/Sonderpädagogische%20Schulung_beschluss20152015_03_12_Schule-der-Vielfalt.pdf) / Robinson, S. (2023, April 6). GPTAvatar (GitHub-Repository). GitHub. <https://github.com/SarahRobinson/GPTAvatar> / Schmidt, D. A., Baran, E., Thompson, A. D., Mehra, P., Kretzer, M. J., & Shin, Y. S. (2009). Technological Pedagogical Content Knowledge (TPACK): The Development and Validation of an Assessment Instrument for Preservice Teachers. *Journal of Research on Technology in Education*, 42(3), 123–149. <https://doi.org/10.1080/15393020.2009.10526262> / Zellner, J., Ebenbeck, N., & Gebhardt, M. (2024). Entwicklung digitaler Simulationsumgebungen zur Förderung der diagnostischen Entscheidungskompetenzen in der sonderpädagogischen Lehrkräfteausbildung. *Qualifizierung für Inklusion*, 6(2). <https://doi.org/10.25558/623524> / Zellner, J., Hagmann-von Arx, P., Koch, J., Tönnissen, L., Link, P.-C., & Gebhardt, M. (2025). Simulationen zur Förderung der diagnostischen Kompetenzen in der heil- und sonderpädagogischen Hochschullehrkräftebildung: Strukturiertes Klickspiel oder Einsatz GPTAvatar – ein Vergleich mit Noviz:innen und Expert:innen. SGBF-SGL-Jahreskongress 2025, Pädagogische Hochschule Luzern

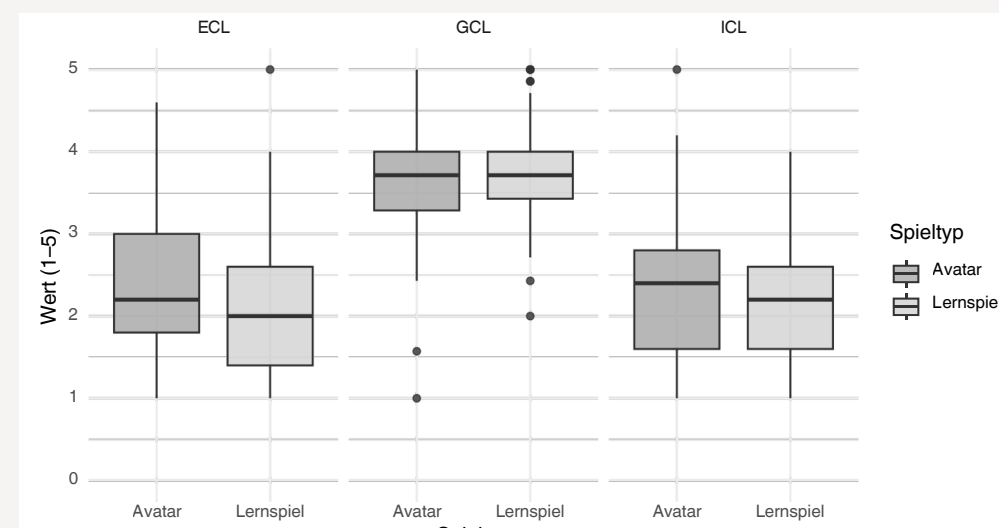


### Mixed-ANOVA für Cognitive Load Subskalen mit TPACK als Kovariate

Skala	Effekt	df1	df2	F	p
ECL	Gruppe × TPACK	1	69	4.04	.048*
ECL	Gruppe	1	69	3.98	.050
ICL	Gruppe	1	69	0.31	.582
GCL	Gruppe	1	69	0.00	.951

Mixed-ANOVA mit TPACK als Kovariate zeigte für die Extraneous Load (ECL) einen signifikanten Interaktionseffekt von Gruppe × TPACK,  $F(1,69) = 4,04, p = .048, \eta^2 = .034$ . Höhere technologische Kompetenz ging insbesondere bei Studierenden mit geringerer extrinsischer Belastung einher; bei Dozierenden zeigte sich dieser Zusammenhang nicht. Zudem berichteten Dozierende tendenziell geringere ECL-Werte ( $p = .050$ ). Für Intrinsic (ICL) und Germane Load (GCL) ergaben sich keine signifikanten Effekte. Insgesamt lagen ICL und ECL im niedrigen bis mittleren, GCL im hohen Bereich, was auf gut bewältigbare und kognitiv aktivierende Lernaufgaben hinweist.

### Cognitive Load nach Simulation und CL Bereich



Die Abbildung zeigt die kognitive Belastung differenziert nach Trainingsformat und Load-Dimension. Intrinsic Load (ICL) und Extraneous Load (ECL) liegen in beiden Formaten im niedrigen bis mittleren Bereich, während die Germane Load (GCL) durchgehend hohe Werte erreicht. Hohe GCL-Werte sind lernförderlich, da sie auf eine intensive, lernrelevante kognitive Verarbeitung hinweisen. Insgesamt waren beide Formate gut bewältigbar und zugleich kognitiv aktivierend gestaltet.

### Testen Sie den Avatar:

Alle Forschungsprojekte welche mit dem Avatar (GPTAvatar; Robinson, 2023) durchgeführt werden sind zentral auf <https://avatar-research.com/> zu finden. Hier können Sie zusätzlich einen eigenen Avatar erstellen und weitere Avatare testen.

