Nordic Psychology, 2025
Vol. 77, No. 1, 39–52, https://doi.org/10.1080/19012276.2023.2247572 RESEARCH ARTICLE

Does deliberate practice surpass didactic training in learning empathy skills? – A randomized controlled study

JOHANNES LARSSON[1], DAVID WERTHÉN[1], JAN CARLSSON[1], OSAME SALIM[1], EDVIN DAVIDSSON[1], ALEXANDRE VAZ[2], DANIEL SOUSA[2] & JOAKIM NORBERG[1]

Corresponding address: Johannes Larsson School of Behavioral, Social and Legal Sciences, Örebro University, Örebro, Fakultetsgatan 1, SE 701 82, Sweden. Email: johannes.larsson@oru.se

Abstract

A large body of research identifies therapist expressed empathy as one of the most important predictors of psychotherapy outcome. Deliberate practice (DP) is an effective method to improve skills in many fields. We asked if DP also can be used to enhance the skill of expressing empathy. *Objective:* The aim was to compare the efficacy of DP to didactical learning methods (DLM) in enhancing the skill of empathic expression in students. *Method*: A repeated measures randomized controlled group design was used. Novice students ($N = 36$) from psychologist-, medicine-, social work-, and nursing programs received two training sessions of either DP ($n = 21$) or DLM ($n = 15$). Participants' skills in empathic expression were assessed with the Measure of Expressed Empathy (MEE) on three occasions: before the first, between, and after the last training session. *Results:* Participants in the DP-group showed improved empathic expression, whereas participants in the DLM-group did not. *Conclusions:* These findings suggest that DP is an effective training method for therapeutic skills such as empathic expression and holds implications for the future development of educational practices to incorporate active skill training methods.

KEYWORDS: Deliberate practice, empathy, randomized controlled trial, repeated measures, therapeutic skills

A large body of research identifies empathy as one of the most important interpersonal skills in psychotherapy (Elliott et al., 2011; Norcross, 2010; Wampold, 2015) and as much as 9% of the variance in client outcomes may be explained by the therapist's ability to express empathy (Elliott et al., 2011, 2018). However, there is considerable heterogeneity between therapists and their level of empathic skills (Elliott et al., 2011, 2018; Norcross, 2010; Wampold, 2015). Some therapists are more empathic and consequently have been found to form better therapeutic alliances (Anderson et al., 2016) and serve a lower risk of both client relapse and drop-out compared to their less empathic peers (Moyers & Miller, 2013).

[1]School of Behavioral, Social and Legal Sciences, Örebro University, Örebro, Sweden;
[2]Clinical Psychology Department, ISPA University Institute, Lisboa, Portugal

Deliberate practice

The adverse impacts of low levels of empathic skills together with the fact that empathy could be considered as a cornerstone in the therapeutic process highlights the need for developing effective training methods for improving therapists' empathic skills. Today's practice of interpersonal skills mainly take place within the framework of supervision, personal therapy, and didactics (a passive learning format with activities such as seminars, workshops and readings; Rousmaniere et al., 2017). All of which there have been a paucity of research confirming the effectiveness (Geller et al., 2005; Rønnestad & Orlinsky, 2005; Rousmaniere et al., 2016). Of these learning methods didactics is the most practiced and is included in continuing education and educational programs all around the world (Rousmaniere et al., 2017). However, in the field of medicine, didactics have been found to only have minor effects on on physicians' practice (Bloom, 2005) and although few studies have evaluated the effects of didactics on psychotherapy outcomes, the same result may hold there (Neimeyer et al., 2009).

In contrast, Goldberg et al. (2016), found that therapists who engaged in training principles with ongoing outcome monitoring, consultation, feedback processes and purposeful training, increased their interpersonal skills and client outcomes with an effect size of $d = .035$ per year, a seemingly small effect but promising when accrued over time. These principles have been summarized by Ericsson and colleagues under the name of deliberate practice (DP; Ericsson et al., 1993). For an ideal learning situation Ericsson et al. (1993) argues that the following conditions need to be met: first, the trainee needs to be motivated to engage in activities with the aim of improving performance. Second, the learning task should be designed based on the trainee's current level of performance, so that the trainee after a brief instruction can comprehend the task and proceed with subsequent training. In other words, learning of new skills happens when practice is on a level just above onés present capacity but still within reach with the help of guidance, which requires intense cognitive effort and sustained concentration (Chow et al., 2015; Ericsson et al., 1993; Ericsson & Lehmann, 1996). Third, when the trainee has completed the learning task, instant feedback on the performance should be provided. The feedback is commonly delivered by a teacher or a coach. Last, the trainee should commit to successive refinement by extended, repetitious practice of the learning task at hand or another task at a similar level of difficulty.

The use of DP has been found to be important in several domains, such as music (Ericsson et al., 1993), chess (Gobet & Charness, 2006), sports (Côté et al., 2005), and business (Sonnentag & Kleine, 2000). Likewise, in the field of medicine and surgery, DP has proved to facilitate development towards enhanced performance (Hashimoto et al., 2015; Norman et al., 2006), and improved patient outcomes (McGaghie & Kristopaitis, 2015). However, the complexity of psychological treatment and psychotherapy has left some questions regarding the utility and effectiveness of DP in that specific field (Clements-Hickman & Reese, 2020). Still, a growing body of studies suggest the use of DP as an effective method for improving therapists' results (e.g. Anderson et al., 2020; Goldberg et al., 2016; Hill et al., 2020; Perlman et al., 2020; Westra et al., 2021).

Deliberate practice for developing empathic expression

Studies have shown that DP can be used as an addition to didactics to improve therapists' ability to express empathy. Westra et al. (2021) studied how DP could enhance therapists'

skills in Motivational Interviewing (MI) and found that therapists who engaged in DP were rated higher by observers in relevant skills, including empathic expression, compared to therapists who engaged in a traditional workshop based solely on didactics. Similarly, Li et al. (2019) found that nurses who engaged in a combination of DP and didactics improved their self-rated empathic expression more than those who received didactics only.

These results have also been shown in therapy trainees. Anderson et al. (2020) found that trainees who engaged in DP with video clips relevant to therapy improved more on observer ratings of facilitative interpersonal skills (FIS), including empathic expression, than students who studied video clips unrelated to therapy and with less opportunity to practice. In a similar line, Perlman et al. (2020) demonstrated that trainees who practiced DP in combination with didactics reached higher improved observer rated therapeutic skills than trainees who engaged in didactics only.

The present study

Although some studies (Li et al., 2019; Perlman et al., 2020; Westra et al., 2021) support DP as an effective method for improving therapeutic skills, none of these studies have evaluated the effects of DP exclusively without the influence of other learning sources. The aim of the current study was to compare the isolated effects of DP to didactical learning methods (DLM), based on lectures, group discussions and educative video recordings, in the acquisition of empathic skills.

Following Anderson et al. (2020) we decided to use observer ratings to assess empathic skills (see methods for details). Although client ratings of therapist empathy may be a better predictor of client outcomes, it is closely followed by observer ratings (Barrett-Lennard,1981; Elliott et al., 2011). Moreover, observer ratings—which are usually performed by other therapists—commonly capture a more nuanced picture of therapist empathy in comparison to client ratings that are more global (Bachelor, 1988; Decker et al., 2014).

Method

Participants

Following Anderson et al. (2020), we included novice trainees as participants who had not yet started clinical training in their respective program. They were all registered at a master's or bachelor's program in psychology, social work, nursing, or medicine (i.e. the four most common professions for someone who later in their professional career would apply for becoming a psychotherapist in Sweden; Socialstyrelsen, 2017). Participants were excluded if they had started any training focusing on enhancing therapeutic skills, so that other types of ongoing clinical training could not confound the results. Participants also had to be fluent in English. See Figure 1 for flow of participants at each stage of the trial.

A total of 72 students showed interest in the study and were randomized to either Deliberate Practice (DP) or Didactical Learning Methods (DLM). They were then invited to an introductory meeting (separately for each group), which a total of 43 participants

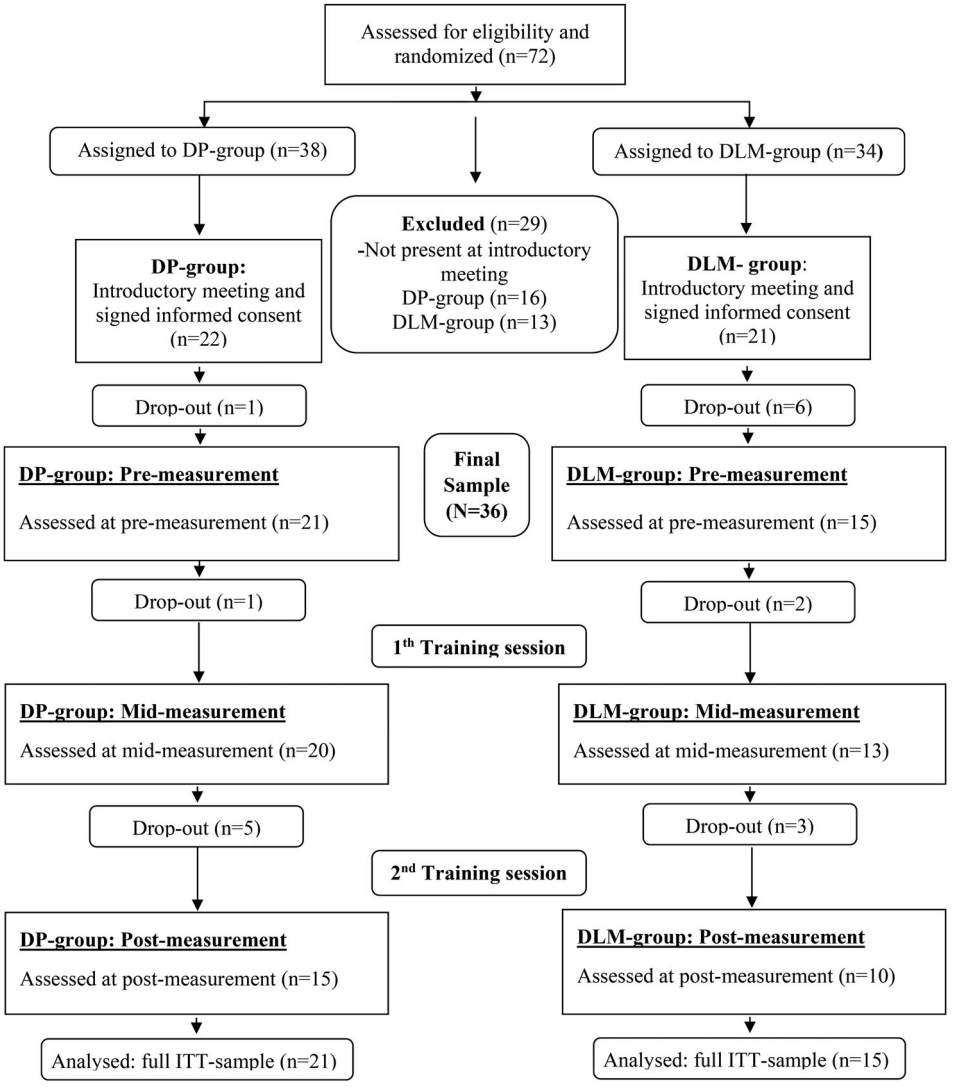

Figure 1. Flow chart.
Note: Flow of participants through the trial. DP = deliberate practice. DLM = didactical learning methods. ITT = intention-to-treat.

attended (22 = DP, 21 = DLM). After the introductory meeting, seven participants (1 = DP, 6 = DLM) declined further participation in the study, all due to scheduling conflicts. This resulted in a final sample of 36 participants (DP; $n = 21$, DLM; $n = 15$). In agreement with the intention-to-treat (ITT) principle (Kazdin, 2014), we included all 36 participants who performed the pre-measurement in the analyses. 11 participants interrupted their participation (DP; $n = 6$, DLM = 5) between pre- and post-measurement, also all due to scheduling conflicts. For demographic characteristics of the sample see Table 1.

Table 1. Demographic characteristics of the sample.

Characteristics	DP-group ($n = 21$)		DLM-group ($n = 15$)		Total ($N = 36$)	
Age: m (SD)	24.7	(4.5)	26.0	(5.1)	25.2	(4.7)
Gender: *n* (%)						
Male	11	(52.4)	3	(20.0)	14	(38.9)
Female	10	(47.6)	12	(80.0)	22	(61.1)
Programme: *n* (%)						
Psychology	10	(47.6)	8	(53.3)	18	(50.0)
Social work	6	(28.6)	7	(46.7)	13	(36.1)
Nursing	3	(14.3)	–	(–)	3	(8.3)
Medicine	2	(9.5)	–	(–)	2	(5.6)

No significant differences on any demographic characteristics were found between the DP- and DLM-groups.
Note: SD = Standard deviation. N = Participants.

Procedure

Prior to the enrolment of the study, participants from each group were called for two separate introductory meetings. During the introductory meetings both groups received a demonstration of the digital training platform that were to be used during the subsequent training (only DP-group) and assessment (both DP- and DLM-group), as well as an oral presentation of the concept of empathy and its relevance for psychotherapy outcomes (Elliott et al., 2011, 2018; Norcross, 2010; Wampold, 2015). The participants in the DLM-group received a brief overview of the content of the two upcoming didactical learning sessions while the participants in the DP-group were introduced to the concept of deliberate practice and allocated to subgroups, each with three to four participants.

For assessments of both groups, and training in the DP-group, we used the digital training platform Theravue (https://www.theravue.com). Six videos of English-speaking actors depicting patients in therapy, were chosen from the Theravue database. For each assessment, participants logged on to the platform and recorded video responses of them responding empathically to two 2–3-min videos portraying clients in challenging therapy situations (e.g. a depressed client feeling unmotivated). The same videos were used for both groups. They could make as many recordings as they wanted before submitting their responses for review by raters and DP-instructors.

Interventions

Both DP- and DLM-interventions (training sessions) started one week after the introductory meetings and lasted for two weeks and consisted of two sessions each (one session per week) conducted by the first two authors of this study, who were master's level psychology students.

Deliberate practice (DP)

DP-sessions were based on the principles of DP (Ericsson, 1993), including observation of current level of performance, feedback, establishment of learning goals, and rehearsal.

Following Perlman et al. (2020), training was done in subgroups of three to four participants. After a brief introduction about the purpose of practicing empathic expression with DP, each participant was invited to demonstrate one or both of his or her recorded video-responses from the digital platform. The DP-instructors then provided individualized feedback and learning goals for each participant. The learning goals served as a basis for the next step where participants were encouraged to engage in behavioral rehearsal whilst watching the same video stimuli again and rehearsing a new response. The new response was also rehearsed two to three times in role-plays with the other participants in the group to further facilitate learning. Finally, the instructors encouraged the participant to write down the learning goals and use them for practice at home. The procedure was then repeated with the remaining participants. Each DP-session lasted for approximately 60 min.

Didactical learning methods (DLM)

The DLM group was designed to mimic the DP-group (Kazdin, 2014). Therefore, the DLM-group also received two sessions (both lasting for approximately 60 min) over a two-week period with the same instructors as in the DP-group. The DLM-instructors designed the DLM-sessions based on materials such as videos, group discussions, and other sources of information on empathy described below, but without providing opportunities for any behavioral rehearsal.

In the first session the *"The still face experiment"* (UMass Boston, 2007) was presented, followed by a discussion of verbal and nonverbal expressions of empathy. Participants were introduced to empathy as an interpersonal process (Barrett-Lennard, 1981) and five general principles of becoming empathic as a therapist from the book *Emotion focused therapy: the process-experiential approach to change* (Elliott et al., 2004). The second session was based on the ten items of the MEE-rating form (Watson & Prosser, 1999). Examples of poor empathic skills as expressed by verbal and nonverbal behavior was retrieved from YouTube and shown to the participants. Additionally, a recorded session of Carl Rogers (Hill, 2017) working with a client was shown as a demonstration of a therapist that is considered to have great empathy skills.

Outcome measure

Expressed empathy

Empathic skills were assessed at three times with the Measure of Expressed Empathy (MEE; Watson & Prosser, 1999): before the first training session, between the first and second, and finally after the last. MEE is an observer-rated measure designed to assess therapists' verbal and nonverbal expressions of empathy. The MEE-scale has demonstrated high internal consistency ($\alpha = .88$; Malin, 2016) and correlate significantly with client ratings of therapist empathy ($r = .66$, $p < .01$; Watson & Prosser, 1999). MEE consists of ten items, each corresponding to a specific dimension of expressed empathy. Each item is rated on a 9-point Likert-scale (0 = never, 8 = all the time), representing the extent or frequency of behaviors in that dimension (higher scores indicating higher frequency of empathic behaviors). However, because inter-rater reliability has varied across previous studies (.51 − .85; Elliott et al., 2011) we reduced the response options to a 5-point Likert-scale (1 = never, 5 = all the

time). According to Simms et al. (2019) a scale over 6 responses may be more unclear and difficult to answer and shows no improvements in psychometric precision. In this study, we assessed empathic expression on a global level by summing items on the MEE for a total score (maximum score = 50).

Each participant response was rated independently by two master's level psychology students who were masked to group affiliation (i.e. they were different from the trainers). We used the mean of these two ratings. The raters had ten hours of practice on fictitious cases prior to rating participants' responses. ICC for the pre-, mid-, and post assessment were .67 with a 95% confidence interval from .561 to .753 (F (187, 187) = 3.035, $p < .001$) implying a moderate degree of reliability (Koo & Li, 2016).

Statistical analyses

Statistical analyses were conducted in Statistical Package of Social Sciences (SPSS) version 26 (IBM Corp, 2019) with an alpha level of .05 for all statistical tests. Prior to the primary analysis, pre-intervention differences between groups and missing data were investigated with independent t-tests and chi-square tests. For the primary analyses, mixed ANOVAs were used to assess mean differences between and within the two groups over time. Furthermore, paired sample t-tests were conducted for within-group calculations of effect sizes for each point of measurement. P-values were adjusted by using the Bonferroni correction to account for multiple comparisons.

Pre-intervention differences and missing data analysis

Independent t-tests showed that there were no significant differences on age, program, or educational level between the groups. Regarding gender, there was a statistically significant difference between the groups (t(32.39) = 2.10, p = .04). Specifically, the DLM group had a higher percentage of female participants (80.00%) compared to the DP-group (47.60%). Furthermore, a statistically significant difference between the groups was found at pre-measurement MEE scores (t(32.65) = 2.58, p = .02), the DLM group achieved a higher mean level of MEE scores (M = 30.65, SD = 2.46) compared to the DP-group (M = 27.71, SD = 4.33). Results from independent t-tests showed that there was a significant effect for gender at pre-measurement (t(34) = 2.50, p = .02), females attained higher scores (M = 30.15, SD = 3.33) than males (M = 27.04, SD = 4.11). Given the observed differences, gender was included as a factor in a separate analysis. The analysis on missing data showed no statistically significant differences on any demographics (gender, age, program, semester, and pre-measurement of observer rated empathy) between missing and complete cases. To account for the missing data, the single imputation model of mean substitution was used (Kazdin, 2014).

Ethics

The study was approved by the Swedish Ethical Review Authority (2020-00609). All participants were informed about the study and signed informed consent prior to participation. Further, participants were informed that they could choose to suspend participation at any time without stating cause.

Results

To assess mean changes over time within and between the DP- and DLM-group, a 2×3 mixed ANOVA was performed, with group (DP, DLM) and time (pre-, mid-, post measurement) as factors. The results showed a significant main effect of time ($F(2, 68) = 15.12$, $p < .001$, $\eta p^2 = .308$) but not group ($F(1, 34) = 0.01$, $p = .93$, $\eta p^2 < .001$) on participants' MEE scores.

Importantly, there was a statistically significant interaction between time and group ($F(2, 68) = 10.02$, $p < .001$, $\eta p^2 = .228$), indicating that participants' MEE scores differed between groups over time. Simple main effects analyses showed that participants from the DP group had a significant increase in MEE scores over time ($F(2, 40) = 24.15$, $p < .001$, $\eta p^2 = .548$), but participants in the DLM group did not ($F(2, 28) = 1.67$, $p = .21$, $\eta p^2 = .106$).

When comparing the groups at each time point (pre-, mid- and post measurement), the results showed that the DLM-group had statistically significantly higher scores on MEE compared to the DP-group at pre-measurement ($F(1, 34) = 5{,}58$, $p = .02$, $\eta p^2 = .141$). During mid-measurement, the DP-group surpassed the DLM-group, performing significantly better than the DLM-group ($F(1, 34) = 4.62$, $p = .04$, $\eta p^2 = .120$). At post-measurement there were no significant differences between the DLM-group and the DP-group (See Figure 2 for *M* and *SD)*. In other words, participants in the DP-group went from having significantly lower MEE scores to significantly higher compared to the DLM-group. Yet, at post-measurement both groups ended with similar scores.

The mean differences between pre- and mid measurement for the DP-group were significant ($t(20) = 5.46$, $p < .001$, $\eta p^2 = 0.278$) and the within-group effect size was large, indicating that participants in the DP-group made significant increases in empathic expression after the first DP-session. Mean differences between mid- and post measurement were not statistically significant for the DP-group implying that the rapid increase in MEE-scores stalled after the second DP-session. The overall effect between pre- and post-measurement

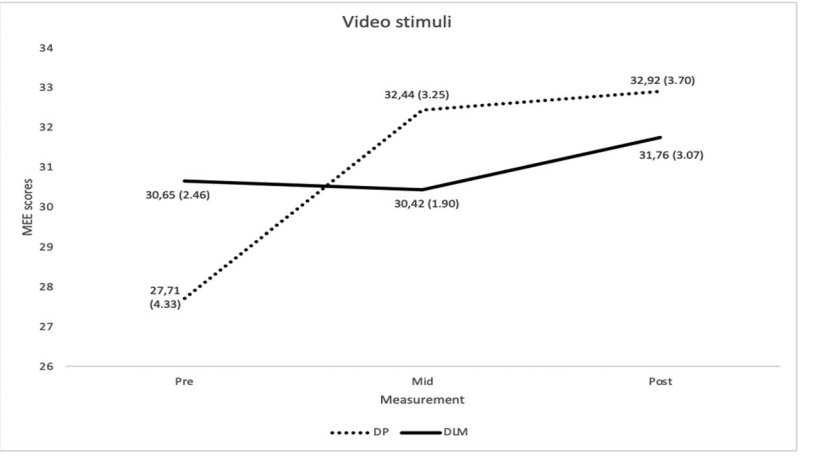

Figure 2. Repeated measures of MEE-scores for each group (DP and DLM).
Note: The vertical axis displays mean values (M) of MEE scores with Standard deviations (SD) in brackets. The horizontal axis represents the three points of measurement (pre-, mid- and post). Dashed lines denote the DP-group and solid-colored lines the DLM-group.

was large ($t(20) = 6.30$, $p < .001$, $\eta p^2 = 0.294$), indicating that participants in the DP-group made notable improvements in their abilities to express empathy altogether. No statistically significant changes were observed for the DLM-group in any of the comparisons. Which suggests that participants in the DLM-group did not significantly improve their abilities in empathic expression during the intervention.

A 2x3x2 mixed ANOVA was conducted, to investigate whether gender would influence the association between time and groups on MEE scores. The analysis included the factors of group (DP vs DLM), time (pre, mid, post measurement) and gender (female and male). The results of the 2x3x2 mixed ANOVA analysis showed that there was no statistically significant interaction between group, time and gender ($F(2, 64) = 0.66$, $p = .52$, $\eta p^2 < .020$) indicating that males and females reached similar results over the course of the interventions (Figure 2).

Discussion

The aim of the current study was to compare the effects of deliberate practice (DP) to didactical learning methods (DLM) in the acquisition of empathic skills. The results showed that there was a significant increase in MEE-scores over time within the DP-group, but not in the DLM-group. However, the results should be interpreted with caution due to the small sample size, high variability in MEE-scores between groups at pre-measurement and the relatively short intervention period.

The capacity to express empathy increased significantly more within the DP-group compared to the DLM-group. This finding is consistent with results from previous research were the incorporation of behavioral components such as modelling, practice and feedback (similar to the behavioral principles of DP; Ericsson, 1993) have been found to enhance the effect of training on empathy (Li et al., 2019; Teding van Berkhout & Malouff, 2016) and helping skills (Hill & Lent, 2006). As in Perlman et al. (2020), effect sizes for the DP group were large between pre and mid measurement, while there was no effect for didactics between any of the three measurements. The lack of improvement seen in the DLM-group may reflect the difficulties of establishing a link between passive didactic learning methods and improved client outcomes (Bloom, 2005; Neimeyer et al., 2009; Rønnestad & Orlinsky, 2005; Rousmaniere et al., 2016). The present finding, that active skills training (DP) may be a more effective method for acquiring empathic skills compared to DLM, is also in line with Li et al. (2019), were the combination of traditional teachings and DP generated significantly better results than traditional teachings exclusively. The results thus support the notion that didactic learning alone may be insufficient for improving clinical skills including empathy (e.g. Perlman et al., 2020).

At a closer examination of the results, we can see that the DLM-group performed significantly better than the DP-group at pre-measurement but did not improve over time. In contrast, the DP-group improved rapidly between pre- and mid-measurement (after the first DP-session). According to Ericsson (2004) it is not uncommon to make quick progress on a new area even with minimal training, which could explain the accelerated progress in the DP-group after only receiving one DP-session. Between mid- and post measurement (after the second DP-session), however, no sign of further improvement was observed. A possible explanation for the decline in progress could be that the DP-group reached a

plateau where more effort would be required for further improvement. For example, in DP it has been hypothesized that after the initial phase of learning a new skill most reach a plateau after around 50 h of training, which has been repeatedly demonstrated in fields like sports and music (Ericsson et al., 2009). To overcome such a plateau, it is recommended to spend a lot of time and effort on adequate training (Chow et al., 2015; Ericsson, 2006). As such, more training may have been needed to see further improvements within the DP-group.

To conclude, the results of the present study suggest that DP is more efficient than DLM for the acquisition of empathic skills. Nonetheless, several limitations should be borne in mind.

Limitation, strengths and future research

Firstly, gender and pre-measurement MEE-scores were not evenly distributed between the groups, more specifically the DLM-group had more females (80%) than the DP-group (52%). The inequality in gender distribution might be a plausible explanation to the initial differences in MEE-scores since research suggests that females, on average, are more empathic than males (Rueckert & Naybar, 2008). Yet at post-intervention, the gender differences were no longer significant indicating that with training, males and females reached a similar level of performance. This is in line with previous research showing that training can reduce potential initial gender differences in empathy (Ickes, 1997).

Secondly, missing values constituted a substantial portion of the sample (12.96%) and since the sample size was relatively small missing data was substituted through mean substitution to avoid reduction of power (Kang, 2013), a method found to reflect the data quite well if the missing values do not exceed 20% of the sample (Downey & King, 1998). When comparing the results with and without substituted mean scores, the results were still the same and not distorted. Most of the missing data were due to drop-out before post measurement. A possible explanation could be a potential diminished motivation attributable to the busy lives of students and the absence of subsequent incentives from instructors. Efforts were made to avoid non-responding as much as possible by mailings and reminders. Future research might further aim at overcoming non-responding by offering incentives, which was not possible in the current study due to regulations in university guidelines.

Thirdly, it is possible that the instructors' expectancies may have caused differences across the two intervention groups (Kazdin, 2014). To prevent this from happening we made sure that instructors put in the same effort and spent the same amount of time preparing for both interventions. Further, to avoid diffusion effects across the treatment groups we carefully separated the groups by having separate introductory meetings and encouraging non-disclosure of information. However, future studies would benefit from including one instructor per group with equal expertise in their respective fields.

Fourthly, while both groups started on a different level in measured empathic expression both the DLM-group and the DP-group ended with similar MEE-scores, indicating equal performances in expression of empathy between the groups at post-measurement. Clients would clearly see both groups as being equally empathic after the training. However, the result showed a significant difference for time suggesting that empathy can in fact be

taught, or more specifically empathic expressing can be increased. Future research should, if possible, match the groups initial level of measured empathic expression to clearer distinguish the effects of DLM and DP.

Lastly, even though an increasing amount of research points to the benefits of deliberate practice in psychotherapy there are several barriers to the implementation of DP (Clements-Hickman & Reese, 2020). Creating a supportive environment with coaches and supervisors willing to give feedback requires a great deal of time and resources, which is often a scarce commodity for professional psychotherapists. Also, engagement in DP involves sustained effort in challenging tasks over a longer period and training routines needs to be implemented in everyday practice for DP to be efficient. In everyday settings therapists are vulnerable to emotional exhaustion (Malinowski, 2013) and the time-consuming demands of DP can seem insuperable. Furthermore, a self-critical attitude, often referred to as professional self-doubt (Clements-Hickman & Reese, 2020) is embedded in the framework of DP. Professional self-doubt can however easily evoke defensive attitudes resulting in behaviors such as blaming the client or withholding information in supervision instead of reflecting on performance (Clements-Hickman & Reese, 2020). Moreover, continuously monitoring outcomes and evaluating interventions to identify learning tasks may cause fear and anxiety instead of inspiration and motivation if not implemented carefully in the daily routines. Nevertheless, professional self-doubt seems to be linked to positive therapy outcome when it stimulates reflection on the professional performance rather than triggering inhibitory self-blame (Nissen-Lie et al., 2017).

Notwithstanding these limitations, this study also has several strengths. Firstly, the randomized design with three repeated measures and two independent raters (who were masked to group affiliation) lends support for a causal interpretation of the results, namely that increased levels of empathic expression was due to DP-training. Furthermore, considering the limited time span, a significant effect of DP on empathic expression could be seen as promising and future studies should conduct training over a longer period to preclude novelty effects and do follow-up assessments to investigate any lasting long-term effects. To address motivational differences in the often tedious and demanding task of engaging in DP (Chow et al., 2015), future studies would benefit from including measures on motivation and professional self-doubt, two factors known to positively influence successful implementation of DP (Chow et al., 2015; Clements-Hickman & Reese, 2020). Additional measures on empathy, such as self- and client-ratings of empathic expression should preferably also be added to account for the multidimensionality of empathy as a construct. In comparing different training methods for acquiring the skill of empathic expression future studies should extend their scope by comparing DP to other types of training currently used in clinical training programs, such as regular supervision.

Implications of the results

The result of the present study supports previous findings in that the skill to express empathy can be improved with training (Teding van Berkhout & Malouff, 2016) and adds to the growing body of research suggesting that active skills training as DP may be more efficient than didactical learning methods for improving therapeutic skills (Li et al., 2019; Perlman et al., 2020; Westra et al., 2021). Even if more research is needed, this holds

implications for the future development of educational practices for psychotherapists to incorporate DP as a training method for therapeutic skills such as empathic expression.

Disclosure statement

No potential conflict of interest was reported by the author(s).

REFERENCES

Anderson, T., Crowley, M. E. J., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2016). Therapist facilitative interpersonal skills and training status: A randomized clinical trial on alliance and outcome. *Psychotherapy Research*, *26*(5), 511–529. https://doi.org/10.1080/10503307.2015.1049671

Anderson, T., Perlman, M. R., McCarrick, S. M., & McClintock, A. S. (2020). Modeling and deliberate practice of interpersonal skills improves facilitative interpersonal. *Journal of Clinical Psychology*, *76*(4), 659–675. https://doi.org/10.1002/jclp.22911

Bachelor, A. (1988). How clients perceive therapist empathy: A content analysis of "received" empathy. *Psychotherapy*, *25*(2), 227–240. https://doi.org/10.1037/h0085337

Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, *28*(2), 91–100. https://doi.org/10.1037/0022-0167.28.2.91

Bloom, B. S. (2005). Effects of continuing medical education on improving physician clinical care and patient health: A review of systematic reviews. *International Journal of Technology Assessment in Health Care*, *21*(3), 380–385. https://doi.org/10.1017/s026646230505049x

Chow, D., Miller, S., Seidel, J., Kane, R., Thornton, J., & Andrews, W. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy*, *52*(3), 337–345. https://doi.org/10.1037/pst0000015

Clements-Hickman, A. L., & Reese, R. J. (2020). Improving therapists' effectiveness: Can deliberate practice help? *Professional Psychology*, *51*(6), 606–612. https://doi.org/10.1037/pro0000318

Côté, J., Ericsson, K., & Law, M. (2005). Tracing the development of athletes using retrospective interview methods: A proposed interview and validation procedure for reported information. *Journal of Applied Sport Psychology*, *17*(1), 1–19. https://doi.org/10.1080/10413200590907531

Decker, S., Nich, C., Carroll, K., & Martino, S. (2014). Development of the therapist empathy scale. *Behavioural and Cognitive Psychotherapy*, *42*(3), 339–354. https://doi.org/10.1017/S1352465813000039

Downey, R. G., & King, C. V. (1998). Missing data in likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*(2), 175–191. https://doi.org/10.1080/00221309809595542

Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Evidence-based responsiveness* (2nd ed., pp. 132–152). Oxford University Press. 10.1093/acprof:

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. *Psychotherapy*, *55*(4), 399–410. https://doi.org/10.1037/pst0000175

Elliott, R., Watson, J. C., Goldman, R. N., & Greenberg, L. S. (2004). *Emotion focused therapy; the process-experiential approach to change*. American Psychological Association.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*(Supplement), S70–S81. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Ericsson, K., & Lehmann, A. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Ericsson, K., Krampe, R., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Ericsson, K., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: Attaining superior performance through deliberate practice. *Annals of the New York Academy of Sciences*, *1172*(1), 199–217. https://doi.org/10.1196/annals.1393.001

Geller, J. D., Norcross, J. C., & Orlinsky, D. E. (2005). *The psychotherapist's own psychotherapy: Patient and clinician perspectives*. *Oxford University Press*.

Gobet, F., & Charness, N. (2006). Expertise in chess. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance*. (pp. 523–538). Cambridge University Press. https://doi.org/10.1017/9781316480748.031

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, *63*(1), 1–11. https://doi.org/10.1037/cou0000131.supp

Hashimoto, D. A., Sirimanna, P., Gomez, E. D., Beyer-Berjot, L., Ericsson, K. A., Williams, N. N., Darzi, A., & Aggarwal, R. (2015). Deliberate practice enhances quality of laparoscopic surgical performance in a randomized controlled trial: From arrested development to expert performance. *Surgical Endoscopy*, *29*(11), 3154–3162. https://doi.org/10.1007/s00464-014-4042-4

Hill, A. (2017, January 26). *Carl Rogers Councils an Individual on Anger*. [Video file]. Retrieved from https://www.youtube.com/watch?v=uRCD3anKsa0&t=1105s

Hill, C. E., Kivlighan, D. M., Rousmaniere, T., Kivlighan, D. M., Jr., Gerstenblith, J. A., & Hillman, J. W. (2020). Deliberate practice for the skill of immediacy: A multiple case study of doctoral student therapists and clients. *Psychotherapy*, *57*(4), 587–597. https://doi.org/10.1037/pst0000247

Hill, C. E., & Lent, R. W. (2006). A narrative and meta-analytic review of helping skills training: Time to revive a dormant area of inquiry. *Psychotherapy*, *43*(2), 154–172. https://doi.org/10.1037/0033-3204.43.2.154

IBM Corp. Released. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. IBM Corp.

Ickes, W. J. (Ed.). (1997). *Empathic accuracy*. The Guilford Press.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

Kazdin, A. (2014). *Research design in clinical psychology* (4th ed., Pearson new international ed.) Pearson.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Li, J., Li, X., Gu, L., Zhang, R., Zhao, R., Cai, Q., Li, Y., Wang, H., Meng, Q., & Wei, H. (2019). Effects of simulation-based deliberate practice on nursing students' communication, empathy, and self-efficacy. *The Journal of Nursing Education*, *58*(12), 681–689. doi: https://doi.org/10.3928/01484834-20191120-02

Malin, A. J. (2016). *Therapist expressed empathy across experiential treatment for depression: Its growth and relationship to other psychotherapy processes* [Doctoral thesis]. York University. Retrieved from https://yorkspace.library.yorku.ca/xmlui/bitstream/handle/10315/32705/Malin_Ashley_J_2016_PhD.pdf?sequence=2&isAllowed=y.

Malinowski, A. (2013). Characteristics of job burnout and humor among psychotherapists. *Humor*, *26*(1), 117–134. https://doi.org/10.1515/humor-2013-0007

McGaghie, W. C., & Kristopaitis, T. (2015). Deliberate practice and mastery learning: Origins of expert medical performance. In J. Cleland and S. J. Durning (Eds.), *Researching Medical Education*. (eds). Wiley Online Library. https://doi.org/10.1002/9781118838983.ch19

Moyers, T. B., & Miller, W. R. (2013). Is low therapist empathy toxic? *Psychology of Addictive Behaviors*, *27*(3), 878–884. https://doi.org/10.1037/a0030274

Neimeyer, G. J., Taylor, J. M., & Wear, D. W. (2009). Continuing education in psychology: Outcomes, evaluations and mandates. *Professional Psychology*, *40*(6), 617–624. https://doi.org/10.1037/a0016655

Nissen-Lie, H. A., Rønnestad, M. H., Høglend, P. A., Havik, O. E., Solbakken, O. S., Stiles, T. C., & Monsen, J. T. (2017). Love yourself as a person, doubt yourself as a therapist? *Clinical Psychology & Psychotherapy*, *24*(1), 48–60. 10.1002/cpp.1977

Norcross, J. C. (2010). The therapeutic relationship. In B. L. Duncan, S. D., Miller, B. E., Wampold, & M. A., Hubble (Eds.), *The heart and soul of change: Delivering what works in therapy* (pp. 113–142). American Psychological Association. https://doi.org/10.1037/12075-004

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 339–354) Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.019

Perlman, M., Anderson, T., Foley, V., Mimnaugh, S., & Safran, J. (2020). The impact of alliance-focused and facilitative interpersonal relationship training on therapist skills: An RCT of brief training. *Psychotherapy Research*, *30*(7), 871–884. https://doi.org/10.1080/10503307.2020.1722862

Rousmaniere, T., Swift, J., Babins-Wagner, R., Whipple, J., & Berzins, S. (2016). Supervisor variance in psychotherapy outcome in routine practice. *Psychotherapy Research* , *26*(2), 196–205. https://doi.org/10.1080/10503307.2014.963730

Rousmaniere, T., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (2017). (red.) *The cycle of excellence: Using deliberate practice to improve supervision and training*. Wiley Blackwell. https://doi.org/10.1002/9781119165590

Rueckert, L., & Naybar, N. (2008). Gender differences in empathy: The role of the right hemisphere. *Brain and Cognition*, *67*(2), 162–167. (). (), -. https://doi.org/10.1016/j.bandc.2008.01.002

Rønnestad, M., & Orlinsky, D. (2005). Therapeutic work and professional development: Main findings and practical implications of a long-term international study. *Psychotherapy Bulletin*, *40*, 27–32. https://doi.org/10.1037/e565712006-008

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, *31*(4), 557–566. https://doi.org/10.1037/pas0000648

Socialstyrelsen. (2017). *Statistik om hälso- och sjukvårdspersonal*. [National Board of Health and Welfare. Statistics on health- and hospital staff]. Retrieved August 08, 2020, from https://www.socialstyrelsen.se/statistik-och-data/statistik/statistikamnen/halso-och-sjukvardspersonal/

Sonnentag, S., & Kleine, B. (2000). Deliberate practice at work: A study with insurance agents. *Journal of Occupational and Organizational Psychology*, *73*(1), 87–102. https://doi.org/10.1348/096317900166895

Teding van Berkhout, E., & Malouff, J. M. (2016). The efficacy of empathy training: A meta-analysis of randomized controlled trials. *Journal of Counseling Psychology*, *63*(1), 32–41. https://doi.org/10.1037/cou0000093

UMass Boston. (2007). *Still Face Experiment: Dr. Edward Tronick*. [Video file]. Retrieved from: https://www.youtube.com/watch?v=apzXGEbZht0.

Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, *14*(3), 270–277. https://doi.org/10.1002/wps.20238

Watson, J. C., & Prosser, M. C. (1999). *Measure of expressed empathy* (Unpublished Manuscript). University of Toronto.

Westra, H. A., Norouszian, N., Poulin, L., Hara, K., Coyne, A., Constantino, M. J., Hara, K., Olson, D., & Antony, M. M. (2021). Testing a deliberate practice workshop for developing appropriate responsivity to resistance markers. *Psychotherapy*, *58*(2), 175–185. https://doi.org/10.1037/pst0000311

