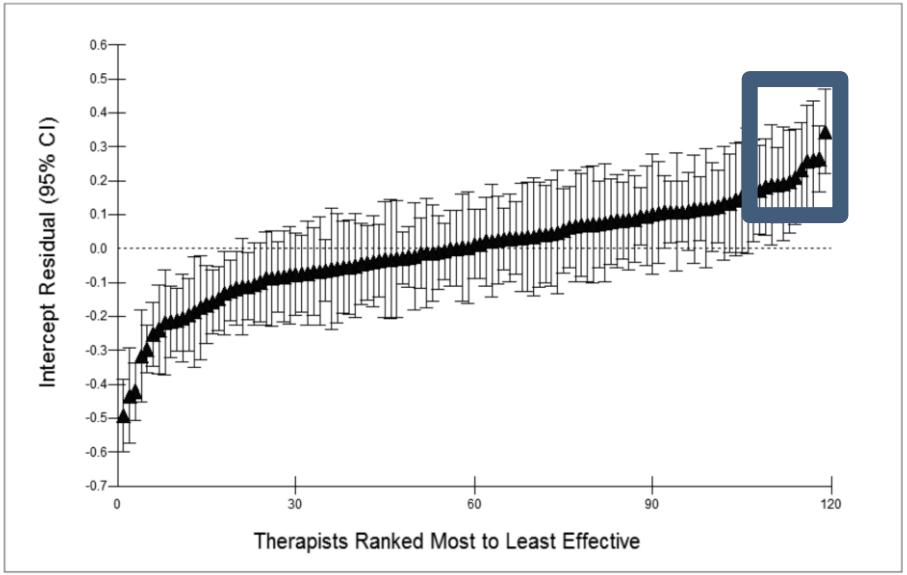
Otto-Friedrich-Universität Bamberg

Wirksamkeit von Feedback in menschlicher vs. KI-basierter Deliberate Practice – Eine randomisiert-kontrollierte Machbarkeitsstudie

Stefan Blümel, Nour Krischker, Frank Lörsch & Sabine Steins-Löber
Lehrstuhl für Klinische Psychologie und Psychotherapie
Otto-Friedrich-Universität Bamberg

10.20378/irb-116801

Wonach suchen wir?



(Saxon & Barkham, 2012)

„Supershrinks“

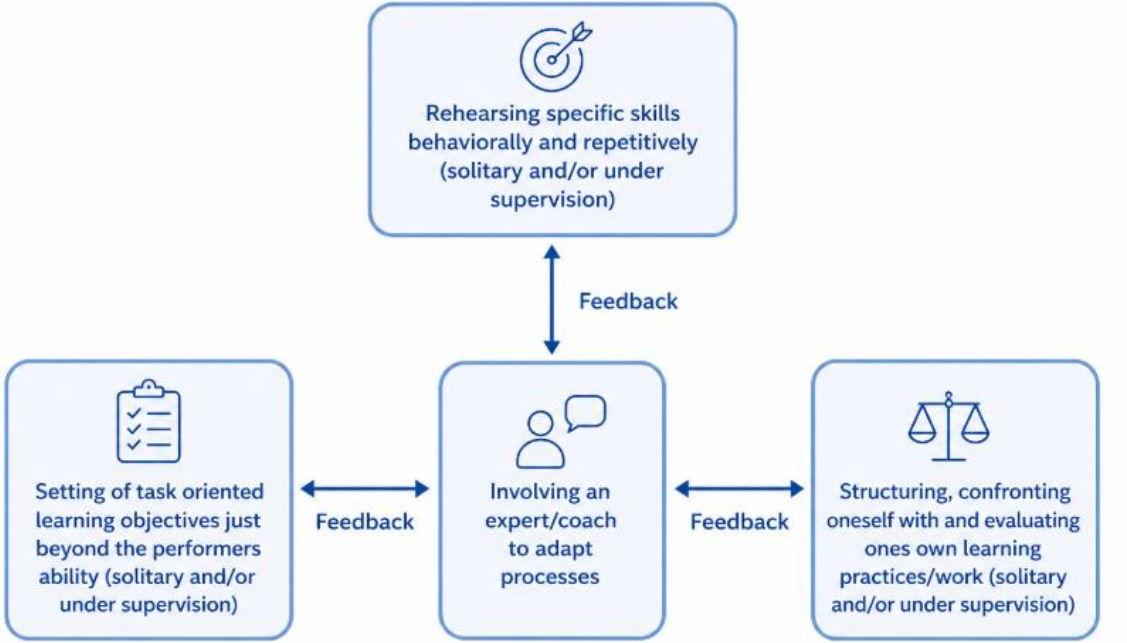
Was macht eine:n „supershrink“ aus?

Was macht eine:n „supershrink“ aus?

(Vielleicht…)

Wie entwickelt man Expertise: Deliberate Practice

(Chow et al., 2015; Miller et al., 2020; Rousmaniere, 2024; Diamond et al., 2025)



Herausforderungen beim Einsatz und Durchführung von DP

- Zeitaufwand

- Konsequente und adhärente Umsetzung aller DP-Komponenten

- Herausfordernde und sich nicht angenehm anfühlende Lernsituationen

- Schaffung eines sicheren Lernumfelds

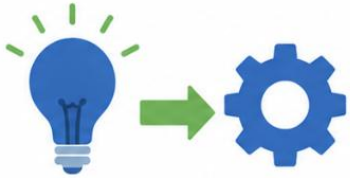
- ...

Wie sollte gutes Feedback aussehen?

(Caspar, 2025a; Caspar, 2025b; Mahon, 2023; Miller et al., 2020)

- Spezifisch

- Konstruktiv



- Verbesserungspotenziale aufzeigen

- Fürsorglich

- Präzise

- Zeitnah

- Auf zukünftige Aufgaben bezogen

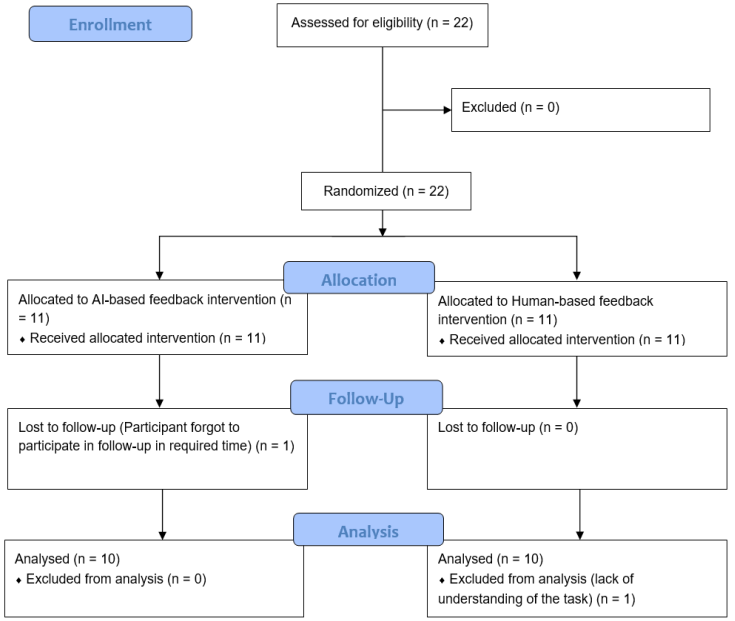
- An Lernkontext angepasst

Forschungsfragen:

1. Inwieweit sind LLM-basiertes Feedback zur Entwicklung interpersoneller Fertigkeiten mit menschlichem Feedback vergleichbar...
2. Und inwiefern steht die Akzeptanz von KI im Zusammenhang mit dieser?

Analysierte Stichprobe

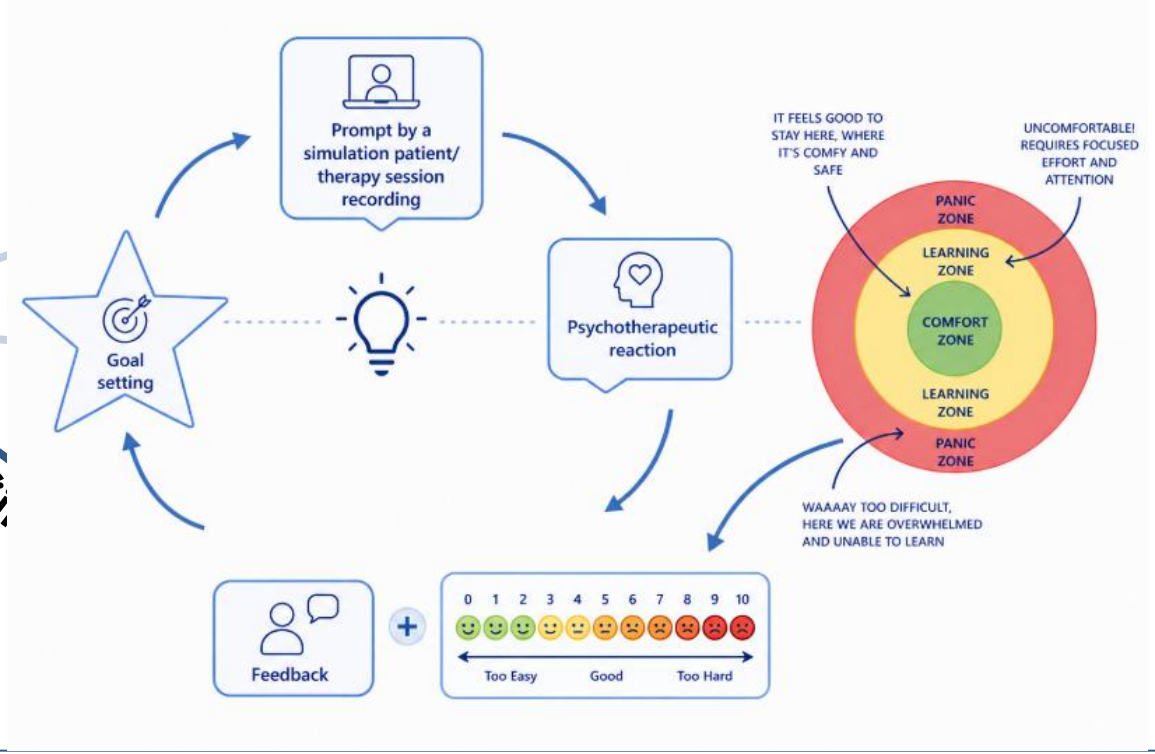
- $N = 20$ ($n_{männlich} = 8$, $n_{weiblich} = 12$)
- Alter: $M = 22.95$, $SD = 2.65$
- Fortgeschrittene Bachelorstudierende = 19
- Masterstudierende = 1
- **Teilnahmevoraussetzung: Vorerfahrung mit DP und Fertigkeitserwerb**



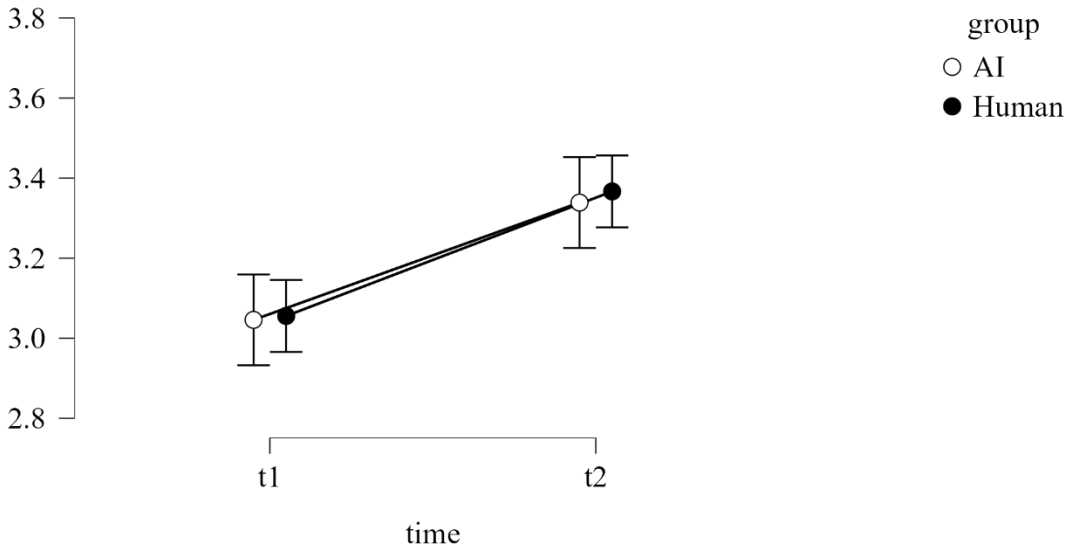
Ablauf des Trainings

Ablauf des Trainings: Schematische Darstellung eines DP-Zyklus zum Micro-Skilltraining

Austausch durch KI (ChatGPT 4.o)

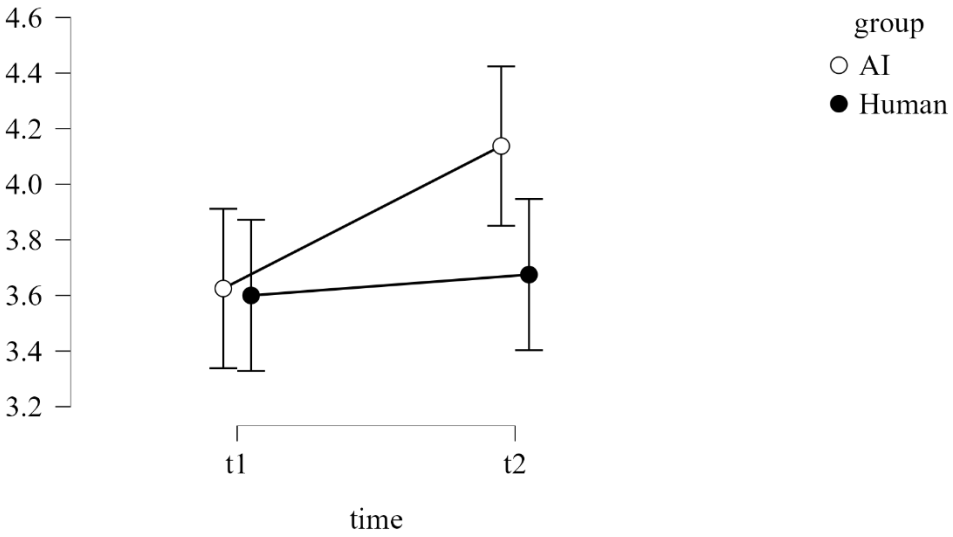


Ergebnisse: fremdeingeschätzte FIS



Within-Effekt: $F(1,18) = 44.64$, $p < .01$, $\eta^2 = .16$, $\eta_p^2 = 0.71$
Interaktions-Effekt: $F(1,18) = 0.04$, $p = 0.84$

Ergebnisse: selbsteingeschätzte FIS



Within-Effekt: $F(1,18) = 5.66$, $p = .03$, $\eta^2 = .09$, $\eta_p^2 = 0.24$
Interaktions-Effekt: $F(1,18) = 3.14$, $p = 0.09$

Weitere Ergebnisse

- **Kein signifikanter** Zusammenhang zwischen KI-Akzeptanz und Fertigkeitsentwicklung in der KI-Gruppe
- **Keine signifikanten** Unterschiede zwischen den Gruppen hinsichtlich
  - **Subjektive Wirksamkeit** des Trainings
  - **Zufriedenheit** mit dem Training
  - Empfundener **Stress** durch das Training

Studienmaterial/Prompts

- Stellen wir gerne zur Verfügung!
  → stefan.bluemel@uni-bamberg.de

Ethik

- Vertretung einer **deontologischen Perspektive** → es sollte Grenzen **(„rote Linien“)** des Einsatzes von KI geben, die nicht durch Mehrnutzen aufgewogen werden können
- Ansonsten mglw. Anwendung von Prinzipienethik (Beauchamp & Childress, 1979)
  - Autonomie
  - Non-Malefizienz
  - Benefizienz
  - Gerechtigkeit

Fazit

1. Ähnliche fremdeingeschätzte Fertigkeitsentwicklung unabhängig von Feedbackquelle
2. Möglicherweise Überschätzung subjektiver Fertigkeitsentwicklung aufgrund von KI-Sycophancy
3. Generierung von Trainingsmaterial durch KI stark skalierbar

Nächste Schritte zu Supershrinks

- Replikation!
- Vergleich mehrerer LLMs und Expert:innen
- Erhöhung der Feedback-Freiheitsgrade während der DP-Zyklen
- Fertigkeitstraining mit KI-Avataren
- ...

Danke für Ihre und Eure Zeit und Aufmerksamkeit!

Ich freue mich über jegliche Fragen und Kommentare!

Kontakt:
stefan.bluemel@uni-bamberg.de