

こんにちは

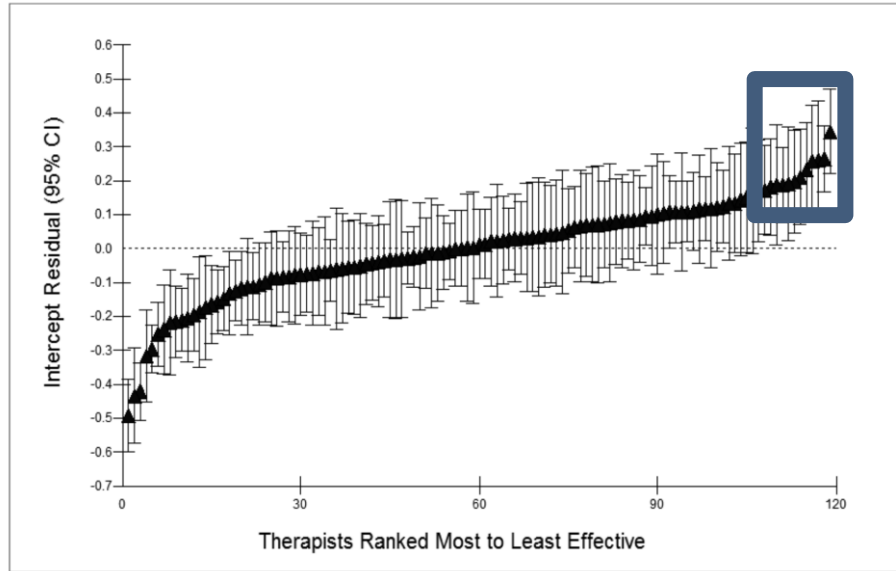
Feedback Effectiveness in Human vs. AI Based Deliberate Practice – A Randomized-Controlled Feasibility Study

Stefan Blümel, Nour Krischker, Frank Lörsch & Sabine Steins-Löber
Department of Clinical Psychology and Psychotherapy
University of Bamberg, Germany

10.20378/irb-116802



What are we looking for?



(Saxon & Barkham, 2012)

„Supershrinks“



What characterizes a „supershrink“?



What characterizes a „supershrink“?

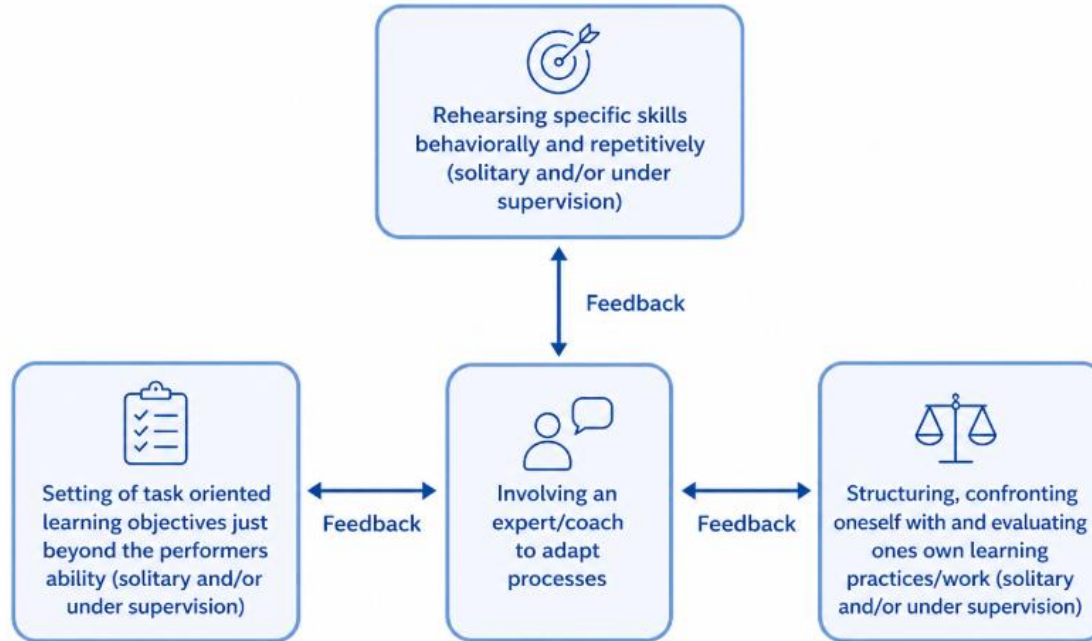


Deliberate Practice

(Maybe...)

Wie entwickelt man Expertise: Deliberate Practice

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



Implementation Challenges for DP in Psychotherapy

- Time investment



- Consistent and comprehensive implementation of all DP components



- Challenging and unpleasant learning situations



- Creating a safe learning environment



• ...



Good feedback!



What is good feedback?

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

- Specific



- Constructive



- Show improvement potential



- Supportive



- Precise



- Timely



- Focused on future tasks



- Adapted to the learning context



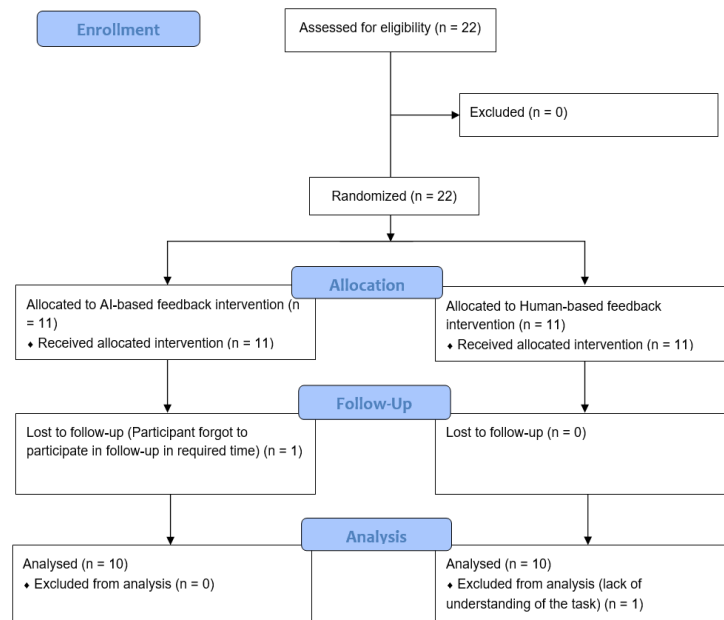
Research Questions:



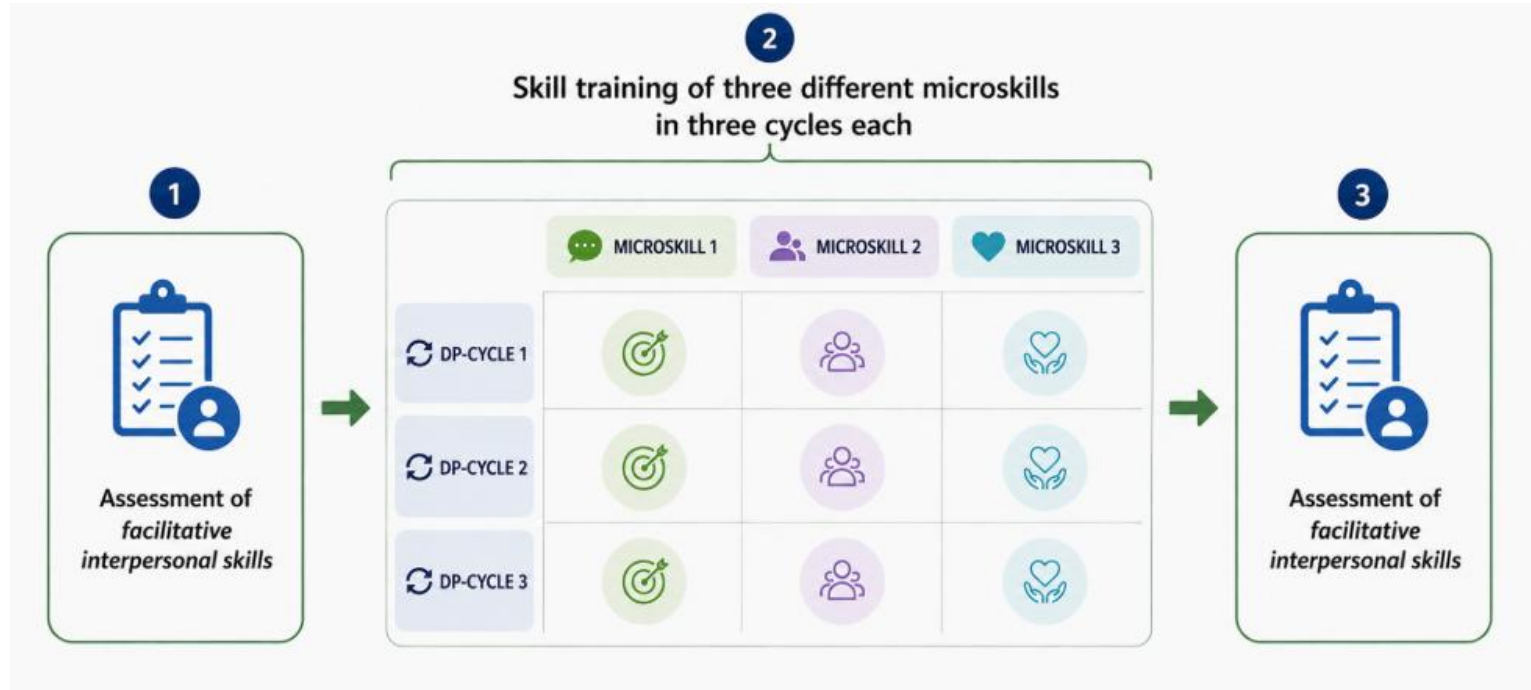
1. To what extent is LLM-based feedback for the development of interpersonal skills comparable to human feedback...
2. And to what extent is the acceptance of AI related to this?

Sample

- $N = 20$ ($n_{\text{male}} = 8$, $n_{\text{female}} = 12$)
- Alter: $M = 22.95$, $SD = 2.65$
- Undergrad students = 19
- Grad student = 1
- **Prerequisite for participation:**
Prior experience with DP and skill acquisition exercises



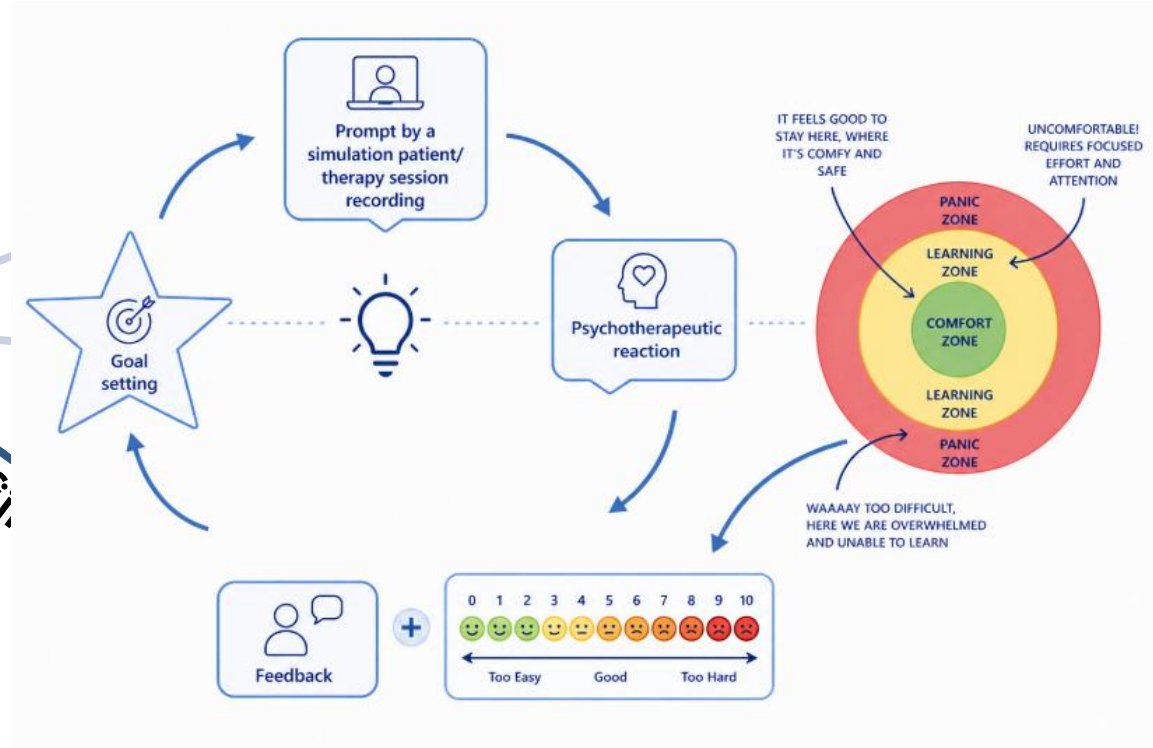
Training Procedure



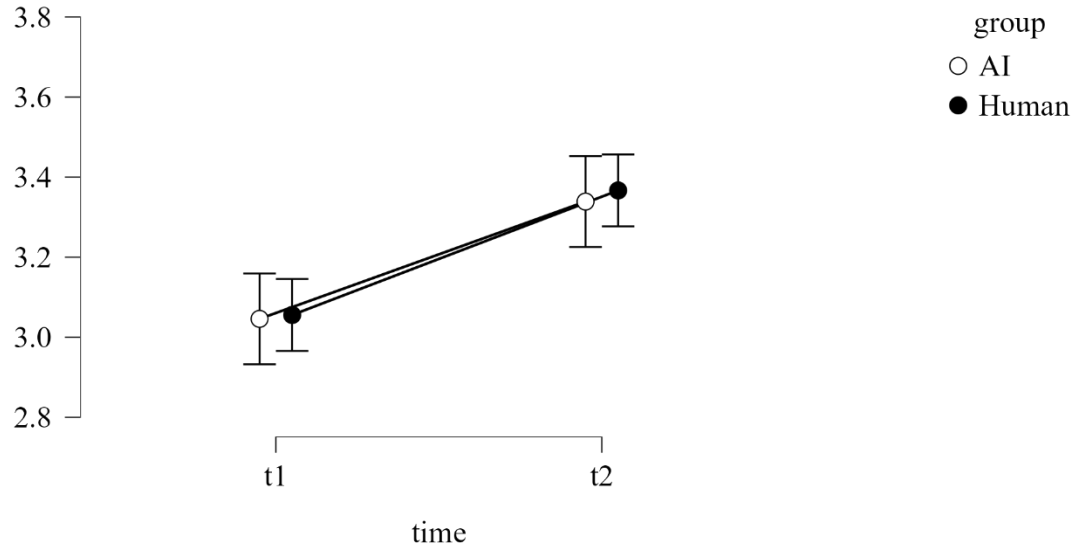
Training procedure: DP-Cycle

Austausch durch KI
(ChatGPT 4.o)

Nur für



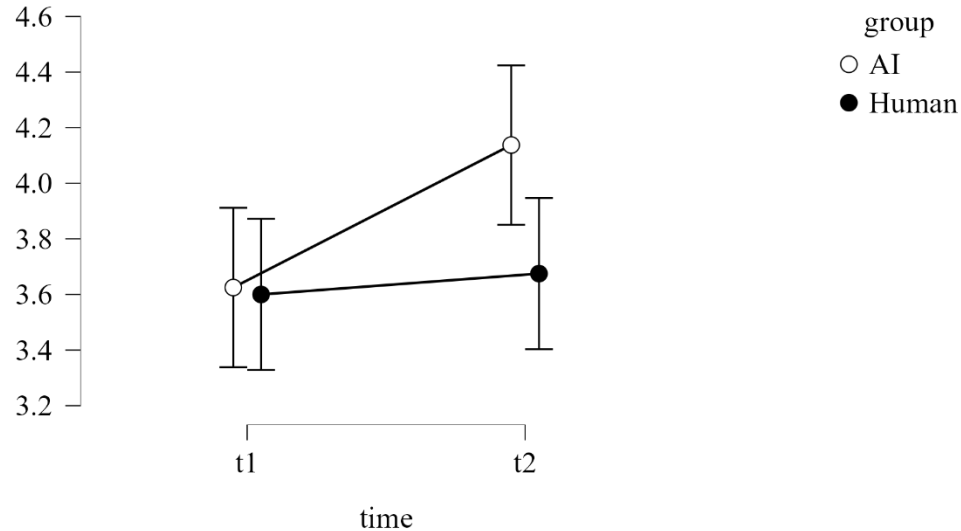
Results: observer-rated FIS



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

Interaction-Effect: $F(1,18) = 0.04, p = 0.84$

Results: self-rated FIS



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

Interaction-Effect: $F(1, 18) = 3.14$, $p = 0.09$

Additional Results



- **No significant correlation** between AI acceptance and skill development in the AI group
- **No significant differences** between the groups in terms of
 - **Perceived effectiveness** of the training
 - **Subjective satisfaction** with the training
 - **Perceived stress** from the training

Materials/Prompts



- Happy to share them!
→ stefan.bluemel@uni-bamberg.de

Ethics



- **Deontological perspective** → there should be borders („red lines“) for the use of AI in psychotherapy, no matter the potential benefits
- Application of Principles of Biomedical Ethics (Beauchamp & Childress, 1979)?
 - Autonomy
 - Non-Maleficence
 - Beneficence
 - Justice

Conclusion



1. Similar observer-rated skill development independent of the feedback source
2. Maybe over-estimation of self-rated skill development due to AI sycophancy
3. Generation of training materials using AI highly scalable

Next Steps on our way to „supershrinks“

- Replication!
- Comparison of different LLMs and experts
- Increasing the degrees of freedom for feedback during DP cycles
- Skill training with AI avatars
- ...



Thank you for your time and attention!
さようなら

I am happy to discuss your questions and comments!

email:

stefan.bluemel@uni-bamberg.de