ISPA

INSTITUTO UNIVERSITÁRIO

CIÊNCIAS PSICOLÓGICAS, SOCIAIS E DA VIDA

The application of a Deliberate Practice training in the improvement of students' and in-training psychotherapists' empathic skills

Orientador de Dissertação:

Professor Doutor Daniel Cunha Monteiro de Sousa

Coordenador de Seminário de Dissertação:

Professor Doutor Daniel Cunha Monteiro de Sousa

Tese submetida como requisito parcial para a obtenção do grau de:

Mestre em psicologia

Especialidade em Psicologia Clínica

2020

Dissertação de Mestrado realizada sob a orientação de Professor Doutor Daniel Cunha Monteiro de Sousa apresentada no ISPA – Instituto Universitário para obtenção de grau de Mestre na especialidade de Psicologia Clínica.

Abstract

Contrary to other professions, research suggests that psychotherapists do not improve their performance with experience. The application of Deliberate Practice into psychotherapeutic training has been suggested as one way to improve psychotherapists' performance. **Objective:** Study 1: Compare the impact of a deliberate practice training and an expositive training (without feedback) on the empathic quality of undergraduates' therapeutic responses to clinical simulation videos. Study 2: Follows the same objective as study 1, controlling for practice and individualized feedback effects, with experienced in-training therapists. **Method**: Study 1: Psychology undergraduates (N = 36) were randomly assigned to the Deliberate Practice or the Training As Usual conditions. They recorded empathic responses to two videos at three different time points, over the course of three weeks (once before having received any training and twice after the DP or TAU interventions). Study 2: This follows the same procedure as Study 1 but the sample consists of experienced in-training therapists (N = 11). To control feedback and practice effects, participants in both conditions received individualized feedback and responded to 6 videos. In both studies, blind raters evaluated the empathic quality of the responses using the Measure of Expressed Empathy (Watson, 1999). **Results:** Study 1: Undergraduate students in the DP condition improved their empathic responses compared to participants under the TAU condition. Additionally, they were the only group that improved over the two interventions. Study 2: When using a sample constituted by more experienced therapists and controlling for both practice and feedback effects, the effects of DP are inconclusive.

Keywords: Deliberate Practice; Randomized Control Trial (RCT); Empathy; Psychotherapist Training; Expertise.

Resumo

Contrariamente a outras profissões, os psicoterapeutas não melhoram a sua performance com a experiência. A aplicação prática deliberada à formação de psicoterapeutas tem sido proposta como uma ferramenta para melhorar a performance dos psicoterapeutas. **Objetivo:** Estudo 1: Comparar o impacto de uma formação de prática deliberada com uma formação teórica (sem feedback individualizado), na qualidade empática de respostas terapêuticas de estudantes de licenciatura. Estudo 2: Segue o mesmo objetivo do estudo 1 mas pretende controlar efeitos de prática e de feedback, em terapeutas com experiência clinica. **Método:** Estudo 1: Estudantes de licenciatura (N = 36) foram alocados, aleatoriamente, à condição de Prática Deliberada ou Formação Teórica e gravaram respostas empáticas a dois vídeos em 3 momentos diferentes (Uma vez antes de qualquer intervenção e duas vezes após a formação de prática deliberada ou teórica). Estudo 2: segue o mesmo procedimento do primeiro, mas a amostra é constituída por terapeutas com experiência clínica (N = 11) que respondem a 6 vídeos diferentes e ambas as condições recebem feedback individualizado. Avaliadores cegos cotaram a qualidade das respostas através da *Measure of Expressed Empathy* (Watson, 1999) em ambos os estudos. **Resultados:** Os participantes na condição de Prática Deliberada melhoraram a qualidade das suas respostas empáticas quando comparados com os participantes na condição de Treino Teórico e foram o único grupo a melhorar as respostas, após duas intervenções. Contudo, ao usar uma amostra constituída por terapeutas mais experientes e controlar os efeitos da prática e do feedback, os resultados tornam-se inconclusivos.

Keywords: Prática Deliberada; Ensaio Clínico Aleatório; Empatia; Formação de psicoterapeutas; Expertise.

List of Tables and Figures

Tables

Study 1

Study 2

Figures

Study 1

Study 2

Index

Introduction

Psychotherapy has been proven to be effective treating several psychological disorders, in all age groups (Cuijpers et al., 2016; Kamenov et al., 2016; Lambert, 2013a, 2013b; Lee et al., 2016; Perry et al., 1999; Reynolds et al., 2012; Wampold, 2013). This effect appears to be similar across different psychotherapeutic approaches, both for adults (Benish et al., 2008; Gloster et al., 2020; Imel et al., 2008; Luborsky et al., 1975; Seligman, 1995; Smith & Glass, 1977; Wampold, 1997) and children/adolescents (Miller et al., 2008). Hence, "specific ingredients" seem to account little for therapy efficacy (Ahn & Wampold, 2001; Bell et al., 2013), existing transversal factors to all psychotherapeutic approaches that account for therapy outcomes: "common factors" (Rosenwieig, 1936). Systematic reviews on the topic, suggest that the therapist accounts for 5% of the variability of the therapy outcomes (Chow et al., 2015; Jonhs et al., 2019). Also, the amount of variance attributed to the therapist is consistently greater than the one attributed to differences in psychotherapeutic approaches (Miller et al., 2013). On the other hand, gender, educational qualifications and the degree of theoretical integration of therapists are not predictors of clients' outcomes (Chow et al., 2015; Goldberg et al., 2016a). Additionally, it seems that therapists do *not* improve with experience (Anderson et. al., 2009; Brown et al., 2005; Chow et al., 2015; Tracy et al., 2014; Wampold, 2015; Wampold & Brown, 2005), with some even getting worse over time (Goldberg et al., 2016a). This effect might occur due to the lack of systematic training and feedback that therapists receive (Tracy et al., 2014). However, there are therapists who are consistently better than their peers (Baldwin & Imel, 2013; Jonhs et al., 2019). Hence, it is crucial to understand how therapists' performance can be improved.

Deliberate Practice

In contrast to psychotherapy, professionals in areas such as chess, mathematics and accounting (Shanteau, 1992) have been shown to improve their performance over time (Shanteau, 1992). Deliberate practice (DP), <<Individualized training activities specially designed by a coach or teacher to improve specific aspects of an individual's performance through repetition and successive refinement. >> (Ericsson & Lehmann, 1996, pp. 278–279), has been proposed to be one of the causes for performance improvement. Research in this area suggests that the number of hours spent practising a

certain domain predicts the practitioner's performance more than innate characteristics such as “talent” (Ericsson et al., 1993). Deliberate practice seems to improve mental representations in areas such as chess. By improving their mental representations, chess experts can analyse possible moves and its consequences faster, allowing them to make better decisions, and consequently, improve their performance (Ericsson, 2006; Ericsson & Pool, 2016)

Despite Ericsson and colleagues’ (1993) proposal that differences in performance occur mainly due to practice, rather than innate characteristics, some authors have argued that the effects of DP are overly announced (Hambrick et al., 2014a/2014b; Macnamara et. al., 2014). Hambrick and colleagues (2014b) published a meta-analysis showing that only one-third of the participants’ performance is explained by DP. Additionally, Macnamara and colleagues (2014) published another meta-analysis (including 88 studies) that suggest that DP explains 21% of the performance variance in sports, 18% in music, 4% in education and less than 1% in professions. Lastly, contrary to Ericsson and colleagues’ (1993) results, that defend that the sooner a person starts practising a certain domain the better it will become in its mastery, current data has shown that high performers do not start to practice sooner than normal performers (Gobet & Ereku, 2014; Macnamara et al., 2016). Nonetheless, Miller and colleagues (2018) repeated Macnamara and colleagues (2014) meta-analysis, including only studies that measured deliberate practice hours, instead of just hours which could be interpreted as DP, and found that in these studies (18 focused on education) the correlation coefficient for DP practice effects increased from .21 to .40**.** These results suggest that DP might explain greater variance in educational matters than previously thought. However, even if DP explained *only* 4% of the variance, it is important to remember that studies have assigned only 1% of variance to specific ingredients in psychotherapy (Baldwin & Imel, 2013; Miller et al., 2013) and these are widely taught in the training of psychotherapists. Nonetheless, it is necessary to stress that the quality increase of the participants’ performance in a domain such as “education” is not the same as the clients’ outcome.

Deliberate practice in psychotherapy

As previously mentioned, deliberate practice has been shown to be effective in areas such as music, sports, chess and medicine (Ericsson & Pool, 2016; McGaghie et

al., 2011). However, until recently, nobody thought of its application to psychotherapeutic training. The impact of DP on psychotherapy, emerged with the study of *supershrinks*- therapists that constantly have better results than their peers (Miller et al., 2007). This study proposes that the only difference between *supershrinks* and normal therapists is that the first work harder and are more attentive to feedback. Hence, it was suggested that to become a better therapist it is necessary to: a) determine the baseline of effectiveness, b) engage in deliberate practice and c) get feedback (Miller et al., 2007). The DP applied to psychotherapy training has five components: a) the observation of current performance; b) individualized and immediate feedback; c) the creation of small learning goals within the trainee's current capacity; d) behavioural training focusing on the learning objectives identified previously; and e) performance evaluation over time (Miller et al., 2017; Rousmaniere, 2017). Moreover, it is crucial that the learning environment is perceived, by the trainee, as safe and non-judgmental. In contrast with traditional supervision, the DP training focuses on specific skills of the trainee rather than case conceptualization or patients' diagnosis (Rousmaniere, 2017).

Research concerning DP is scant. However, in past years, a great improvement has been done in this area. The literature has demonstrated that DP might have some positive effect in psychotherapists' performance (Anderson et al., 2019; Chow et al., 2015; Golberg et al., 2016b; Hill et al., 2019; Westra et. al., *In press*). The first attempts to understand the impact of DP in psychotherapy were dominated by studies with natural samples where DP was not an isolated factor. Namely, one study showed that the number of hours that a therapist recalls spending in solitary practice predicts better outcomes (Chow et al.,2015). That is, therapists who recall spending more time in solitary deliberate practice have better results (measured in clients' outcomes), than the ones recalling spending less time in solitary deliberate practice activities (Chow, et al., 2015). However, DP only accounted for 0.3% of the variance in clients' outcomes and the study relied on the ability of therapists to recall the time spent in solitary practice, as well as, distinguishing between mere solitary practice and deliberate solitary practice (Clements-Hickman & Reese, 2020). Additionally, a case study concerning a clinical agency where, for seven years, therapists monitored their patients' outcomes and had regular meetings where DP was performed, showed that therapists improved their clinical performance (Goldberg et al., 2016b). Nevertheless, it is important to notice that these therapists not only engaged in DP training but also had concrete feedback on how

to work/ deal with difficult cases. The feedback offered to the therapists aims to increase their responsiveness to the patient (Shimokawa et al., 2010). A meta-analysis comprising 140 studies in different health professions, showed that feedback improved adherence to good practices and participants' clinical outcomes (Ivers et al., 2012). The feedback has shown to be especially effective when: a) the professional is not having a good initial performance, b) the feedback provider is a supervisor or a colleague, and c) the feedback is continuous and contains action plans (Ivers et al., 2012). In psychotherapy, in particular, research suggests that receiving feedback improves therapists' outcomes (Miller, et al., 2006; Østergård et al., 2020; Tracy et al., 2014). This way, despite feedback being a component of DP, it is important to evaluate its isolated impact.

In order to address limitations due to naturalistic settings of the previously mentioned studies, recent studies have used quasi-experimental designs to test the effectiveness of DP (Anderson et al., 2019; Hill et al., 2019; Pearlman et al., 2020; Westra et al., *In press*). One study, consisting of two interventions, addressing the impact of a modelled training in improving *Facilitative interpersonal skills (FIS),* suggests that rehearsal and repetition of specific skills improve the quality of participants intervention, without individualized feedback (Anderson et al., 2019). This study showed a marginal increment (p = .051), after the first intervention, in the *FIS* of participants allocated to the modelled practice group. However, after the second intervention, the significant difference (p = .042) between experimental and control groups, appears to be due to a decrease in participants' performance in the control group, rather than an improvement in the experimental group. Hence, these results have to be carefully read since they might generate hasty conclusions. Additionally, despite being an important contribution to the research in DP, the "modelled training" only addresses two components of DP training, namely: a) structured and repeated training, and b) focus on a specific goal, leaving other DP components untested.

A recent study compared the impact of an alliance-focused and FIS training (with DP components), and a demonstrating training (DT) in therapeutic skills. The results suggest that participants in the DP training increase their observer-rated scores for empathy, alliance, bond capacity and alliance rupture repair responsiveness when compared to participants in the DT (Pearlman et al., 2020). This effect did not occur for

verbal fluency, hope/positive expectations, persuasiveness, expressed emotion and warmth/acceptance/understanding. This might suggest that some skills, but not all skills, improve with DP training. However, this study may address too many skills at once. Hence, it is important to study one skill at a time to guarantee that the participants are focusing their improvement in a targeted skill rather than a set of skills.

Until today, only two studies have compared a DP training (with all de components proposed by Rousmaniere, 2017) to a control condition. Hill and colleagues (2019) applied DP training to therapist's immediacy to understand whether DP affected trainees' self-efficacy, working-alliance perception and client ratings of working-alliance. The results suggest that trainees felt more effective and perceived a small increment in their working-alliance with their patients. However, no effects emerged for patients working-alliance ratings. Lastly, a study that measured the impact of a DP training in participants' interviewer skills (measured by interviewees feedback) and its impact over time (four months follow-up), had positive results. Participants under the DP training improved all skills (better responding to simulate-videos addressing resistant patients, lower levels of observed-rate resistance and higher levels of empathy measured by interviewees) when compared with participants in the control group. Despite a decrease in both groups, DP participants maintained their gains after 4 months (measured in community interviewees empathy ratings), compared to control participants (Westra et. al., *In press*).

Limitations of Deliberate Practice application to psychotherapy training

DP requires that the field in question is competitive and well developed, where the difference between an expert and a trainee is highly evident (Ericsson & Pool, 2016). DP has better results in highly predictable activities (Macnamara et al., 2014), it has been studied in systematic, structured and well-defined domains, such as sports and music, where the difference between a good and bad performance is less subjective and guidelines on how to achieve a better performance are explicit (Clements-Hickman & Reese, 2020). However, in psychotherapy, not only these guidelines are non-existent, but also the specific concept of therapist expertise, and how to measure it, are subject of debate (for a detailed review see: Goodyear et al., 2017; Hill et al., 2017). Additionally, clinical psychology is an area where its therapists have dynamic and subjective objects to evaluate (human behaviour), have little feedback and higher expectations to commit

no mistakes (Shanteau, 1992). Hence, the application of DP on psychotherapy is complex and its principles cannot be directly transposed to psychotherapy.

DP training is based on the idea that part of the therapists' characteristics are malleable and therefore can be improved. However, the malleability of intra and interpersonal skills is unclear (Clements-Hickman & Reese, 2020). Studies have shown that DP can improve the therapists' perspective on the improvement of their skills (Hill et al, 2019; Westra et al., *In press*). Nevertheless, part of this effect might occur due to self-assessment bias, since this self-perspective has not shown to be congruent to clients' perspectives in all studies. Hence, it is important to access clients' perspectives and/or observer's ratings, to control for self-assessment bias. Therefore, more research is needed to understand to what extent are interpersonal skills such as empathy, subject to malleability. Despite not being part of the research hypothesis, the present study can shed some light on this question. If participants improve their responses, then, it is possible to assume that, at least to some extent, therapists can improve such characteristics.

Several authors have stressed that, as it is a new training system, it is crucial to test and access performance's improvement based on skills that predict better outcomes (Miller et al., 2018; Perlman et al., 2020; Tracey et al., 2015), and that are common to all psychotherapeutic approaches, rather than skills specific to one approach (Perlman et al., 2020). This way, and since facilitative interpersonal skills have been associated with positive clients' outcomes (Anderson et al., 2009), at least in short duration therapy (less than eight sessions) (Anderson et al., 2015a; Anderson et al., 2015b), it is important to conduct DP research on these particular skills (Anderson et al., 2009; Pearlman et al., 2020; Rousmaniere, 2017; Wampold, 2017). Since empathy has shown to be one of the common ingredients with great impact in therapeutic outcomes (Wampold, 2015b), the present study uses empathy as its dependent variable.

The present studies

These studies intend to test if a DP training improves the quality of empathic responses of psychology students and experienced therapists in-training when compared to a usual and theoretical training (Training As Usual condition). This way, the study contributes not only for the development of scientific research on the

application of deliberate practice to psychotherapy training but also to the effectiveness of current teaching practices.

Simulation-based learning has shown to be effective in areas such as medicine and nursing (Cant & Cooper, 2010; McGaghie et al., 2011). Hence, the effect of DP is studied using videos simulating psychotherapeutic moments. Additionally, the use of videos in research concerning psychotherapy performance is effective (Anderson et al., 2019). Therefore, in these studies, the participants gave empathic responses to simulation-videos, over the course of three weeks. In the first study, undergraduate students recorded their responses to two clinical-simulation videos over the course of three weeks (one time each week) and received either DP or TAU (without individualized feedback) interventions. In the second study experienced in-training therapists recorded their responses to six different videos over the course of three weeks (two videos each week) and received individualized feedback in both the PD and TAU conditions.

The researchers intend to analyse if (H1): Participants under the DP training (DP condition) improve the quality of their empathic responses, in the second and third evaluation moments, when compared to a group in the TAU training. It is expected that, at the baseline, the quality of responses is the same for participants in both the DP and TAU groups and that, after the interventions, the responses in the DP group are significantly more empathic than the responses of the TAU group. Regarding the second hypothesis, it is expected that (H2): Participants in the DP group improve the quality of their responses throughout the second and third evaluation moments (after the DP training), while participants in TAU maintain the quality of their responses in the three moments.

As individualized feedback has shown to increase therapist's performance (Miller et al., 2006; Østergård et al., 2020; Tracy et al., 2014) and in the first study a practice effect might occur (due to the video repetition), a second study was carried out to control possible threats to the internal validity of the first study - the individualized feedback and the practice effect. Thus, while in the first study the participants respond to 2 videos over the course of three weeks, in the second study, participants respond to 6 different videos (two videos each time), during the same amount of time (three weeks). Additionally, in the second study, both conditions receive individualized feedback, to ensure that the effect occurs due to the DP training

as a whole and not just the individualized feedback. The results presented in the second study of this paper are the preliminary results of an ongoing experiment. Hence, these are not our final findings and results may change as new data is collected.

Study 1

Method

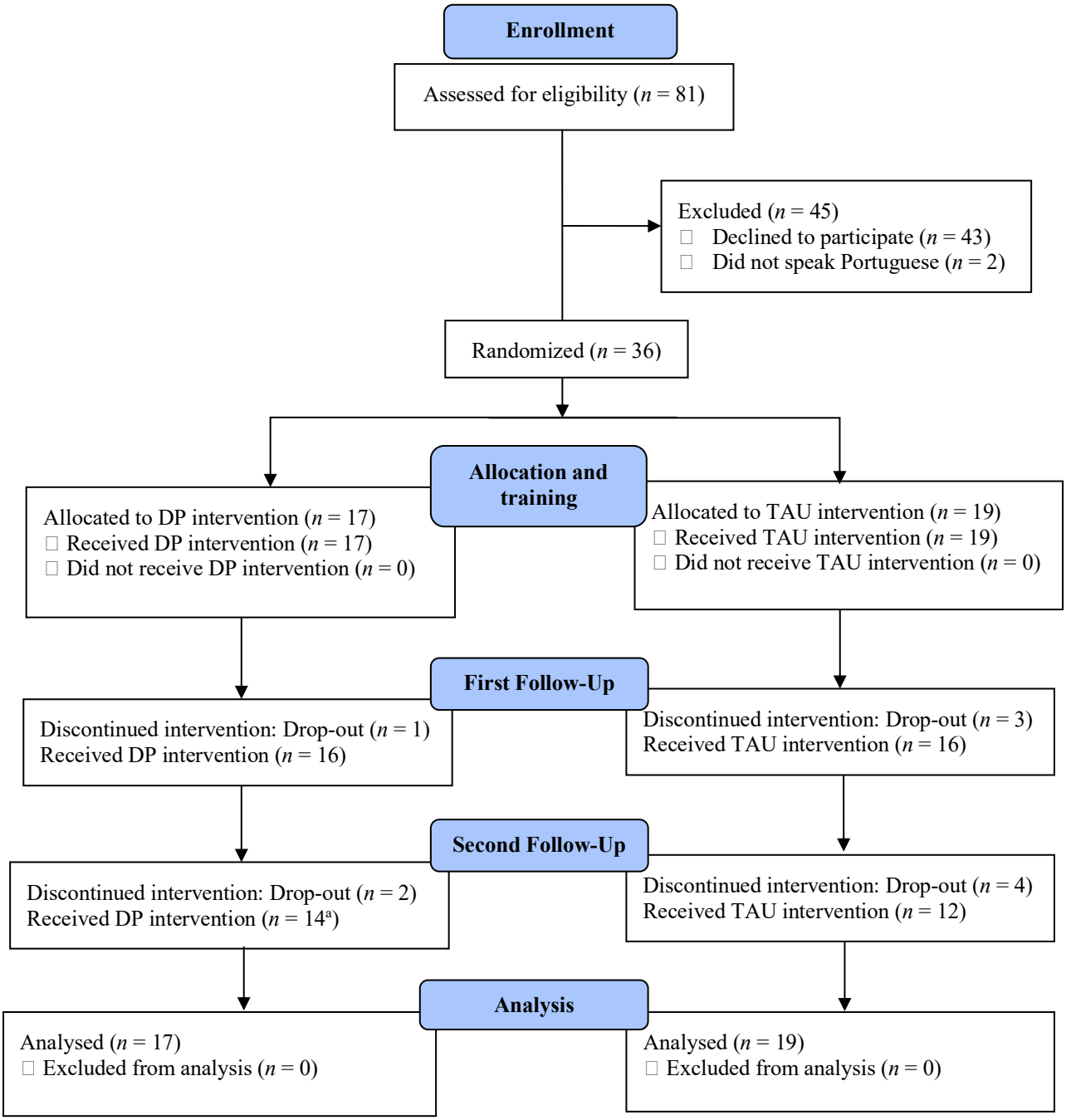
Overview and design

The present study is a randomized control trial (RCT) following a mixed design. The data was collected on three occasions, with a week interval between collections using the *Theravue* platform (March, 2020). Once the data was collected, the responses were evaluated and rated by two independent raters that were blind to the participants' conditions. The interrater agreement was accessed using the intraclass correlation coefficient (ICC).

Participants

Data collection took part in two academic years (2018-2019 and 2019-2020). Participants were students from a Psychological Sciences bachelor's degree at a Portuguese university. The overall sample was constituted by 36 participants (*Mage* = 25.67, *SDage* = 10.61, Range age=19 – 56; Table 1). However, the study suffered from subjects loss during the three weeks (Figure 1) and only 26 participants took part in all evaluation moments (Figure 1). None of the participants completed clinical hours professionally. However, two participants were trained in specific psychotherapy approaches (Table 1). All the participants signed an informed consent and the study was approved by the ISPA Research Ethics Committee.

Figure 1 CONSORT follow chart of participants allocation



[a] One participant responded to Video A but not B. Hence, video A n=15; Video B n=14

Table 1 Condition allocation and Demographics of the study sample [a]

	2018	*2019*	*Total*
Condition			
Deliberate Practice	*8*	*9*	*17*
Training As Usual	*9*	*10*	*19*
Gender			
Female	*6*	*14*	*20*
Male	*2*	*2*	*4*
Nationality			
Portuguese	*7*	*15*	*22*
Brazilian	*1*	*1*	*2*
Ethnicity [b]			
Caucasian	*6*	*15*	*21*
Black	-	*1*	*1*
Highest Educational Level			
High School	*5*	*12*	*17*
Bachelor's degree	*2*	*1*	*3*
Postgraduate degree	*1*	*2*	*3*
Master's degree	-	*1*	*1*
Previous training in psychotherapy [c]			
Cognitive-behavioural Therapy	*1*	-	*1*
Systemic Family Therapy	-	*1*	*1*

Note. N=36

[a] *only 24 participants provided their demographic data*

[b] *2 participants refused to answer this question*

[c] *Only two participants had previous psychotherapeutic training*

Materials

We used *Theravue*, an online deliberate practice system, for the video presentation and recording of the participants' responses. The demographic survey was implemented in Qualtrics software, XM (March, 2020) (see Appendices 4 and 5). Additionally, the *Measure of Expressed Empathy (MEE)* (Watson, 1999) was used to evaluate the quality of the empathic responses given by the participants (see Appendix 2).

The *Measure of Expressed Empathy (MEE)* (Watson, 1999) (Appendix 2) is a 10-item observer-rated measure that addresses the time length in which therapists

maintain empathic verbal and nonverbal behaviours. It is built to evaluate five-minute segments of therapy sessions constituted with a 9-point scale (from 0 = 'never' to 8 = 'all the time'), which addresses the amount of time during the segment that the therapist showed a certain behaviour. Hence, if the therapist showed a behaviour (e.g: the therapist looked concerned) 50% of the segments' time, he would receive a 4 for the item addressing that behaviour. After the evaluation of the 10 items, a global empathy score is calculated by the mean of the 10 items. This measure shows a high (α = .88) internal consistency and construct validity (r = 0.66, p < 0.01) measured by Barrett-Lennard Relationship Inventory (BLRI; Barrett-Lennard, 1962) (Malin, 2016).

Video A is 59 seconds long and depicts a woman with PTSD symptoms that feels misunderstood by her husband. She describes her symptoms and her inability to stop thinking about something that concerns her. She says she understands why she is in therapy and that her husband should be in therapy as well.

Video B is 47 seconds long and shows a young man with a conflictual relationship with his sister that frustrates him and negatively influences the relationship with his nephew.

Procedure

The study and data collection took part in a Portuguese university during two academic years (2018/ 2019 and 2019/2020). Participants were invited to participate in a study investigating the impact of different training methods in key clinical skills. Their participation was voluntary and there was no reward for participating in the study. Students and coaches met three times during the study. In the first meeting, coaches explained that the study implied two training sessions with the coaches and the self-recording of three responses to two videos showing psychotherapeutic moments, at home. The schedule was explained to the participants and the importance of responding to all the videos after each training session and before the next was stressed. After a brief lecture concerning the importance of empathy in clinical outcomes, the use *of Theravue* platform was explained. At the end of the first meeting, participants were told the group they were allocated to. However, no explanation about the differences of the groups was disclosed, in order to maintain the internal validity of the study. Additionally, no discussion of deliberate practice was conducted.

After the first meeting, participants received an email with their credentials to log in their *Theravue* account and they were requested to record the first response to both videos before the second meeting with the coaches. Until this point, the procedure was the same for both the experimental (*Deliberate Practice*) and the control *(Training As Usual)* groups. From the second meeting onwards the participants' activities differed depending on the group to which they are allocated.

In the *Deliberate Practice* meeting, the participants were told that they would get individualized feedback of their responses and then practice specific weaknesses. They were also told, by the end of the first session, that homework would be presented to them and that they ought to practice 3 times a week before responding to the next video. The coach invited the first participant to show one of their responses to the videos. After watching the response, objective and precise performance feedback was given to the participant. Afterwards, the coaches proposed objective exercises, concerning verbal and/or nonverbal tasks, that were at the edge of the trainees' current capacity. The participant then re-watched the video and tried to give a new response taking into consideration the feedback and the proposed exercises, which were rehearsed at least 3 times for each response. Except for the overtime performance evaluation, due to the nature of the study, all Rousmaniere's (2017) assumptions of deliberate practice were met in this training.

In the *Training As Usual* meeting, a lecturer gave a presentation about empathy and its importance for clinical outcomes. The lecturer initiated a debate on the importance of empathy for psychotherapy and some theoretical papers and books regarding this topic were shown. This condition aimed to mimic the environment of a traditional expositive lecture.

These processes were repeated two times during the study: The participants responded to the videos and had their first *Deliberate Practice* or *Training As Usual* session and then all the students responded to the videos again. In the following week, they had their second and last session (either *Deliberate Practice* or *Training As Usual*) and then responded for the third time to the videos.

Data analysis

The interrater agreement was accessed through a Two-Way mixed effect intraclass correlation for the absolute agreement of multiple raters (k=2). This index is considered appropriated to analyse the exact agreement between two independent raters for ordinal variables from randomized samples (Field, 2005, Koo & Li, 2016, McGraw & Wong, 1996).

As both studies present ordinal variables, non-parametric tests were chosen over parametric tests since these do not assume a specific data distribution and are more appropriate to apply in non-continuous data and small samples (Marôco, 2007). Thus, a Mann-Whitney test was conducted to access differences between the empathic quality of the control and the experimental group's responses. To analyse the degree in which participants improved the empathic quality of their responses, an empathy score comprising the scores for both video A and B using the *Measure of Expressed Empathy (MEE)* (Watson, 1999) was created for each participant. The scores were compared between the *Deliberate Practice* and *Training As Usual* condition at the three evaluation moments, separately. Differences within both *Deliberate Practice* and *Training As Usual* conditions were accessed through a Friedman's ANOVA. All the analyses were conducted in SPSS statistical package version 24.

Results

Interrater agreement

The Intraclass correlation analysis (ICC) yield a good agreement for participants' empathy scores, specifically: the average ICC was .824 with a 95% CI [.653, .910] (F(35,35) = 5.549, p<.001), for the first responses; the average ICC was .749 [.483, .878] (F(31,31) = 3.902, p<.001) for the second responses and .768 [.486, .896] (F(25,25) = 4.279, p<.001), for the third responses.

Responses' empathic quality between groups

The comparison of empathic scores between the experimental (DP) and control group (TAU), accessed through a Mann-Whitney test, partially support the first research hypothesis.

The *Deliberate Practice* (*Mdn = 3.65)* and *Training As Usual* (*Mdn = 3.66)* conditions did not differ at baseline $U = 126$, $z = -1.13$, $p = .271$, $r = .19$. After the first intervention, there was no difference between *Deliberate Practice* (*Mdn* = 4.33) and *Training As Usual* (*Mdn* = 3.71) $U = 86$, $z = -1.58$, $p = .119$, $r = .28$. However, at the last intervention, the *Deliberate Practice* group (*Mdn* = 3.83) performed significantly better than the *Training As Usual* group (*Mdn* = 4.90) $U = 45$, $z = -2.01$, $p = .046$, $r = .39$.

Figure 2 Empathy scores at baseline and postinterventions

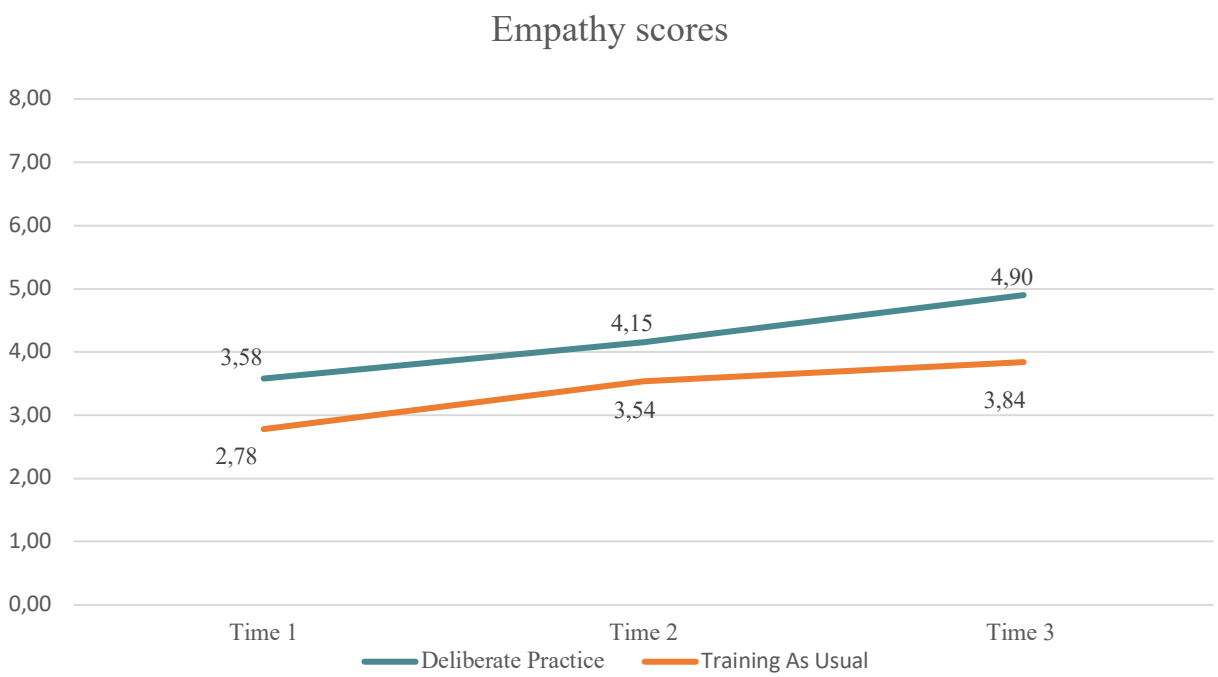


Table 2 Mann-Whitney results for between groups comparisons

	Deliberate Practice	*Training As Usual*				
	Mdn	*Mdn*	*U*	*Z*	*p*	*r*
Time 1	3.58	2.78	126	-1.13	.271	.19

Time 2	4.15	3.54	86	-1.58	.119	.28
Time 3	4.90	3.84	45	-2.01	.046	.39

Responses' empathic quality within groups

Friedman's ANOVA was used to test if the participants' performance increased within groups. That is, if the participants' performance on the *Deliberate Practice* group and the *Training As Usual* group increased, separately. For participants in the *Deliberate Practice* group, there was a significant increase in the responses' empathic quality, X^2 (2) = 9.143, p = .010, W = .327. Dunn-Bonferroni post hoc tests were carried out to analyse the differences between the three measures and this yield a significant difference between the baseline (time 1) and the last intervention (time 3) X^2 (2) = -1.143, p = .007. There were no significant differences in the control (training as usual) group for the different measures X^2 (2) = 2.167, p = .338, W = .090. The results partially support the second hypotheses since participants in the experimental group performed better after two DP interventions and participants in the control condition did not improve their responses.

Table 3 Friedman's ANOVA for within groups comparisons at the three times

	Time 1	Time 2	Time 3			
	Mdn	*Mdn*	*Mdn*	X^2	*df*	*W*
Deliberate practice	3.58	4.15	4.90	9.143*	2	.327
Training As Usual	2.78	3.64	3.84	2.167	2	.090

**p=0.01*

Table 4 Dunn-Bonferroni post hoc for DP condition

	X^2	*df*	*p*
Time 1 – Time 2	-.571	2	.392
Time 2 – Time 3	-.571	2	.392
Time 3 – Time 1	-1.143	2	.007

Discussion

The present findings partially support both hypotheses, since a) Participants under the DP training gave significantly better empathic responses than participants under the TAU condition, after the second intervention; and b) Participants under the DP training improved their empathic responses, after the second intervention, while participants under TAU condition did not improve the quality of their performance. It seems that, despite both conditions tending to increase after the first intervention, after the second intervention the DP condition continues to increase while the TAU condition tends to flatten. The effect size after the second DP intervention was moderate (r = .39; Cohen, 1992), which suggests that the manipulation explained, after two interventions, 15.21% of the effect variance. Nonetheless, to ensure that these tendencies would be maintained over time, it would be necessary to continue monitoring the performance.

Longitudinal data analysis was performed for both DP and TAU conditions, separately. This analysis yields significant improvement for participants in the DP condition, after two interventions, and no significant improvement for participants in the TAU condition. This reinforces the effectiveness of a DP training, since not only DP participants performed better than TAU participants, but also, they were the only participants who improved with the interventions (participants in the TAU did not improve nor worsen their median responses scores). These results may clarify the impact of methods often used in continuing learning, such as expository classes and theoretical supervision on trainees' skills. Continuing learning (e.g: supervision, postgraduate degrees, workshops…) tends to have a theoretical rather than a practical nature (Rousmaniere et al., 2017) and no individualized feedback (Ericsson & Pool, 2016). This kind of training is important for the development of knowledge in the area but may not suppress the need for improvement in the performance of psychotherapists (Rousmaniere, 2017).

Just as in expository classes, supervision has more passive interventions, such as case discussion, than active ones, such as role-play and feedback (Kühne et al., 2019). A systematic review concluded that the effects of supervision in psychotherapy are blurred

since studies lack methodological quality and control for alternative explanations (Kühne et al., 2019). This review also suggested that the supervision satisfied the supervisees expectations, but that supervisees satisfaction does not mean effectiveness. Another systematic review advocates that supervision might have some positive impact on the supervisee (in terms of self-awareness, skills and self-efficacy) and clients' outcomes (Wheeler & Richards, 2007). On the other hand, one study using a sample with 6521 clients, 175 trainee therapists and 23 supervisors, suggested that trainee's supervision explains less than 1% of clients' outcomes (Rousmaniere et al., 2014). Our findings suggest that, such as supervision, an expository training might have little impact on the clinical skills of psychology undergraduates. Nonetheless, it is important to stress that these students were not licensed practitioners yet, which makes impossible to generalize the results to licensed therapists. Hence, more research needs to be conducted in this area to understand which impact (if any) mere expository training has on therapists' clinical skills. Moreover, it is important to replicate this experiment in licensed therapists and address the impact of theoretical learning on clients.

Contrary to what Ericsson and colleagues (1993) and Ericsson and Pool (2016) postulate, and congruent with Macnamara et al (2014) results, our results suggest that DP in group is effective. This might be promising for the application of DP in Continuing Learning since this is often practised in group. Moreover, as suggested by Westra and colleagues (*In press*), the group sessions may be helpful for trainees since they can watch other trainees' performance and feedback. The effectiveness of DP in group reduces the impact of certain limitations addressed at DP, for both the trainee and the coach. Personal DP sessions might be too expensive for trainees and too time-consuming for coaches. Hence, the suggested effectiveness of group DP training might help solve these two big DP constrains. We stress that, despite being a group training, it is necessary to ensure both a non-judgmental and safe environment, as well as to avoid mere non-purposeful and non-individualized repetition (Ericsson et al, 1993; Ericsson & Pool, 2016; Rousmaniere, 2017). Nevertheless, it is crucial to conduct research on the impact of private DP sessions on trainees' skills and to compare them with the existing literature concerning group DP trainings. To our knowledge, no empirical studies have been done concerning private DP sessions.

Deliberate practice applied to psychotherapy is a recent area of study and therefore it lacks scientific literature. Nevertheless, the present results advocate that interpersonal skills such as empathy might be improved by DP training, at least in participants with no previous clinical skills. These findings go along with Pearl and colleagues' (2020) and Westra and colleagues' (*In Press*) results. In both these studies, participants under a training with DP components increased their empathy levels comparing with the control condition. Additionally, there are four studies where participants who were under a DP condition (containing individualized feedback) either improved comparing to participants in TAU (Pearl et al., 2020; Westra et al., *In Press*), or improved their clients' outcomes (Chow, et al., 2015; Goldberg, et. al., 2016a).

Contrary to the DP positive impact suggested by our study, in Hill and colleagues' (2019) study DP training did not impact clients' outcomes. Additionally, in Anderson and colleagues' (2020) study the significant difference between the experimental and the control conditions, occurs due to the decrease in the control group's performance. In that study, the authors provided a good and a bad response and gave *no* individual feedback to participants. As the experimental group did not have individualized feedback, it is possible to hypothesize that individual feedback is crucial for participants improvement. Taking into account that the feedback therapists usually receive is focused on the patient's outcome rather than the therapists' performance (Tracy et al., 2014), and it is shown to improve therapists' performance (Miller, et al., 2006; Østergård et al., 2020; Tracy et al., 2014), it is not possible to know to what extend feedback concerning the therapist's behaviour might improve his/her performance. Thus, the possibility that feedback alone may be responsible for the positive impact of DP training, cannot be excluded. Therefore, it is necessary to conduct research controlling for the feedback effect. This effect may be controlled creating a study with three conditions where one group receives DP training, the other one receives TAU plus feedback and the last one receives TAU with no feedback (control).

The fact that the effect of DP is only significant after the second intervention goes along with the DP literature that states that DP training has no immediate results (Ericsson et al., 1993; Ericsson & Pool, 2016; Tracey et al., 2015; Miller et al., 2018). In fact, DP training is a highly mental and physically demanding activity, which requires extreme focus and effort for short periods of time, over long-term practise

(Ericsson et al., 1993; Ericsson & Pool, 2016). Consequently, trainees need to focus on long-term achievements and gains in order to sustain their practice (Ericsson et al., 1993; Ericsson & Pool, 2016). This might be a constrain for DP training, especially for solitary practice, since psychotherapy trainees report feeling too challenged when practising alone and do not enjoy homework assignments, preferring to work within the presence of a coach or supervisor (Hill et al., 2019). Regarding the present study, it is important to stress that even though homework assignments were given to the participants in the DP condition, the researchers did not control if the participants had done it. Given the case the assigned homework was not carried out, these results could indicate that the DP training might improve participants' performance even without training between sessions. However, this is just a hypothesis, since we do not know if participants completed their assignments or not.

Limitations and future research

Despite the promising results, it is important to point out that this study had some limitations. As previously mentioned, participants responded to the same two videos three times. This enables a practice effect where the improvement in participants' performance occurs due to the repetition of the same task and not the experimental manipulation (American Psychological Association, 2020). Since participants in the DP condition got feedback and rehearsed their responses during the training, they might have had mimicked their previously rehearsed responses. Which, in return, might have improved their responses to that specific video, but not their empathic skills in response to all clinical cases. Moreover, participants in the DP condition got individualized feedback and participants in the control condition did not. Thus, the improvement may occur due to the individualized feedback alone and not to the DP training overall. These two threats to the internal validity of the study will be controlled in the second study.

Regarding the study procedure, and as previously mentioned, researchers did not control the assigned homework. Additionally, participants were aware that their recordings would be evaluated. Despite the results not suggesting this effect, the recordings may have contributed to a feeling of unsafeness and a sense of judgment, which is unproductive to DP training (Ericsson et al., 1993; Rousmaniere, 2017). Another limitation of the study is that the *Measure of Expressed Empathy*

(MEE) (Watson, 1999) is an unpublished scale which might arise some concerning regarding its validity. This scale was created to evaluate five-minutes segments of psychotherapy videos. However, the participants' responses of these study lasted on average ≈35 seconds, which falls under the time for which the scale was created. Besides, the scale contains items addressing relational and process-focused moments such as item 5. <<Is the therapist responsively attuned to the client's inner world moment by moment in the session?>> and 7. <<Is the therapist responsive to the client?>>. These items are difficult to evaluate only through the therapist's responses, without seeing the ongoing relationship between the therapist and the client. Furthermore, despite the coders being blind to participants conditions, the coaches that delivered both the DP and TAU interventions were not blind to the participants' conditions, nor the research hypotheses. Hence, the experiment might suffer from an experimenter effect (Rosenthal, 1963). Lastly, since training with practical activities is uncommon, and undergraduate students are used to expository lectures, our results might have suffered from a novelty effect. This effect happens when an innovation is introduced (e.g: a new treatment) and increases participants' motivation and is, consequently, the cause for the effect rather than the experimental manipulation. In the first study, the TAU condition suffered from greater subject loss ($n = 7$) than the DP ($n = 3$), which may indicate that participants in the DP condition were more motivated than participants in the TAU condition. This way, it is important to interpret our results carefully since not all alternative explanations where controlled.

Study 2

Method

Participants

This experiment was constituted by 11 trainees from an existential psychotherapeutic training (*Mage = 44.11, SDage = 8.67,* range age=*28 - 53*) (Table 5). From the 11 participants, only 8 participants completed the entire study (Figure 3). Participants' clinical experience ranged from 2 to 10 years ($M = 3.13$; $SD = 2.80$). Only one participant had previous psychotherapeutic training, in psychoanalysis and brief psychotherapies. All the participants signed an informed consent and the study was approved by the ISPA Research Ethics Committee.

Figure 3 CONSORT follow chart of participants allocation

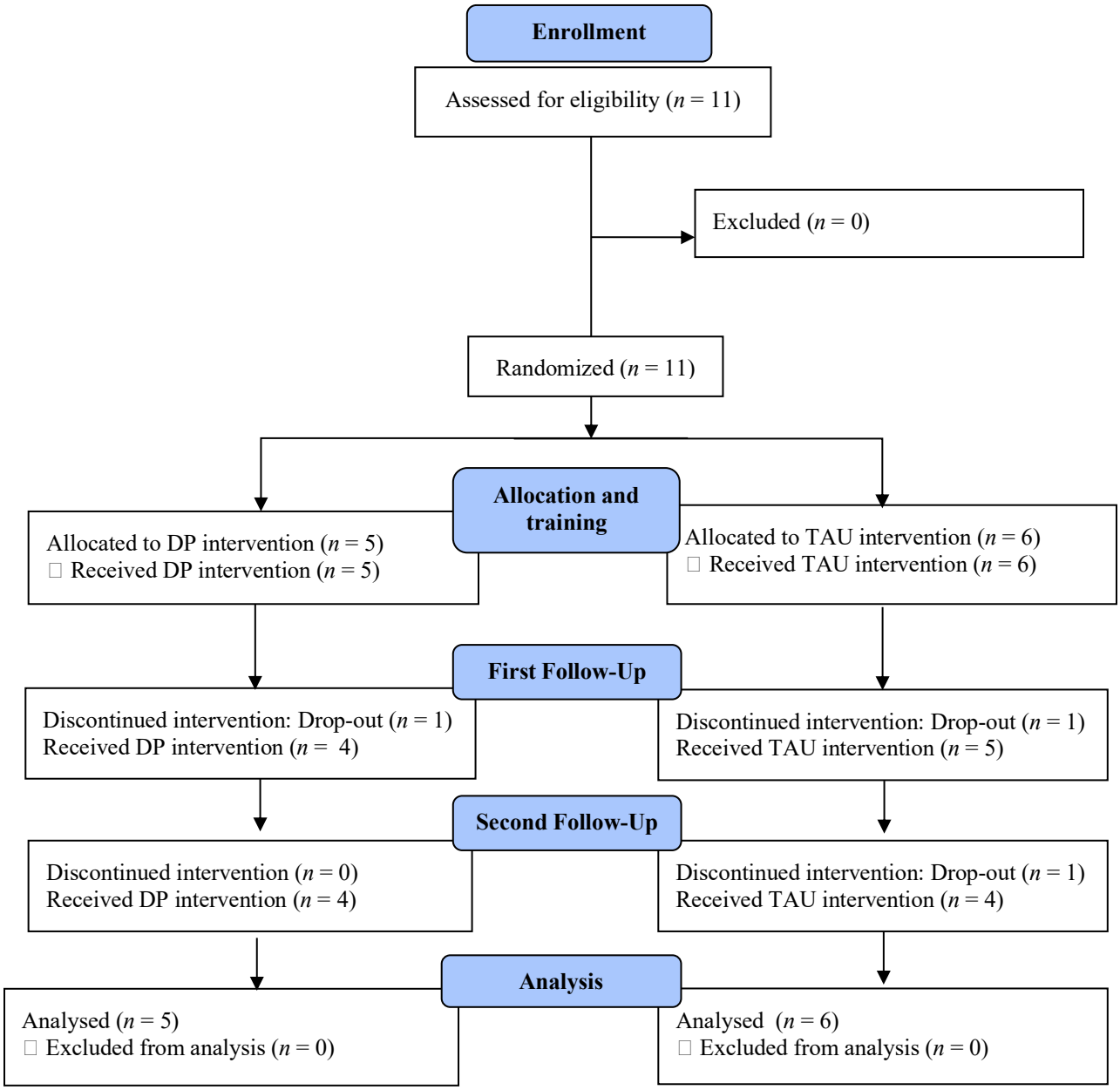


Table 5 Condition allocation and Demographics of the study sample [a]

Condition	Frequency
Deliberate Practice	5
Training As Usual	6
Gender	
Female	*9*
Male	*0*
Nationality	
Portuguese	*8*
Brazilian	*1*
Ethnicity	
Caucasian	*9*
Highest Educational Level	
Bachelor's degree	*3*
Master's degree	*3*
PhD	*1*
Did not respond	*2*
Previous training in psychotherapy	
psychoanalysis and brief psychotherapies	*1*

[a] *only 9 participants consented to give their demographic information*

Materials

Except for the videos, the materials used in this experiment were the same as the ones used in the previous experiment.

Video A and *Video E* are the same as *Video B* and *Video A* in the previous study.

Video B is 68 seconds long and shows a young man explaining his anxiety symptoms and how they impact his life and routine.

Video C is 22 seconds long and reports a young woman explaining that her family does not understand why she is in therapy.

Video D is 100 seconds long and shows a female student explaining her demotivation and how it is impacting her grades, extracurricular activities and life in general.

Video F is 69 seconds long and shows a depressed woman stating her discontentment with the therapeutic process and her disbelief that therapy can help.

Procedure

This study followed the same procedure as the previous one, with the addition of four new videos and the implementation of one more TAU condition component. To control for practice effect (due to video repetition), participants were shown 6 different videos over the course of three weeks (two each week). As the sample size was too small to counterbalance the videos, in order to control an order effect, the videos were rated by four independent raters (two masters students, a PhD student and a lecturer) for their degree of difficulty. Then, pairs of videos were formed by joining a low-difficulty video with a high-difficulty video. The other difference between this and the previous study is that participants in the TAU condition had individualized feedback as well as participants in the DP condition. This was done to analyse if the positive impact of DP training, found in the previous study, occurred due to feedback alone instead of the DP.

Preliminary results

Interrater agreement

The Intraclass correlation analysis (ICC) reveals a moderate agreement for participants' empathy scores in the first and second responses and a good agreement for the third response. The average ICC was .587 with a 95% CI [.-606, .890] ($F(10,10)= 2.340$, $p <.098$. for the first responses; .677 [-.273, .930] ($F(8,8)= 6.931$, $p = .006$, for the second responses; and .791 [-.004, .958] ($F(7,7) = 7.126$, $p = .009$), for the third responses.

Responses' empathic quality between groups

There were no differences in the empathic quality of participants under the DP and TAU trainings. The *Deliberate Practice* (*Mdn = 6.45)* and *Training As Usual* (*Mdn = 6.23)* conditions did not differ at baseline $U= 12.50$, $z = -.46$, $p = .662$, $r = .14$.

Additionally, there was no differences between *Deliberate Practice* (*Mdn* = 6.25) and *Training As Usual* (*Mdn* = 5.48) U = 3.50, z = -1.60, p = .111, r = .53, after the first intervention, nor the second intervention - *Deliberate Practice* (*Mdn*=6.74) and *Training As Usual* group (*Mdn*= 6.25) U = 4, z = -1.16, p = .343, r = .41. However, both groups decreased their performance after the first intervention. The reasons for this effect are discussed below.

Figure 4 Empathy scores at baseline and postinterventions

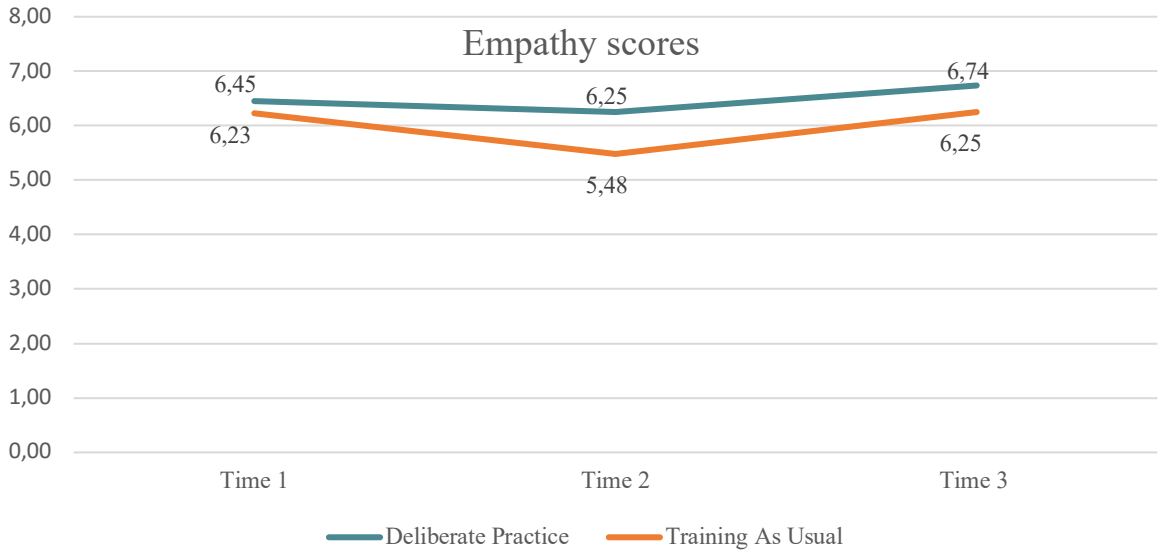


Table 6 Mann-Whitney results for between groups comparisons

	Deliberate Practice	*Training As Usual*				
	Mdn	*Mdn*	*U*	*Z*	*p*	*r*
Time 1	6.45	6.23	12.50	-.46	.662	.14
Time 2	6.25	5.48	3.50	-1.60	.111	.53
Time 3	6.74	6.25	4	-1.16	.343	.41

Responses' empathic quality within groups

Only participants under the TAU condition performed differently in the three evaluation moments. These participants improved from the second to the third evaluation moment X^2 (2) = -1.750, p = .040**.** However, these results seem to occur due

to the decrease in participants' performance from the first (*Mdn* = 6.23) to the second evaluation moment (*Mdn* = 5.48), since they did not improve between the first and the last responses.

Table 7 Friedman's ANOVA for within groups comparisons at the three times

	Time 1	Time 2	Time 3				
	Mdn	*Mdn*	*Mdn*	X^2	*df*	*p*	*W*
Deliberate Practice	6.45	6.25	6.74	2	2	.368	.250
Training As Usual	6.23	5.48	6.25	6.5	2	.039	.812

Table 8 Dunn-Bonferroni post hoc for TAU condition

	X^2	*df*	*p*
Time 1 – Time 2	1.250	2	.231
Time 2 – Time 3	-1.750	2	.040
Time 3 – Time 1	-.500	2	1

Discussion

The results in this study do not support either of the hypotheses since participants in the DP condition did not improve their responses when compared to participants in the TAU condition. Additionally, participants in the DP condition did not improve between interventions. An unpredicted pattern occurred in the responses given that both groups decreased their performance from the first to the second moment of evaluation, and then increased at the third moment. In the control group, the increment from the second to the third response was significant. This pattern suggests that the videos shown in the second evaluation moment (video C and D) might have been harder to respond than the other videos. Despite the fact the videos were rated in terms of difficulty, in an attempt to avoid order effects, the results suggest that this was not successful. It is crucial to replicate this experiment with a larger sample, so as to counterbalance the video's order.

It is important to stress that these are preliminary results and the sample size is extremely small (only 4 participants, in each condition, responded in all evaluation moments). Statistical tests in small sample sizes have low statistical power (Shaughnessy et al., 2012). Hence, it is important to analyse the effect size, which is not directly influenced by the sample size. At the third evaluation moment, the effect size was r = .41 which is considered a moderate effect size (Cohen, 1992) and is equivalent to 16.8% of variance. Additionally, in the second evaluation moment, all the participants decreased their performance, but participants in the control condition had a more abrupt decrease than the ones in the DP condition however, this difference was not significant. Moreover, while in the first study the interrater agreement was good in all three responses, in the present study the interrater agreement was moderated, for the first and second responses. This might occur, once again, due to the small sample size (Gwet, 2014). This effect might influence the results since the raters did not always have the same opinion regarding the participants' performance. Thus, as the effect size is moderate and participants in the control condition have a bigger decrease after the second intervention, it is possible that even with no significant results, participants in the DP training performed better than participants in the TAU condition.

The participants in this sample are experienced in-training therapists, with more than two years of clinical experience, which might impact the results. These results might suggest that when a certain performance level is achieved, further improvements are harder to achieve. Previous research has shown that therapists do not improve with time and may even deteriorate (Goldberg et al., 2016). However, this is incongruent with the DP research that indicates that DP training does improve licensed therapists' results (Chow et al., 2015; Golberg et al., 2016b; Pearlman et al., 2020; Westra et al*., In press*). As previously mentioned, DP is particularly effective in predictable and structured activities, the method results due to the regular practice of a specific domain in order to achieve a certain goal (Ericsson & Pool, 2016). Therefore, it is possible that when changing the target stimuli (the video), participants are unable to generalize the previously practised skill onto a different situation. This might suggest that DP is not effective for subjective domains such as the improvement of interpersonal skills.

The presented results might indicate that the training implementation has not been strict enough, which might have led to a lower coaches' engagement and, consequently, lower impact on participants. It is important to address this limitation in future studies, so as to analyse up to which extent the coach engaged and adhered to DP training. The Deliberate Practice Coach Competency Scale (DPTCS) might be provided to the participants for this purpose (Vaz & Rousmaniere, 2020). This scale was developed to analyse up to what extent the coach adhered and provided competent DP training. It addresses areas concerning the coach tasks, such as: collaboration with the trainees, time management, feedback quality and consistency, and coach interpersonal abilities such as emotional sensitivity, acceptance and self-acceptance promotion. If coaches were not sufficiently engaged in the training, participants might have not adhered to the training. As DP is an extremely difficult training, requiring focus and attention (Ericsson et al, 1993; Ericsson & Pool, 2016; Rousmaniere, 2017), if coaches and participants are not thoroughly engaged in its practice, the manipulation may not be effective.

Conclusions and future research

The results of the first study suggest that DP might have a positive impact on the training of psychology students. However, the findings in the second study do not support this claim, for in-training therapists. Taking both studies into account, it is possible to affirm that more research is needed on how DP applies to psychotherapy training. Despite its limitations, the present studies have strengths in their proposal, namely: a) they control for self-assessment bias since participants performance is addressed by independent raters instead of participants' own perspective; b) they are quasi-experimental rather than correlational, which is a necessary condition to infer causality; and c) to our knowledge these are the first studies to question the impact of feedback alone.

The difference between the designs of each study makes it difficult to understand the origin of the effect found. When controlling the order effect of the first study, by showing six different videos in the second study, and controlling the feedback effect by adding individualized feedback to the control group of the second study, as well as using a more experienced sample, it becomes harder to pinpoint the origin of the discrepant results. That is, it is not possible to conclude if, for example, the effect of the

DP training in the first study, was a consequence of a practice effect due to the video repetition or a consequence of the individualized feedback alone, since these variables were not isolated factors in the second study. Additionally, it is not possible to know whether the discrepant results are caused by a higher level of experience in the sample of the second study. Despite literature suggesting that therapists do not improve with experience and that trainees might have similar results to experienced therapists (Goldberg, 2016a; Budge et al., 2013), it is necessary to note that in the first study the sample is constituted by undergraduate students, that had never contacted with patients. This is different from therapists in training that have some experience. For the participants in the first study, it is possible that the video record is their first attempt to practice clinical skills, which can lead to a greater impact of the DP training, due to the lack of previous skills. Hence, it is possible that the discrepancy between results in the first and second study occurs due to the inexperience of the first study participants.

The sample size in both studies is very small, which might result in an underpowered statistical test in the second study. An A priori power analysis using G*Power 3.1 (Faul et al., 2007) and setting a medium effect size (d =0.5; Cohen, 1988), an alpha level of α = .05 and power level of $1 - \beta$ = .80, yielded a required sample size of 220 participants for the present studies, which is quite far from the sample achieved (Study 1: N = 36 and Study 2: N = 11) . Lastly, the independent raters rated the first study responses first, and then the second study responses. As the participants in the first study are undergraduates and the ones in the second study are experienced psychologists, this experience discrepancy might have inflated the ratings for the second study. For at the baseline, the first study rates around 3 points and the second study around 6 points. This might cause raters to be less sensitive to participants variations in the second study because compared to the previous sample the responses' quality was higher.

For future research, it is important to replicate this study addressing the impact of feedback alone in participants' performance, with a larger sample size to control for practice effects and analyse differences in experienced and non-experienced therapists, as well as collecting data in long-term follow-ups. Additionally, it would be beneficial to create a double-blind study, where neither the participants nor the coaches are aware of the participants' conditions or study's hypothesis. Ideally, it would be created a study

with 3 conditions, namely: *Deliberate Practice, Training As Usual + Individualized feedback* and *Training as usual* (control). Then, the sample size would be large enough to counterbalance the videos presentation and would be constituted by both trainees and licensed psychologists. For it is important to analyse if any interaction effect between the conditions and the therapist experience emerges. Additionally, the performance of these participants would be monitored over time and clients' outcomes would be collected. To guarantee the quality of the DP manipulation, participants should evaluate DP training quality and coaches' competence, through the Deliberate Practice Coach Competency Scale (DPTCS) (Vaz & Rousmaniere, 2020). Also, it is important to replicate this study with individualized DP sessions, to analyse if there are differences in group vs individual training. Lastly, despite the preliminary results of the last study, there is literature supporting DP effectiveness. Hence, it is crucial to start asking << *Why* and *when* is DP effective in therapists training?>> rather than << Is DP effective in therapists training?>>. The answer to this question might be addressed in the impact of DP in therapists' mental representations.

References

Ahn, H., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counselling and psychotherapy. *Journal of Counselling Psychology, 48*(3), 251–257. https://doi.org/10.1037/0022-0167.48.3.251

American Psychological Association (2020). *APA Dictionary of Psychology*. https://dictionary.apa.org/practice-effect.

Anderson, T. Pearlman, M. R., & McCarrick, S. M. (2019) Modelling therapist responses with structured practice enhances facilitative interpersonal skills. *Journal of Clinical Psychology, 76*(1)*, 1-17*. https://doi.org/10.1002/jclp.22911

Anderson, T., Ogles, B., Patterson, C., Lambert, M., & Vermeersch, D. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology, 65*(7), 755–768. https://doi.org/10.1002/jclp.20583

Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behaviour change* (6th ed., pp 258-297). Wiley.

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*(4), 722–736. https://doi.org/10.1037/a0033004

Benish, S. G., Imel, Z. E., & Wampold, B. E. (2008) The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons. *Clinical Psychology Review, 28*(5)*,* 746-758. https://doi.org/10.1016/j.cpr.2007.10.005

Brown, G. S., Lambert, M. J., Jones, E. R., & Minami, T. (2005). Identifying highly effective psychotherapists in a managed care environment. *American Journal of Managed Care, 11*(8), 513-520. https://pubmed.ncbi.nlm.nih.gov/16095437/

Budge, S. L., Owen, J. J., Kopta, S. M., Minami, T., Hansen, M. R., & Hirsch, G. (2013). Differences among trainees in client outcomes associated with the phase model of change. *Psychotherapy, 50*(2), 150–157. https://doi.org/10.1037/a0029565

Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: Systematic review. *Journal of Advanced Nursing, 66*(1), 13-15. https://doi.org/10.1111/j.1365-2648.2009.05240.x

Chow, D. L., Seidel, J. A., Miller, S. D., Kane. R. T., Thornton. A., & Andrews, W. P. (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: Research and Practice.* Advance online publication. https://doi.org/10.1037/pro0000318

Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Cohen, J. (1992) Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101. https://doi.org/10.1111/1467-8721.ep10768783

Cuijpers P., Cristea I. A., Karyotaki E., Reijnders M., & Huibers M. J. (2016). How effective are cognitive behaviour therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *Official Journal of the World Psychiatric Association (WPA),* 15(3), 245–258. https://doi.org/10.1002/wps.20346

De Groot, A. (1978) *Thought and choice in chess* (2nd ed.). Mounton Publishers.

De Maat, S., Dekker, J., Schoever, R., & De Jonghe, F. (2006) Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: A meta-analysis. *Psychotherapy Research, 16*(5), 566-578. https://doi.org/10.1080/10503300600756402

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 book of expertise and expert* performance (pp. 683–704). Cambridge University Press http://doi.org/10.1017/CBO9780511816796.038

Ericsson, K. A., & Harwell, K. W. (2019) Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matter and recommendations for future research. *Frontiers in Psychology, 10.* https://doi.org/10.3389/fpsyg.2019.02396

Ericsson, K. A., & Lehmann, A. C. (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. A., & Pool, R. (2016) *Peak: Secrets from the new science of expertise.* Houghton Mifflin Harcourt.

Ericsson, K. A., Krampe, R. T., & 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.A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: prospects and limits* (pp. 1-38). Cambridge University Press. https://doi.org/10.1002/acp.2350060511

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, *16*(5), 319–324. https://doi.org/10.1037/h0063633

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Field, A. (2005) Intraclass correlation. In Everitt, B. S., & Howell, D. C. (Eds.), *Encyclopaedia of Statistics in Behavioral Science*. John Wiley & Sons. https://doi.org/10.1002/0470013192.bsa313

Friedman, M. A., Detweiler-Bidell, J. B., Leventhal, H. E., Horne, R., Keitner, G. I., & Miller, I. W. (2004). Combined psychotherapy and pharmacotherapy for the treatment of major depressive disorder. *Clinical Psychology: Science & Practice*, *11*(1), 47–68. https://doi.org/10.1093/clipsy.bph052

Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown and Company.

Gloster, A. T. Rinner, M. T. B., Iannou, M., Vilanueva, J., Block, V. J., Ferrari, G., Benoy, C., Bader K. & Karekla, M. (2020) Treating treatment non-responders: A meta-analysis of randomized controlled psychotherapy trials. *Clinical Psychology Revire, 75.* https://doi.org/10.1016/j.cpr.2019.101810

Gobet, F. & Ereku, M. H. (2014) Checkmate to deliberate practice: the case of Magnus Carlsen. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.00878

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, *43*(1), 159–172. https://doi.org/10.1037/0012-1649.43.1.159

Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W. T., Whipple, J. L., Miller, S. D., & Wampold, B. E. (2016b). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy*, *53*(3), 367-375. https://doi.org/10.1037/pst0000060

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

Goodyear, R. K., Wampold, B. E., Tracey, T. J., & Lichtenberg, J. W. (2017). Psychotherapy expertise should mean superior outcomes and demonstrable

improvement over time. *The Counseling Psychologist, 45*(1), 54-65. https://doi.org/10.1177/0011000016652691

Gwet, K., L. (2014) *Handbook of inter-rater reliability* (4th ed). Advance- Analytics.

Hambrick, D. Z, Altmann, E. M., Oswald, F. L., Meinz, E. J. & Gobet, F. (2014a). Facing facts about deliberate practice. *Frontiers in psychology, 5.* https://doi.org/10.3389/fpsyg.2014.00751

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014b). Deliberate practice: Is that all it takes to become an expert? *Intelligence* 45, 34–45. https://doi.org/10.1016/j.intell.2013.04.001

Hatfield, D., McCullough, L., Frantz, S. H., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17*(1), 25-32. https://doi.org/10.1002/cpp.656

Hill, C. E., Kivlighan, D. M., III, Rousmaniere, T., Kivlighan, D. M., Jr., Gerstenblith, J. A., & Hillman, J. W. (2019). Deliberate Practice for the Skill of Immediacy: A Multiple Case Study of Doctoral Student Therapists and Clients. *Psychotherapy*. Advance online publication. https://doi.org/10.1037/pst0000247

Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan Jr, D. M., & Gelso, C. J. (2017). Therapist expertise in psychotherapy revisited. *The Counseling Psychologist, 45*(1), 7-53. https://doi.org/10.1177/0011000016641192

Hubble, M. A., & Chow, D. L. (2017). Professional development: From oxymoron to reality. In T. Rousmaniere, R. K. Goodyear, S. D. Miller, & B. E. Wampold (Eds*.), The cycle of excellence: Using deliberate practice to improve supervision and training* (pp. 23–47). Wiley-Blackwell. https://doi.org/10.1002/9781119165590

Imel, Z. E., Malterer, M. B., McKay, K. M., & Wampold, B. E. (2008) A meta-analysis of psychotherapy and medication in unipolar depression and dysthymia.

Journal of Affective Disorder, 110(3), 197-206. https://doi.org/10.1016/j.jad.2008.03.018

Imel, Z. E., Wampold, B. E., Miller, S. D., & Fleming, R. R. (2008) Distinctions without a difference: Direct comparisons of psychotherapies for alcohol use disorders. *Psychology of Addictive Behaviors, 22*(4), 533-543. https://doi.org/10.1037/a0013171

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M., A., Johasen, M., Grimshaw, J., & Oxman, A. D. (2012). Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews, 6.* https://doi.org/10.1002/14651858.CD000259

Kamenov, K., Twomey, C., Cabello, M., Prina, A. M., & Ayuso-Mateos, J. L. (2016) The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: a meta-analysis. *Psychological Medicine, 47*(3), 414–425. https://doi.org/10.1017/S0033291716002774

Karyotaki, E., Smit, Y., Henningsen, K. H., Huibers, M. J. H., Robays, J., Beurs, D., & Cuijipers, P. (2016) Combining pharmacotherapy and psychotherapy or monotherapy for major depression? A meta-analysis on the long-term effects. *Journal of Affective Disorders, 194, 144-152.* https://doi.org/10.1016/j.jad.2016.01.036

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

Kühne, F., Maas, J., Wiesenthal, S., & Weck, F. (2019) Empirical research in clinical supervision: a systematic review and suggestions for future studies. *BMC Psychology, 7*(1), *54.* https://doi.org/10.1186/s40359-019-0327-7

Lambert, M. J. (2013a). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin & Garfield's Handbook of Psychotherapy and Behavior Change* (6th ed., pp. 169-218.). John Wiley & Sons Inc.

Lambert, M. J. (2013b). Outcome in psychotherapy: the past and important advances. *Psychotherapy, 50*(1), 42-51. https://doi.org/10.1037/a0030682

Lambert, M. J. (2015). Progress Feedback and the OQ-System: The Past and the Future. *Psychotherapy, 52*(4), 381-390. https://doi.org/10.1037/pst0000027

Lee, D. J., Schnitzlein, C. W., Wolf, J. P., Vythilingam, M., Rasmusson, A. M., & Hoge, C. W. (2016). Psychotherapy versus pharmacotherapy for posttraumatic stress disorder: Systemic review and meta-analyses to determine first-line treatments. *Depression and Anxiety, 33*(9), 792-806. https://doi.org/10.1002/da.22511

Luborsky, L., Singer, B., & Luborsky, L. (1975) Comparative studies of psychotherapies: Is it true "Everyone has won and all must have prizes"?. *Archive of General Psychiatry, 32*(8), 995-1008. https://doi.org/10.1001/archpsyc.1975.01760260059004

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychological Science, 25*(8), 1608-1618. https://doi.org/10.1177/0956797614535810

Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2016) The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspectives on Psychological Science*, *11*(3), 333-350. https://doi.org/10.1177/1745691616635591

Malin, J. A. (2016) *Therapist expressed empathy across experiential treatment for depression: its growth and relationship to other psychotherapy processes* [Unpublished doctoral dissertation]. The York University.

Marôco, J. (2007) *Análise estatística com utilização do SPSS* (3rd ed.). Edições Sílabo.

McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review

of the evidence. *Academic Medicine*, *86*(6), 706–711. https://doi.org/10.1097/ACM.0b013e318217e119

McGraw, K. O., & Wong, S. P. (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46. https://doi.org/10.1037/1082-989X.1.1.30

McHugh, R.K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013) Patient preference for psychological vs. pharmacological treatment of psychiatric disorders: A meta-analytic review. *Journal of Clinical Psychiatry, 74*(6), 595-602. *https://doi.org/10.4088/JCP.12r07757*

Méndez-Bustos, P., Calati, R., Rubio-Ramírez, F., Olié, E., Courtet, P., & Lopez-Castroman, J. (2019) Effectiveness of psychotherapy on suicidal risk: A systematic review of observational studies. *Frontiers in Psychology, 10, 277.* https://doi.org/10.3389/fpsyg.2019.00277

Messer, S. B., & Wampold, B. E. (2002) Let´s face facts: Common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice, 9*(1), 21-25. https://doi.org/10.1093/clipsy.9.1.21

Miller, S, D., Chow, D., Wampold, B, E., Hubble, M, A., Del Re, A, C., Maeschalck, C., & Bargmann, S. (2018). To be or not to be (an expert)? Revisiting the role of deliberate practice in improving performance. *High Ability Studies.* https://doi.org/10.1080/13598139.2018.1519410

Miller, S. D., Ducan, B. L., Brown, J. Sorrell, R., & Chalk, M.B. (2006) Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *Journal of Brief Therapy, 5*(1), 5-22. https://www.semanticscholar.org/paper/Using-Formal-Client-Feedback-to-Improve-Retention-Miller-Duncan/8315c783b6c64f95dbcd82cd025599dc7d1205d1

Miller, S. D., Hubble, M. A., Chow, D. L., & Seidel, J. A. (2013). The outcome of psychotherapy: yesterday, today, and tomorrow. *Psychotherapy, 50*(1), 88-97. https://doi.org/10.1037/a0031097

Miller, S. D., Wampold, B., & Varhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, *18*(*1*), 5-14. https://doi.org/10.1080/10503300701472131

Miller, S.D., Hubble, M.A., & Duncan, B.L. (2007). Supershrinks: Learning from the field's most effective practitioners. *The Psychotherapy Networker, 31(6),* 26-3.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomized trials. *Journal of Clinical Epidemiology, 63*(8), 1-37. https://doi.org/10.1016/j.jclinepi.2010.03.004

Munder, T., Flückiger, C., Leichsenring, F., Abbass, A. A., Hilsenroth, M. J., Luyten, P. Rabung, S., Steinert, C. & Wampold, B. E. (2019) Let's Distinguish Relative and Absolute Efficacy to Move Psychotherapy Research Forward. *Zeitschrift Psychosomatic Medicine and Psychotherapy*, *65*(2), 178-182. https://doi.org/10.13109/zptm.2019.65.2.178

Okiishi, J., Lambert, M. J., Nielsen, S. L., & Ogles, B. M. (2003). Waiting for supershrink: An empirical analysis of therapist effects. *Clinical Psychology & Psychotherapy, 10*(*6*), 361-373. https://doi.org/10.1002/cpp.383

Østergård, O. K., Randa, H., & Hougaard, E. (2020) The effect of using the Partners for Change Outcome Management System as feedback tool in psychotherapy—A systematic review and meta-analysis. *Psychotherapy Research, 30*(2), 195-212. https://doi.org/10.1080/10503307.2018.1517949

Owen, J., Wampold, B., Kopta. M., Rousmaniere, T., & Miller, S. (2016) As good as it gets? Therapy outcomes of trainees over time. *Journal of Counseling Psychology, 63*(1), 12-19. https://doi.org/10.1037/cou0000112

Padro-Abril, J., Gimeno-Peón., A., Inchausti, F., & Sánchez-Reales, S. (2019) Expertise, Therapist effects and deliberate practice: The cycle of excellence.

Papeles del Psicólogo, 40(2), 89-100. https://doi.org/10.23923/pap.psicol2019.2888

Perlman, M. R., Anderson, T., Victoria, K. F., Mimnaugh, S., & Safran, J. D. (2020) The impact of alliance-focused and facilitative interpersonal relationship training on therapist skills: An RCT of brief training. *Psychotherapy Research,* 1-14. https://doi.org/10.1080/10503307.2020.1722862

Perry, J. C., Banon, E., & Ianni, F. (1999) Effectiveness of psychotherapy for personality disorders. *American Journal of Psychiatry, 15*(9)*,* 1312-1321.

Reynolds, S., Wilson, C., Austin, J., & Hooper, L. (2012) Effects of psychotherapy for anxiety in children and adolescents: A meta-analytic review. *Clinical Psychology Review, 32*(4)*,* 251-262. https://doi.org/10.1176/ajp.156.9.1312

Rosenthal, R. (1963). On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist, 51,* 268–283.

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry, 6*(3)*,* 412–415. https://doi.org/10.1111/j.1939-0025.1936.tb05248.x

Rousmaniere, T. (2014). Using technology to enhance clinical supervision and training. In C. E. Watkins, Jr. & D. L. Milne (Eds.), *The Wiley international handbook of clinical supervision* (p. 204–237). Wiley-Blackwell.

Rousmaniere, T. (2017) Deliberate practice for psychotherapists: A guide to improving clinical effectiveness. Routledge.

Rousmaniere, T., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (Eds.). (2017*). The cycle of excellence: Using deliberate practice to improve supervision and training.* John Wiley & Sons.

Rousmaniere, T., Swift, J. K., Babins-Wagner, R., Whipple, J. L., & 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., Swift, K. J., Banins-Wagner, R. & Berzins, S. (2014) Supervisor variance in psychotherapy outcome in routine practice. *Psychotherapy Research.* https://doi.org/10.1080/10503307.2014.963730

Seligman, M. E. P. (1995) The effectiveness of psychotherapy: The consumer reports study. *American Psychologist, 50*(12), 965-974. https://doi.org/10.1037/0003-066X.50.12.965

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*(2), 252–266. https://doi.org/10.1016/0749-5978(92)90064-E

Shaughnessy, J. J., Zechmeister, E. B. & Zechmeister, J. S. (2012) *Reserach methods in psychology* (9th ed.). McGraw-Hill.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298–31. https://doi.org/10.1037/a0019247

Singh, N., & Reece, J. (2013) Psychotherapy, pharmacotherapy, and their combination for adolescents with major depressive disorder: A meta-analysis. *The Australian Educational and Developmental Psychologist, 31*(1), 47-65. https://doi.org/10.1017/edp.2013.20

Smith, M. L., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*(9), 752–760. https://doi.org/10.1037/0003-066X.32.9.752

Spielmans, G. I., Berman, M. I., & Usitalo, A. N. (2011) Psychotherapy versus second-generation antidepressants in the treatment of depression: A meta-analysis. T*he Journal of Nervous and Mental Disease, 199*(3), 142-149. https://doi.org/10.1097/nmd.0b013e31820caefb

Swift, J. K., Callahan, J. L., & Vollmer, B. M. (2010) Preferences. *Journal of Clinical Psychology, 67*(2)*,* 155-165. https://doi.org/10.1002/jclp.20759.

Theravue (2020) Theravue.com. Available at: https://www.theravue.com/ (accessed March 2020).

Tracey, T. Wampold, B., Lichtenberg, J. W., & Goodyear, R. K. (2014) Expertise in psychotherapy: An elusive goal?. *American Psychologist, 69*(3), 218-229. https://doi.org/10.1037/a0035099

Vaz, A., & Rousmaniere, T. (2020). Deliberate Practice Coach Competency Scale. *Unpublished manuscript. Retrieved from ttps://www.dpfortherapists.com/dp-forms*

Walfish, S., McAlister, B., O'donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological Reports, 110*(2), 639-644. https://doi.org/10.2466/02.07.17.PR0.110.2.639-644

Wampold, B. E. (2005a). Establishing specificity in psychotherapy scientifically: Design and evidence issues. *Clinical Psychology: Science & Practice, 12*(2)*,* 194–197. https://doi.org/10.1093/clipsy.bpi025

Wampold, B. E. (2005b) How important are common factors in psychotherapy? Un update. *World Psychiatry,14*(3)*,* 270-277. https://doi.org/10.1002/wps.20238

Wampold, B. E. (2007) Psychotherapy: The humanistic (and effective) treatment. *American Psychologist, 62(8),* 857–873. https://doi.org/10.1037/0003-066X.62.8.857

Wampold, B. E. (2013). The good, the bad, and the ugly: a 50-year perspective on the outcome problem. *Psychotherapy*, *50*(1), 16-24. https://doi.org/10.1037/a0030570

Wampold, B. E. (2017). What Should We Practice? In T. Rousmaniere, R. K. Goodyear, S. D. Miller, & B. E. Wampold (Eds.), *The Cycle of Excellence: Using Deliberate Practice to Improve Supervision and Training.* (p. 49-65). John Wiley & Sons.

Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology, 73*(5)*,* 914–923. https://doi.org/10.1037/0022-006X.73.5.914

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work* (2nd ed.). Routledge.

Wampold, B. E., Flückiger C., Del Re, A. C., Yulish, N, E., Frost, N. D., Pace, B. T., Goldberg, S.B., Miller, S.D., Baarsdeth, T.P., Laska, K.M., & Hilsenroth, M.J. (2017) In pursuit of truth: A critical examination of meta-analyses of cognitive behavior therapy. *Psychotherapy Research* 27(1), 14–32. https://doi.org/10.1080/10503307.2016.1249433

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin, 122*(3)*,* 203–215. https://doi.org/10.1037/0033-2909.122.3.203

Wheeler, S., & Richards, K. (2007) The impact of clinical supervision on counsellors and therapists, their practice and their client: A systematic review of the literature. *Counselling and Psychotherapy Research*,*7*(1)*,* 54-65. https://doi.org/10.1080/14733140601185274

Appendices

Appendix 1) Enquadramento Teórico

Eficácia em psicoterapia

A literatura demonstra que a psicoterapia é eficaz (Lambert, 2013a, 2013b; Wampold, 2013). Apesar de alguns autores ponderarem a ambiguidade dos resultados relativamente a outros tratamentos (e.g: psicofarmacoterapia), a eficácia da psicoterapia quando comparada com uma condição sem tratamento, é um efeito estabelecido (Lambert, 2013a; Munder et al., 2019; Wampold, 2007). Em média, 65% a 80% das pessoas a receber tratamento em psicoterapia, têm níveis de saúde mental superior à amostra não tratada (Lambert, 2013a; Wampold, 2013). Isto é, a eficácia absoluta da psicoterapia está comprovada (relativo à relevância clínica da psicoterapia para o paciente), a eficácia relativa (referente à relevância clínica da psicoterapia quando comparada com outras intervenções) tem resultados, por vezes ambíguos (Munder et al., 2019). Esta pode demonstrar taxas de eficácia iguais ao tratamento farmacológico (De Maat et al., 2006; Imel et. al.,2008; Spielmans et al., 2011) à junção de ambos (psicoterapia e tratamento farmacológico) (Seligman, 1995; Singh & Reece, 2013) ou, por vezes, resultados inferiores (Luborsky et al., 1975). Contudo, são vários os estudos que demonstram que a intervenção psicoterapêutica tem resultados mais duradouros do que a intervenção psicofarmacológica (De Maat et al., 2006; Imel et al., 2008; Lambert, 2013a; Wampold, 2013).

A psicoterapia é eficaz em adultos, crianças, adolescentes e pessoas com diferentes diagnósticos, como é o caso de: perturbações do humor (Lambert, 2013a), perturbações da personalidade (Perry et al., 1999), pacientes em risco de suicídio (Mendéz-Bustos et al., 2019), pacientes com depressão (Kamenov et. al., 2016) e depressão Major (Cuijpers et. al., 2016), pacientes com stress pós traumático (Lee et al., 2016), perturbação de ansiedade em adultos (Cuijpers et al., 2016) e em crianças e adolescentes (Reynolds et. al., 2012). Quando comparados com amostra em condições de placebo ou lista de espera, os pacientes em tratamento de psicoterapia têm melhorias significativas que se mantém após 2 e 3 anos de *follow-up* (Lambert, 2013a). A investigação demonstra também que a combinação de terapêutica medicamentosa e psicoterapia é mais eficaz que cada uma em separado (Luborsky et al., 1975), para

pacientes com depressão (Cuijpers et. al., 2015; Kamenov et. al., 2016) e depressão major (Friedman et al., 2004) sendo que, em pacientes com depressão major estes resultados ocorrem em fase aguda de tratamento e em tratamento de manutenção (e.g: sessões mensais cujo objetivo é a prevenção de recaída) (Karyotaki et al.,2016). Deste modo, tendo em conta os resultados do tratamento psicoterapêutico em diversas amostras, independentemente da sua eficácia ser aumentada com a junção de tratamento psicofarmacológico, pode-se afirmar que a psicoterapia por si, é um método eficaz para a intervenção em diversas perturbações psicológicas.

As pessoas preferem recorrer a um tratamento psicológico ao invés de tratamentos psicofarmacológicos e caso esta questão se direcione especificamente ao tratamento para ansiedade ou depressão, 75% das pessoas prefere tratamento psicológico (McHugh et al., 2013). Sabe-se que a vontade de iniciar um tratamento psicológico, melhora os resultados do mesmo (Seligman, 1995) e que ter em conta as preferências do paciente (e.g: idade, género e etnia do terapeuta, ou abordagem terapêutica) tem uma componente preditiva dos resultados terapêuticos (Norcross & Wampold, 2013; Swift et al., 2010). Adicionalmente, os tratamentos psicológicos têm menores taxas de *droupout* do que os psicofarmacológicos, menores recaídas na avaliação de *follow-up,* havendo uma taxa de recaídas duas vezes maior para pacientes em tratamento psicofarmacológico (De Maat et al., 2006). O primeiro passo da investigação em psicoterapia foi estabelecer a eficácia das intervenções terapêuticas. Após se estabelecer que a psicoterapia é eficaz, o rumo da investigação em psicoterapia alterou-se para tentar compreender o porquê da psicoterapia ser eficaz e qual a abordagem mais eficaz (Miller et al., 2008; Seligman, 1995; Smith & Glass, 1977; Wampold et al., 1993; Wampold et al., 1997; Wampold & Imel, 2015).

Diversas abordagens para um mesmo objetivo

O debate sobre a eficácia das diferentes abordagens psicoterapêuticas remonta aos anos 50, altura em que Eysenck, inicia um debate sobre a eficácia das diferentes abordagens de psicoterapia, que dura até os dias de hoje. Em 1952, o autor publica um artigo intiludado *The effects of psychoterapy: An evaluation,* que defende que a psicoterapia (especificamente, psicanálise, psicoterapia de orientação psicanalítica e eclética) não é eficaz e pode ser prejudicial, pelo menos para pacientes neuróticos (Eysenck, 1952). O seu artigo de 1952 criou uma onda de debate, cujo um dos

principais impulsionadores foi Struup (1964) com o seu artigo *The outcome problem in psychoterapy: A reply.* Este debate gerou uma enorme produção de investigação acerca da eficácia da psicoterapia, e mais precisamente, da superioridade de uma abordagem psicoterapêutica. Surge assim, a primeira meta-análise que tem em conta diversas abordagens psicoterapêuticas comportamentais e não comportamentais (e.g: psicodinâmica, Gestalt, Racional-emotiva, entre outras), que demonstra que a eficácia da psicoterapia não difere consoante a sua abordagem (Smith & Glass, 1977). Estes resultados, têm vindo a ser replicados em diversas meta-análises até aos dias de hoje, com amostras de adultos (Benish et. al., 2008; Gloster et al., 2020; Imel et al., 2008; Luborsky et al., 1975; Seligman, 1995; Wampold, 1997;), e adolescentes/crianças (Miller et al., 2008).

O fenómeno de igualdade de eficácia entre psicoterapias foi primeiramente abordado por Rosenzweig (1936), que defendia existirem fatores transversais a todas as abordagens terapêuticas que tornavam a terapia eficaz. Este ficou cunhado, após uma revisão de literatura de Luborsky e colaboradores (1975) de *O veredito do pássaro Dodo.* No seu artigo de 1936, Rosenzweig propõe existirem fatores "não reconhecidos" durante o processo terapêutico, que podem ser mais importantes no processo do que aqueles usados propositadamente. Nasce assim a questão << *What do these therapies actually have in common that makes them equally successful?*>>. O autor propõe que um fator necessário à eficácia da terapia é a existência de um terapeuta eficiente que domine um conjunto de técnicas e conceitos que possa adaptar à problemática do paciente, defendendo que este fator tem mais importância que a abordagem de que o mesmo terapeuta faz uso. Atualmente, a investigação sugere que as técnicas de cada abordagem, conhecidas por "ingredientes específicos" não são responsáveis pelos resultados do tratamento (Ahn & Wampold, 2001; Bell et al., 2013), o que vai ao encontro da proposta de Rosenwieig (1936) de que existem fatores transversais a todas as abordagens terapêuticas "fatores comuns" que são responsáveis por parte dos *outcomes* em psicoterapia. Hoje sabe-se que existem vários fatores a influenciar os resultados da psicoterapia (Wampold, 2003). Lambert e Barley (2001) propõe quatro fatores que predizem resultados em psicoterapia, a saber: a) fatores extra-terapia, da vida do cliente, b) técnicas especificas de cada tratamento, c) expectativas e d) fatores comuns. Destes, os fatores extra-terapia explicam 40% da variância dos resultados da intervenção, os fatores comuns a todas as terapias 30%, a expectativa do paciente 15% e

as técnicas especificas de cada abordagem 15%. Assim, podemos concluir que dos fatores que estão ao alcance do terapeuta, são os fatores comuns os que mais explicam os *outcomes* dos paciente, tendo estes mais impacto nos *outcomes* do que os fatores específicos (Messer & Wampold, 2002).

Fatores comuns - A pessoa do terapeuta

Destes fatores ou "ingredientes ativos", fazem parte as variáveis referentes à pessoa do terapeuta, que é um fator fulcral para o resultado da psicoterapia. Estas incluem: a aliança terapêutica, o melhor e mais estudado preditor dos resultados em psicoterapia (Wampold & Imel, 2015), a adesão do terapeuta à abordagem teórica, o efeito de placebo ou expectativa e a competência do terapeuta (Messer & Wampold, 2006). Os efeitos de terapeutas explicam entre 3 a 7% de variabilidade dos *outcomes* da terapia (Baldwin & Imel, 2013; Chow et al., 2015; Jonhs et al., 2019). Este efeito varia entre 0.2 e 29.9% nos 20 estudos incluídos na revisão de Jonhs e colaboradores (2019). No entanto, a variabilidade explicada por efeitos do terapeuta é consistentemente mais alta do que a explicada por diferentes abordagens psicoterapêuticas (Miller et al., 2013). Deste modo, é extremante importante investir na investigação e, consequentemente, no treino e formação dos psicoterapeutas para melhorar o efeito da psicoterapia.

Pode afirmar-se que existem terapeutas que são consistentemente melhores que os seus pares (Baldwin & Imel, 2013; Jonhs et al., 2019; Miller et al., 2007) e que os anos de prática não garantem melhorias na eficácia do terapeutas, podendo até, resultar na deterioração dos resultados (Goldberg et al., 2016a). Também o género, habilitações literárias e grau de integração teórica dos terapeutas não são preditores dos *outcomes* reportados pelo cliente (Chow et al., 2015; Goldberg et al., 2016a). Isto é, terapeutas mais experientes não são, à partida, terapeutas mais eficazes (Brown et al., 2005). Adicionalmente, a maior parte dos terapeutas sobre estima a sua eficácia, sendo que 25% dos terapeutas consideram enquadra-se nos 10% melhores terapeutas (Walfish et al., 2012). Estes têm dificuldade a sinalizar os clientes que estão a piorar durante a terapia (Hatfield et al., 2010). No entanto, sabemos que existem terapeutas que são extremamente eficazes, podendo ter resultados duas vezes melhores que os terapeutas menos eficientes (Wampold & Brown, 2005). Estes tendem a conseguir criar uma boa aliança terapêutica com diferentes pacientes (Baldwin et. al., 2007). Além da capacidade para formar e manter uma boa aliança terapêutica, estes terapeutas mostram

capacidades facilitadoras da relação- conhecidas por *Facilitative Interpersonal Skills* (FIS)- Como é o caso da criação de expectativas positivas, empatia, criação de um racional convincente para o paciente, flexibilidade na sessão e identificação e reparação nas roturas da aliança. Adicionalmente, estes terapeutas tendem a pôr os seus resultados e eficácia em causa e a monitorizar o progresso clínico dos pacientes (Wampold, 2017). Deste modo, é possível concluir que existem terapeutas que, por serem consistentemente melhores no seus resultados, podem ser considerados terapeutas *experts* - *Supershrinks.*

Expertise

De Groot (1946) recria, através da observação de situações reais em jogos de xadrez entre profissionais, um evento de contexto natural no contexto do laboratório, tornando possível a avaliação das *skills* de jogadores profissionais em contexto controlado (Ericsson, 2006). Para compreender e medir a performance dos participantes no seu domínio, é necessário criar uma medida em que seja possível comparar a performance dos mesmos e selecionar aqueles que têm uma performance *expert* (Ericsson & Smith, 1991). O estudo de performance de jogadores de xadrez pode ser operacionalizado através de problemas criados para o jogador resolver, a performance de médicos através de casos de pacientes cujos médicos devem criar um diagnóstico podendo depois confirmar ou não a sua precisão (Pattel et al. 1994), e a dos músicos através excertos musicais que os músicos devem repetir (Ericsson & Lehmann, 1996). Contudo, devido à falta de consenso relativo à expertise em psicoterapia, torna-se difícil operacionalizar a mesma.

A falta de operacionalização de expertise em psicoterapia representa um obstáculo ao estudo de eficácia dos psicoterapeutas (Hill et al., 2017) e divide a comunidade científica na sua definição. Assim, existem autores a defender que a expertise é resultante da prática ao longo do tempo, que leva a melhor performance e, consequentemente, a melhores resultados (Goodyear et al., 2017). Por outro lado, existem autores que defendem que a expertise é a manifestação de altos níveis de competências e que está inserida num continuum que vai de "altamente inexpert a totalmente expert". Defendem também que para aceder à expertise se dever ter em conta, primeiramente, a performance, processamento cognitivo e *outcomes* do clientes, e de um modo mais secundário a experiência, qualidades pessoais e relacionais,

habilitações, reputação e auto-avaliação (Hill et al., 2017). Apesar das diferentes abordagens ao tema, tanto Hill e colaboradores (2017) como Goodyear e colaboradores (2017) propõem que o feedback através da medição dos *outcomes* dos clientes, é fulcral para a melhoria da performance. No entanto, há que ter em conta que 40 % da variância dos *outcomes* dos clientes são explicados por fatores dos próprios, sendo que o terapeuta só explica 5% dos mesmos (Lambert & Barley, 2001), o que pode representar um obstáculo à avaliação da performance real dos terapeutas através dos *outcomes* dos clientes. Contudo, a presença de feedback pode levar a melhorias na performance de terapeutas (Miller et al., 2006; Østergård et al., 2020; Tracy et al., 2014). Assim, o treino em prática deliberada, que inclui o feedback de um *coach* pode ser um processo essencial para atingir a expertise em psicoterapia (Goodyear et al.,2017; Hill et al., 2017).

Prática deliberada

A investigação em prática deliberada surge da necessidade de desenvolver uma explicação científica e, consequentemente, um treino eficaz para a aprendizagem optimal e desenvolvimento de performance excecionais, fora do laboratório (Ericsson, et al., 1993; Ericsson & Harwell, 2019). Ericsson e colaboradores (1993) propõe que uma performance de topo se deve, não (apenas) à presença de qualidades inatas como proposto por Galton (1969), mas também ao número de horas e esforço deliberado na prática da atividade em questão. Isto é, ao número de horas despendido em <<atividades estruturadas criadas especificamente para melhorar a performance num domínio específico>>. No seu estudo sobre performance, os autores descrevem que uma das diferenças entre violinistas "experts", "profissionais" e "bons" é o número de horas que despendem em prática deliberada, sendo que o primeiro grupo dispende em média 10,000 horas, o segundo menos de 8, 000 e o terceiro menos de 5, 000, impulsionando o que mais tarde viria a ser nomeado por Gladwell (2008), a regra das 10,000 horas. Esta defende serem necessárias 10,000 horas para atingir sucesso em qualquer domínio. Assim, os autores concluem que <<diferenças individuais no desempenho final podem ser explicadas *em grande parte* (itálico adicionado pelo autor) por diferentes quantidades de níveis de prática passados e atuais>> (p.392) (Ericsson et al., 1993). Propõem ainda que a idade é preditora da performance. Isto é, quanto mais cedo os participantes têm a primeira apresentação ao domínio a ser treinado, mais cedo

começaram a ter treinos de prática deliberada e, consequentemente, melhor é a sua performance (Ericsson et al., 1993).

Ericsson e Lehmann (1996) propõe quatro condições necessárias à prática deliberada, a saber: 1) tem objetivos de aprendizagem individualizados; 2) os *trainees* têm feedback contínuo sobre desempenho e aprendizagem; pois a mera presença de motivação não garante melhoria de desempenho (Ericsson et., al.,1993**)**, 3) há o envolvimento de um treinador/*coach* mais competente na tarefa; uma vez que, sujeitos numa condição de tutoria ou apoio individualizado tendem a ter melhor performance que numa condição de ensino "tradicional" 4) melhoria sucessiva através da repetição maioritariamente conduzida sozinha, pois a prática solitária é percepcionada, pelos *trainees*, como a condição mais importante para o desenvolvimento da performance (Ericsson et al., 1933). Sugerem ainda que, após dez anos de prática deliberada, a expertise pode ser atingida. Definem prática deliberada como:

> <<Individualized training activities especially designed by a coach or teacher to improve specific aspects of an individual's performance through repetition and successive refinement. To receive maximal benefit from feedback, individuals have to monitor their training with full concentration, which is effortful and limits the duration of daily training. >> (Ericsson & Lehmann, 1996, pp. 278–279)

Esta é uma atividade especialmente criada para a melhoria do nível de performance atual (Ericssson et al., 1993) e manutenção de competências previamente adquiridas (Chow et al., 2015). Pode ser distinguida de atividades laborais que as pessoas tendem a desenvolver no seu dia-a-dia e que não estão associadas a melhor performance. A existência de recompensas externas, competidores, performance pública e custos associados ao erro nas atividades laborais prejudicam a possibilidade de melhoria da performance durante a execução de tarefas. Estes custos levam a que o sujeito faça uso de estratégias previamente existentes para a execução das suas tarefas, o que o impede de conseguir criar ou explorar estratégias necessárias à melhoria de performance, criando-se assim, um ambiente propicio à estagnação da qualidade do desempenho do sujeito (Ericsson et al., 1993). Sabe-se hoje que o número de horas de prática psicoterapêutica não prediz melhores resultados do psicoterapeuta (Hambrick et al., 2014b; Miller et al., 2015; Tracey et al., 2014; Tracey et al., 2015). Deste modo,

torna-se urgente a criação de um espaço e método seguros, em que os psicoterapeutas possam treinar e errar sem pôr em causa os resultados terapêuticos (Rousmaniere, 2017). A prática deliberada tem custos e constrangimentos associados (e.g: tempo despendido no treino, custos monetários associados à necessidade de um *coach*). Assim, o *trainee* deve forcar-se na função instrumental da prática deliberada- a melhoria de performance a longo prazo- pois, a curto prazo, este treino não gera recompensas e não é inerentemente agradável, tendo de ser praticada diariamente por tempo limitado, durante longos períodos temporais, de modo a evitar exaustão (Ericssson et al., 1993). Esta prática implica períodos curtos de grande esforço e atenção, ao longo de vários anos. Assim, é necessário ter em conta a recuperação total após o treino de modo a permitir que o indivíduo consiga manter um nível constante de prática. Caso se mantenha um nível de prática acima daquele que o indivíduo consegue recuperar diariamente, este pode entrar em exaustão mental (e, no caso de treino em psicoterapia, emocional) (Clements-Hickman & Reese, 2020). Assim, a prática deliberada optimal tem de manter um equilíbrio entre esforço e recuperação (Ericsson & Pool, 2016).

Apesar de Ericsson e colaboradores (1993) proporem que as diferenças na performance podem ser explicadas maioritariamente pelas diferenças dos níveis de prática passados e atuais dos sujeitos, ao invés de características inatas do sujeito, alguns autores têm vindo a afirmar que os efeitos da prática deliberada são excessivamente anunciados, tendo resultados mais modestos do que os expectados (Hambrick et al., 2014a/2014b; Macnamara et. al., 2014). Uma meta-análise realizada por Hambrick e colaboradores (2014b), demonstra que a prática deliberada explica *apenas* um terço da variância na performance de participantes nos domínios da música e do xadrez, sendo que dois terços da variância é explicada por outros fatores**.** Adicionalmente, os resultados das horas de prática deliberada parecem variar entre indivíduos com o mesmo nível de performance ao invés de manterem uma relação positiva entre o número de horas de prática e o grau de performance, proposta por Ericsson e colaboradores (1993). O número de horas que um jogador de xadrez pode demorar para atingir o estatuto de "master" pode variar entre 3,016 e 23,608 horas (Gobet & Campitelli, 2007), o que vai contra a ideia de que são precisas 10,000 de treino para atingir *expertise*. Gobet e Ereku (2014) reforçam a ideia de que a prática deliberada é uma condição necessária mas não suficiente para atingir níveis de *expertise*, através do caso do campeão mundial de xadrez: Magnus Carlsen, que tem em

média menos 6.6 anos de prática que os dez melhores jogadores seguintes e que refere só treinar através de atividades que lhe são prazerosas, evitando atividades que não gosta de fazer -uma das condições preconizadas por Ericsson e colaboradores (1993), que referem que a prática deliberada implica a repetição de tarefas monótonas e, por vezes pouco, prazerosas. Por último, enquanto que Ericsson e colaboradores (1993) propõem que quanto mais cedo for introduzido ao domínio em estudo, maior os níveis de prática deliberada e, consequentemente, melhor a performance, os dados atuais demonstram que os participantes com resultados de alta performance não começaram a treinar mais cedo do que os participantes com menor performance e que a idade com que se começa a praticar não prediz maiores níveis de prática deliberada (Gobet & Ereku, 2014; Macnamara et al.,, 2016).

Apesar do tom negativo com que os resultados de Hambrick e colaboradores (2014b) são apresentados, explicando a prática deliberada *apenas* um terço (30%) da variância da performance, é importante ter em conta que conseguir explicar um terço da variância da performance de um terapeuta seria uma excelente ferramenta para os mesmos e para os resultados e eficácia em psicoterapia. Torna-se ainda necessário reforçar que os estudos têm atribuído apenas 1% de variância às técnicas específicas de psicoterapia (Baldwin & Imel, 2013; Miller et al., 2013) e, no entanto, estas são largamente estudadas e ensinadas ao longo da formação de um psicoterapeuta. Mais uma vez, quando se tem em conta o fator idade na iniciação ao domínio em estudo, os dados que "contrariam" os resultados de Ericsson e colaboradores (1993), podem ser benéficos para quem pretende utilizar a prática deliberada para melhorar as competências terapêuticas (Macnamara et al., 2016). Os autores postulam que começar a treinar numa tenra idade é preditor de sucesso do sujeito (Ericsson et al., 1993). No entanto, os resultados têm refutado esta afirmação, mostrando que a idade não está correlacionada com o sucesso da performance ou horas de prática deliberada (Gobet e Ereku, 2014; Macnamara et al., 2016). Tendo em conta que as atividades e tarefas em psicoterapia só começam a ser treinadas com a entrada para a universidade, parece benéfico que a iniciação ao domínio em treino não tenha de ocorrer em idade prematura para o seu sucesso.

Em 2014, uma meta-análise publicada por Macnamara e colaboradores contendo diferentes áreas de estudo: jogos, música, desporto, educação e profissões volta a pôr

em causa a proposta de Ericsson e colaboradores (1993) de que a prática deliberada explica grande parte da performance de topo. Estes resultados demonstram que a relação entre prática deliberada e performance, varia consoante o domínio em estudo. Enquanto que a variância da performance explicada pela prática deliberada nas áreas de jogos (26%), música (21%) e desporto (18%) é alta, esta é baixa em domínios como a educação (4%) e profissões (<1%) em que os resultados não são estatisticamente significativos. Conclui também que o efeito da prática deliberada é maior em atividades altamente previsíveis, como é o caso da corrida, do que em atividade imprevisíveis. Adicionalmente, demonstra que a condição de prática deliberada solitária não é melhor preditora da performance dos participantes do que prática deliberada em grupo. O efeito da prática deliberada parece também variar consoante o nível de performance a ter em conta, enquanto que a prática deliberada parece ter em conta 19% da variância da performance de atletas de sub-elite, esta variância diminui para 1% quando a amostra é constituída apenas por atletas de elite (Macnamara et. al., 2016), sugerindo que outros fatores (talvez talento) sejam responsáveis pela performance da elite. Estes resultados podem ter implicações para a aplicação da prática deliberada em psicoterapia, a saber: a) O efeito da prática deliberada parece ser maior em atividades previsíveis do que atividades imprevisíveis. O *setting* de psicoterapia é um contexto altamente imprevisível em que o terapeuta não consegue antecipar que "ferramentas" serão necessárias a cada sessão (Clements-Hickman & Reese 2020; Tracey et al., 2014); b) A variância atribuída à amostra de sub-elite é muito superior à amostra de elite**.** Contudo, se tivermos em conta que a prática deliberada será um método utilizado para melhorar a performance dos terapeutas e que, mais do que criar *supershrinks*, se quer criar terapeutas competentes, então é necessário que se consiga intervir no maior número de *trainees* possível, para atingir resultados "suficientemente bons", o que reduz a importância do (menor) impacto da prática deliberada em amostras de elite; c) O facto de não existirem diferenças nas condições prática deliberada solitária vs em grupo podem facilitar a aplicação destes métodos em programas de formação de psicoterapia (licenciaturas, sociedades de psicoterapia...) que são, normalmente, lecionados em grupo. Adicionalmente, estudos preliminares têm demonstrado que a aplicação de um treino de PD em grupo pode melhorar as *skills* dos terapeutas (Westra et al., *In press*).

Miller e colaboradores (2018) vêm refutar os resultados de Macnamara e colaboradores (2014), justificando que estes autores incluem indiscriminadamente

estudos que têm em conta prática deliberada e prática "tradicional". Após a exclusão dos artigos que foram incluídos na meta-análise de Macnamara e colaboradores, que não medem prática deliberada, mas sim, horas de treino ou estudo ditos "comuns", Miller e colaboradores consideraram 18 estudos (ao invés de 88), todos da área da educação, que demonstram que a correlação entre horas de prática deliberada e performance é estatisticamente superior à correlação entre horas de prática (não deliberada) e performance. Assim, pode concluir-se que apesar de não garantir, à partida, que uma pessoa se torne um expert, a prática deliberada pode levá-la a melhorar a sua competência (Gobet & Ereku, 2014; Hambrick et al.,2014b; Macnamara et al., 2016; Miller et al., 2018).

Tendo em conta as vantagens e desvantagem acima referidas, torna-se urgente aprofundar a investigação acerca destas práticas em áreas como a psicoterapia. Apesar de poder não explicar a maioria da variância da performance (Hambrick et al., 2014a/2014b; Macnamara, et. al., 2014), e de, como Galton (1969) preconizou, poder haver um impacto das capacidades inatas de cada indivíduo na performance de determinado domínio, parece que a aplicação de um método para melhorar a eficácia dos psicoterapeutas pretende, mais do que criar experts, permitir que cada pessoa consiga melhorar o seu desempenho, tendo em conta as características e condições que lhe são próprias**.** É aqui que propomos que a prática deliberada pode manter papel revolucionário no ensino da psicoterapia.

Prática deliberada em psicoterapia

Como acima descrito, a prática deliberada tem-se mostrado eficaz para a melhoria de várias áreas profissionais, como é o caso da música, desporto, xadrez e medicina (McGaghie et al., 2011). Os resultados destas áreas têm melhorias ao longo do tempo (Ericsson, 1999; Plat, 1966). No entanto, a eficácia dos psicoterapeutas tem vindo a mostrar-se estagnada (Wampold, 2015; Miller et al., 2013; Tracy et al., 2014). Efeito que se pode dever, em parte, à falta treino sistematizado e feedback adequado que os terapeutas, especialmente os recém-licenciados, apresentam (Tracy et al., 2014). O feedback oferecido ao terapeuta tem como objetivo o aumento da sua responsividade ao paciente e à sua problemática (Shimokawa et al., 2010). Este pode ser dado por um *coach*/ supervisor, ou então através de instrumentos de medição do progresso dos pacientes- sistemas de feedback. Uma meta-análise que abrange 140 estudos na área de

cuidados de saúde demonstrou que o feedback promove a aderência dos profissionais de saúde a boas práticas profissionais e melhora os resultados clínicos dos mesmos (Ivers et al., 2012). Estes métodos mostram-se especialmente eficazes quando a) o profissional não está a ter uma boa performance inicial, b) a pessoa que dá o feedback é um supervisor ou um colega e o c) feedback é continuo e inclui planos de ação e objetivos (Ivers et al., 2012). Na área da psicoterapia, os dados demonstram que receber feedback do progresso dos pacientes melhora os *outcomes* dos mesmos (Tracy et al., 2014), sendo que este efeito ocorre em *counselling*, mas não em psiquiatria (Østergård et al., 2020). Adicionalmente, o feedback reduz falhas no tratamento de pacientes que estão em deterioração *"not-on-track patients"* (Shimokawa et al., 2010). Os resultados de Østergård e colaboradores (2020) não corroboram os resultados de Shimokawa e colaboradores (2010), pois a meta-análise dos primeiros autores demonstra que não existem diferenças significativas no uso de feedback em "*not-on-track patients"*. Deste modo, o efeito do feedback é, por vezes, ambíguo na literatura.

A aplicação da prática deliberada em psicoterapia surge numa tentativa de quebrar este estado de *plateau* em que a área se encontra (Tracy et al., 2014). Atualmente, a necessidade de melhoria da performance dos psicoterapeutas é suprimida pela supervisão, educação continuada, disseminação de tratamentos baseados em evidências e sistemas de feedback de resultados (Rousmaniere, et al., 2017). Contudo, a supervisão parece explicar menos de 1% da variância dos *outcomes* da psicoterapia (Rousmaniere, 2014), a formação continua tende a ter um cariz teórico em vez de prático (Rousmaniere, et al., 2017) que é importante para o desenvolvimento do conhecimento na área, mas pode não suprimir as necessidades de melhoria na performance dos psicoterapeutas e, como anteriormente mencionado, as práticas baseadas na evidência cientifica (com suporte empírico) têm mostrado explicar apenas 0 a 4% da variância dos resultados terapêuticos (Wampold, 2005a). Assim, a prática deliberada parece ser um meio promissor para melhorar a competência (e talvez atingir a expertise) clínica (Miller et al., 2005; Miller et al., 2013).

A prática deliberada em psicoterapia apesenta 5 princípios, a saber: a) observação do trabalho e desempenho atual; b) recepção de feedback de desempenho individualizado; c) criação de pequenas metas de aprendizagem dentro da capacidade atual do *trainee*; d) ensaio comportamental com foco nos objetivos de aprendizagem

identificados anteriormente; e) avaliação do desempenho ao longo do tempo. Adicionalmente, é importante que o ambiente em que a prática deliberada ocorre seja um ambiente seguro em que o *trainne* sinta que pode errar e expressar as suas dúvidas e ansiedades, pois só assim conseguirá melhorar o seu desempenho (Rousmaniere, 2017). Contrariamente ao processo de supervisão, em que se costuma debater o caso em questão, a prática deliberada é focada em *skills* específicas do terapeuta e não apenas na conceptualização e diagnóstico dos casos (Rousmaniere, 2017).

Scott Miller e colaboradores são os primeiros investigadores a levantar a questão: <<Porque é que ainda não existe prática deliberada em psicoterapia?>> (Miller et al., 2007; Miller et al., 2013; Rousmaniere, 2017). Atualmente, a investigação tem demonstrado que a prática deliberada melhora o desempenho dos terapeutas (Chow et al., 2015; Goldberg et al., 2016b) e que o número de horas que os terapeutas dispendem fora do trabalho em prática deliberada solitária prediz melhores resultados terapêuticos, contrariamente ao número de horas despendido em atividades solitárias e atividades não terapêuticas, que não prediz resultados terapêuticos (Chow et al., 2015). Adicionalmente, os terapeutas mais eficazes dispendem, em média, mais do dobro do tempo em prática deliberada do que os terapeutas menos eficazes (Chow et al., 2015). Estes percepcionam mais desgaste cognitivo na visualização das suas próprias sessões durante a prática deliberada (Chow et al., 2015). Também em estudantes universitários o treino repetido parece melhorar a qualidade de intervenção dos mesmos. Um estudo realizado por Anderson e colaboradores (2019) em que estudantes universitários dão respostas (com intuito terapêutico) a vídeos em que atores falam de uma problemática como numa sessão de psicoterapia, demonstra que a repetição de respostas melhora a qualidade das intervenções..

Limitações da investigação em prática deliberada em psicoterapia

Apesar de existirem alguns artigos teóricos que sublinham a necessidade e importância da prática deliberada no ensino e treino de psicoterapeutas (Goodyear et al., 2017; Hill et al., 2017; Prado-Abril et al., 2019; Tracey et al., 2014), existem dois estudos correlacionais (Chow et al., 2015; Goldberg et al., 2016b) e três estudos quase-experimentais (Hill et al., 2019; Pearlman et al., 2020; Westra et. al., *In press*) a abordar este tema. Os delineamentos correlacionais (Chow et al., 2015) e estudo de caso (Goldberg et. al., 2016b) acarretam algumas limitações relativas à generalização e

inferência de causalidade dos resultados (Shaughnessy et al., 2012). Os estudos com um delineamento quase-experimental, apesar de necessários para inferir causalidade, são muito escassos na área da prática deliberada (Miller et al., 2018). Contudo, os treinos de prática deliberada têm demonstrado melhorias nas *skills* dos participantes (Hill et al., 2019; Pearlman et al., 2020; Westra et al., *In press*), sendo que estas nem sempre têm impacto nos *outcomes* dos pacientes (Hill, et. al., 2019).

Recentemente, Anderson e colaboradores (2019) publicaram um estudo com delineamento quase-experimental. Este aborda o impacto de *Modelling* e prática sistemática de *Facilitative interpersonal skills (FIS)* (fluência verbal, esperança/ expectativas positivas, persuasão, emoção expressa, aceitação e compreensão, empatia, capacidade para formar aliança teraprutica e capacidade para reparar roturas na aliança terapêutica, responsividade) na melhoria da performances dos psicoterapeutas. Apesar de representar um avanço na investigação de prática deliberada em psicoterapia, este estudo não aborda a prática deliberada na sua totalidade. Como referem os autores, o estudo apenas assegura duas das principais componentes da Prática deliberada, a saber: a) treino estruturado e repetido e b) foco num objetivo específico (por exemplo, melhorar apenas o FIS em vez de melhorar a intervenção terapêutica em geral). Para assegurar um treino de prática deliberada é importante garantir: uma condição em que os participantes a) observem seu próprio trabalho, b) tenham objetivos de aprendizagem pequenos e específicos (micro objetivos) que envolvam um certo grau de desafio (nem demasiado difícil, que leve os participantes a não conseguir trabalhar sobre ele, nem tão fácil que não permita novas aprendizagens aos participantes), c) um treino estruturado e repetido sobre a habilidade/*skill* específica d) tenham feedback de um especialista e e) monitorização sistemática e continua do desempenho (Rousmaniere, 2017). Os estudos experimentais têm demonstrado eficácia da prática deliberada na auto-percepção de competência dos *trainees* (Hill et al, 2019; Westra et. al., *In press*) e na avaliação de juízes independentes (Anderson et al., 2020; Pearlman et. al., 2020). Contudo, nem sempre esta tem impacto nos outcomes dos pacientes (Hill et al, 2019) o que pode sugerir a ocorrência de *self-seving bias.* Deste modo, é fulcral expandir a investigação em prática Deliberada aplicada ao treino de psicoterapeutas com *trainees* em diferentes estádios de aprendizagem/ experiência, isolando apenas uma variável em análise que seja transversal às diferentes abordagens e contribua para o *outcome* dos pacientes.

Appendix 2) Measure of Expressed Empathy (MEE)

Measure of Expressed Empathy (MEE)

(Watson, J. C., 1999)

Unpublished Measure, Department of Adult Education & Counselling Psychology
OISE/University of Toronto, Ontario, Canada

Five-minute segments should be rated. Each segment is given a global rating on a nine- point scale on therapist's behaviours that reflects aspects of expressed empathic communication. To score the measure add the items and calculate the mean.

1. Does the therapist's voice convey concern?

Listen for high energy, colour (expressive of the emotions that it is trying to convey, flexible, musical), soft resonance that matches the verbal expression of concern; calmness, a grounded, open quality to the therapist's voice. The voice should not sound rigid)

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

2. Is the therapist's voice expressive?

(Listen for high energy, colour, varied pitch; is it expressive where it needs to be?)

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

3. Does the therapist's vocal tone or response match the intensity of the client's feelings?

(Listen for high energy, colour, emphasis, pitch variation that matches intensity of client's feelings). Note: There are neutral states and in that case the therapist would match that state – doesn't necessarily have to be highly emotional or field with intense feeling. (The vocal tone should convey a sense that therapist can meet the client at the same level of intensity; voice should show that therapist can handle the intensity and can hold client's feelings e.g. show comfort when client is depressed; A score of 0 = nonchalant, non-caring attitude captured in

vocal tone or complete mismatch between the subject matter that the client is conveying and the therapists response (e.g. vocal tone worried or flat if client excited).

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

4. Does the therapist convey warmth and an atmosphere of safety?

(Does the therapist smile, maintain eye contact, convey softness, and appear receptive to the client's concerns (receptiveness is not involvement; more low key respectful, open); (0 = "cold fish"; blank); (Does the therapist communicate an atmosphere of safety, of "holding" for the client?)

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

5. Is the therapist responsively attuned to the client's inner world moment by moment in the session?

(Does the therapist provide moment-to-moment acknowledgements, not let things go by; pick up the live edges of the client's experience; fine-tune their responses to fit with their client's? Is the therapist attuned to client's facial and/or non-verbal behaviour that may be different from the content of client's responses? Is the therapist attentive to nuances of meaning and feeling (doesn't package what was said and just reiterate it back?). Responses are not just a reflection of surface content but show an understanding of the client's inner world. (Inner world is defined as client's feelings, perceptions, memories, construal, bodily sensations (felt sense, and core values).

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

6. Does the therapist look concerned?

(Does the therapist look engaged and involved and maintain eye contact, or does the therapist look bored, disengaged, blank, and listless? Being attentive is an aspect of concern)

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

7. Is the therapist responsive to the client?

(Does he or she adjust his/her responses to follow the client's track?

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

8. Do the therapist's responses convey an understanding of the client's feelings, and inner experience?

(Do the therapist's responses show a sensitive appreciation and gentle caring for the client's feelings and inner world? Do the therapist's responses convey an emotional understanding of the client's inner world, for example – "so you're just like a little girl in the corner?". Does the therapist convey the emotional meaning and emotional significance of events? Feelings are not just labels of anger, sadness, etc. but can also be metaphors. Keep in mind that if the therapist hasn't said much 5-minute segment that may be appropriate.

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

9. Do the therapist's responses convey an understanding of the client's cognitive framework and meanings?

(It is expected that most therapists will show an understanding of what their clients are saying. To score 0 one person would have to be saying the sky is blue and the other talking about loud music so that there is no overlap in content or continuity between the participants). Ask yourself "Are they on the same page?". Is there a back and forth quality to the interaction? Is the therapist following what the client is saying? To score highly the therapist captures the client's construal/or idiosyncratic perception.

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

10. Is the therapist accepting the clients feeling and inner experience?

(8 = sincere i.e. conveying that you mean what you say – being authentic, open, prizing, genuine; 0 = invalidating of the client's experience and dismissing their perspective or being insincere, putting on an act; trying to appear empathic but coming across as inauthentic.).

0.....2.....4.....6.....8

Never 25% Half the time 75% All the time

Study 1

Appendix 3.1.1) Interrater Agreement

Two-Way mixed effect intraclass correlation coefficient (ICC) for the absolute agreement of multiple raters (k=2).

Scale: ICC Empathy Score Response 1

Case Processing Summary

		N	%
Cases	Valid	36	94,7
	Excluded[a]	2	5,3
	Total	38	100,0

a. Listwise deletion based on all variables in the procedure

Reliability Statistics

Cronbach's Alpha	N of Items
,820	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,700[a]	,485	,835	5,549	35	35	,000
Average Measures	,824[c]	,653	,910	5,549	35	35	,000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Scale: ICC Empathy Score Response 2

Case Processing Summary

		N	%
Cases	Valid	32	84,2
	Excluded[a]	6	15,8
	Total	38	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,744	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,598[a]	,318	,782	3,902	31	31	,000
Average Measures	,749[c]	,483	,878	3,902	31	31	,000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Scale: ICC Empathy Score Response 3

Case Processing Summary

		N	%
Cases	Valid	26	68,4
	Excluded[a]	12	31,6
	Total	38	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,766	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,623[a]	,321	,811	4,279	25	25	,000
Average Measures	,768[c]	,486	,896	4,279	25	25	,000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Appendix 3.1.2) Descriptive Statistics

Case Processing Summary

	1-PD 2-TAU	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Empathy Score Response 1	PD	17	100,0%	0	0,0%	17	100,0%
	TAU	19	100,0%	0	0,0%	19	100,0%
Empathy Score Response 2	PD	16	94,1%	1	5,9%	17	100,0%
	TAU	16	84,2%	3	15,8%	19	100,0%
Empathy Score Response 3	PD	14	82,4%	3	17,6%	17	100,0%
	TAU	12	63,2%	7	36,8%	19	100,0%

Descriptives

	1-PD 2-TAU			Statistic	Std. Error
Empathy Score Response 1	PD	Mean		3,5279	,24113
		95% Confidence Interval for Mean	Lower Bound	3,0168	
			Upper Bound	4,0391	
		5% Trimmed Mean		3,5172	
		Median		3,5750	
		Variance		,988	
		Std. Deviation		,99420	
		Minimum		1,48	
		Maximum		5,78	
		Range		4,30	
		Interquartile Range		1,03	
		Skewness		,075	,550
		Kurtosis		1,081	1,063
	TAU	Mean		2,9961	,29480
		95% Confidence Interval for Mean	Lower Bound	2,3767	
			Upper Bound	3,6154	
		5% Trimmed Mean		3,0026	
		Median		2,7750	
		Variance		1,651	
		Std. Deviation		1,28502	
		Minimum		,70	
		Maximum		5,18	

		Range		4,48	
		Interquartile Range		1,78	
		Skewness		,011	,524
		Kurtosis		-,977	1,014
Empathy Score Response 2	PD	Mean		4,0891	,26062
		95% Confidence Interval for Mean	Lower Bound	3,5336	
			Upper Bound	4,6446	
		5% Trimmed Mean		4,1128	
		Median		4,1500	
		Variance		1,087	
		Std. Deviation		1,04247	
		Minimum		2,08	
		Maximum		5,68	
		Range		3,60	
		Interquartile Range		1,74	
		Skewness		-,152	,564
		Kurtosis		-,778	1,091
	TAU	Mean		3,4375	,28729
		95% Confidence Interval for Mean	Lower Bound	2,8252	
			Upper Bound	4,0498	
		5% Trimmed Mean		3,4181	
		Median		3,5375	
		Variance		1,321	
		Std. Deviation		1,14917	
		Minimum		1,20	
		Maximum		6,03	
		Range		4,83	
		Interquartile Range		1,60	
		Skewness		,255	,564
		Kurtosis		,792	1,091
Empathy Score Response 3	PD	Mean		4,5625	,22623
		95% Confidence Interval for Mean	Lower Bound	4,0738	
			Upper Bound	5,0512	
		5% Trimmed Mean		4,5597	
		Median		4,9000	
		Variance		,717	
		Std. Deviation		,84647	
		Minimum		3,23	
		Maximum		5,95	
		Range		2,73	
		Interquartile Range		1,51	

		Skewness		-,322	,597
		Kurtosis		-1,029	1,154
	TAU	Mean		3,7208	,34420
		95% Confidence Interval for Mean	Lower Bound	2,9633	
			Upper Bound	4,4784	
		5% Trimmed Mean		3,6856	
		Median		3,8375	
		Variance		1,422	
		Std. Deviation		1,19234	
		Minimum		1,63	
		Maximum		6,45	
		Range		4,83	
		Interquartile Range		1,36	
		Skewness		,534	,637
		Kurtosis		2,115	1,232

Appendix 3.1.3) Differences between groups - Mann-Whitney Test

Mann-Whitney Test

Ranks

	1-PD 2-TAU	N	Mean Rank	Sum of Ranks
Empathy_Scores_R1	PD	17	20,59	350,00
	TAU	19	16,63	316,00
	Total	36		
Empathy_Scores_R2	PD	16	19,13	306,00
	TAU	16	13,88	222,00
	Total	32		
Empathy_Scores_R3	PD	14	16,29	228,00
	TAU	12	10,25	123,00
	Total	26		

Test Statistics[a]

	Empathy_Scores_R1	Empathy_Scores_R2	Empathy_Scores_R3
Mann-Whitney U	126,000	86,000	45,000
Wilcoxon W	316,000	222,000	123,000
Z	-1,126	-1,584	-2,006
Asymp. Sig. (2-tailed)	,260	,113	,045
Exact Sig. [2*(1-tailed Sig.)]	,271[b]	,119[b]	,046[b]

a. Grouping Variable: 1-PD 2-TAU

b. Not corrected for ties.

Appendix 3.1.4) Differences within groups - Friedman's ANOVA

Deliberate Practice

Friedman Test

Ranks

	Mean Rank
Empathy_scores_R1	1,43
Empathy_scores_R2	2,00
Empathy_scores_R3	2,57

Test Statistics[a]

N	14
Chi-Square	9,143
df	2
Asymp. Sig.	,010

a. Friedman Test

Kendall's W Test

Test Statistics

N	14
Kendall's W[a]	,327
Chi-Square	9,143
df	2
Asymp. Sig.	,010

a. Kendall's Coefficient of Concordance

Training As Usual

Friedman Test

Ranks

	Mean Rank
Empathy_scores_R1	1,67
Empathy_scores_R2	2,25
Empathy_scores_R3	2,08

Test Statistics[a]

N	12
Chi-Square	2,167
df	2
Asymp. Sig.	,338

a. Friedman Test

Kendall's W Test

Test Statistics

N	12
Kendall's W[a]	,090
Chi-Square	2,167
df	2
Asymp. Sig.	,338

a. Kendall's Coefficient of Concordance

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	14
Test Statistic	9,143
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	,010

Pairwise Comparisons (Deliberate Practice Group)

	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Empathy_Scores_R1-Empathy_Scores_R2	-,571	,378	-1,512	,131	,392
Empathy_Scores_R1-Empathy_Scores_R3	-1,143	,378	-3,024	,002	,007
Empathy_Scores_R2-Empathy_Scores_R3	-,571	,378	-1,512	,131	,392

Study 2

Appendix 3.2.1) Interrater Agreement

Two-Way mixed effect intraclass correlation coefficient (ICC) for the absolute agreement of multiple raters (k=2).

Scale: ICC Empathy Score Response 1

Case Processing Summary

		N	%
Cases	Valid	11	16,4
	Excluded[a]	56	83,6
	Total	67	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,573	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,415[a]	-,233	,802	2,340	10	10	,098
Average Measures	,587[c]	-,606	,890	2,340	10	10	,098

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Scale: ICC Empathy Score Response 2

Case Processing Summary

		N	%
Cases	Valid	9	13,4
	Excluded[a]	58	86,6
	Total	67	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,856	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,511[a]	-,120	,869	6,931	8	8	,006
Average Measures	,677[c]	-,273	,930	6,931	8	8	,006

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Scale: ICC Empathy Score Response 3

Case Processing Summary

		N	%
Cases	Valid	8	11,9
	Excluded[a]	59	88,1
	Total	67	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,860	2

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,654[a]	-,002	,919	7,126	7	7	,009
Average Measures	,791[c]	-,004	,958	7,126	7	7	,009

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Appendix 3.2.2) Descriptive Statistics

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	Condição	N	Percent	N	Percent	N	Percent
Empathy_Scores_R1	PD	5	100,0%	0	0,0%	5	100,0%
	TAU	6	100,0%	0	0,0%	6	100,0%
Empathy_Scores_R2	PD	4	80,0%	1	20,0%	5	100,0%
	TAU	5	83,3%	1	16,7%	6	100,0%
Empathy_Scores_R3	PD	4	80,0%	1	20,0%	5	100,0%
	TAU	4	66,7%	2	33,3%	6	100,0%

Descriptives

				Statistic	Std. Error
Empathy_Scores_R1	PD	Mean		5,9650	,54173
		95% Confidence Interval for Mean	Lower Bound	4,4609	
			Upper Bound	7,4691	
		5% Trimmed Mean		6,0347	
		Median		6,4500	
		Variance		1,467	
		Std. Deviation		1,21135	
		Minimum		3,85	
		Maximum		6,83	
		Range		2,98	
		Interquartile Range		1,74	
		Skewness		-1,981	,913
		Kurtosis		4,061	2,000
	TAU	Mean		6,1750	,17488
		95% Confidence Interval for Mean	Lower Bound	5,7255	
			Upper Bound	6,6245	
		5% Trimmed Mean		6,1736	
		Median		6,2250	
		Variance		,184	
		Std. Deviation		,42837	
		Minimum		5,55	
		Maximum		6,83	
		Range		1,28	
		Interquartile Range		,64	

		Skewness		,052	,845
		Kurtosis		,763	1,741
Empathy_Scores_R2	PD	Mean		6,2875	,32105
		95% Confidence Interval for Mean	Lower Bound	5,2658	
			Upper Bound	7,3092	
		5% Trimmed Mean		6,2833	
		Median		6,2500	
		Variance		,412	
		Std. Deviation		,64210	
		Minimum		5,60	
		Maximum		7,05	
		Range		1,45	
		Interquartile Range		1,24	
		Skewness		,247	1,014
		Kurtosis		-2,143	2,619
	TAU	Mean		5,4700	,22310
		95% Confidence Interval for Mean	Lower Bound	4,8506	
			Upper Bound	6,0894	
		5% Trimmed Mean		5,4750	
		Median		5,4750	
		Variance		,249	
		Std. Deviation		,49887	
		Minimum		4,88	
		Maximum		5,98	
		Range		1,10	
		Interquartile Range		,99	
		Skewness		-,110	,913
		Kurtosis		-2,595	2,000
Empathy_Scores_R3	PD	Mean		6,4938	,38221
		95% Confidence Interval for Mean	Lower Bound	5,2774	
			Upper Bound	7,7101	
		5% Trimmed Mean		6,5208	
		Median		6,7375	
		Variance		,584	
		Std. Deviation		,76441	
		Minimum		5,40	
		Maximum		7,10	
		Range		1,70	
		Interquartile Range		1,37	
		Skewness		-1,500	1,014
		Kurtosis		2,128	2,619

	TAU	Mean		6,0750	,31689
		95% Confidence Interval for Mean	Lower Bound	5,0665	
			Upper Bound	7,0835	
		5% Trimmed Mean		6,0944	
		Median		6,2500	
		Variance		,402	
		Std. Deviation		,63377	
		Minimum		5,18	
		Maximum		6,63	
		Range		1,45	
		Interquartile Range		1,15	
		Skewness		-1,402	1,014
		Kurtosis		2,114	2,619

Appendix 3.2.3) Study 1: Differences between groups - Mann-Whitney Test

Rank

	Condição	N	Mean Rank	Sum of Ranks
Empathy_Scores_R1	PD	5	6,50	32,50
	TAU	6	5,58	33,50
	Total	11		
Empathy_Scores_R2	PD	4	6,63	26,50
	TAU	5	3,70	18,50
	Total	9		
Empathy_Scores_R3	PD	4	5,50	22,00
	TAU	4	3,50	14,00
	Total	8		

Test Statistics[a]

	Empathy_Scores_R1	Empathy_Scores_R2	Empathy_Scores_R3
Mann-Whitney U	12,500	3,500	4,000
Wilcoxon W	33,500	18,500	14,000
Z	-,457	-1,599	-1,155
Asymp. Sig. (2-tailed)	,647	,110	,248
Exact Sig. [2*(1-tailed Sig.)]	,662[b]	,111[b]	,343[b]

a. Grouping Variable: Condição

b. Not corrected for ties.

Appendix 3.2.4) Study 1: Differences within groups - Friedman's ANOVA

Deliberate Practice

Friedman Test

Ranks

	Mean Rank
Empathy_scores_R1	2,00
Empathy_scores_R2	1,50
Empathy_scores_R3	2,50

Test Statistics[a]

N	4
Chi-Square	2,000
df	2
Asymp. Sig.	,368

a. Friedman Test

Kendall's W Test

Test Statistics

N	4
Kendall's W[a]	,250
Chi-Square	2,000
df	2
Asymp. Sig.	,368

a. Kendall's Coefficient of Concordance

Training As Usual

Friedman Test

Ranks

	Mean Rank
Empathy_scores_R1	2,25
Empathy_scores_R2	1,00
Empathy_scores_R3	2,75

Test Statistics[a]

N	4
Chi-Square	6,500
df	2
Asymp. Sig.	,039

a. Friedman Test

Kendall's W Test

Test Statistics

N	4
Kendall's W[a]	,813
Chi-Square	6,500
df	2
Asymp. Sig.	,039

a. Kendall's Coefficient of Concordance

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	4
Test Statistic	6,500
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	,039

Pairwise Comparisons (Training As Usual)

	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Empathy_scores_R2-Empathy_scores__R1	1,250	,707	1,768	,077	,231
Empathy_scores__R2-Empathy_scores__R3	-1,750	,707	-2,475	,013	,040
Empathy_scores_R1-Empathy_scores_R3	-,500	,707	-,707	,480	1,000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Appendix 4) Demographic Questionnaire Study 1

Obrigada por participar neste estudo. Pedimos-lhe que preencha um curto questionário demográfico. As suas respostas não serão partilhadas com pessoas externas à equipa responsável pelo estudo.

Eu compreendi todas as informações acima descritas e **consinto** em preencher este questionário

Eu **não consinto** em preencher este questionário

→

Indique o seu género

Masculino

Feminino

Outro/ prefiro não dizer

→

Indique a sua idade (por favor preencha os espaço apenas com **números**)

→

Indique a sua nacionalidade

→

Indique a sua etnia

→

Indique a as suas habilitações literárias e área de estudo/ especialização

→

Há quantos anos faz prática clinica? (Por favor, preencha o espaço utilizando apenas **números**)

→

Obrigado por responder ao questionário.

→

Appendix 5) Demographic Questionnaire Study 2

Obrigada por participar neste estudo. Pedimos-lhe que preencha um curto questionário demográfico. As suas respostas não serão partilhadas com pessoas externas à equipa responsável pelo estudo.

Eu compreendi todas as informações acima descritas e **consinto** em preencher este questionário

Eu **não consinto** em preencher este questionário

→

Indique o seu género

Masculino

Feminino

Outro/ prefiro não dizer

→

Indique a sua idade (por favor preencha os espaço apenas com **números**)

→

Indique a sua nacionalidade

Indique a sua etnia

Indique a as suas habilitações literárias e área de estudo/ especialização

Além da formação na SPPE, tem algum treino, posterior à faculdade, em psicoterapia noutra abordagem (e.g: Cognitivo-Comportamental, Psicanáise ...)

Sim, já fiz a formação completa noutra abordagem

Não, a formação na SPPE é a minha primeira formação numa abordagem de psicoterapia específica

→

Há quantos anos faz prática clinica? (Por favor, preencha o espaço utilizando apenas **números**)

→

Encontra-se a fazer estágio profissional (psicólogo Junior)?

Sim

Não

→

Obrigado por responder ao questionário.

→