

ORIGINAL RESEARCH

Practice of therapeutic skills with memory *versus* video-based self-reflection: a mixed-method randomized controlled trial

Klara Eisert¹, Franziska Kühne¹, Florian Weck¹, Anna Schimmrigk¹ and Ulrike Maaß¹

Department for Clinical Psychology and Psychotherapy, University of Potsdam, Germany

Corresponding author: Klara Eisert; Email: klara.eisert@uni-potsdam.de

(Received 13 November 2024; revised 5 August 2025; accepted 5 August 2025)

Abstract

Self-reflection is central to the development of psychotherapeutic competence. Given the positive reports of video analysis use in psychotherapy training, we suggest that self-reflection based on video analysis may be particularly effective. The aim of this study was to test whether video-based structured self-reflection (VSR) is superior to memory-based structured self-reflection (MSR) in terms of its effect on students' psychotherapeutic competence and the therapeutic alliance. As part of a university seminar within a Master's program, $N = 34$ psychology students ($M = 25$ years; $n = 32$ identifying as female) were randomly assigned to 4 weeks of practice with either VSR ($n = 16$) or MSR ($n = 18$). Independent raters assessed students' psychotherapeutic competence and the alliance before and after the practice phase (pre- and post-assessment). Students additionally rated their own competence during the practice phase. The written self-reflections were analysed using qualitative content analysis and frequency analysis. A repeated measures MANOVA revealed no significant differences between the study groups in the development of externally rated psychotherapeutic competencies from pre- to post-assessment. An analysis of students' self-reported competencies during the practice phase revealed a significant time effect ($\eta^2_G = 0.12$). Analysis of the written self-reflections showed that students focused mainly on the positive aspects of their behavior. The assumption that VSR is particularly beneficial was not confirmed. It seems that self-reflection requires additional guidance, feedback, and a sufficient time frame to effectively improve trainee skills.

Key learning aims

- (1) To find out whether video-based structured self-reflection is an effective means of developing students' psychotherapeutic skills and the therapeutic alliance.
- (2) To identify the main characteristics of students' self-reflections and the benefits and challenges they perceive during the self-reflection process.
- (3) To draw methodologically valid conclusions for the implementation of self-reflection in the university context.

Keywords: competencies; role-play; self-reflection; training

Introduction

As research in the field of therapist education moves towards evidence-based training for psychotherapists, a growing body of literature focuses on which methods are most effective in

© The Author(s), 2025. Published by Cambridge University Press on behalf of British Association for Behavioural and Cognitive Psychotherapies. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

improving therapist skills (Bennett-Levy *et al.*, 2009a; Hill *et al.*, 2017; Henrich *et al.*, 2023; Rakovshik and McManus, 2010). In this context, experience-based techniques, such as role-playing, modelling, and self-reflective practice, are increasingly coming to the fore (Bennett-Levy *et al.*, 2009a). The present study focuses on self-reflection, as it is considered to be a fundamental element in the development of psychotherapeutic skills (Bennett-Levy, 2006; Jennings and Skovholt, 1999; Prasko *et al.*, 2023; Rønnestad and Skovholt, 2003; Schön, 1983; Thwaites *et al.*, 2014).

Bennett-Levy *et al.* (2009b; p. 121) define self-reflection as a ‘specific form’ of reflection, in which the focus is directed towards one’s own thoughts, emotions and behaviours or personal history. Applied to the clinical context, this means that self-reflection enables the therapist to pay attention to his or her own internal processes and behaviours during therapy sessions, and to analyse them in a way that leads to an increased understanding of the therapeutic process (Griffith and Frieden, 2000). According to Bennett-Levy and Thwaites (2007), this is especially important when dealing with complex interpersonal interactions, which is why self-reflection is seen as a key element for the development of interpersonal skills and psychotherapeutic skills in general (Bennett-Levy, 2006).

Self-reflection within training programs for psychotherapeutic skills

The value of self-reflection as a method to improve psychotherapeutic competence has mostly been studied within the framework of broader training programs, especially within self-practice/self-reflection (SP/SR) programs (Bennett-Levy *et al.*, 2001; Bennett-Levy *et al.*, 2003; Scott *et al.*, 2021). SP/SR is embedded in the framework of cognitive behavioural therapy (CBT) and consists of two main components, a *self-practice* component, which includes the actual practising of CBT-techniques on oneself (e.g. completing thought records or conducting behavioural experiments), and a *self-reflection* component, which refers to reflecting on the experiences made during self-practice (Bennett-Levy *et al.*, 2003). Several studies emphasize the beneficial effects of SP/SR as a training method (Chaddock *et al.*, 2014; Chigwedere *et al.*, 2020; Gale and Schröder, 2014; Prasko *et al.*, 2012). For example, Bennett-Levy *et al.* (2001) conducted a study in which 19 trainees participated in a one-semester course in CBT, which included SP/SR as a formal course requirement. Using a qualitative approach, the authors found that SP/SR led to a ‘deeper sense of knowing’, that is, a better understanding of the role of the therapist, the cognitive model, and the change process. Other positive effects reported by trainees were improvements in therapeutic self-concept and of self-reported therapeutic skills. In a follow-up study, Bennett-Levy *et al.* (2003) found that the positive effects of SP/SR, such as a refinement of specific psychotherapeutic skills, increased attention devoted to the therapeutic relationship, and increased therapeutic flexibility, were reported even 1–5 months after participation in SP/SR activities (Bennett-Levy *et al.*, 2003). Because this follow-up study focused on practising therapists, it supports the notion that SP/SR may be helpful not only at the beginning of one’s career, but also after a certain level of competence has been reached (Thwaites *et al.*, 2014).

Video analysis and self-reflection

Given the promising reports of the positive effects of video analysis in psychotherapy and medical training, it seems worthwhile to consider combining self-reflection with video analysis. For example, in their article, Gonsalvez *et al.* (2016) report on two specific video techniques in CBT supervision and outline the benefits of using video feedback in supervision. They argue that video feedback allows supervisees to effectively observe and reflect on their competencies and improve potential deficiencies. In this way, they link the use of video with deeper self-reflection to enhance learning. The benefits of analysing videotaped therapy sessions were also demonstrated by Lane and Gottlieb (2004). They conducted a study of 60 medical students who participated in a training

program to improve their interviewing and self-assessment skills. Part of the program included two reviews of videotaped patient encounters, one week apart, with one or two faculty members. The results showed that, from the faculty member perspective, the improvement in students' overall performance from one video review session to the next was large ($d = 1.0$). In addition, the concordance between students' and faculty members' ratings increased significantly from the first to the second measurement point (i.e. within one week). Although this study did not explicitly include self-reflection, the results support the notion that reviewing videotapes of one's own patient encounters can facilitate skill improvement and possibly enhance the accuracy of self-evaluation (Aafjes-van Doorn *et al.*, 2022).

By contrast, recalling a therapy session based on memory alone carries the risk of not only forgetfulness (Gonsalvez *et al.*, 2016), but also assessment bias and distortion (e.g. Haggerty and Hilsenroth, 2011; Yourman and Farber, 1996). These risks could be minimized or even avoided by using video recordings. Another important aspect is that video analysis shows aspects of behaviour of which therapists are not necessarily aware, such as automated non-verbal behaviour (Briggle *et al.*, 2016; Haggerty and Hilsenroth, 2011). According to Haggerty and Hilsenroth (2011), awareness of one's own non-verbal behaviour can lead to increased empathy and improved therapeutic alliance.

Nevertheless, it is important to note that not all studies report clear beneficial effects of video analysis on skill development. For example, a recent study by Weck *et al.* (2024) did not find significant differences in psychotherapeutic competence between students who received supervision based on video analysis and students who received supervision based on verbal reports alone. However, this study was not about self-reflection but about feedback from a supervisor. It is quite possible that video analysis may have a different effect in the context of self-reflection.

Self-assessment vs independent assessment of competence

Several studies have shown a low level of agreement between self-assessment and independent assessments of competence. For example, therapists and trainees tend to over-estimate their competence compared with independent observers (Brosan *et al.*, 2008; Longley *et al.*, 2023; Parker and Waller, 2015). However, Loades and Myles (2016) found a significant correlation between the reflective ability of mental health professionals and the level of agreement between self-assessments and independent assessments of competence. Thus, when examining the role of self-reflection in competence development, it seems important to consider both perspectives and to gain greater insight into how self-reflection influences both the therapist's own judgements and individually perceived development as well as externally observable skills.

Our study

Overall, there is evidence that self-reflection can be an enriching method for the therapeutic development of trainees and experienced therapists. Based on the literature, it also seems plausible that video analysis may even enhance its beneficial effects. However, most studies on the effects of self-reflection rely only on qualitative analysis and/or self-reports of trainee and therapist psychotherapeutic competencies. Accordingly, there is a need to expand research using quantitative methods and independent competence ratings (McGillivray *et al.*, 2015).

In this study, we investigated the effect and characteristics of structured self-reflection as a method for practising and enhancing therapeutic skills in a university seminar with undergraduate psychology students, using both quantitative and qualitative methods. Firstly, we investigated whether structured self-reflection is more effective when based on video recordings of therapeutic role-plays than on memory recall alone. Other studies have shown that students with minimal clinical experience can improve their skills after only a few hours of training (e.g. Maaß *et al.*, 2024;

Weck *et al.*, 2024), even without feedback, but with other interventions such as modelling (e.g. Kühne *et al.*, 2022). In this sense, we believe that in our study, self-reflection combined with other active learning strategies typically associated with significant learning success (role-playing, video analysis; Knox and Hill, 2021) can lead to skill improvement even without the external input of a trainer. Secondly, we explored the characteristics of the content of students' written self-reflections. More specifically, we investigated the aspects on which students reflected, whether the content of the self-reflections differed between students who practised video-based structured self-reflection (VSR) and students who practised memory-based structured self-reflection (MSR), and whether the content changed over the course of the seminar. Finally, by examining a focus group discussion that took place at the end of the training, we analysed how the students experienced the self-reflection process and what advantages and disadvantages they perceived. Our research questions were as follows:

- (1) Do two types of self-reflection differ in terms of their effect on therapeutic competence and therapeutic alliance?

Hypothesis: Students practising video-based self-reflection (VSR) improve their therapeutic skills significantly more than students practicing memory-based self-reflection (MSR), as judged by (a) independent raters (primary outcomes) and (b) student self-reports (secondary outcome).

- (2) What are the characteristics of students' self-reflections?
- (3) Does the content of the self-reflections differ between students practising VSR and students practising MSR?
- (4) Does the content on which students reflect change during 4 weeks of practice?
- (5) What are students' experiences with self-reflection? What advantages and disadvantages do they perceive?

Method

This study is a mixed-method randomized controlled trial (RCT), comparing VSR with MSR in a university seminar for undergraduate psychology students. The main topic of the seminar was *interpersonal therapeutic competencies*, and it consisted of a 6-week theory phase, followed by a 6-week practice phase. Two independent and trained raters assessed the students' psychotherapeutic competence and the therapeutic alliance before (pre-assessment; t_1) and after the practice phase (post-assessment; t_6) based on their performance in videotaped role-plays with standardized patients. During the practice phase (t_2 – t_5), the students carried out peer role-plays and completed a structured self-reflection questionnaire, either based on video (VSR) or based on memory (MSR), and provided a self-assessment of their psychotherapeutic skills. After the post-assessment, there was a guided focus group discussion in which students were asked to report their experiences during the practice phase, and what advantages and disadvantages they perceived (t_7).

Ethical approval and pre-registration

The study was pre-registered with the Open Science Framework in June 2021 (<https://osf.io/r5wdy>). Ethical approval was obtained from the University Ethical Committee (reference number 60/2021). The study conformed to the Declaration of Helsinki, and informed consent was obtained from all participants.

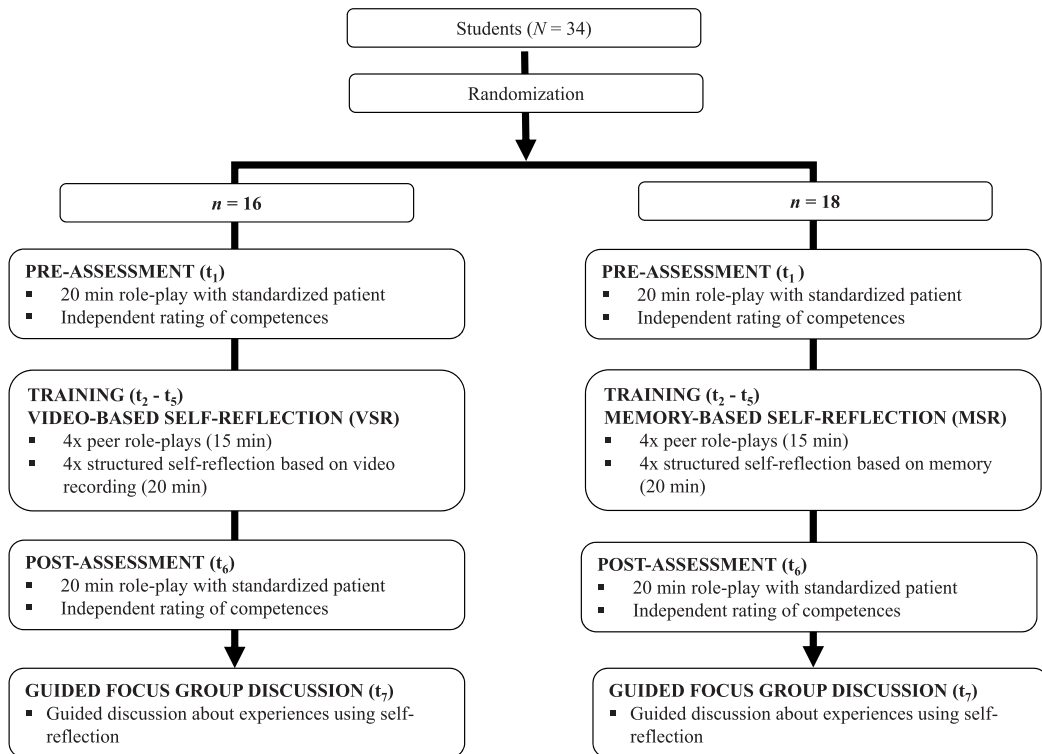


Figure 1. Participant flow chart.

Eligibility criteria

Eligibility criteria for study participation were assessed during the first two seminar sessions: (a) study of psychology (MSc), (b) participation in the seminar, and (c) a signed informed consent. Participation in the seminar was not compulsory for graduation, and participation in the seminar was possible without participation in the study (i.e. provision of study data was voluntary).

Randomization and blinding

We used simple randomization without extra stratification, and a 1:1 allocation ratio to assign participants to the study groups (VSR vs MSR). The presentation of the videotapes to the raters was also randomized to reduce recall bias or halo effects. An independent researcher, who was not involved in the project, programmed all randomization codes using the R package *randomizr* (RStudio Team, 2015; Ternovski and Coppock, 2017). Study group allocation was disclosed to participants after all had provided signed informed consent. All standardized patients, raters, and data analysts were blinded to the study group.

Participants

Thirty-six students enrolled in the seminar, two of whom did not consent to participate in the study (both from the VSR-group; see Fig. 1). The final sample consisted of $N = 34$ students (32 female) with a mean age of 25.0 years ($SD = 3.08$; VSR: $M = 24.6$, $SD = 3.44$; MSR: $M = 25.3$, $SD = 2.78$; see Table 1). Participants indicated their level of prior helping experiences on eight items (Maaß *et al.*, 2024), with a 5-point rating scale (1 = 'never' to 5 = 'at least once a week'; example item: 'Within the last five years, I completed internships in a consulting or therapeutic

Table 1. Mean age and prior helping experience of students

	VSR (<i>n</i> = 16)		MSR (<i>n</i> = 18)		Total sample (<i>N</i> = 34)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Age	24.6	3.44	25.3	2.78	25.0	3.08
Prior helping experience (1–5)	2.79	0.48	2.70	0.66	2.74	0.57

VSR, video-based self-reflection; MSR, memory-based self-reflection.

setting’). The students’ average level of prior helping experience was moderate ($M = 2.74$, $SD = 0.57$; VSR: $M = 2.79$, $SD = 0.48$; MSR: $M = 2.70$, $SD = 0.66$). The differences in age and prior helping experience between the groups were not significant (age: $t_{28.89} = 0.60$, $p = .551$; prior helping experience: $t_{30.8} = -0.45$, $p = .658$).

Standardized patients

Two undergraduate psychology students who worked as student assistants in our department acted as standardized patients (SPs) during the pre- and post-assessments (one female: MSc student, 23 years old; one male: BSc student, 21 years old). We ensured that SPs and participants did not know each other. The SPs were prepared for their task by attending a total of 8 hours of training. SP authenticity was evaluated by the independent raters using the Authenticity of Patient Demonstrations Scale (APD; Ay-Bryson *et al.*, 2022), which includes 10 items on a 4-point rating scale (1 = ‘strongly disagree’ to 4 = ‘strongly agree’; example item: ‘The way of speaking (pitch, tempo) is appropriate in the context of the disorder’). The average authenticity rating across time points and raters was good ($M = 3.64$, $SD = 0.38$, range: 2.80–4.0), and reliability at pre- and post-measurement was excellent (t_1 : $\alpha = .96$; t_6 : $\alpha = .97$).

Measures

Primary outcomes

The primary outcomes were the students’ psychotherapeutic skills, and the therapeutic alliance as assessed at pre- and post-assessment (t_1 and t_6) by two independent raters (both female, advanced psychotherapy trainees, clinical experience of at least 2 years). In order to counteract rater drift (i.e. changes in rater performance over time), the common understanding of the rating scales was discussed not only during the rater training (8 hours), but also after the first 18 videos had been rated. The intra-class correlation (ICC) was initially only fair, $ICC_{2,2} < .60$. We suspect that the low inter-rater reliabilities were a consequence of the difficulties in rating associated with the online setting. We had initially trained our raters to assess the behaviour of therapists in ‘real life’ therapy sessions, i.e. sitting across from their patients in the same room. Due to the COVID-19 pandemic, the practice phase of the seminar had to take place online from the beginning, which meant that the raters could not see a ‘regular’ therapy setting, but only two frontal views of the faces superimposed in separate windows. As a result, they lacked both spatial and non-verbal information (e.g. eye contact and gestures related to one another, such as leaning forward or nodding, posture), which made the situation more ambiguous and possibly made it more difficult to reach a consistent judgement. It was then decided that the raters should come to a common assessment for a random subsample of $k = 24$ videos within a consensus rating procedure. The final ICCs reported in the section below therefore refer to the average of the consensus and independent ratings. The independent raters used the following standardized rating scales to assess the therapeutic alliance, students’ CBT skills, and communication skills:

Therapeutic alliance

The German Helping Alliance Questionnaire (HAQ; Luborsky, 1984; German version: Bassler *et al.*, 1995) assesses the collaborative and affective bond between therapist and patient with 11 items (1 = 'strongly disagree' to 6 = 'strongly agree'). We used the independent rater's version of the HAQ in German, which was developed by Richtberg *et al.* (2016). An example item is: 'I believe the patient is working together with the therapist in a joint effort'. The intra-class correlation in this study was $ICC_{2,2} = 0.70$.

CBT skills

The German version of the Cognitive Therapy Scale (CTS; Weck *et al.*, 2010; Young and Beck, 1980) was used to assess competence in delivering CBT skills. Although the seminar focused on interpersonal competencies and the therapeutic relationship, we were also interested in whether the process of self-reflection would influence a broader range of therapeutic skills. Therefore, we included the CTS as an additional measure of CBT skills. The CTS consists of 14 items that each represent a specific CBT skill (e.g. 'agenda setting', 'guided discovery', 'resource activation'). Skills are assessed on a 7-point rating scale (0 = 'poor', 6 = 'excellent'). Five items (i.e. Item 7: 'reviewing previous homework'; Item 11: 'rationale'; Item 12: 'selection of appropriate strategies'; Item 13: 'appropriate implementation of techniques'; Item 14: 'homework setting') were excluded because they were considered inappropriate for the specific task in this study (conducting an initial interview with the patient). The intraclass correlation was $ICC_{2,2} = 0.90$.

Communication skills

The Clinical Communication Skills Scale – Short Version (CCSS-S; Maaß *et al.*, 2022) assesses professional clinical communication skills with 14 items (1 = 'extremely inadequate' to 4 = 'extremely adequate'). A sample item is: 'The therapist summarizes interim results'. If any particular item cannot be judged, the option 'not applicable' can be used. The intraclass correlation in this study was $ICC_{2,2} = 0.74$.

Secondary outcomes

The secondary outcomes include both qualitative as well as quantitative data. The qualitative data refers to the written self-reflections. The quantitative data represents self-ratings of students' CBT skills during practice (t_2 – t_5).

CBT skills

As part of the self-reflections during the practice phase (t_2 – t_5), students judged their own CBT skills by rating the following three items from the CTS: 'interpersonal effectivity', 'feedback', and 'focusing on key cognitions and behaviors' ($\alpha = .50$ at the first measurement point).

Role-plays

In all role-plays, the students' task was to conduct an initial interview with a patient with the aim of exploring the patient's main problem. However, there were two types of role-plays: with a SP and with a peer. The role-plays with an SP were conducted at pre- and post-assessment, were videotaped, lasted 20 minutes, and formed the basis for the independent raters' assessment of the primary outcomes. The peer role-plays were conducted during the practice phase, lasted 15 minutes, and formed the basis for analysing the secondary outcomes (written self-reflections and self-assessments of competence), as students completed a self-reflection questionnaire and provided self-assessments after each peer role-play. From practice session to practice session,

students alternated between the roles of therapist and patient in their peer role-plays. In the VSR group, peer role-plays were videotaped, whereas in the MSR group, they were not.

Self-reflection

The structured self-reflection questionnaire mentioned above included the following aspects: (1) own behaviour (e.g. empathy, collaboration, appreciation), (2) SP behaviour during the session (e.g. shy, dominant), (3) own reaction to SP (e.g. own body language, facial expression, feelings, and thoughts; e.g. 'nervous', 'stern look'), and (4) suggestions for improvement. Students were asked to write individual words or short sentences in response to the questions. In addition, they were asked to rate their own performance on three items of the CTS ('interpersonal effectivity', 'feedback', and 'focusing on key cognitions and behaviors'; see secondary outcomes).

Procedure

The undergraduate seminar focused on interpersonal skills and the therapeutic relationship. It was part of a module called 'Prevention and Treatment' in a Master's program at our university specializing in clinical psychology, psychotherapy, and counselling psychology. The aim of the seminar was to combine scientific theoretical input with hands-on practice, thus promoting a close link between research and practice. During a theory phase, students familiarized themselves with current research on the therapeutic relationship. During a following practice phase, they developed their interpersonal skills through self-reflection in practical role-plays.

Theoretical input

During a 6-week theory phase, students and teachers discussed current evidence on the therapeutic relationship, and students gave presentations on different aspects of the topic, which were later graded by the teachers. Topics included the measurement of therapeutic alliance, strategies for establishing a strong therapeutic alliance, CBT and psychodynamic approaches, and alliance ruptures. Theoretical models were presented, and current research was discussed.

Practice with peer role-plays

Students were then familiarized with the procedure for the following practice phase, which consisted of four weekly practice sessions. Each practice session included two peer role-plays and structured self-reflection. There were different role scripts for the patient role for each practice session. This means that, since there were two peer role-plays per practice session, students used the same standardized therapy situation for two consecutive peer role-plays in one session, and a different situation for the next session.

Students remained in the same role (i.e. 'therapist' or 'patient') during one practice session, but switched roles during the next practice session one week later. This means that if a student played the 'therapist' in two peer role-plays during the first practice session, he or she would play the 'patient' in the next session. Then, he or she would play the 'therapist' again in the third session and the 'patient' again in the fourth session. Ultimately, each student acted as both the 'therapist' and the 'patient' in two practice sessions, each of which included two peer role-plays. Thus, each student played each role four times. After the first peer role-play, 'therapists' in the VSR group watched the videotape of their role-play and completed a structured self-reflection questionnaire. 'Therapists' in the MSR group completed the same questionnaire. This was done from memory, as their sessions were not video recorded. The time frame for the self-reflection unit in both groups was 20 minutes (i.e. students in the VSR group were not given extra time to watch the video). After a short break, students repeated their peer role-play (15 minutes of peer role-play followed by 20 minutes of self-reflection).

Pre- and post-assessment

The pre- and post-assessment role-plays were scheduled outside of the seminar time and took place before and after the practice phase.

Group discussion

During the last session of the seminar, a guided group discussion was conducted in each study group and recorded on audiotape (t_7 ; Fig. 1). The discussion was guided by the following questions: (1) what went well with the self-reflection?, what was difficulty?; (2) what did you learn?; (3) what are the main advantages of self-reflection?; (4) what are the main disadvantages?; and (5) were there any side-effects?, i.e. did it hurt you in any way?

Data analysis

All statistical analyses were performed using RStudio (RStudio Team, 2015).

Sample size

We expected a moderate effect size for the difference between the two study groups. The *a priori* power analysis for the repeated measures ANOVA resulted in a total sample size of $N = 34$ (G^* power; Faul *et al.*, 2009; power = 0.80, $d = 0.50$, $\alpha = .05$).

Changes in therapeutic skills and therapeutic alliance

To answer research question (RQ) 1 and to test the corresponding hypothesis of VSR superiority, we conducted two analyses. First, we analysed the primary outcomes with a repeated measures MANOVA (i.e. independent ratings of the CTS, CCSS-S, and HAQ from pre- to post-assessment, t_1 vs t_6). The modified ANOVA-Type Statistic (MATS) from the R package MANOVA.RM (RStudio Team, 2015; Friedrich *et al.*, 2023) was used to interpret the results. We chose MATS because we had distribution violations in the data, and MATS is less sensitive to distribution violations (Friedrich and Pauly, 2018). Secondly, we analysed the secondary outcome with a repeated measures ANOVA (i.e. students' self-assessments during the practice phase, t_2 – t_5). In both analyses, the group variable (i.e. VSR = 1 vs MSR = 0) was the independent variable. *Post hoc* comparisons between measurement points were performed using *t*-tests with Bonferroni-Holm correction. Generalized η^2 ($\eta^2_G > .01$ = small effect, $\eta^2_G > .05$ = moderate effect, $\eta^2_G > .14$ = large effect) and Cohen's d ($d < 0.50$ = small effect, $0.50 < d < 0.80$ = moderate effect, $d > 0.80$ = large effect) were used to interpret the effect sizes (Cohen, 1988).

There were no missing data for the primary outcomes (i.e. rater-based data). For the secondary outcome (i.e. students' self-ratings), missing values on individual items (three in total) were imputed using the student's mean score on the remaining items of the CTS. If all of a student's values were missing at a given measurement point, the student was excluded from the analysis of the corresponding measurement point. This was the case for four questionnaires.

Written self-reflections

To answer RQs 2 to 4 (characteristics, group differences, and changes over time of the self-reflections), we analysed the written self-reflections from each peer role-play, using qualitative content analysis and frequency analyses, and compared the content and frequency of each topic in the self-reflections, both between groups (VSR vs MSR) and over time (t_2 to t_5). The qualitative content analysis was performed using MAXQDA 2020 (VERBI Software, 2019) and following an inductive-deductive procedure (Kuckartz and Rädiker, 2019). This means that the main content categories were pre-determined by the structure of the self-reflection questionnaire, while the subcategories were derived inductively, based on the specific content of the individual text segments. We were particularly interested in students' reflections on their 'own behaviour', as this

Table 2. Descriptive statistics of the primary outcomes (independent ratings of therapeutic skills and alliance)

	VSR (<i>n</i> = 16)				MSR (<i>n</i> = 18)			
	Pre		Post		Pre		Post	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
HAQ	3.05	0.29	3.04	0.28	3.04	0.35	3.12	0.34
CTS	3.05	0.45	3.04	0.52	3.18	0.59	3.20	0.45
CCSS-S	2.97	0.26	3.00	0.32	2.99	0.29	3.08	0.24

VSR, video-based self-reflection; MSR, memory-based self-reflection; HAQ, Helping Alliance Questionnaire; CTS, Cognitive Therapy Scale; CCSS-S, Clinical Communication Skills Scale – Short Version.

is the category most closely related to observable performance and competence. Therefore, we will focus on analyses of the ‘own behaviour’ category. A preliminary system of subcategories of ‘own behaviour’ was created and discussed by two researchers after having analysed three self-reflection sheets. Inter-rater agreement was then calculated from another 14 randomly selected self-reflection sheets ($\kappa = .74$, 11% of data material). Of course, narrowing the focus in this way and analysing only the ‘own behaviour’ category has its drawbacks (e.g. limiting the breadth of responses to our qualitative research questions). At the same time, the seminar focused on improving students’ (observable) skills, and students’ thoughts about this aspect are best reflected in their descriptions of their external behaviour. Therefore, to follow a clear thematic line, we decided to focus on the category of ‘own behaviour’ and leave out the other categories.

Students’ experiences with self-reflection and their perceived advantages and disadvantages

To answer RQ 5, we analysed the key points of the focus group discussions using a procedure based on Braun and Clark’s (2006) instructions for thematic analysis. Prior to the analysis, both discussions were transcribed using MAXQDA 2020 (VERBI Software, 2019). All analysis steps (familiarization with the data, generation of codes, collation of codes into themes and subthemes) were carried out by the first author, and applied across the two study groups, rather than separately for each group. Again, an inductive-deductive approach was used, as the main themes were pre-determined by the questions that guided the discussion (‘advantages’, ‘disadvantages’, ‘learning progress’), and the subthemes were deduced inductively from the data itself.

Deviations from pre-registration

Due to the COVID-19 pandemic, we had to switch to an online setting from the fourth week of the seminar. This meant that all role-plays were conducted using Zoom software (Zoom, 2021). As an additional analysis, we have included in this paper the quantitative examination of students’ self-assessments on the three items of the CTS, which was not previously planned. Also, instead of applying qualitative content analysis to the transcripts of the group discussions, we analysed them using thematic analysis, as this enabled a more flexible approach to the data.

Results

The descriptive statistics for the primary and secondary outcomes are displayed in Tables 2 and 3.

Changes in therapeutic skills and alliance

Primary outcomes

The repeated measures MANOVA for the primary outcomes (i.e. independent ratings of CTS, CCSS-S, HAQ) yielded no significant differences for the main effects of time (pre- vs post-assessment, MATS = 0.85, $p = .582$), or group (VSR vs MSR; MATS = 2.20, $p = .551$), or for the time $\times$ group interaction (MATS = 0.54, $p = .741$).

Table 3. Descriptive statistics of the secondary outcome (students’ self-rated CBT skills)

	VSR				MSR				Total sample			
	RP1 (<i>n</i> = 16)	RP2 (<i>n</i> = 16)	RP3 (<i>n</i> = 15)	RP4 (<i>n</i> = 14)	RP1 (<i>n</i> = 18)	RP2 (<i>n</i> = 17)	RP3 (<i>n</i> = 16)	RP4 (<i>n</i> = 17)	RP1 (<i>n</i> = 34)	RP2 (<i>n</i> = 33)	RP3 (<i>n</i> = 31)	RP4 (<i>n</i> = 31)
<i>M</i>	3.34	3.80	3.72	3.96	3.76 (0.52)	4.24	3.59	4.14	3.56 (0.53)	4.03	3.65	4.06
(<i>SD</i>)	(0.45)	(0.48)	(0.51)	(0.52)		(0.47)	(0.69)	(0.78)		(0.52)	(0.61)	(0.67)

VSR, video-based self-reflection; MSR, memory-based self-reflection; RP, role-play.



Figure 2. Development of students' self-assessed skills during training.

Table 4. *Post hoc comparisons between measurement points for the secondary outcome (students' self-rated CBT skills)*

Comparison			<i>t</i> (d.f.)	<i>p</i>	<i>d</i> [95% CI]
Role-play 1	vs	Role-play 2	-6.67 (32)	<.001	1.16 [0.71, 1.60]
	vs	Role-play 3	-1.13 (30)	1.00	0.20 [-0.16, 0.56]
	vs	Role-play 4	-3.77 (30)	.004	0.68 [0.28, 1.06]
Role-play 2	vs	Role-play 3	2.82 (30)	.050	-0.51 [-0.88, -0.13]
	vs	Role-play 4	-0.52 (29)	1.00	0.10 [-0.26, 0.45]
Role-play 3	vs	Role-play 4	-4.20 (29)	.001	0.77 [0.35, 1.17]

Secondary outcomes

The repeated measures ANOVA for the secondary outcome (i.e. students' self-ratings of the CTS items during the practice phase) revealed a significant time effect, $F_{1,92,53.77} = 9.94$, $p < .01$, $\eta^2_G = 0.12$, but no significant group effect, $F_{1,28} = 1.62$, $p = .214$, $\eta^2_G = 0.03$, and no significant time $\times$ group interaction, $F_{1,92,53.77} = 3.13$, $p = .054$, $\eta^2_G = 0.04$ (see Fig. 2). *Post hoc* comparisons are displayed in Table 4. All students self-reported improved CBT skills from the first to the fourth peer role-play (t_2-t_5 , $d = 0.68$), and within a practice session, that is, from the first to the second peer role-play (t_2-t_3 , $d = 1.16$), and from the third to the fourth peer role-play (t_4-t_5 , $d = 0.77$). In addition, there was a significant decrease in skill performance from the second to the third role-play (t_3-t_4), that is, from one session to the next with a 2-week break in between (see Fig. 2).

Written self-reflections

Characteristics of the written self-reflections

Students reflected on their behaviour regarding six categories, which were mentioned with the following frequencies relative to the total number of comments in the dataset: (1) empathy (reflections on validation, empathy, and understanding, 30%), (2) appreciation (reflections on praise, respect, and appreciation of the patient, 23%), (3) collaboration (reflections on

Table 5. Frequencies of the individual subcategories of ‘own behaviour’

	Positive				Negative			
	Example	<i>n</i>	Frequency (%)		Example	<i>n</i>	Frequency (%)	
Empathy	‘I validated the patient a lot’	195	24		‘I need to show more understanding’	50	6	
Appreciation	‘I was respectful’	169	21		‘I could have been more appreciative’	20	2	
Collaboration	‘I have assured my support’	74	9		‘I was too directive’	9	1	
Encouragement	‘I have communicated hope for change’	59	7		‘I have not encouraged enough that change is possible’	34	4	
Normalization	‘I took away her worries about going crazy’	12	1		—	0	0	
Communication Skills	‘I managed to find a good start to the conversation’	132	16		‘I don’t have a good wording to start the conversation’	57	7	
Total		641	79			170	21	

n refers to the absolute number of text elements within the whole dataset.

cooperation, support, and collaboration, 10%), (4) encouragement (reflections on encouragement and conveying confidence and hope, 11%), (5) normalization (reflections on normalization and eradicating the fear of being crazy, 1%) and (6) communication (reflections on structuring the conversation, summarizing and exploring, 23%). We divided each category into a ‘positive’ subcategory, including all text segments that emphasized positive aspects of one’s own behaviour, and a ‘negative’ subcategory, including all text segments that emphasized negative aspects. Frequency analysis revealed that, across both groups and across all measurement points, students reflected more on positive aspects of their behaviours than on the negative ones (79% vs 21%, see Table 5).

Differences between the two study groups

In addition, we examined whether the content of the self-reflections differed between the VSR and the MSR groups. Our analyses showed that almost all of the subcategories derived in the first step had similar frequency distributions in both study groups, with discrepancies occurring only in the categories of appreciation and collaboration. Specifically, the frequency distributions in both study groups were as follows: empathy: 30% (VSR and MSR), appreciation: 19% (VSR) vs 27% (MSR), collaboration: 15% (VSR) vs 6% (MSR), encouragement: 11% (VSR) vs 12% (MSR), normalization: 2% (VSR) vs 1% (MSR), communication: 23% (VSR) vs 24% (MSR; see Supplementary material). However, it was noticeable that the self-reflections of the students from the MSR group tended to be more extensive, as indicated by a greater number of text segments coming from the MSR group than from the VSR group [$n = 464$ (MSR) vs $n = 398$ (VSR)]. We also compared the two study groups in terms of the ratio of positive to negative reflections, and found that the MSR group reflected more on positive aspects than the VSR group, VSR: $N_{\text{positive}} = 275$ (74%), $N_{\text{negative}} = 98$ (26%); MSR: $N_{\text{positive}} = 366$ (84%), $N_{\text{negative}} = 72$ (16%).

Changes over time

Finally, we examined whether the frequencies of the subcategories changed over time. Except for the fact that the frequency of the communication category decreased considerably from the first to the fourth role-play (29% vs 19%) and the collaboration category increased by 6% from the first to the fourth role-play (7% vs 13%), this was not the case (see Fig. 3). In terms of the ratio of positive

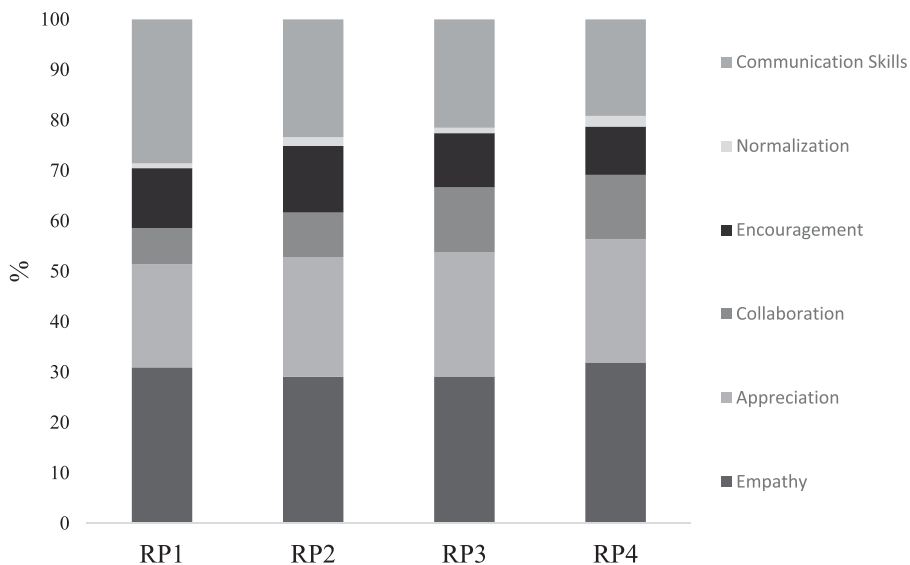


Figure 3. Frequencies of the individual subcategories of 'own behavior' over time.

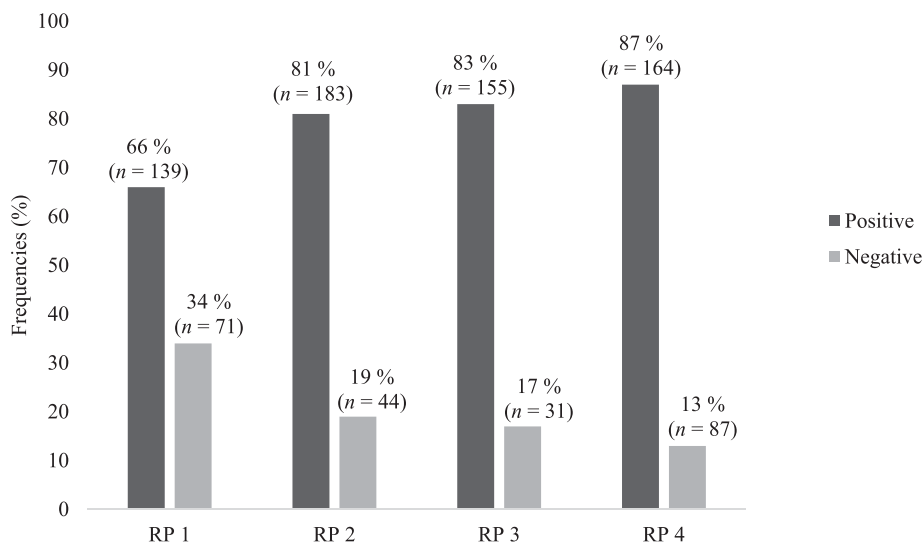


Figure 4. Frequencies of positive and negative reflections of 'own behavior' over time.

to negative reflections, the proportion of positive reflections increased over the course of the practice phase in both study groups, as shown in Fig. 4.

Experience of self-reflection and perceptions of advantages and disadvantages

The thematic analysis of the focus group discussions provided deeper insights into student perceptions of the advantages, disadvantages, and learning progress associated with self-reflection. With regard to the main theme of 'advantages', students mainly mentioned the 'helpful aspects of the self-reflection sheet' (subtheme 1) and of the 'video analysis' (subtheme 2). Regarding the self-reflection sheet, they appreciated the fact that it gave them structure and drew their attention to

aspects that would otherwise have been overlooked, such as facial expressions and posture. Students also felt that they became more comfortable answering the questions as the practice phase progressed. Students in the VSR group found it helpful that the videos enabled them to perceive immediate (non-verbal) reactions (e.g. facial expressions). With regard to the main theme of 'disadvantages', many arguments concerned 'difficulties associated with the self-reflection sheet' (subtheme 1), e.g. a lack of clarity of the questions and lack of an open-ended response field for free reflection. Students also pointed out the 'difficulties associated with video analysis' (subtheme 2), which some of them found demanding in the short time available. Other disadvantages related to 'difficulties associated with other structural aspects of the practice phase' (subtheme 3), i.e. the lack of external feedback and the brevity of the self-reflection units, which was perceived as overwhelming. Students also pointed out the brevity of the practice phase as a whole as a major disadvantage. Students in the VSR group indicated that the self-reflection process generally increased their awareness of their own weaknesses. With regard to the main theme of 'learning progress', students reported improvements in self-evaluation, structuring a therapy session, and dealing with difficult situations.

Discussion

The aim of this study was to investigate whether video-based structured self-reflection (VSR) and memory-based structured self-reflection (MSR) differed in terms of their effect on students' psychotherapeutic competence and on the therapeutic alliance. Contrary to our expectations, there were no significant differences between the two study groups. This means that VSR was not more effective than MSR in improving students' CBT and communication skills and the therapeutic alliance. In fact, according to the judgement of the independent raters, structured self-reflection did not improve students' skills and the alliance from pre- to post-assessment at all. There were also no differences between the two study groups in terms of students' self-assessed skills during the practice phase, but students reported significant improvements in their skills from one role-play to the next and from the first to the last role-play. The qualitative analysis of the written self-reflections showed that students reflected on various aspects of their behaviour during the standardized therapy sessions, such as empathy, appreciation, collaboration, encouragement, normalization and communication. Both study groups focused significantly more on the positive aspects of their behaviour than on the negative aspects, with the proportion of positive reflections actually increasing over the course of the practice phase. Finally, the analysis of the focus group discussions showed that the students perceived both advantages and disadvantages of the self-reflection practice, particularly with regard to the self-reflection sheet and the video analysis. Important difficulties from the students' perspective were the time constraints (both for the individual self-reflection units and for the practice phase as a whole) and the lack of external feedback. Students in the VSR group, but not the MSR group, indicated that the self-reflection process increased their awareness of their own weaknesses.

Several possible explanations exist for the lack of differences between MSR and VSR in the quantitative outcomes. Firstly, in order to keep the time factor constant, we did not give the students in the VSR group extra time to watch the video, which in turn may have created time pressure and excessive demands that prevented the students from benefiting from the video analysis. This is reflected not only in the arguments that occurred in the group discussion, where the students in the VSR group mentioned that they had great difficulty in combining self-reflection with video analysis within the given time frame, but also in the written self-reflections, where the number of reflections in the MSR group exceeded that of the VSR group. Secondly, the students were not given any instructions on how to analyse the video. It is possible that a brief introduction to important aspects to pay attention to when watching the video would have helped the students to benefit from the video analysis. However, despite the non-significant results

regarding the quantitative outcomes, the qualitative results of our study provide evidence that VSR offers advantages that MSR does not. Although the two study groups reflected on similar content, the tendency to focus on the positive aspects of one's behaviour was less pronounced in the VSR group, indicating that students in the VSR group appeared to have a more critical view of their own performance. This suggests that watching videotapes of one's own performance could help to focus on one's weaknesses and thus recognize more easily what still needs to be improved. This is consistent with the results of the group discussion, where students in the VSR group noted that the videos allowed them to see immediate non-verbal reactions and that the self-reflection process generally made them more aware of their weaknesses.

Looking at the effect of self-reflection in general, regardless of the study group, the current findings regarding the independent raters' perspective are somewhat at odds with the literature, which emphasizes the positive impact of self-reflection on professional (and personal) development (e.g. Bennett-Levy *et al.*, 2001; Bennett-Levy *et al.*, 2003; Chigwedere *et al.*, 2020; Scott *et al.*, 2021). One explanation for the discrepancy between the findings may be that methodological aspects of our study design may have limited the benefits of the self-reflection practice for the students. For example, the structured self-reflection sheet may have been too long and extensive for a 20-minute self-reflection unit. This is also revealed in the feedback from the students in the focus group discussions, who pointed out the time limitations during the self-reflection units. Furthermore, the students only received instructions during the theoretical phase of the seminar and were not given further guidance during the following practice phase. They also reported in the group discussions that they missed external feedback. This is consistent with the recommendations of Collard (2024), who advocates combining SP/SR with expert feedback. In line with this, Henrich *et al.* (2023) conclude in their review that additional supervision or guidance by an expert has a significant incremental benefit on competence development. This suggests that external feedback is a necessary factor in the development of psychotherapeutic competence, and that the lack of it may have been an important reason for no significant increase in externally rated competence among the students in our study.

Another important aspect may be that our participants had too little practical experience to really engage in self-reflection in a way that would benefit them. According to Bennett-Levy's (2006) three-systems (DPR) model of therapist skill development, the reflective system compares past, present, or future experiences with existing knowledge, and adjusts the existing information if necessary. It is considered the central system through which experienced therapists refine their skills, as they no longer need to learn skills from scratch, but must apply existing knowledge to new contexts. With this in mind, it seems possible that self-reflection is most effective when a certain level of competence already exists, which then needs to be refined and adapted to different contexts. At the beginning of a career, there may not be enough experience for the reflective system to draw on, which could reduce the impact of self-reflection.

Also, when attempting to explain the non-significant results on the primary outcomes, it is important to note that the literature has mainly looked at self-reflection in combination with self-practice components and has not isolated the effect of self-reflection. Furthermore, it has mainly focused on self-reports of psychotherapeutic skills. In our study, too, the students themselves noticed a significant improvement in CBT skills from one role-play to the next within one practice session and from the first to the last role-play. On the one hand, this finding is not surprising, given that the same standardized therapy situation was repeated twice within a practice session. Other studies also point to a training effect, especially in training sessions in which difficult situations can be repeated (Kühne *et al.*, 2022; Maaß *et al.*, 2024; Weck *et al.*, 2024). On the other hand, the discrepancies in judgements between the independent observers and the students are consistent with the general finding in the literature that the two perspectives often differ. One interpretation is that the self-perceived progress indicates a positive self-assessment bias, i.e. an over-estimation of one's own abilities relative to the judgement of independent observers. As mentioned in the introduction, such a bias is well documented in the literature (Longley *et al.*,

2023; Probst *et al.*, 2022; Walfish *et al.*, 2012). Another explanation might be that the self-reported progress indicates an increase in self-awareness and self-confidence rather than objectively measurable skills. This improvement may be a direct result of the self-reflective task completed by all students, regardless of their study group. From this perspective, the results are consistent with the qualitative responses as well as the literature on SP/SR, which shows that self-reflection is often associated with increased self-awareness (e.g. Bennett-Levy, 2019). Although such improvements are inconsistent with observers' ratings of students' skills, they remain an important milestone in therapist development (Pieterse *et al.*, 2013). However, there was a decline in skill performance from the second to the third role-play (t_3 – t_4). We hypothesize that this decline was due to the 2-week break for the therapist role and the associated change in the therapy scenario, suggesting that students found they could not maintain the level of competence they had achieved in the first practice session and were unable to generalize their learning experiences to the new context of the second practice session 2 weeks later. This is consistent with the notion that learning is often highly context-specific and requires extra effort and strategies to transfer new insights and skills to new situations (Butler *et al.*, 2017; Rivière *et al.*, 2019). For our study, this means that explicit practice of different scenarios would likely have been needed to maintain the perceived positive effects of self-reflection and to further strengthen the development of self-awareness and self-confidence that began during the first practice session.

The qualitative analysis of the written self-reflections shows that the students focused mainly on the positive aspects of their performance. Furthermore, 74% of the comments in the written self-reflections belong to subcategories that reflect skills that are important for building a sustainable therapeutic relationship and motivating the patient (e.g. empathy, appreciation, collaboration, encouragement). This is certainly due in large part to the specific task in the role-play (conducting an initial interview) and the focus of the seminar (therapeutic relationship and interpersonal skills), but it may also indicate that students do not pay attention to other skills (e.g. guided discovery, identifying dysfunctional thoughts, etc.). In this context, it may be relevant to note that the frequency of the communication subcategory decreased by 10% from the first to the last role-play. This subcategory entailed more structuring skills (e.g. summarizing the conversation) compared with relationship-related subcategories, and seems to be one of the areas of competence that students tend to lose track of unless their attention is explicitly drawn to it, and with which they may still be struggling. Overall, however, the themes in the self-reflections did not change considerably over time, but the focus on strengths became more entrenched and the basic tone became more and more positive. Explicit feedback and guidance would probably have been needed to counteract the tendency to over-emphasize strengths over weaknesses.

Strengths and limitations

This RCT contributes to research in the field of therapist development by examining the effects of structured self-reflection on competence, using not only qualitative, but also quantitative methods, and relying not only on self-ratings of competence but also on competence ratings by independent observers. In addition, it took place in the naturalistic setting of a university seminar, which facilitates converting conclusions into practice, and it has isolated the effect of self-reflection, in contrast to other studies examining multi-component training programs. This approach enhances our understanding of the effects of individual training elements and may help develop effective training programs. However, there are several limitations to be aware of. First of all, our sample consisted of undergraduate students with little therapeutic experience. While examining competence at all levels (novice, advanced, expert, etc.) may help establish benchmarks (see, e.g. Hill and Knox, 2023) for therapist competence, it is plausible that the lack of experience of our participants may have limited the potential benefits of self-reflection, and that with more experienced therapists, the self-reflection practice in our study may have resulted in greater changes in competence. With regard to our study design, it is important to note that all the role-

plays, including the pre- and post-assessments, had to take place online, which led to difficulties with the independent competence ratings. The self-ratings of the students during the practice phase consisted of an average value of three items of the CTS, which resulted in a low measurement reliability. Furthermore, as already mentioned above, the self-reflection sheet we used may have been too extensive, and we should perhaps have either reduced the number of guiding questions or provided more time for the single self-reflection units. The time aspect also refers to the duration of the practice phase in total. Four weeks of practice with four role-plays in the therapist role may not have been sufficient to improve skills significantly. Finally, the generalizability of our findings to training programs is limited because our program is likely to be significantly different from most other training programs due to the lack of external input from a trainer.

Implications for practice and future research

As mentioned above, self-reflection in general, and video-based self-reflection in particular, must probably meet certain requirements in order to be effective in visibly improving psychotherapeutic skills at the beginning of a therapist's professional career. Most importantly, self-reflection may need to be accompanied by external input and guidance from a trainer. This includes individualized competency feedback and, with respect to the self-reflection process, specific instructions on what exactly to reflect on, how to analyse video recordings, how to focus specifically on negative aspects of one's performance, and how to draw conclusions and suggestions from this. In addition, the focus of self-reflection should be narrowed and directed to specific individual aspects. It is possible that some participants in our study were overwhelmed by the range of competencies to which they had to pay attention. Finally, our results support the notion that learning self-reflection is time-consuming (Bennett-Levy *et al.*, 2001; Spafford and Haarhoff, 2015). A series of multiple practice sessions is probably necessary for self-reflection to have beneficial effects, with self-reflection units ideally lasting more than 20 minutes. Together with the finding that a trainer is needed for effective self-reflection, this leads to the conclusion that self-reflection is not as time- and resource-efficient as one might initially think.

It is important to note that our results do not allow conclusions about the validity of the DPR model, as we examined SR as an isolated skill. According to the DPR model, an integration of all three systems (declarative, procedural, and reflective) is necessary to promote change. For novice therapists with limited experience, engaging in self-reflection alone may not be sufficient to promote psychotherapeutic skills. Therefore, in addition to written self-reflections, the students in our study may have needed the sharing of theoretical knowledge and direct instruction on how to practice certain skills.

What can be learned from our study for future research is that future studies should combine self-reflection with theoretical input and direct feedback and instructions in order to investigate the effect on competence development. For example, to determine whether self-reflection has an incremental benefit on students' development of psychotherapeutic competence, future studies could compare training consisting of feedback and self-reflection with training consisting of feedback alone. In addition, studies should be designed with longer time frames for the self-reflection units, as well as studies that examine not only changes in skills but also changes in self-awareness and self-confidence as primary outcomes, using not only qualitative but also quantitative measures [e.g. the Counselor Activity Self-Efficacy Scales-Revised (CASES-R; Hahn *et al.*, 2021)]. Certainly, larger samples should be used, and competence development should be evaluated monitoring changes over time to investigate the sustainability of potential training effects. Finally, the effects of different methods of promoting self-reflection should be compared between samples of inexperienced trainees and licensed therapists, and open formats for self-reflection should be evaluated.

Conclusion

In summary, the results of this RCT suggest that VSR, as it was conducted in our study, was not more effective than MSR in improving students' therapeutic skills. Doubts have also been raised about whether self-reflection alone, without further guidance and feedback, can visibly improve students' psychotherapeutic competencies. However, it is important to note that students reported progress in their perceived skills (quantitative measures) and mainly reported positive aspects of their behaviour in their self-reflections (qualitative measures), suggesting a potential increase in self-awareness. This leads to the conclusion that well-prepared self-reflection, combined with more instruction and external feedback, could increase students' psychotherapeutic competence.

Key practice points

- (1) For self-reflection to be effective, students may need to be extensively prepared and trained (e.g. on important aspects to reflect on, how to analyse video recordings, and how to focus specifically on negative aspects of one's performance).
- (2) It may be necessary to guide the self-reflection process and provide performance feedback.
- (3) Self-reflection can quickly become overwhelming for students, and it is therefore necessary that they take sufficient time to adequately process and reflect.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1754470X25100287>

Data availability statement. This study design and its analysis were pre-registered at the Open Science Framework (7 October 2021). All analysis codes are available at <https://osf.io/r5wdy/>. The datasets are available from the corresponding author on request.

Acknowledgements. We thank our student assistants for supporting the study and Brian Bloch (University of Münster, Germany) for his English language editing of the article.

Author contributions. **Klara Eisert:** Data curation (equal), Formal analysis (equal), Writing - original draft (lead), Writing - review & editing (lead); **Franziska Kühne:** Conceptualization (equal), Funding acquisition (equal), Investigation (equal), Methodology (equal), Project administration (equal), Writing - review & editing (equal); **Florian Weck:** Resources (equal), Writing - review & editing (equal); **Anna Schimmrigk:** Formal analysis (supporting); **Ulrike Maaß:** Conceptualization (equal), Data curation (equal), Formal analysis (equal), Funding acquisition (equal), Investigation (equal), Methodology (equal), Project administration (equal), Supervision (lead), Writing - review & editing (equal).

Financial support. U.M. and F.K. received funding for this study from a program of the University of Potsdam for innovative teaching projects (grant number: 2021-03-15).

Competing interests. The authors declare none.

Ethical standards. Ethical approval was obtained from the University Ethical Committee (reference number 60/2021). The study conformed to the Declaration of Helsinki, and informed consent was obtained from all participants.

Further reading

Bennett-Levy, J., & Thwaites, R. (2007). Self and self-reflection in the therapeutic relationship: a conceptual map and practical strategies for the training, supervision and self-supervision of interpersonal skills. In *The Therapeutic Relationship in the Cognitive Behavioral Psychotherapies* (pp. 255–282). Routledge.

Bennett-Levy, J., Thwaites, R., Chaddock, A., & Davis, M. (2009). Reflective practice in cognitive behavioural therapy: the engine of lifelong learning. In R. Dallos, & J. Stedmon (eds), *Reflective Practice in Psychotherapy and Counselling* (pp. 115–135). New York: Open University Press.

Thwaites, R., Bennett-Levy, J., Davis, M., & Chaddock, A. (2014). Using Self-Practice and Self-Reflection (SP/SR) to Enhance CBT Competence and Metacompetence. *How to Become a More Effective CBT Therapist: Mastering Metacompetence in Clinical Practice* (pp. 239–254). <https://doi.org/10.1002/9781394266487.ch16>

References

- Aafjes-van Doorn, K., Liu, A., & Kamsteeg, C. (2022). Videorecorded treatment sessions for professional development. *Counselling Psychology Review*, 37, 4–20. [10.53841/bpspr.2022.37.1.4](https://doi.org/10.53841/bpspr.2022.37.1.4)
- Ay-Bryson, D. S., Weck, F., & Kühne, F. (2022). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*, 16, 20–27. <https://doi.org/10.1037/tep0000349>
- Bassler, M., Potratz, B., & Krauthauser, H. (1995). Der “Helping Alliance Questionnaire” (HAQ) von Luborsky [The “Helping Alliance Questionnaire” (HAQ) by Luborsky]. *Psychotherapeutics*, 40, 23–32.
- Bennett-Levy, J. (2006). Therapist skills: a cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy*, 34, 57–78. <https://doi.org/10.1017/S1352465805002420>
- Bennett-Levy, J. (2019). Why therapists should walk the talk: the theoretical and empirical case for personal practice in therapist training and professional development. *Journal of Behavior Therapy and Experimental Psychiatry*, 62, 133–145. <https://doi.org/10.1016/j.jbtep.2018.08.004>
- Bennett-Levy, J., Lee, N., Travers, K., Pohlman, S., & Hamernik, E. (2003). Cognitive therapy from the inside: enhancing therapist skills through practising what we preach. *Behavioural and Cognitive Psychotherapy*, 31, 143–158. <https://doi.org/10.1017/S1352465803002029>
- Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009a). Acquiring and refining CBT skills and competencies: which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, 37, 571–583. <https://doi.org/10.1017/S1352465809990270>
- Bennett-Levy, J., & Thwaites, R. (2007). Self and self-reflection in the therapeutic relationship: a conceptual map and practical strategies for the training, supervision and self-supervision of interpersonal skills. In Gilbert, P., & Leahy, R. (eds), *The Therapeutic Relationship in the Cognitive Behavioral Psychotherapies* (pp. 255–282). Routledge.
- Bennett-Levy, J., Thwaites, R., Chaddock, A., & Davis, M. (2009b). Reflective practice in cognitive behavioural therapy: the engine of lifelong learning. In R. Dallos, & J. Stedmon (eds), *Reflective Practice in Psychotherapy and Counselling* (pp. 115–135). New York: Open University Press.
- Bennett-Levy, J., Turner, F., Beaty, T., Smith, M., Paterson, B. & Farmer, S. (2001). The value of self-practice of cognitive therapy techniques and self-reflection in the training of cognitive therapists. *Behavioural and Cognitive Psychotherapy*, 29, 203–220. <https://doi.org/10.1017/S1352465801002077>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101
- Briggie, A. M., Hilsenroth, M. J., Conway, F., Muran, J. C., & Jackson, J. M. (2016). Patient comfort with audio or video recording of their psychotherapy sessions: relation to symptomatology, treatment refusal, duration, and outcome. *Professional Psychology: Research and Practice*, 47, 66. <http://doi.org/10.1037/a0040063>
- Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy*, 36, 581–587. <https://doi.org/10.1017/S1352465808004438>
- Butler, A. C., Black-Maier, A. C., Raley, N. D., & Marsh, E. J. (2017). Retrieving and applying knowledge to different examples promotes transfer of learning. *Journal of Experimental Psychology: Applied*, 23, 433–446. <https://doi.org/10.1037/xap0000142>
- Chaddock, A., Thwaites, R., Bennett-Levy, J., & Freeston, M. H. (2014). Understanding individual differences in response to self-practice and self-reflection (SP/SR) during CBT training. *the Cognitive Behaviour Therapist*, 7, e14. <https://doi.org/10.1017/S1754470X14000142>
- Chigwedere, C., Bennett-Levy, J., Fitzmaurice, B., & Donohoe, G. (2020). Personal practice in counselling and CBT trainees: the self-perceived impact of personal therapy and self-practice/self-reflection on personal and professional development. *Cognitive Behaviour Therapy*, 1–17. <https://doi.org/10.1080/16506073.2020.1846608>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edn). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Collard, J. (2024). Use of self-practice/self-reflection (SP/SR) exercises for competency-based training and assessment in CBT. *the Cognitive Behaviour Therapist*, 17, e1. doi: [10.1017/S1754470X23000375](https://doi.org/10.1017/S1754470X23000375)
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyzes using G*Power 3.1: tests for correlation and regression analyzes. *Behavior Research Methods*, 41, 1149–1160.
- Friedrich, S., Konietzschke, F., & Pauly, M. (2023). MANOVA.RM: Resampling-Based Analysis of Multivariate Data and Repeated Measures Designs. <https://cran.r-project.org/web/packages/MANOVA.RM/index.html>
- Friedrich, S., & Pauly, M. (2018). MATS: inference for potentially singular and heteroscedastic MANOVA. *Journal of Multivariate Analysis*, 165, 166–179. <https://doi.org/10.1016/j.jmva.2017.12.008>
- Gale, C., & Schröder, T. (2014). Experiences of self-practice/self-reflection in cognitive behavioural therapy: a meta-synthesis of qualitative studies. *Psychology and Psychotherapy: Theory, Research and Practice*, 87, 373–392. <https://doi.org/10.1111/pa.pt.12026>
- Gonsalvez, C. J., Brockman, R., & Hill, H. R. M. (2016). Video feedback in CBT supervision: Review and illustration of two specific techniques. *the Cognitive Behaviour Therapist*, 9, e24. <https://doi.org/10.1017/S1754470X1500029X>

- Griffith, B. A., & Frieden, G. (2000). Facilitating reflective thinking in counselor education. *Counselor Education and Supervision*, 40, 82–93. <https://doi.org/10.1002/j.1556-6978.2000.tb01240.x>
- Haggerty, G., & Hilsenroth, M. J. (2011). The use of video in psychotherapy supervision. *British Journal of Psychotherapy*, 27, 193–210. <https://doi.org/10.1111/j.1752-0118.2011.01232.x>
- Hahn, D., Weck, F., Withthöft, M., & Kühne, F. (2021). Assessment of counseling self-efficacy: validation of the German Counselor Activity Self-Efficacy Scales-Revised. *Frontiers in Psychology*, 12, 780088. <https://doi.org/10.3389/fpsyg.2021.780088>
- Henrich, D., Glombiewski, J. A., & Scholten, S. (2023). Systematic review of training in cognitive-behavioral therapy: summarizing effects, costs and techniques. *Clinical Psychology Review*, 101, 102266. <https://doi.org/10.1016/j.cpr.2023.102266>
- Hill, C. E., & Knox, S. (2023). Psychotherapy training and supervision with undergraduate and graduate students. In L. G. Castonguay & C. E. Hill (eds), *Becoming better psychotherapists: Advancing training and supervision* (pp. 31–53). American Psychological Association. <https://doi.org/10.1037/0000364-003>
- Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan, D. M., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist*, 45, 7–53. <https://doi.org/10.1177/0011000016641192>
- Jennings, L., & Skovholt, T. M. (1999). The cognitive, emotional, and relational characteristics of master therapists. *Journal of Counseling Psychology*, 46, 3.
- Knox, S., & Hill, C. E. (2021). Training and supervision in psychotherapy: what we know and where we need to go. In Barkham, M., Lutz, W., & Castonguay, L. G. (eds), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change* (pp. 327–349). Hoboken: John Wiley & Sons.
- Kuckartz, U., & Rädiker, S. (2019). *Analyzing Qualitative Data with MAXQDA: Text, Audio, and Video*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-15671-8>
- Kühne, F., Heinze, P. E., Maaß, U., & Weck, F. (2022). Modeling in psychotherapy training – a randomized controlled proof-of-concept trial. *Journal of Consulting and Clinical Psychology*, 90, 950–956. <https://doi.org/10.1037/ccp0000780>
- Lane, J. L., & Gottlieb, R. P. (2004). Improving the interviewing and self-assessment skills of medical students: is it time to readopt videotaping as an educational tool? *Ambulatory Pediatrics*, 4, 244–248. <https://doi.org/10.1367/A03-122R1.1>
- Loades, M. E., & Myles, P. J. (2016). Does a therapist's reflective ability predict the accuracy of their self-evaluation of competence in cognitive behavioural therapy? *the Cognitive Behaviour Therapist*, 9, e6.
- Longley, M., Kästner, D., Daubmann, A., Hirschmeier, C., Strauß, B., & Gumz, A. (2023). Prospective psychotherapists' bias and accuracy in assessing their own facilitative interpersonal skills. *Psychotherapy*, 60, 525–535. <https://doi.org/10.1037/pst0000506>
- Luborsky, L. (1984). *Principles of Psychoanalytic Psychotherapy. Manual for Supportive-Expressive Psychotherapy*. New York: Basic Books.
- Maaß, U., Kühne, F., Ay-Bryson, D. S., Heinze, P. E., & Weck, F. (2024). Efficacy of live-supervision regarding skills, anxiety and self-efficacy: a randomized controlled trial. *The Clinical Supervisor*, 43, 1–21. <https://doi.org/10.1080/07325223.2023.2267528>
- Maaß, U., Kühne, F., Heinze, P. E., Ay-Bryson, D. S., & Weck, F. (2022). The concise measurement of clinical communication skills: validation of a short scale. *Frontiers in Psychiatry*, 13, 977324. <https://doi.org/10.3389/fpsyg.2022.977324>
- McGillivray, J., Gurtman, C., Boganin, C., & Sheen, J. (2015). Self-practice and self-reflection in training of psychological interventions and therapist skills development: a qualitative meta-synthesis review. *Australian Psychologist*, 50, 434–444. <https://doi.org/10.1111/ap.12158>
- Parker, Z. J., & Waller, G. (2015). Factors related to psychotherapists' self-assessment when treating anxiety and other disorders. *Behaviour Research and Therapy*, 66, 1–7. <https://doi.org/10.1016/j.brat.2014.12.010>
- Pieterse, A. L., Lee, M., Ritmeester, A., & Collins, N. M. (2013). Towards a model of self-awareness development for counselling and psychotherapy training. *Counselling Psychology Quarterly*, 26, 190–207. <https://doi.org/10.1080/09515070.2013.793451>
- Prasko, J., Mozný, P., Novotný, M., Slepecky, M., & Vyskocilova, J. (2012). Self-reflection in cognitive behavioural therapy and supervision. *Biomedical Papers of the Medical Faculty of Palacky University in Olomouc*, 156, 377–384. <http://doi.org/10.5507/bp.2012.027>
- Prasko, J., Ociskova, M., Abeltina, M., Krone, I., Kantor, K., Vanek, J., Slepecky, M., Minarikova, K., Monzy, P., Piliarova, M., & Bite, I. (2023). The importance of self-experience and self-reflection in training of cognitive behavioral therapy. *Neuroendocrinology Letters*, 44.
- Probst, T., Humer, E., Jesser, A., & Pieh, C. (2022). Attitudes of psychotherapists towards their own performance and the role of the social comparison group: the self-assessment bias in psychodynamic, humanistic, systemic, and behavioral therapists. *Frontiers in Psychology*, 13, 966947. <https://doi.org/10.3389/fpsyg.2022.966947>
- Rakovshik, S. G., & McManus, F. (2010). Establishing evidence-based training in cognitive behavioral therapy: a review of current empirical findings and theoretical guidance. *Clinical Psychology Review*, 30, 496–516. <https://doi.org/10.1016/j.cpr.2010.03.004>
- Richtberg, S., Jakob, M., Höfling, V., & Weck, F. (2016). Assessment of patient interpersonal behavior: development and validation of a rating scale. *Psychotherapy Research*, 26, 106–119. <https://doi.org/10.1080/10503307.2014.947391>

- Rivière, E., Jaffrelot, M., Jouquan, J., & Chiniara, G. (2019). Debriefing for the transfer of learning: the importance of context. *Academic Medicine*, 94, 796–803. [10.1097/ACM.0000000000002612](https://doi.org/10.1097/ACM.0000000000002612)
- Rønnestad, M. H., & Skovholt, T. M. (2003). The journey of the counselor and therapist: research findings and perspectives on professional development. *Journal of Career Development*, 30, 5–44. <https://doi.org/10.1023/A:1025173508081>
- RStudio Team (2015). *RStudio: Integrated development for R* [computer software]. R. RStudio, Inc. <http://www.rstudio.com/>
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Basic Books.
- Scott, J., Yap, K., Bunch, K., Haarhoff, B., Perry, H., & Bennett-Levy, J. (2021). Should personal practice be part of cognitive behaviour therapy training? Results from two self-practice/self-reflection cohort control pilot studies. *Clinical Psychology & Psychotherapy*, 28, 150–158. <https://doi.org/10.1002/cpp.2497>
- Spafford, S., & Haarhoff, B. (2015). What are the conditions needed to facilitate online self-reflection for cognitive behaviour therapy trainees? *Australian Psychologist*, 50, 232–240. [10.1111/ap.12076](https://doi.org/10.1111/ap.12076)
- Ternovski, J., & Coppock, A. (2017). *RANDOMIZR: Stata module to implement random assignment procedures* [computer software]. Boston College Department of Economics. <https://EconPapers.repec.org/RePEc:boc:bocode:s458403>
- Thwaites, R., Bennett-Levy, J., Davis, M., & Chaddock, A. (2014). Using self-practice and self-reflection (SP/SR) to enhance CBT competence and metacompetence. *How to Become a More Effective CBT Therapist: Mastering Metacompetence in Clinical Practice*, pp. 239–254. <https://doi.org/10.1002/9781394266487.ch16>
- Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: an elusive goal? *American Psychologist*, 69, 218–229. <https://doi.org/10.1037/a0035099>
- VERBI Software (2019). *MAXQDA 2020* [computer software]. Berlin, Germany. Available from: maxqda.com
- Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports*, 110, 639–644. <https://doi.org/10.2466/02.07.17.PR0.110.2.639-64>
- Weck, F., Hautzinger, M., Heidenreich, T., & Stangier, U. (2010). Erfassung psychotherapeutischer Kompetenzen: Validierung einer deutschsprachigen Version der Cognitive Therapy Scale. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 39, 244–250. <https://doi.org/10.1026/1616-3443/a000055>
- Weck, F., Maaß, U., Paunov, T., Heinze, P. E., & Kühne, F. (2024). Clinical supervision based on video vs. verbal report: a randomized controlled trial. *Cognitive Behaviour Therapy*, 1–19. <https://doi.org/10.1080/16506073.2024.2434016>
- Young, J., & Beck, A. T. (1980). *Cognitive Therapy Scale. Rating Manual* [unpublished manuscript]. Center for Cognitive Therapy.
- Yourman, D. B., & Farber, B. A. (1996). *Nondisclosure and Distortion in Psychotherapy Supervision*.
- Zoom (2021). Version 5.8.4 (1736) [software]. <https://zoom.us/de>