BABCP
www.babcp.com

ORIGINAL RESEARCH

Assessing procedural knowledge in clinical psychology: development and evaluation of a brief video-based measure

Tatjana Paunov🔵, Florian Weck🔵 and Franziska Kühne🔵

Department of Clinical Psychology and Psychotherapy, University of Potsdam, Potsdam, Germany
Corresponding author: Tatjana Paunov; Email: tatjana.paunov@uni-potsdam.de

Abstract

Objective: Research on the acquisition of psychotherapeutic competencies attributes an important role to procedural knowledge. We developed and evaluated a brief video-based measure of procedural knowledge in clinical psychology (the Pro CliPs Task).
Method: The Pro CliPs Task consists of seven distinct case vignettes, each presented in four brief videos, and eight open-ended questions pertaining to each case vignette. The instrument was developed and validated in a multi-stage process, which included conceptual development, a feasibility study and an evaluation study. In the feasibility study ($n = 66$ psychology students), the item characteristics and reliability of the coding method for the open-ended answers, authenticity of the case vignettes and general applicability of the measure were assessed. The following evaluation study ($n = 68$ psychology students, $n = 65$ postgraduate psychotherapy trainees) was conducted to investigate the validity of the instrument.
Results: The item characteristics of the Pro CliPs Task were largely found to be within the expected range and inter-rater reliability was excellent. The case vignettes used were perceived as highly authentic, and the Pro CliPs Task was generally evaluated positively by participants. In the evaluation study, the Pro CliPs Task was strongly associated with declarative knowledge and therapeutic self-efficacy. Postgraduate psychotherapy trainees outperformed psychology students in the Pro CliPs Task.
Conclusion: The Pro CliPs Task holds promise as a valid and brief measure of procedural knowledge in clinical psychology with potential for use in research, training and teaching contexts.

Key learning aims

(1) To gain insight into the role of procedural knowledge in clinical psychology and the limitations regarding its assessment.
(2) To learn about the theoretical basis and conceptual development of the Pro CliPs Task, a recently developed measure to assess procedural knowledge in clinical psychology.
(3) To evaluate the Pro CliPs Task regarding its item characteristics, inter-rater reliability, applicability, authenticity and validity.

Introduction

The transfer of theoretical knowledge into practice is considered a crucial aspect of therapist competence development (Bennett-Levy, 2006; Muse *et al.*, 2022). Derived from information processing theory, Bennett-Levy (2006) proposed a model of therapist skill development that focuses on three related systems: the declarative, the procedural and the reflective (DPR) system. The declarative system encompasses factual knowledge, for example knowing symptoms of a depressive episode. Declarative knowledge is assumed to be acquired by reading, listening to lectures or observational learning (Bennett-Levy, 2006; Bennett-Levy *et al.*, 2009). As training progresses, it is assumed that declarative knowledge will transform into more practical abilities, namely procedural knowledge. Procedural knowledge refers to the actual application of knowledge in practice and is defined as 'knowledge of how to and when to rules' (Bennett-Levy, 2006; p. 59) or 'practical understanding of application of cognitive-behavioural therapy' (Muse and McManus, 2013; p. 488). Novice therapists are assumed to build their procedural knowledge by didactic learning, modelling and feedback (Bennett-Levy *et al.*, 2009). Bennett-Levy (2006) suggests that strategies for building competence change at different stages of training. While declarative and procedural processes are especially relevant for novice therapists, the reflective system becomes more relevant for more experienced therapists. Reflection is defined as a metacognitive process, encompassing 'the observation, interpretation and evaluation of one's own thoughts, emotions and actions, and their outcomes' (p. 60). The DPR model has since been widely applied in psychotherapy competence research, and has been adapted to different modalities and contexts (e.g. Churchard, 2022).

In accordance with the DPR model, it is assumed that procedural knowledge is based on declarative knowledge. However, only few studies have investigated the empirical relationship between different aspects of competence. In an experimental study, the communication and therapy skills of 69 psychology students were trained and compared in terms of their theoretical and procedural knowledge (Heinze *et al.*, 2023). The results indicated that procedural knowledge was significantly correlated with communication and therapy skills and predicted the level of skills after a training intervention. Similarly, in a naturalistic study, 219 psychotherapy trainees were assessed over a 3-year period and compared with a control group of psychologists without postgraduate psychotherapy training (Evers *et al.*, 2022). Over the course of the study, the trainees demonstrated an increase in procedural knowledge, along with declarative knowledge. However, it was only in procedural knowledge that the trainees outperformed the control group over time.

In order to assess procedural knowledge, it is recommended to use an open-response format, such as in short-answer clinical case vignettes or written case reports, as this approach reduces guesswork and allows for more complex answers (Muse *et al.*, 2022; Muse and McManus, 2013). For example, Eells and colleagues (1998) provide an elaborate coding method for evaluating comprehensive case conceptualisations, a method that is frequently employed in training research (e.g. Evers *et al.*, 2022). However, this method necessitates considerable resources, both in terms of writing and the subsequent evaluation. In contrast, shorter case vignettes provide standardised information and facilitate more efficient assessment and evaluation (Muse *et al.*, 2022; Muse and McManus, 2013).

Few measures have been developed to assess procedural knowledge. Heinze and colleagues (2023) introduced written case vignettes with multiple-choice answer options but reported improvable indices for item difficulty and item discrimination, suggesting that the items were either too easy to solve or that participants with different knowledge levels could not be distinguished sufficiently. Based on these findings, the authors propose to use video-based vignettes with an open response format instead (Heinze *et al.*, 2023). In a similar attempt, situational judgement tests (SJTs), written or video-based, present participants with hypothetical clinical scenarios. Participants indicate how they would behave in these situations based on the given answer options (Campion *et al.*, 2014). Another exemplary assessment tool was developed

by Myles and Milne (2004), the so-called Video Assessment Task. It involves a short (simulated) patient video and open questions about the reported symptoms, problems and suitable treatment techniques. The authors report good inter-rater reliabilities but no other psychometric data (Myles and Milne, 2004).

In summary, the existing measures for the reliable assessment of procedural knowledge in clinical psychology, contrary to recommendations, frequently rely on pre-defined answer options and often lack sufficient psychometric reporting (Frank *et al.*, 2020; Muse and McManus, 2013).

Objective

We thus aimed to develop and validate a psychometrically sound instrument for the assessment of procedural knowledge in clinical psychology (the Pro CliPs Task), that would allow for a resource-efficient application and assessment. This project aimed to:

- Evaluate the Pro CliPs Task regarding its inter-rater reliability, item characteristics, applicability and authenticity (feasibility study), and to
- Examine the instrument's validity (evaluation study).

Method

Data were collected in a multi-stage process, which included conceptual development (pre-tests 1 and 2, expert survey), a feasibility study and an evaluation study. The study project was approved by the university's ethics committee (55/2023) and pre-registered prior to data collection (Paunov *et al.*, 2024b).

Conceptual development of the Pro CliPs Task

The structure of our instrument was derived from tasks commonly used for the assessment in medical education and research (Clark *et al.*, 2017; Daniel *et al.*, 2019; Epstein *et al.*, 2008; Peabody *et al.*, 2000; Schmidmaier *et al.*, 2013). These tasks entail the concise presentation of a clinical case vignette followed by questions about the vignette. In the field of medicine, these questions typically encompass areas such as (differential) diagnosis, necessary steps in the evaluation process, therapeutic measures and medication or correct behaviour in specific situations. In the context of clinical psychology, these elements are reflected in the formulation of a case conceptualisation. A case conceptualisation (or formulation) consists of hypotheses about the causing and maintaining influences of a person's problems and serves to guide therapy in identifying treatment goals and suitable interventions (Kendjelic and Eells, 2007).

A common framework for case conceptualisations is the trans-theoretical '5 Ps model' (Dudley and Kuyken, 2013; Kendjelic and Eells, 2007; Polipo *et al.*, 2024; Weerasekera, 1993). The framework asks therapists to describe the patient's 'presenting problem' and complement their analysis with 'precipitating factors' (which triggered the current symptoms), 'predisposing factors' (which increased the patient's vulnerability to the development of their problems), 'perpetuating factors' (which maintain the patient's problems in the present) and 'protective factors' such as the patient's resources (Weerasekera, 1993). Given its generic nature and widespread acceptance (Bucci *et al.*, 2016; Polipo *et al.*, 2024), we decided to base the information in the case vignettes and corresponding questions largely on the '5 Ps model'.

Construction of videos and questions

First, written case vignettes were developed, providing the information possibly needed to answer the questions following the '5 Ps model'. The scripts were based on fictitious patient histories,

Table 1. Characteristics of groups involved in the construction process

	Group	Age	Education	Clinical experience
Pre-test 1	Psychotherapy trainees ($n = 5$)	$M = 29.6$ ($SD = 3.19$)	MSc in psychology	$M = 1.7$ ($SD = 2.04$) years in in-patient settings
Pre-test 2	Psychology students ($n = 6$)	$M = 25.83$ ($SD = 2.14$)	BSc or MSc in psychology	—
Expert survey	Psychotherapists ($n = 7$)	$M = 38.29$ ($SD = 3.45$)	MSc or PhD in psychology, licensed psychotherapists	$M = 7.86$ ($SD = 4.74$) years after licensure

diagnostic criteria (World Health Organization, 2009) and casebooks (World Health Organization, 2012). To cover a broad spectrum of mental disorders, multiple case vignettes exemplifying the most prevalent mental disorders (Jacobi and Kessler-Scheil, 2013) were developed, with corresponding open-ended questions. The scripts and questions were discussed within the team of researchers and refined in an iterative process to ensure coherence and authenticity. After the construction was completed, a first pre-test using the written case vignettes was conducted with five psychotherapy trainees (see Table 1 for characteristics of all groups involved in the construction process). The pre-test aimed to assess comprehensiveness, duration and usability of the task. As a result, the case vignettes and questions were refined.

Subsequently, all case vignettes were videotaped with simulated patients (SPs), portraying the content of the written case vignettes. The majority of SPs were actors, female (71%) and with a mean age of $M = 47.29$ years ($SD = 15.41$, range 27–65). They memorised the script and received a 1-hour training session on the study project, rules for their portrayal and details on their role. All SPs received an expense allowance of €50.

Afterwards, a second pre-test was conducted using the video-based case vignettes. Six psychology students (see Table 1) evaluated the authenticity of the videos in addition to the feasibility, comprehensiveness and usability of the instrument. Therefore, think-aloud and probing techniques were used. Participants were further invited to provide open feedback. Based on their feedback, minor changes were made to the wording of the questions in order to make them more specific and easier to understand.

Construction of the coding method

For the evaluation of the responses to the open-ended questions, a unique coding system was developed. The objective was to devise a coding system that would facilitate a simple and efficient evaluation of the answers, while enabling the differentiation between incorrect (0), mostly incorrect (1), mostly correct (2) and completely correct (3) answers. For this purpose, seven experienced psychotherapists (see Table 1) completed the case vignettes. Their responses were then combined with information from clinical guidelines and disorder-specific manuals (e.g. Bandelow *et al.*, 2021; Hautzinger, 2021) and summarised in a coding sheet giving examples of incorrect, mostly incorrect, mostly correct and completely correct responses for each case vignette. The coding sheet aims to provide guidance for the raters when evaluating the open-ended responses. Similar to the case formulation content coding method proposed by Eells and colleagues (1998), our coding system takes into account not only the accuracy of the content but also the use of terminology, the diversity of solution ideas, and the level of elaboration and abstraction. This was assumed to be relevant for distinguishing between different knowledge levels (Eells *et al.*, 1998).

The Pro CliPs Task

As a result of the construction process, the final Pro CliPs Task consists of a case vignette presented in four brief videos and eight open-ended questions pertaining to each case vignette
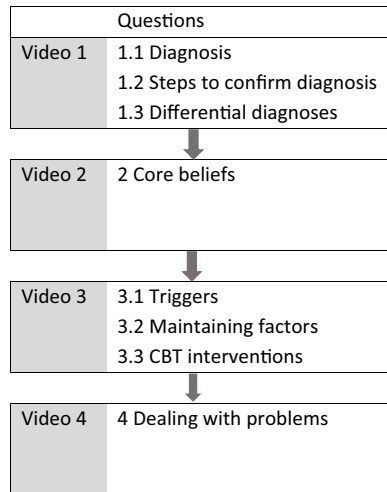
	Questions
Video 1	1.1 Diagnosis 1.2 Steps to confirm diagnosis 1.3 Differential diagnoses
Video 2	2 Core beliefs
Video 3	3.1 Triggers 3.2 Maintaining factors 3.3 CBT interventions
Video 4	4 Dealing with problems

Figure 1. Schematic structure of the Pro CliPs Task. Every case vignette consists of four videos. The structure and questions apply to all case vignettes.

(see Fig. 1). Each video lasts approximately half a minute ($M = 28.11$ s, $SD = 10.42$ s). In total, we developed seven different case vignettes covering a spectrum of the most prevalent mental disorders (e.g. depression, panic disorder, obsessive-compulsive disorder). For assessment purposes, we recommend to use one of the provided seven case vignettes. The completion of a single case vignette requires approximately 10–15 minutes. Once participants have answered all eight questions of one case vignette, their responses are then coded in accordance with the provided coding sheet. Subsequently, a sum score can be calculated (up to 3 points per question, with a maximum total score of 24 per case vignette). This sum score represents the primary outcome of the Pro CliPs Task. An example video vignette and the pertaining questions of the Pro CliPs Task are available in English (see Paunov, 2024). The German version of the Pro CliPs Task, including the coding system, is available upon request from the first author.

The Pro CliPs Task assesses procedural knowledge in clinical psychology and is grounded in the trans-theoretical, generic '5 Ps model', yet it also encompasses CBT-specific aspects, such as core beliefs and CBT interventions.

Feasibility study

The principal objective of the feasibility study was to examine the item characteristics and inter-rater reliability of the Pro CliPs Task in a student sample. In addition, the general applicability of the Pro CliPs Task and the authenticity of the case vignettes were assessed using the Authenticity of Patient Demonstrations Scale (APD; Ay-Bryson *et al.*, 2022). We hypothesised good inter-rater reliability (ICC$\geq$.75; following Koo and Li, 2016) and adequate item characteristics (item difficulty Pi .20 $< Pi <$ .80, item variance $\sigma^2 >$.16, item discrimination $r >$.30; following Moosbrugger and Kelava, 2012). We expected high applicability ratings and all case vignettes to be considered authentic ($M = 2.0$ on the APD, ranging from 0 to 3). A moderate correlation between the Pro CliPs Task and the participants' level of education was expected ($r = .30$ to .50), with Master's students achieving significantly higher sum scores than Bachelor's students ($p < .05$).

Statistical analysis and ratings

Following power analysis for the inter-rater reliability, we aimed at a sample size of $n = 62$ participants answering two complete vignettes (minimum acceptable ICC $= .75$, power 0.80, based

on data of two raters). We computed intraclass correlation coefficients, psychometric indices, Pearson correlation coefficients and a *t*-test to address our hypotheses. All analyses were performed using RStudio version 2023.12.1.402 (R Core Team, 2024). Two independent raters were trained in the use of the coding system. Their training lasted 10 hours. The raters worked independently from one another and convened after completing the ratings for one case vignette to discuss their ratings. The ratings were not changed after discussion. Both raters were female and one was the first author and a psychotherapist in training. The second rater was a Master's student in clinical psychology.

Participants and procedure

Sixty-six Bachelor's and Master's psychology students participated in the feasibility study. Their mean age was $M = 25.75$ years ($SD = 5.02$). The majority was female (77%) and Bachelor's psychology students (71%). They were recruited through the university's participant pool for the participation in scientific studies. The students participated in a 1-hour online study and received course credit for their participation. The university's survey platform was used, and participation was anonymous and voluntary. All participants gave informed consent prior to study completion. Each participant saw two out of seven case vignettes in a randomised balanced order. Based on the results of the feasibility study, minor changes were made to the wording of one question.

Evaluation study

The principal objective of the evaluation study was to examine the validity of the Pro CliPs Task. We sought to examine the association between procedural knowledge, assessed with the Pro CliPs Task and (a) declarative knowledge, as assessed by 15 true/false statements derived from the German psychotherapy licensure test (convergent validity), (b) self-reported grades in clinical psychology (criterion validity), and (c) the reflective system, as assessed by therapeutic self-efficacy with the Counsellor Activity Self-Efficacy Scales Revised-Basic Skill (CASES-R; Hunsmann et al., 2024), which consists of 15 items on a scale from 0 to 9. Moderate correlations with declarative knowledge, self-reported grades in clinical psychology and therapeutic self-efficacy were expected ($r = .30$ to $.50$). We also compared the Pro CliPs Task sum scores of psychology students with those of postgraduate psychotherapy trainees (having obtained a Master's degree in psychology) in order to ascertain whether the Pro CliPs Task was sensitive to different training levels. We hypothesised that postgraduate psychotherapy trainees would achieve higher sum scores in the Pro CliPs Task than psychology students ($p < .05$, $d = .50$ to $.80$).

Statistical analysis and ratings

Following power analysis for the group comparison, we aimed at a sample size of $N = 128$ ($n = 64$ in each group; $d = .50$, power .80, $\alpha = .05$) with each participant answering one complete vignette. To address the hypotheses, a *t*-test and Pearson correlation coefficients were computed. The answers to the Pro CliPs Task were rated by the first author. As inter-rater reliability had been established in the feasibility study, this rater coded all the data and 25% of the answers were then coded by a second rater. The raters were blinded to group affiliation, and did not change their ratings after discussion.

Participants and procedure

One hundred thirty-three individuals participated in the evaluation study. In the student sample ($n = 68$), the mean age was $M = 25.43$ ($SD = 5.10$). The majority was female (78%) and Bachelor's

students (69%). In the trainee sample ($n = 65$), the mean age was $M = 31.94$ ($SD = 6.76$). Trainees were mostly female (80%) and had been in postgraduate psychotherapy training for $M = 3.29$ years ($SD = 1.63$). The majority were training in CBT (89%), and some were training in psychodynamic (6%) or systemic (5%) therapy. Participants were recruited via email lists and newsletters across German universities for the student sample and across institutions offering postgraduate psychotherapy training for the trainee sample. Participation was anonymous, voluntary, and took approximately 30 minutes. All participants provided informed consent prior to study completion.

Results

Feasibility study

The average inter-rater reliability for the coded answers of the Pro CliPs Task was considered excellent for the vignette sum scores ($ICC_{(2,2)} = 0.97$–0.99) and good to excellent for the single items across vignettes ($ICC_{(2,2)} = .83$–$.98$). The mean score across all vignettes was $M = 10.27$ ($SD = 4.98$, range 1.5–21.5). Descriptive results, ICCs and item characteristics for all items across case vignettes of the Pro CliPs Task are displayed in Table 2. Item characteristics were largely within the expected range, with a few exceptions. These included high item difficulty values for question 1.1 regarding the diagnosis in two case vignettes and low item discrimination values for five different questions in different case vignettes (see Supplementary material for the results on the case vignette-specific level). The correlation between sum scores in one vignette and another was high ($r = .78$).

In accordance with our hypothesis, the case vignette sum scores were also highly correlated with participants' level of education ($r = .60$), indicating that more advanced students achieved higher Pro CliPs Task sum scores. The mean Pro CliPs sum score for Master's students ($M = 14.95$, $SD = 4.1$) was significantly higher than that for Bachelor's students ($M = 8.37$, $SD = 3.97$; $t_{64} = -6.03$, $p < .001$, $d = 1.6$).

In accordance with our hypothesis, the case vignettes were perceived as authentic ($M = 2.25$, $SD = 0.47$; on a scale from 0 to 3) and the applicability of the Pro CliPs Task was rated as excellent (see Table 3).

Evaluation study

For the Pro CliPs Task, the mean score across all vignettes was $M = 10.06$ ($SD = 4.22$, range 1–19) in the student sample and $M = 16.03$ ($SD = 3.54$, range 5–24) in the trainee sample, with significantly higher scores in the trainee group ($t_{131} = -8.83$, $p < .001$, $d = 1.5$). In the overall sample, there was a strong positive correlation between procedural knowledge and both declarative knowledge ($r = .55$) and therapeutic self-efficacy ($r = .53$). Additionally, there was a moderately negative correlation between self-reported grades in clinical psychology and procedural knowledge ($r = -.34$), indicating that better grades were associated with higher sum scores. The results were therefore in accordance with our hypotheses. Furthermore, the inter-rater reliability was computed for 25% of the data based on two case vignettes and considered excellent ($ICC_{(2,2)} = .93$).

Discussion

The objective of the present study was to develop and evaluate a measure of procedural knowledge in clinical psychology. We thus constructed the Pro CliPs Task, consisting of a case vignette presented in four brief videos, and eight open-ended questions pertaining to the case vignette. In total, the Pro CliPs Task provides seven different case vignettes. We evaluated the Pro CliPs Task

Table 2. Item characteristics of all questions across case vignettes in the feasibility study

Item	M	SD	ICC$_{(2,2)}$	Item difficulty P_i	Item variance σ^2	Item discrimination r
1.1 Diagnosis	2.23	1	.98	0.54–0.96	0.21–1.56	0.19–0.62
1.2 Steps to confirm diagnosis	1.2	0.87	.89	0.30–0.53	0.38–1.10	0.57–0.77
1.3 Differential diagnoses	0.94	0.98	.94	0.22–0.69	0.32–1.25	0.19–0.78
2 Core beliefs	1.06	1.01	.90	0.28–0.47	0.48–1.49	0.47–0.74
3.1 Triggers	1.39	1.12	.94	0.33–0.67	0.34–1.56	0.20–0.73
3.2 Maintaining factors	0.98	0.82	.91	0.25–0.40	0.42–0.85	0.44–0.73
3.3 CBT interventions	1.13	1.03	.92	0.25–0.49	0.69–1.35	0.56–0.81
4 Dealing with problems	1.31	1.02	.83	0.32–0.60	0.77–1.74	0.48–0.84

$n = 132$ per item/question in the Pro Clips Task. Ratings on a scale from 0 to 3. Ranges across all case vignettes are reported for the psychometric indices.

Table 3. Applicability of the Pro CliPs Task as rated in the feasibility study

Item	M	SD
Comprehensibility of the task	2.53	0.61
Comprehensibility of the questions	2.48	0.66
Difficulty of the questions	1.45	0.71
Distress by watching videos	0.20	0.47
Questions ask for memorized knowledge only	0.88	0.77
Need to call on existing knowledge to answer questions	2.18	0.70
Learning through the task	1.97	0.72
Interestingness of the task	2.65	0.57
Demand of the task	2.32	0.66
Video quality	2.65	0.54
Technical problems with videos	0.15	0.47

$n = 66$. Ratings on a scale from 0 to 3.

in a feasibility study regarding its item characteristics, inter-rater reliability, applicability and authenticity of the Pro CliPs Task. In addition, we examined its validity in a second study.

Overall, the Pro CliPs Task demonstrated adequate psychometric characteristics. With two exceptions; all item difficulty values were within the expected range of $.20 < Pi < .80$. Regarding item discrimination, the values for five different questions in different case vignettes were lower than recommended, yet still acceptable. Furthermore, all questions were deemed to be content-theoretically relevant for the measure and were thus not excluded following the recommendations by Moosbrugger and Kelava (2012). It is noteworthy that none of the seven case vignettes nor any of the open-ended questions proved problematic regarding their psychometric data. The inter-rater reliability for the coded answers was excellent for the vignette sum scores and good to excellent for the single items across vignettes. In comparison to similar measures (e.g. Eells *et al.*, 1998; Myles and Milne, 2004) and instruments of competence in which observed therapist competence is rated (such as the CTS; Young and Beck, 1980), the results are particularly encouraging. As inter-rater reliability was established based on the ratings of the first author and a trained student, these results indicate that with proper training and supervision individuals with less experience can also assess the Pro CliPs Task. The case vignettes were judged highly authentic and the applicability of the task was generally considered excellent. These are important aspects for the resource-efficient and realistic assessment of clinical competencies in general (Muse *et al.*, 2022; Muse and McManus, 2013).

The evaluation study yielded evidence supporting the instrument's validity. The Pro CliPs Task was able to effectively differentiate between varying levels of study as well as between psychology students and psychotherapy trainees. Furthermore, it was strongly associated with declarative knowledge in clinical psychology and with therapeutic self-efficacy, as well as moderately

correlated with grades in clinical psychology. The evidence for its convergent and criterion validity is therefore robust, and the Pro CliPs Task may be considered an easy-to-use, additional measure alongside more comprehensive rating scales, such as the Cognitive Therapy Scale (Young and Beck, 1980). In terms of the DPR model (Bennett-Levy, 2006), we investigated the basic relationships between the procedural system and the declarative and reflective systems. The results also indicate an improvement in procedural abilities over the course of training. This is consistent with previous research indicating that declarative knowledge and case formulation skills are dependent on psychotherapists' experience (Eells and Lombart, 2003; Eells *et al.*, 2011) and may improve over the course of training (Evers *et al.*, 2022) in a manner analogous to the growth of reflexive capacity over the course of therapeutic development (Campbell-Lee *et al.*, 2024).

Limitations and future research

Firstly, it should be noted that the sampling of participants in our studies was not entirely random. It would be beneficial for further studies to explore the generalisability of the Pro CliPs Task to other populations, such as expert clinicians, or in different settings, such as before and after interventions focusing on the improvement of case conceptualisation competences. Secondly, although we conducted two studies that were consistent with the previous power analyses, it would be beneficial to include more diverse samples for further evaluation. Thirdly, as our primary objective was the preliminary evaluation of the Pro CliPs Task, the study design was not suitable for investigating the relationships between declarative, procedural and reflective systems in greater detail. Although declarative knowledge is frequently regarded as a pre-requisite for procedural and reflective skills (Heinze *et al.*, 2023; Muse and McManus, 2013), a more comprehensive investigation of the inter-relationships between the different competence systems is warranted. Further research is also needed to address the lack of information on divergent validity of the Pro CliPs Task. Similarly, further research is required to examine the convergent validity with other measures of procedural knowledge. As such, case conceptualisation is *one* approach to operationalising procedural knowledge and other dimensions of this concept (i.e. standardised role-plays, real therapist–patient interactions) should also be investigated in order to gain a more thorough understanding of the role of procedural knowledge in competence development. Further research may also compare the Pro CliPs Task with other measures of case conceptualisation, such as the Collaborative Case Conceptualisation Rating Scale (Padesky *et al.*, 2011). Fourthly, the Pro CliPs Task has a focus on CBT and in the evaluation study, the trainees were predominantly from a CBT background. This limits its suitability for assessing clinical knowledge in psychodynamic or systemic therapy orientations. Nevertheless, the instrument could be adapted to assess other psychotherapy orientations. Finally, it should be noted that the high inter-rater reliability is also a result of rater training and the fact that one rater was the primary developer of the instrument. As previous research has demonstrated (Paunov *et al.*, 2024a), inter-rater reliability is likely to be contingent upon rater training and the provision of ongoing supervision. Consequently, researchers intending to utilise the measure in future studies should ensure that rater training is provided at the beginning and throughout the rating process.

Conclusion

In conclusion, the Pro CliPs Task is a psychometrically sound measure of procedural clinical knowledge. This study provides psychometric and reliability data that are rarely reported in competency research (Frank *et al.*, 2020). The Pro CliPs Task requires approximately 10–15 minutes for most respondents to complete one case vignette, making it a time-efficient assessment tool. As there are few instruments for the reliable assessment of procedural knowledge, the Pro CliPs Task represents a valuable contribution to the measurement of psychotherapeutic competencies. The Pro CliPs Task is a pragmatic tool that can be utilised in research and teaching to assess and improve

procedural knowledge in clinical psychology, particularly in CBT. Our findings indicate excellent psychometric properties and reliability, strong convergent and criterion validity, as well as applicability for measuring clinical procedural knowledge with the Pro CliPs Task. Consequently, it has the potential to serve as a comparatively brief and reliable competence measure.

Key practice points

(1) The Pro CliPs Task is a brief measure of procedural knowledge in clinical psychology. It consists of a case vignette presented in four brief videos, and eight open-ended questions pertaining to the case vignette. In total, the Pro CliPs Task provides seven different case vignettes, each exhibiting comparable psychometric properties. For assessment purposes, we recommend using one of the seven case vignettes.

(2) The Pro CliPs Task can be used in research, training and teaching contexts. It requires respondents 10–15 minutes to answer one case vignette, thus making it a time-efficient assessment tool.

(3) The evidence for the role of procedural knowledge in the development of therapeutic competence remains limited. However, theoretical frameworks such as the DPR model offer a foundation for understanding its importance in translating declarative knowledge into practice.

Further reading

Bennett-Levy, J. (2006). Therapist skills: a cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy*, *34*, 57–78. https://doi.org/10.1017/S1352465805002420

Muse, K., Kennerley, H., & McManus, F. (2022). The why, what, when, who and how of assessing CBT competence to support lifelong learning. *the Cognitive Behaviour Therapist*, *15*. https://doi.org/10.1017/S1754470X22000502

Myles, P. J., & Milne, D. L. (2004). Outcome evaluation of a brief shared learning programme in cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, *32*, 177–188. https://doi.org/10.1017/s1352465804001183

References

Ay-Bryson, D. S., Weck, F., & Kühne, F. (2022). Can simulated patient encounters appear authentic? Development and pilot results of a rating instrument based on the portrayal of depressive patients. *Training and Education in Professional Psychology*, *16*, 20–27. https://doi.org/10.1037/tep0000349

Bandelow, B., Wiltink, J., Alpers, G. W., Benecke, C., Deckert, J., Eckhardt-Henn, A., Ehrig, C., Engel, K., Falkai, P., Geiser, F., Gerlach, A. L., Harfst, T., Hau, S., Joraschky, P., Kellner, M., Köllner, V., Kopp, I., Langs, G., Lichte, T., … & Beutel, M. E. (2021). *German S3 guideline for the treatment of anxiety disorders.* https://madoc.bib.uni-mannheim.de/37583

Bennett-Levy, J. (2006). Therapist skills: a cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy*, 34, 57–78. https://doi.org/10.1017/S1352465805002420

Bennett-Levy, J., McManus, F., Westling, B. E., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, 37, 571–583. https://doi.org/10.1017/S1352465809990270

Bucci, S., French, L., & Berry, K. (2016). Measures assessing the quality of case conceptualization: a systematic review. *Journal of Clinical Psychology*, 72, 517–533. https://doi.org/10.1002/jclp.22280

Campbell-Lee, D., Barton, S., & Armstrong, P. (2024). Higher-order CBT skills: are there differences in meta-competence between trainee and experienced therapists? *the Cognitive Behaviour Therapist*, 17. https://doi.org/10.1017/S1754470X24000047

Campion, M. C., Ployhart, R. E., & MacKenzie, W. I. (2014). The state of research on situational judgment tests: a content analysis and directions for future research. *Human Performance*, 27, 283–310. https://doi.org/10.1080/08959285.2014.929693

Churchard, A. (2022). How can psychotherapists improve their practice with service users from minoritised ethnicities? An application of the Declarative-Procedural-Reflective (DPR) model of clinical skill development. *the Cognitive Behaviour Therapist*, 15. https://doi.org/10.1017/S1754470X21000386

Clark, G. I., Hanstock, T. L., & Clark, L. H. (2017). Psychological treatment of co-occurring anxiety disorders in clinical practice: a vignette study. *Australian Psychologist*, 52, 480–490. https://doi.org/10.1111/ap.12214

Daniel, M., Rencic, J., Durning, S. J., Holmboe, E., Santen, S. A., Lang, V., Ratcliffe, T., Gordon, D., Heist, B., Lubarsky, S., Estrada, C. A., Ballard, T., Artino, A. R., Da Sergio Silva, A., Cleary, T., Stojan, J., & Gruppen, L. D. (2019). Clinical reasoning assessment methods: a scoping review and practical guidance. *Academic Medicine: Journal of the Association of American Medical Colleges*, 94, 902–912. https://doi.org/10.1097/ACM.0000000000002618

Dudley, R., & Kuyken, W. (2013). Case formulation in cognitive behavioural therapy: a principle-driven approach. In L. Johnstone & R. Dallos (eds), *Formulation in Psychology and Psychotherapy*. Routledge.

Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of a content coding manual. *Journal of Psychotherapy Practice and Research*, 7, 144–153.

Eells, T. D., & Lombart, K. G. (2003). Case formulation and treatment concepts among novice, experienced, and expert cognitive-behavioral and psychodynamic therapists. *Psychotherapy Research*, 13, 187–204. https://doi.org/10.1093/ptr/kpg018

Eells, T. D., Lombart, K. G., Salsman, N., Kendjelic, E. M., Schneiderman, C. T., & Lucas, C. P. (2011). Expert reasoning in psychotherapy case formulation. *Psychotherapy Research*, 21, 385–399. https://doi.org/10.1080/10503307.2010.539284

Epstein, S. A., Hooper, L. M., Weinfurt, K. P., DePuy, V., Cooper, L. A., Harless, W. G., & Tracy, C. M. (2008). Primary care physicians' evaluation and treatment of depression: results of an experimental study using video vignettes. *Medical Care Research and Review*, 65, 674–695. https://doi.org/10.1177/1077558708320987

Evers, O., Schröder-Pfeifer, P., Möller, H., & Taubner, S. (2022). The competence development of German psychotherapy trainees: a naturalistic, longitudinal and multidimensional outcome study. *Psychotherapy Research*, 32, 539–553. https://doi.org/10.1080/10503307.2021.1950939

Frank, H. E., Becker-Haimes, E. M., & Kendall, P. C. (2020). Therapist training in evidence-based interventions for mental health: a systematic review of training approaches and outcomes. *Clinical Psychology*, 27, e12330. https://doi.org/10.1111/cpsp.12330

Hautzinger, M. (2021). *Kognitive Verhaltenstherapiebei Depressionen [Cognitive-Behavioural Therapy for Depression]*. Beltz.

Heinze, P. E., Weck, F., Maaß, U., & Kühne, F. (2023). The relation between knowledge and skills assessments in psychotherapy training: secondary analysis of a randomized controlled trial. *Training and Education in Professional Psychology*. https://doi.org/10.1037/tep0000463

Hunsmann, J. J., Ay-Bryson, D. S., Kobs, S., Behrend, N., Weck, F., Knigge, M., & Kühne, F. (2024). Basic counseling skills in psychology and teaching: validation of a short version of the counselor activity self-efficacy scales. *BMC Psychology*, 12, 32. https://doi.org/10.1186/s40359-023-01506-7

Jacobi, F., & Kessler-Scheil, S. (2013). Epidemiologie psychischer Störungen [Epidemiology of mental disorders]. *Psychotherapeut*, 58, 191–206. https://doi.org/10.1007/s00278-013-0962-z

Kendjelic, E. M., & Eells, T. D. (2007). Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy*, 44, 66–77. https://doi.org/10.1037/0033-3204.44.1.66

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

Moosbrugger, H., & Kelava, A. (Eds.). (2012). *Testtheorie und Fragebogenkonstruktion [Test Theory and Questionnaire Construction]* (2nd edn). Springer.

Muse, K., Kennerley, H., & McManus, F. (2022). The why, what, when, who and how of assessing CBT competence to support lifelong learning. *the Cognitive Behaviour Therapist*, 15. https://doi.org/10.1017/S1754470X22000502

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33, 484–499. https://doi.org/10.1016/j.cpr.2013.01.010

Myles, P. J., & Milne, D. L. (2004). Outcome evaluation of a brief shared learning programme in cognitive behavioural therapy. *Behavioural and Cognitive Psychotherapy*, *32*, 177–188. https://doi.org/10.1017/s1352465804001183

Padesky, C. A., Kuyken, W., & Dudley, R. (2011). The Collaborative Case Conceptualization Rating Scale & Coding Manual. Version 5. Retrieved from: https://www.padesky.com/wp-content/uploads/2023/07/CCCRS_Coding_Manual_v5_web.pdf

Paunov, T. (2024, June 19). *Video-based assessment of clinical procedural skill.* Retrieved from: osf.io/f6k52

Paunov, T., Weck, F., Heinze, P. E., Maaß, U., & Kühne, F. (2024a). Competence ratings in psychotherapy training – a complex matter. *Cognitive Therapy and Research*, *48*, 500–510. https://doi.org/10.1007/s10608-023-10445-x

Paunov, T., Weck, F., & Kühne, F. (2024b). *Development and validation of a measure assessing procedural knowledge in clinical psychology and psychotherapy.* https://doi.org/10.23668/psycharchives.14070

Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*, *283*, 1715–1722. https://doi.org/10.1001/jama.283.13.1715

Polipo, N. F., Willemsen, J., Hustinx, M., & Bazan, A. (2024). Developing psychoanalytic case conceptualization skills through didactic teaching: a randomized controlled trial. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *34*, 379–397. https://doi.org/10.1080/10503307.2023.2241623

R Core Team (2024). *R: A language and environment for statistical computing [computer software].* R Foundation for Statistical Computing, Vienna, Austria.

Schmidmaier, R., Eiber, S., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2013). Learning the facts in medical school is not enough: which factors predict successful application of procedural knowledge in a laboratory setting? *BMC Medical Education*, *13*, 28. https://doi.org/10.1186/1472-6920-13-28

Weerasekera, P. (1993). Formulation: a multiperspective model. *Canadian Journal of Psychiatry*, *38*, 351–358.

World Health Organization (2009). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.* World Health Organization.

World Health Organization (2012). *WHO Case Book on ICD-10: Chapter V (F): Mental and behavioural disorders – Case presentations of adults.* Huber.

Young, J., & Beck, A. (1980). *Cognitive Therapy Scale: Rating Manual. Unpublished manuscript.* University of Philadelphia, Philadelphia, USA.