ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-09-10

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Sousa, D., Santos, J., Rathenau, S. & Vaz, A. (2026). "Is it really deliberate practice?" A new measure for assessing deliberate practice in psychotherapy. Journal of Psychotherapy Integration, N/A. https://doi.org/10.1037/int0000406

Further information on publisher's website:
10.1037/int0000406

Publisher's copyright statement:
This is the peer reviewed version of the following article: Sousa, D., Santos, J., Rathenau, S. & Vaz, A. (2026). "Is it really deliberate practice?" A new measure for assessing deliberate practice in psychotherapy. Journal of Psychotherapy Integration, N/A. https://doi.org/10.1037/int0000406, which has been published in final form at https://dx.doi.org/10.1037/int0000406. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

"Is it Really Deliberate Practice?"

- **A New Measure for Assessing Deliberate Practice in Psychotherapy**

Daniel Sousa
ISPA – Instituto Universitário, Portugal
Daniel.sousa@ispa.pt
https://orcid.org/0000-0001-8213-4616

João Santos
Centro de Investigação e Estudos em Sociologia, ISCTE – Instituto Universitário de Lisboa
joao_marques_santos@iscte-iul.pt
https://orcid.org/0000-0001-6414-6891

Sara Rathenau
Clínica ISPA, Instituto Universitário, Portugal
sara.marques.rathenau@gmail.com
https://orcid.org/0000-0001-5067-8029

Alexandre Vaz
Sentio University, Los Angeles, California, USA
alexmagvaz@gmail.com
https://orcid.org/0000-0001-8443-0064

Corresponding Author:

Daniel Sousa
ISPA – Instituto Universitário, Portugal
Daniel.sousa@ispa.pt
https://orcid.org/0000-0001-8213-4616

Abastract

Objective: This study aims to develop and validate a new measure for assessing deliberate practice in psychotherapy, addressing the ambiguous components currently identified in the literature.

Method: A total of 442 psychotherapists participated in an online survey, providing data on their engagement in activities related to deliberate practice. Exploratory Factor Analysis (EFA) was conducted on the training sample to identify the underlying structure of the measure, followed by Confirmatory Factor Analysis (CFA) on a holdout sample to validate the instrument.

Results: The final model yielded four distinct factors: Goals and Training, Recording, Feedback, and Monitoring, demonstrating high reliability (Composite Reliability > .70) and strong validity (Average Variance Extracted > .50). While mean scores indicated a general tendency towards high engagement in "Goals and Training" and "Collaborative Feedback and Guidance," lower scores in "Recording" suggest underutilization of session reviews among therapists.

Conclusion: The newly developed measure provides a comprehensive tool for assessing deliberate practice in psychotherapy, highlighting important areas for improvement in training programs. By incorporating direct observation and structured feedback mechanisms, this instrument can enhance therapist development, improve therapeutic practices, and ultimately benefit client outcomes.

Key-words: deliberate practice; psychotherapy; training; outcomes, measure

Clinical and methodological significance of the article: This study offers a significant contribution to both the clinical practice and methodological rigor of psychotherapy training. Clinically, it introduces a psychometrically validated measure of deliberate practice, providing therapists with a tool to assess and enhance their skill development across critical dimensions such as goal setting, session recording, feedback utilization, and client outcome monitoring. Methodologically, the research developed and validated a comprehensive measurement scale, ensuring that future studies can benefit from a robust, reliable assessment instrument. This advancement enables more accurate comparisons and a clearer understanding of the effects of deliberate practice on therapist competence and client outcomes, paving the way for targeted interventions and enhanced training methodologies.

Public Health Statement

This study provides a new, reliable way to measure how psychotherapists engage in deliberate practice, a key method for improving skills and client outcomes. By identifying

strengths and weaknesses in therapists' training, this tool can inform targeted improvements in training programs, ultimately enhancing the quality of mental health services and benefiting the public.

Introduction

Introduction

Deliberate practice (DP) has been identified as a major concept in the study of expertise development across multiple professional domains (Ericsson, 2018). Developed through Ericsson and colleagues' research across a wide range of occupational groups and methodological approaches, DP has been proposed as a way of characterizing a learning process that may be especially important for the development of high-level professional competence (Ericsson, 2018). In psychotherapy, DP has increasingly been proposed as an important framework for improving therapist performance through structured, goal-oriented activities designed to refine clinical skills (Rousmaniere, 2024). In this context, DP has been described as involving activities such as behavioral rehearsal, the identification of specific areas for improvement, and the use of feedback to support ongoing skill development (Barata et al., 2025). One of the central features that distinguishes DP from broader educational or professional development approaches is its emphasis on procedural learning rather than solely on conceptual understanding (Mahon, 2022; Vaz & Rousmaniere, 2022; Vaz et al., 2023). Although conceptual and cognitive learning remain important, their value within DP is thought to be enhanced when paired with repeated skill practice and immediate feedback (Mahon, 2022; Vaz & Rousmaniere, 2022).

At the same time, the broader expertise literature has not treated DP as an uncontested construct (Hambrick et al., 2020). Although the general idea of deliberate practice has been highly influential, researchers have continued to debate both the strength of the claims made for it and the conceptual and methodological difficulties involved in defining and measuring it. In particular, critics have argued that the literature has contained important inconsistencies in what counts as DP, including shifts over time in whether it must be teacher-designed, whether it must be solitary rather than group-based, and where the boundary lies between deliberate practice and related forms of effortful learning such as purposeful practice (Hambrick et al., 2020). Relatedly, the standards used to decide what should count as evidence for DP have also been questioned, making it difficult to compare findings across studies or determine whether researchers are in fact studying the same construct (Hambrick et al., 2020). This matters because, without close examination of what practitioners are actually doing, it can be difficult to know whether they are engaging in DP or in some other kind of learning process. Moreover, findings from broader expertise research suggest that, while DP is often an important predictor of performance, it may not by itself account for most individual differences in expertise, which has led some researchers to argue for more multifactorial models that also consider other experiential, developmental, and ability-related influences (Hambrick et al., 2020).

This broader uncertainty is important to acknowledge because the psychotherapy literature on DP represents only one part of a much larger and more debated body of expertise research.

Within psychotherapy specifically, the application of DP remains conceptually unsettled. The literature has evolved considerably, but there is still substantial disagreement regarding what should count as DP in psychotherapy training and practice (Barata et al., 2024; Nurse et al., 2024; Faber et al., 2021). In particular, the field has been divided into two main approaches (Diamond et al., 2025). One emphasizes skill-based practice focused on identifying deficits in specific clinical performances and addressing them through behavioral rehearsal with immediate coaching or feedback (Rousmaniere, 2024; Rousmaniere & Vaz, 2021). The other places greater emphasis on broader processes associated with effective therapy, such as the therapeutic alliance and client expectancy, and seeks to improve these through exploratory learning projects informed by routine outcome monitoring (Chow et al., 2015; Miller et al., 2023). This divergence has made it more difficult to establish a coherent and widely accepted definition of DP in psychotherapy.

Related conceptual ambiguities concern the core ingredients of DP itself. Authors have differed in the emphasis they place on structured feedback and iterative learning, on repeated skill performance, and on the role of expert guidance in shaping improvement (Mahon, 2022). Similarly, the definition of "observable clinical practices" and the methods used to identify and assess areas for improvement often remain unclear, contributing to variation in how DP is understood and implemented (Ericsson & Pool, 2016; Rousmaniere et al., 2017). In addition, although DP is typically framed as primarily procedural, some work suggests that conceptual understanding may also be necessary for more complete professional development, thereby complicating simpler distinctions between procedural and conceptual learning (Vaz & Rousmaniere, 2022).

Empirical findings on DP in psychotherapy are promising, but the evidence base remains limited. Some studies suggest that psychotherapists who engage in structured practice activities may show enhanced effectiveness in clinical work, with possible benefits for client outcomes (Karow et al., 2020). Other studies have linked reflective practices to greater efficacy in forming therapeutic alliances, a variable consistently associated with positive treatment outcomes (Tomás et al., 2024; Anderson et al., 2022). However, the empirical literature is marked by substantial heterogeneity. A recent systematic review found that among 20 studies, only three met the contemporary definition of DP, with considerable variation in how the construct was operationalized (Diamond et al., 2025). Moreover, the studies included in that review displayed important methodological shortcomings, limiting confidence in the conclusions that can be drawn about the impact of DP in psychotherapy (Vaz et al., 2025).

These limitations have important implications for the field. Inconsistent operationalization makes it difficult to compare studies, synthesize findings, and determine whether researchers are examining the same phenomenon (Barata et al., 2024; Diamond et al., 2025; Miller et al., 2023; Rousmaniere, 2024). More broadly, these challenges mirror the larger debate within expertise research, in which the definition and measurement of DP remain central points of disagreement. Methodological weaknesses, including small samples, lack of control groups, and reliance on subjective self-report measures, further weaken the psychotherapy evidence base (Diamond et al., 2025). As a

result, although the literature suggests that DP may hold promise for therapist development and client care, the current evidence remains too preliminary to strongly support or refute its broader implementation in psychotherapy training.

This state of the literature points to a central problem: the field still lacks sufficiently robust measurement tools for assessing DP in psychotherapy. Existing measures often do not capture the construct in adequate detail and may fail to represent the multidimensional nature of psychotherapeutic work (Clements-Hickman & Reese, 2020; Jones & Taylor, 2021). Traditional evaluation methods may focus on variables such as session frequency or client feedback without adequately assessing the strategies therapists use to actively improve their performance, which can result in superficial feedback that does not foster meaningful growth (Lim & Nguyen, 2023). More broadly, researchers and practitioners continue to face difficulties implementing evidence-based practices when effective tools for assessing therapist competence and professional development are lacking (Cochrane et al., 2022; Khanna et al., 2023). The absence of clear and standardized evaluation criteria may also contribute to variability in therapist training and treatment outcomes (Norcross & Wampold, 2018), thereby constraining professional development and the quality of training available to emerging therapists (Bennett-Levy et al., 2020). In parallel, the American Psychological Association has emphasized the need for more measurable outcomes in training and practice, highlighting the importance of moving beyond subjective feedback toward more standardized assessment approaches (APA, 2021).

A more precise and psychometrically sound measure of DP could therefore help address one of the field's central challenges: inconsistent operationalization. If a measure were developed that adequately captured the core elements of DP as described in the literature, it could provide a useful benchmark for evaluating the fidelity of DP-oriented interventions and for improving comparability across future studies (Vaz & Rousmaniere, 2022). Such a measure may also help address, at least in part, the broader concern that it is often difficult to determine whether practitioners are truly engaging in DP rather than in more general learning or reflective activities. While there remains a significant lack of clarity regarding the specific components of DP in psychotherapy (Vaz et al., 2025), several elements are frequently identified as central to the construct, including direct observation, immediate feedback, the definition of micro-goals, behavioral rehearsal, and assessment (Rousmaniere, 2019; Faber et al., 2021). For the purposes of the present study, DP is understood as a cyclical process involving: (1) observing clinical practices to identify areas needing improvement; (2) obtaining expert direct feedback to facilitate learning; (3) establishing small, realistic goals that present an appropriate challenge; (4) practicing behaviors through structured repetition of targeted tasks; and (5) critically evaluating performance in order to refine skills (Rousmaniere et al., 2017). Conceptualized in this way, DP is not only a means of enhancing specific therapist competencies, but also a framework for ongoing professional development.

At the same time, the present study should not be taken to imply that the field has already completed the kind of broad dialogue that would justify viewing any single measure as a definitive or widely accepted representation of DP in psychotherapy. Rather, the rationale for the present instrument is more modest: given the current conceptual and methodological heterogeneity, an initial theory-informed measure may help make the construct more explicit, more examinable, and more comparable across studies. Although the development process drew on the authors' expertise and consultation with experts, it did not constitute a wide-ranging co-design process with the broader professional

community, practitioners, or service users. Accordingly, the present measure is best understood as an initial contribution to an ongoing process of clarification and refinement, one that future research should extend through broader consultation, collaborative development, and further testing across diverse psychotherapy contexts.

In response to the limitations of existing measures and the broader need for more rigorous assessment tools in psychotherapy training (Miller et al., 2021; Jones & Taylor, 2021; Cochrane et al., 2022; Khanna et al., 2023; APA, 2021), the present study pursued three aims: (1) to develop a comprehensive and psychometrically sound measure of deliberate practice in psychotherapy that captures its multifaceted nature; (2) to validate this measure using rigorous statistical techniques (EFA and CFA) to assess its reliability and validity for measuring therapist engagement in deliberate practice behaviors; and (3) to provide a descriptive account of the extent to which psychotherapists report using deliberate practice principles in their clinical work. By pursuing these aims, this study seeks to provide a useful tool for therapist development, improved training methods, and, ultimately, better client care.

Method

Participants

Potential participants were contacted by the authors through e-mail (personal contacts and networks from the authors) and social networks (Linkedin, ResearchGate, etc.), with an invitation to participate in this study by means of an online survey. Data collection was ongoing from January 2024 until April 2024. The target minimum sample size was 300 participants in line with the recommendations of Comrey & Lee (2013). A total of 442 participants – all psychotherapy practitioners - accepted to participate in this study. Prior to proceeding with the survey itself, they were required to read and accept an informed consent form which informed them of the purposes of the study and their rights as participants. The participants had on average 50.31 years of age (SD = 13.75), and were mostly females (N = 260; 58.8%), with 175 being males (39.6%), 3 identifying as non-binary / third gender (0.7%) and 4 preferred not to disclose their gender (0.9%). The majority held a MSc degree (N = 282; 63.8%), followed by doctorates (N = 91; 20.6%), while the remaining held either a Bachelor degree (N = 28; 6.3%), or a graduate certificate (N = 41; 9.3%). The sample was global in nature; however, the most represented countries were the United States (N = 127; 28.7%), Portugal (N = 114; 25.8%), the United Kingdom (N = 32; 7.2%), Australia (N = 30; 6.8%), Brazil (N = 26; 5.9%), and Canada (N = 23; 5.2%), with the remaining participants hailing from a myriad of other countries.

In terms of psychotherapeutic approach, they were equally diverse – Integrative was the most represented (N = 141; 31.9%), followed by Cognitive-Behavioral (N = 72; 16.3%), Psychodynamic (N = 56; 12.7%), Existential (N = 40; 9%), Systemic (N = 37; 8.4%), Humanist (N = 36; 8.1%), Behavioral (N = 4; 0.9%), with the remaining participants indicating other approaches (N = 56; 12.7%).

No missing data was noted, and as such imputation was not required. The final sample was randomly split into two sub-samples (Johnson & Stevens, 2001) – a training data set, with roughly 30% of the participants, to be used in the Exploratory Factor Analysis (EFA); and a holdout data set comprised of the remaining 70% of the participants, to be used in the Confirmatory Factor Analysis (CFA) stage.

Question drafting

To develop this instrument, an initial pool of 27 questions related to the topic of deliberate practice were drafted by the authors. These were created while keeping in mind the theoretical dimensions that are identified by the literature in regards to deliberate practice. These questions were then circulated between two experts, to obtain qualitative feedback on them. Following this, and after carrying out numerous changes, a quantitative pre-test was carried out with a small sample of 29 colleagues – this exercise was aimed simply at early identification of numeric issues with the question pool before deployment of the survey to a larger sample. Again, several refinements were done based on this; the number of questions remained unchanged from the initial draft, even though the contents changed from the initial formulation. All questions were Likert-style on the 1-5 range and convered various domains identified in the literature as being relevant to deliberate practice, including: monitoring, observation of one's own clinical performance, identifying clinical challenges and areas for improvement, and defining goals.

Procedure

Data were analyzed using SPSS, version 29, and AMOS, version 29. The analytical roadmap was as such: we began by conducting an EFA on the training data set, in order to perform a preliminary triage of potentially problematic items, as well as obtaining a tentative factorial structure to be tested at the CFA stage. Following this, we specified this initial structure in the CFA and iteratively refined upon it until achieving the optimal fit for the model. We employed Maximum Likelihood for model estimation due to its known robustness (Arbuckle, 2007; Marôco, 2010). For model fit, we employed multiple indicators in line with statistical best practices (Hu & Bentler, 1999). These were – the χ^2 goodness-of-fit test (Barrett, 2007), the X^2 statistic (Bentler, 2007), the X^2/df index (Arbuckle, 2007), the Comparative Fit Index – CFI (Bentler, 1990), the Parsimony-adjusted CFI – PCFI (Marôco, 2010), and the Root-Mean-Square Error of Approximation – RMSEA (Steiger et al., 1985). Once this model was identified, we proceeded with computing validity (measured through three distinct facets – factorial validity, convergent validity, and discriminant validity), reliability (measured through composite reliability), and sensitivity (evaluated through skewness and kurtosis). For this purpose, we used the StatsToolsPackage macro (Gaskin, 2016), which provides a global validity assessment. Finally, we compute descriptive statistics of the instrument's scores across degree and gender.

Transparency and Openness

We report how we determined our sample size and all measures in the study. All data are available at http://doi.org/10.6084/m9.figshare.29835140. This study's design and its analysis were not pre-registered. To ensure data transparency and avoid duplicate or piecemeal publication, we confirm that no other studies have been published, are currently in press, or are under review that utilize the same dataset as the one presented in this manuscript. This submission is the sole work derived from this specific dataset.

Results

Exploratory Factor Analysis

In this first analysis, we conducted an Exploratory Factor Analysis (EFA) using all of the initial items, and employing the training sample. The goal of this exercise was two-fold: first, to identify potentially problematic items; and second, to establish a possible factorial structure to be tested during the Confirmatory Factor Analysis (CFA) stage. To accomplished the proposed goals, we began by inputting all of the initial variables, eliminating those which presented potential issues, re-estimating the model, and repeating

this process in an iterative manner until we reach a model without any evidence of problematic items. The exclusion criteria for items, at this stage, were the presence of cross-loadings across factors, or low factorial weights (under .50).

After repeating this process, we reached a factorial structure which was suitable for CFA. In this final model, the Kaiser-Meyer-Olkin (KMO) indicator yielded a value of .759, meeting the threshold for data adequacy (Kaiser, 1974), and the individual Measures of Sampling Adequacy (MSA) were also all above the .50 threshold (Hair et al., 2014). Finally, Bartlett's Test of Sphericity (Hair et al., 2014; Marôco, 2003) was significant ($\chi^2(91) = 995.523$, $p < .001$), indicating adequacy of the data. Factor extraction was conducted using the principal components method, and Direct Oblimin rotation. We employed two criteria to determine the optimal number of factors: Kaiser's Criterion (Kaiser, 1960), and the percentage of extracted variance (Marôco, 2003). Kaiser's Criterion indicated the extraction of 4 factors explaining a total of 71% of variance, and as such we proceeded with this solution, which is shown in the following table:

Table 1: EFA Factor Loadings

		Component			
Item	Code	1	2	3	4
I use outcome-monitoring tools to track my clients' progress.	MT1	**0.967**	-0.040	0.003	0.003
I use outcome-monitoring tools to identify clients not progressing as expected.	MT2	**0.960**	-0.011	0.020	-0.030
I use outcome-monitoring tools to assess my performance	MT3	**0.940**	-0.024	-0.007	0.049
I record my sessions regularly.	R1	-0.038	**0.922**	-0.023	-0.029
I use psychotherapy session recordings to reflect on the therapeutic process.	R3	-0.013	**0.913**	0.042	0.013
I review my session recordings to identify skills in my approach that need to be improved.	R2	-0.010	**0.859**	-0.028	0.038
I usually ask for feedback from a supervisor or peer to perfect specific therapeutic skills.	BP2	0.030	-0.073	**0.844**	0.052
I seek feedback from supervisors or peers to identify concrete challenges my client presents to the therapeutic process.	CH4	-0.075	-0.126	**0.793**	0.080
I habitually request feedback about the therapeutic process and my performance during supervision or with my peers.	AF3	0.072	0.229	**0.787**	-0.114
I can maintain a regular training routine to improve my clinical skills	BP5	-0.093	-0.017	-0.057	**0.796**
I set specific training goals according to the gaps in my performance.	GOA1	0.081	0.020	-0.060	**0.735**
I seek practical training opportunities to foster the therapeutic skills I need to develop.	GOA2	-0.057	-0.134	0.073	**0.732**
I practice repeated and structured exercises to train a specific therapeutic skill.	BP1	0.264	0.121	-0.050	**0.642**

Notes: Highest loadings for each item indicated in bold.

As no further optimizations or issues were identified within this structure, it was decided to specify it for further testing using Confirmatory Factor Analysis, which is detailed in the following section.

Confirmatory Factor Analysis

At this stage, we specified a model using the structure obtained through EFA, using the hold-out sample. In this section, the various iterations of the measurement model are documented – two in total. The initial model, based on the EFA structure, the diagnostics that were conducted, remedial action that was taken to address detected issues, and then a full report detailing the various psychometric aspects of the final model.

Model I – Initial measurement model

The initial model, including all the items and structures obtained through EFA, exhibited a very good fit ($\chi^2(71) = 94.168$, $p = 0.034$; $X^2/df = 1.326$; CFI = 0.989; PCFI = 0.772; RMSEA = 0.033; P[rmsea ≤ 0.05] = 0.957). No items exhibited low loadings, and due to the good fit it was not necessary to delve into exploration of Modification Indices. However, as we proceeded further into the validity study, we detected an issue with the first dimension, which exhibited an Average Variance Extracted (AVE; Fornell & Larcker, 1981) below the .50 threshold. Although the difference was not substantial, we opted to further optimize the model in order to maximize its psychometric properties. As such, we removed two items from the first dimension – BP4 – "I practice repeated and structured exercises to train a specific therapeutic skill" (with a loading of .64), and BP1 – "I can detect improvements in my therapeutic interventions after repeatedly training in specific therapeutic skills." (with a loading of .62), as these two had somewhat lower loadings than the remaining items within the factor. We specified a new model without these items, which is detailed below.

Model II – Refined measurement model

The refined model with the two removed items further improved the fit, ($\chi^2(48) = 42.950$, $p = 0.679$; $X^2/df = 0.895$; CFI = 1.000; PCFI = 0.727; RMSEA = 0.000; P[rmsea ≤ 0.05] = 0.999), with some of the indicators saturating near their maximum values. As before, all items exhibited very good loadings, indicating factorial validity of the instrument. This is shown below:

Table 2. Standardized factor loadings for the final measurement model.

Item	Item content	Estimate
Goals and Training		
GOA1	I set specific training goals according to the gaps in my performance.	0.700
GOA2	I seek practical training opportunities to foster the therapeutic skills I need to develop.	0.742
BP5	I can maintain a regular training routine to improve my clinical skills.	0.710
Monitoring		
MT1	I use outcome-monitoring tools to track my clients' progress.	0.956

MT2	I use outcome-monitoring tools to identify clients not progressing as expected.	0.927
MT3	I use outcome-monitoring tools to assess my performance	0.906
Recording		
R1	I record my sessions regularly.	0.801
R2	I review my session recordings to identify skills in my approach that need to be improved.	0.729
R3	I use psychotherapy session recordings to reflect on the therapeutic process.	0.819
Feedback		
CH4	I seek feedback from supervisors or peers to identify concrete challenges my client presents to the therapeutic process.	0.890
BP2	I usually ask for feedback from a supervisor or peer to perfect specific therapeutic skills.	0.816
AF3	I habitually request feedback about the therapeutic process and my performance during supervision or with my peers.	0.760

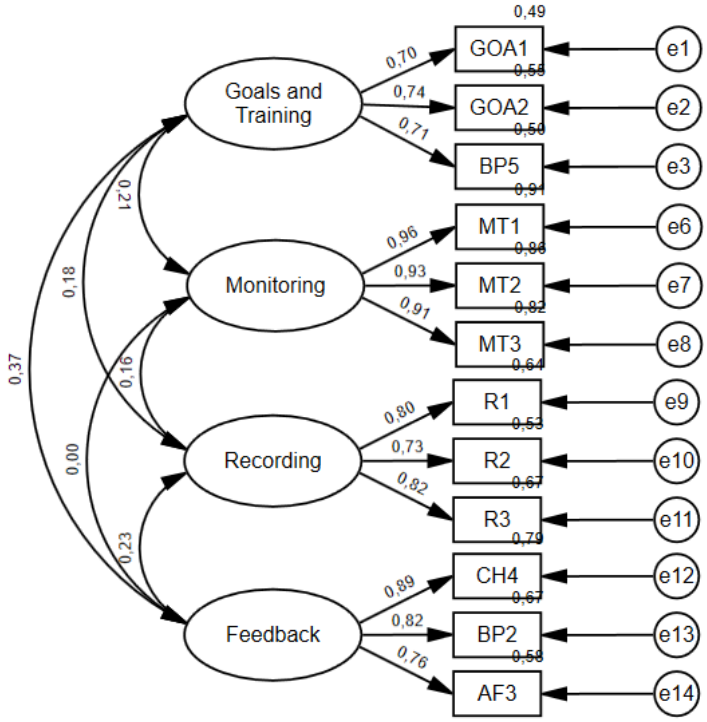


Fig 1. Final measurement model.

As such, we proceeded with a more in-depth psychometric study of the instrument. The global validity and reliability assessment is shown below:

Table 3: Validity and reliability evaluation

					Correlations between factors			
	CR	AVE	MSV	ASV	I	II	III	IV
I. Goals and Training	0.761	0.515	0.136	0.071	0.718			
II. Recording	0.827	0.615	0.055	0.038	0.183	0.784		
III. Feedback	0.863	0.679	0.136	0.064	0.369	0.235	0.824	
IV. Monitoring	0.950	0.865	0.042	0.023	0.205	0.160	0.000	0.930

Note: the diagonal of the correlation matrix indicates the square root of the AVE. CR - Composite Reliability; AVE - Average Variance Extracted; MSV - Mean Shared Variance. AVS - Average Shared Variance.

Validity

Three facets of validity were assessed in this exercise: factorial validity, convergent validity, and discriminant validity (Anastasi & Urbina, 1997; Hair et al., 2014; Marôco, 2010).

Factorial Validity. Factorial validity requires that all individual items of the construct have standardized loadings exceeding .50 (Marôco, 2010). This was already demonstrated during the model specification step, and as such we confirmed factorial validity.

Convergent Validity. To demonstrate convergent validity, items must exhibit overall high loadings into their construct. A common way to demonstrate this is through the Average Variance Extracted indicator (AVE; Fornell & Larcker, 1981). According to the commonly used Fornell-Larcker criterion, AVE for a given factor must exceed .50 in order to demonstrate the convergent validity of that construct. After removing the two items which were threatening convergent validity in Model I, the AVE scores for all constructs in Model II met this threshold, confirming their convergent validity.

Discriminant Validity. Discriminant validity requires low inter-factor correlations and cross-loadings (which we endeavored to eliminate immediately at the EFA stage). One way to determine discriminant validity is by using both the Maximum Shared Variance (MSV) and the Average Shared Variance (ASV). Demonstrating discriminant validity requires that the square root of the AVE exceeds all inter-factorial correlations, and that the AVE for a given factor exceeds both its MSV and ASV (Hair et al., 2014). As shown above, all these criteria are met, indicating the discriminant validity of all constructs.

Reliability

To assess the reliability of our scale, we computed the Composite Reliability (CR) for all factors (Fornell & Larcker, 1981). Reliability is considered acceptable if it exceeds the .70 threshold (Hair et al., 2014). This target is met for all of the various factors, confirming the overall reliability of the scale.

Sensitivity

The final aspect to be assessed was the scale's sensitivity, which is its ability to differentiate between individuals. This is typically done by verifying if each individual item possesses a sufficiently normal distribution (Marôco, 2010). A criteria which is commonly used for this is having the skewness and kurtosis's absolute values under 3 (Kline, 2016). As shown below, this was confirmed for all items, thus also demonstrating the scale's sensitivity.

Table 4: Sensitivity evaluation

Item	Min	Max	M	SD	Sk	Ku
I set specific training goals according to the gaps in my performance.	1	5	3.23	1.057	-0.319	-0.507
I seek practical training opportunities to foster the therapeutic skills I need to develop.	1	5	3.81	0.938	-0.784	0.535
I can maintain a regular training routine to improve my clinical skills	1	5	3.3	1.074	-0.215	-0.632
I use outcome-monitoring tools to track my clients' progress.	1	5	3.19	1.419	-0.223	-1.252
I use outcome-monitoring tools to identify clients not progressing as expected.	1	5	3.04	1.38	-0.095	-1.265
I use outcome-monitoring tools to assess my performance	1	5	2.99	1.433	-0.09	-1.321
I record my sessions regularly.	1	5	1.87	1.191	1.249	0.509
I review my session recordings to identify skills in my approach that need to be improved.	1	5	1.95	1.107	0.845	-0.417
I use psychotherapy session recordings to reflect on the therapeutic process.	1	5	1.78	1.101	1.234	0.437
I seek feedback from supervisors or peers to identify concrete challenges my client presents to the therapeutic process.	1	5	3.61	0.971	-0.509	-0.065
I usually ask for feedback from a supervisor or peer to perfect specific therapeutic skills.	1	5	3.4	1.05	-0.471	-0.209
I habitually request feedback about the therapeutic process and my performance during supervision or with my peers.	1	5	3.48	1.073	-0.531	-0.262

Characterization of the sample's DP

In this section, we provide several descriptive statistics of the various dimensions of our scale, based on a series of participant characteristics. The goal is to provide an overall characterization of the tendency to conduct (or not) deliberate practice; or, in other words, to provide a preliminary characterization of how the scale behaves across commonly examined demographic groups. We begin by showing the DP scores by education, which are shown in Table 5.

Table 5: DP by education

	Bachelor's Degree		Graduate Certificate		Master's Degree		Doctoral Degree / PhD	
	M	SD	M	SD	M	SD	M	SD
Goals and Training	3.298	0.857	3.268	0.779	3.501	0.824	3.575	0.804
Monitoring	3.417	1.428	2.992	1.403	2.921	1.352	3.341	1.222
Recording	2.036	1.215	1.894	1.037	1.846	0.998	2.044	1.028

Feedback	3.714	0.901	3.528	0.891	3.509	0.899	3.414	0.829

With the notable exception of Recording, which is below average, most other facets of DP are shown to be slightly above the average. There are some slight variations across levels of education, but the prevalence of DP seems to be mostly consistent.

Table 6: DP by gender

	Male		Female		Non-binary / Prefer not to say third gender			
	M	SD	M	SD	M	SD	M	SD
Goals and Training	3.286	0.852	3.606	0.766	4.222	1.072	3.417	1.101
Monitoring	3.410	1.290	2.794	1.326	2.889	2.009	3.583	1.167
Recording	2.042	1.049	1.815	0.998	1.333	0.577	2.000	1.277
Feedback	3.280	0.898	3.659	0.851	3.222	0.694	3.500	0.430

In this comparison, we can observe some slight differences – for example, females and non-binaries score higher on the Goals and Training sub-scale, whereas Males tend to score higher on the Monitoring sub-scale. Females also exhibit higher scores on the Feedback scale, and finally, Recording is, as before, overall low.

Discussion

This study significantly enhances our understanding of deliberate practice (DP) within the field of psychotherapy by addressing the urgent need for a robust, psychometrically validated measure that captures its multifaceted nature. As established by prior research, deliberate practice is crucial for skill development across various fields, including psychotherapy (Ericsson, 2018; Barata et. al, 2025). The research aimed to achieve three specific objectives: 1) to develop a comprehensive and psychometrically sound measure for assessing deliberate practice in psychotherapy; 2) to validate this measure through rigorous statistical techniques; and 3) to provide a descriptive understanding of therapists' engagement with deliberate practice principles in their clinical work.

The first goal of developing a comprehensive and psychometrically sound measure was successfully achieved through a carefully structured methodological approach. This development began with the formulation of an initial pool of items grounded in extant literature, capturing the various theoretical dimensions of deliberate practice specific to psychotherapy. Through consultations with experts and subsequent qualitative feedback, the instrument was refined to ensure that it holistically represented key components relevant to practicing psychotherapists. This process aligns with the consensus in the literature that a measure must reflect both the established theoretical frameworks of deliberate practice and the practical realities of clinical work (Rousmaniere, 2019; Faber et al., 2021).

Central to the measure's design were the identification and evaluation of four core dimensions: Goals and Training, Recording, Collaborative Feedback and Guidance, and Monitoring. Each of these components addresses specific aspects of therapist development critical for enhancing clinical effectiveness. For instance, the "Goals and

Training" factor emphasizes the importance of establishing specific, measurable objectives for skill enhancement. This aspect directly reflects the tenets of deliberate practice, which advocate for goal-oriented activities as foundational to improvement. In clinical settings, having concrete goals allows therapists to focus their efforts, track progress, and adapt their strategies to better meet client needs. Moreover, the "Recording" factor addresses an often-overlooked but essential component of deliberate practice: the systematic review of therapeutic sessions. Engaging in self-reflection through the examination of recorded sessions fosters a deeper understanding of therapeutic techniques and facilitates critical self-assessment. The measure's inclusion of this factor underscores its comprehensiveness; by incorporating reflective practices, the instrument ensures that therapists are encouraged to engage in regular evaluation of their clinical skills, an element consistently linked with improved therapeutic outcomes (Anderson et al., 2022).

The “Collaborative Feedback and Guidance” factor further enriches the measure by emphasizing the necessity of obtaining constructive feedback from supervisors and peers. This aligns with the literature advocating for the importance of collaborative learning environments, where therapists can receive guidance and insights that help refine their approaches (Vaz & Rousmaniere, 2022). The measure empowers therapists to create a feedback loop, whereby their practice is continuously informed and enriched by external perspectives - an essential element in fostering professional growth. Finally, the "Monitoring" factor focuses on the critical aspect of tracking client progress and outcomes. The ability to monitor therapeutic effectiveness not only informs therapists about the impact of their interventions but also aligns with evidence-based practice standards, which increasingly underscore the importance of data-driven decision-making in clinical psychology (Karow et al., 2020). By integrating these four dimensions, the developed measure offers a comprehensive framework that captures the multifaceted nature of deliberate practice in psychotherapy.

The development resulted in an instrument that demonstrates satisfactory psychometric properties and practical relevance to contemporary psychotherapy, though we stop short of claiming it definitively establishes the theoretical construct of DP. Psychometric validation via Exploratory and Confirmatory Factor Analyses supported a coherent four-factor structure and provided evidence of factorial, convergent, and discriminant validity, as well as acceptable reliability and sensitivity. However, model fit and internal structure alone do not test the broader conceptualization of deliberate practice; the scale indexes self-reported engagement with proposed DP dimensions and requires multi-method corroboration before being taken as evidence of the underlying theory. Accordingly, while the measure offers a useful tool for assessing therapists’ reported engagement and for informing training emphases, its theoretical claims should be regarded as provisional pending convergent validation with objective behavioral measures and outcome studies. However, this development also highlights the need for the integration of ongoing training and practice tailored to the specificities of each dimension. By addressing the identified areas through structured educational initiatives, training programs can ensure that therapists are better equipped to engage in deliberate practice actively, leading to continuous professional development. Despite the advancement achieved through this measure's development, future research should focus on how to best to implement this tool in various training and clinical settings, determining how it can be effectively utilized to enhance therapist engagement and client-care quality.

Limitations and Recommendations

This study presents several limitations that should be considered when interpreting the findings. First, the instrument captures self-reported engagement in DP-related behaviors rather than objective competence or demonstrated performance, and it does not directly assess fidelity to underlying DP principles. As with all survey-based measures, responses may also be influenced by social desirability bias. Second, the near-perfect model fit observed in the CFA raises the possibility of sample-specific optimization or capitalization on chance; this, replication studies need to be conducted to ensure robustness of findings. Future research should seek to cross-validate the model in new samples and examine measurement invariance across contexts; however, such analyses were not feasible in the present study due to limited per-country sample sizes within the multinational dataset. As such, this scale should be treated as a first version which should be subject to continuous refinement and testing, and readers should employ this with these considerations in mind.

Implications

In terms of implications, this study has two main areas of relevance. First, it has implications for practitioners, as the instrument provides a way to assess their relative engagement with deliberate practice, while the descriptive statistics from the present sample offer a preliminary frame of reference. Second, it has implications for researchers studying deliberate practice, as the development of a quantitative measure creates new opportunities for empirical investigation in psychotherapy. In addition, the instrument may assist both investigators and clinicians in developing more targeted interventions and in clarifying how deliberate practice relates to skill acquisition and performance. A further implication is that it helps address persistent problems in the operationalization of deliberate practice and responds to the methodological challenges identified above. At the same time, this contribution should be situated within the broader expertise literature, in which deliberate practice remains an influential but debated construct, especially because researchers have identified inconsistencies in how it has been defined, shifts in the standards used to evaluate supporting evidence, and ambiguity about how to distinguish it from related forms of effortful. In addition, broader expertise research has raised questions about whether deliberate practice alone can account for individual differences in performance, or whether more multifactorial explanations are needed (Hambrick et al., 2020). Thus, the present measure is not meant to settle those wider theoretical disputes, but rather to improve precision and comparability in psychotherapy research by specifying and assessing core components of deliberate practice more explicitly.

The four DP dimensions can be theorized as mechanism-linked leverage points within integrative models of psychotherapy change. Goals and Training functions as an intentionality mechanism: by specifying proximal, skill-focused objectives it shapes therapist expectancy, task structure, and deliberate use of interventions - processes shown to amplify engagement and promote skill automatization across orientations. Recording operates as a self-observation and error-correction mechanism, making tacit moment-to-moment behaviors explicit, supporting reflective practice, fidelity assessment, and refinement of technique that feed back into both competency development and alliance repair when therapists detect and address ruptures. Feedback serves as a social-learning and calibration mechanism: supervisor/peer input and client responses help recalibrate therapist models of effectiveness, reduce blind spots (a key source of between-therapist variance), and reinforce adaptive practices. Monitoring is an

outcome-sensitivity mechanism that closes the gap between intervention and effect - routine progress tracking both informs case formulation adjustments and operationalizes empirical accountability. Taken together, these mechanisms map onto common-factors (expectancy, engagement, alliance), therapist-effects (stable skills plus dynamic behavioral adjustments), and implementation models (supervision, ROM, fidelity), offering clear targets for mechanistic mediation tests and interventions aimed at reducing therapist variability and improving client outcomes.

We recommend treating deliberate practice in psychotherapy as a construct that functions simultaneously as a trans-theoretical process variable - core in-session and between-session behaviors that facilitate change across orientations; a meta-competence - higher-order capacities (e.g., purposeful goal setting, reflective review, adaptive use of feedback) that scaffold orientation-specific techniques; and a training-and-performance enhancement target - concrete, observable behaviors amenable to assessment, supervision, and deliberate practice interventions. Positioning DP in this way clarifies that the measure captures processes relevant to theory (mechanisms), assessment (competency measurement), and practice (supervision/training targets), rather than privileging one single role. Framing DP in the integrative terms above suggests specific next steps: test mediation models linking each DP dimension to outcomes via alliance, engagement, and adherence; examine DP's contribution to therapist effects in multilevel outcome studies; and evaluate whether interventions (supervision modules, feedback cycles, structured recording/review protocols) that target these dimensions improve therapist performance and client outcomes. Such programmatic work would shift the contribution of this measure from psychometric validation to theory-driven integration and practical improvement in training and clinical care.

References

American Psychological Association. (2021). *Guidelines for training in clinical psychology*. https://www.apa.org. APA

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Pearson Education.

Anderson, C., Smith, J., & Lee, T. (2022). The impact of reflective practices on therapeutic alliance. *Journal of Psychotherapy Research, 46*(2), 89–101. https://doi.org/10.XXXX/xxxxx

Arbuckle, J. L. (2007). *Amos 16.0 user's guide*. SPSS.

Barata, M., Santos, J. M., Sousa, D., Silva, C., Alface, C., & Coelho, P. (2025). We are on the road to somewhere: How deliberate practice can enhance the path of continuing education in mental health. *Practice Innovations, 10*(2), 132–143. https://doi.org/10.1037/pri0000273

Barata, M., Sousa, D., & Vaz, A. (2024). The effects of deliberate practice training on psychology students' empathic skills. *Journal of Expertise, 7*(1). [Note: Complete page numbers or article e-location if available]

Barrett, P. (2007). Structural equation modelling: Judging model fit. *Personality and Individual Differences, 42*(5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825–829. https://doi.org/10.1016/j.paid.2006.09.017

Bennett-Levy, J., Thwaites, R., & Haarhoff, B. (2020). The role of supervision in the development of intentional therapy skills: A review of the literature. *Clinical Psychology Review, 80*, 101867. https://doi.org/10.1016/j.cpr.2020.101867

Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. 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*, *51*(5), 456–467. https://doi.org/10.1037/pro0000318

Cochrane, T. E., Smith, S., & Newell, J. (2022). Addressing the gaps in measuring therapeutic competence: A systematic review of the literature. *Psychotherapy Research, 32*(5), 629–645. https://doi.org/10.1080/10503307.2022.2021204

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

Diamond, G., Wlodek, B., Arthey, S., & Parker, S. (2025). A systematic review of deliberate practice in psychotherapy: Definitions, operationalization, and preliminary outcomes. *Psychotherapy*. Advance online publication. https://doi.org/10.1037/pst0000569

Ericsson, K. A. (2018). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 93*(10), 1420–1422. https://doi.org/10.1097/ACM.0000000000002404

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

Faber, M. J., Smits, M., & van Dulmen, S. (2021). The role of feedback in improving psychotherapeutic practices: A systematic review. *Psychotherapy Research, 31*(5), 591–604. https://doi.org/10.1080/10503307.2020.1796332

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 39–50.

Gaskin, J. (2016). *StatsToolsPackage macro*. https://github.com/gaskinj/StatsToolsPackage

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.

Hambrick, D. Z., Macnamara, B. N., & Oswald, F. L. (2020). Is the deliberate practice view defensible? A review of evidence and discussion of issues. Frontiers in Psychology, 11, 1134. https://doi.org/10.3389/fpsyg.2020.01134

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Johnson, B., & Stevens, J. J. (2001). Exploratory and confirmatory factor analysis of the School Level Environment Questionnaire (SLEQ). *Learning Environments Research, 4*(3), 325–345. https://doi.org/10.1023/A:1011374607778

Karow, A., Huber, C., & Hillemacher, T. (2020). The importance of reflective practice in psychotherapy: A systematic review. *Clinical Psychology Review, 78*, 101820. https://doi.org/10.1016/j.cpr.2020.101820

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford Press.

Lagerberg, H., Boswell, J. F., Constantino, M. J., Andersson, G., & Carlbring, P. (2024). Does practice make perfect? The effects of an eight-week manualized deliberate practice course with peer feedback on patient-rated working alliance in adults: A pilot randomized controlled trial. *Clinical Psychology in Europe, 6*(3), Article e12353. https://doi.org/10.32872/cpe.12353

Lim, D., & Nguyen, T. (2023). Methodological considerations for developing training measures in psychotherapy. *Psychological Assessment, 35*(2), 478–488. https://doi.org/10.1037/pas0001154

Mahon, J. (2022). Deliberate practice in psychotherapy training: Enhancing therapeutic competencies. *Journal of Clinical Psychology, 78*(3), 456–467. https://doi.org/10.1002/jclp.23291

Miller, T. D., Hart, M. A., & Eames, S. (2021). The effectiveness of different methods for assessing deliberate practice in psychotherapy: A systematic review. *Clinical Psychology Review, 82*, 101906. https://doi.org/10.1016/j.cpr.2020.101906

Miller, S. D., Chow, D., Malins, S., & Hubble, M. A. (Eds.). (2023). *The field guide to better results: Evidence-based exercises to improve therapeutic effectiveness*. American Psychological Association.

Marôco, J. (2003). *Análise estatística: Com utilização do SPSS*. ReportNumber, Lda.

Marôco, J. (2010). *Análise de equações estruturais: Fundamentos teóricos, software & aplicações*. ReportNumber, Lda.

Norcross, J. C., & Wampold, B. E. (2018). Evidence-based therapy relationships: A 10-year retrospective. *Psychotherapy: Theory, Research, Practice, Training, 55*(4), 348–352. https://doi.org/10.1037/pst0000147

Nurse, K., O'Shea, M., Ling, M., Castle, N., & Sheen, J. (2024). The influence of deliberate practice on skill performance in therapeutic practice: A systematic review of early studies. *Psychotherapy Research*, 1–15. https://doi.org/10.1080/10503307.2024.2308159

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

Rousmaniere, T., & Vaz, A. (2021). *The essentials of deliberate practice*. American Psychological Association.

Rousmaniere, T. (2024). *Deliberate practice for psychotherapists: A guide to improving clinical effectiveness* (2nd ed.). Routledge.

Tomás, M., Sousa, D., Barata, M., Amaro, I., & Coelho, P. (2025). The Effect of a Deliberate Practice Training on the Delivery of Empathetic Interventions in Clinical Psychology Master Students. *Journal of Expertise*, *8*(1), 72-84.

Vaz, A., McLeod, J. and Nissen-Lie, H.A. (2025), Rethinking Psychotherapy Training and Supervision: The Case for Deliberate Practice. Journal of Clinical Psychology. https://doi.org/10.1002/jclp.23777

Vaz, A., & Rousmaniere, T. (2022). Clarifying deliberate practice for mental health training. https://www.sentio.org