AI Patients for Training CBT Therapists to Develop Clinical Competence: Pilot Study of CBT Trainer

Tianyu Terry Zhang, Rob Saunders, Stephen Pilling, Ciarán O'Driscoll

Submitted to: JMIR Medical Education
on: September 22, 2025

Table of Contents

AI Patients for Training CBT Therapists to Develop Clinical Competence: Pilot Study of CBT Trainer

Tianyu Terry Zhang[1]; Rob Saunders[1]; Stephen Pilling[1, 2]; Ciarán O'Driscoll[1]

[1]Research Department of Clinical, Educational & Health Psychology Centre for Outcomes Research and Effectiveness (CORE) University College London London GB
[2] Camden and Islington NHS Foundation Trust London GB

Corresponding Author:
Tianyu Terry Zhang
Research Department of Clinical, Educational & Health Psychology
Centre for Outcomes Research and Effectiveness (CORE)
University College London
1-19 Torrington Place
London
GB

Abstract

Background: Training challenges in Cognitive Behavioral Therapy (CBT) include limited supervised practice with diverse cases, inconsistent feedback, resource-intensive supervision, and difficulties standardising competence assessment.

Objective: This study evaluated the acceptability and feasibility of CBT Trainer, a mobile application for training CBT practitioners through the use of standardized AI patient interactions and the evaluation of therapist responses against competence frameworks to enable structured skill development grounded in experiential learning and deliberate practice.

Methods: This mixed-methods pilot study employed a two-stage approach. Stage 1 involved usability testing with four participants. Stage 2 included 59 participants from psychological practitioner training programs (a Low Intensity CBT Interventions Programme and a Doctorate in Clinical Psychology) who engaged with the CBT Trainer over one month. Measures of impact included the System Usability Scale (SUS), platform engagement, post-study questionnaire on perceived learning outcomes and qualitative feedback.

Results: CBT Trainer performed well on all pre-specified outcome targets. Participants spent an average of 95.24 minutes (SD=134.58) in roleplays, completed an average of 4.15 role-play sessions (SD=3.55), with 49.69 interactions per session. Platform usability was rated as excellent (mean SUS=82.20, SD=12.93). Self-reported competence improvement was highest in assessment skills (96.7%), followed by information gathering (66.7%). Key advantages over traditional methods included immediate feedback (83.3%) and convenience (73.3%).

Conclusions: This pilot study provides evidence that an AI-based patient simulation shows promise as a supplementary training tool for CBT therapists, particularly regarding accessibility and immediate feedback. Future research should employ randomized controlled designs with objective competence assessments. Clinical Trial: The study protocol was pre-registered with the Open Science Framework (https://osf.io/mskb7).

(JMIR Preprints 22/09/2025:84091)
DOI: https://doi.org/10.2196/preprints.84091

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

An AI-Driven Virtual Patient Platform for Training CBT Practitioners against Competencies: Pilot Study of CBT Trainer

Abstract

Background: Cognitive Behavioral Therapy (CBT) training faces significant challenges including supervised practice with diverse cases, inconsistent feedback, resource-intensive supervision, and difficulties standardizing competence assessment.

Objective: This study evaluated the acceptability and feasibility of CBT Trainer, a mobile application for training CBT practitioners through the use of standardized AI patient interactions and the evaluation of therapist responses against competence frameworks to enable structured skill development in a controlled, repeatable environment that complements traditional training methods.

Methods: This mixed-methods pilot study employed a two-stage approach. Stage 1 involved usability testing with four participants. Stage 2 included 59 participants from psychological practitioner training programs (a Low Intensity CBT Interventions Programme and a Doctorate in Clinical Psychology) who engaged with the CBT Trainer over one month. Measures of impact included the System Usability Scale (SUS), platform engagement, post-study questionnaire on perceived learning outcomes and qualitative feedback.

Results: CBT Trainer performed well on all pre-specified outcome targets. Platform usability was rated as excellent (mean SUS=82.20, SD=12.93). Self-reported competence improvement was highest in assessment skills (96.7%), followed by information gathering (66.7%). Key advantages over traditional methods included immediate feedback (83.3%) and convenience (73.3%).

Conclusions: This pilot study provides evidence that an AI-based patient simulation shows promise as a supplementary training tool for CBT therapists, particularly regarding accessibility and immediate feedback. Future research should employ randomized controlled designs with objective competence assessments.

Trial Registration: OSF Registries https://osf.io/mskb7

Keywords: cognitive behavioral therapy; artificial intelligence; virtual patients; competence assessment; competency-based education; clinical training; psychological therapy training; simulation-based learning; mobile application; digital health

Introduction

The training of mental health professionals represents a critical challenge in addressing the global mental health crisis. Within the National Health Service (NHS) there have been initiatives aimed to improve the quality of care and mental health outcomes through improving access to evidence-based psychological therapies [1]. This requires developing a skilled workforce. In addition, as demand for psychological services continues to rise, there is increasing pressure on training programs to develop competent practitioners efficiently and effectively. Technological innovations, particularly machine learning and natural language processing, offer promising new avenues for enhancing mental health training through virtual patient simulations with competence-based feedback, potentially addressing long-standing challenges in traditional approaches.

Training competent therapists - the role of competence frameworks

Competence-based education has emerged as a dominant paradigm in mental healthcare professional training, with structured frameworks serving as essential scaffolding for skill acquisition, assessment, and professional development [2]. Across therapeutic modalities, the UCL CORE Competence Frameworks provide a comprehensive hierarchical model of competences required for effective practice, differentiating between generic therapeutic competences, basic therapy specific technique (e.g. CBT) competences, and meta-competences [3].

Several competence scales have also been developed to operationalize frameworks or professional standards into measurable assessment tools. For example, the Cognitive Therapy Scale-Revised (CTS-R) [4], the revised form of Cognitive Therapy Rating Scale (CTRS) [5], is the main treatment competence measurement tool used for CBT therapist accreditation internationally, comprising 12 distinct competence domains including agenda setting, feedback, collaboration, pacing, and interpersonal effectiveness [6]. Competence frameworks and scales like these serve multiple functions: providing structured developmental pathways [7,8], facilitating formative assessment [9], supporting summative evaluation [10], and promoting training consistency [11].

However, implementing the training and evaluation of competencies comes with practical challenges across multiple domains. First, the reliability of competence assessment methods has been questioned. While some studies show good inter-rater reliability for widely-used competence measures such as the CTS-R [10,12], systematic reviews reveal variable reliability across studies, with inter-rater agreement ranging from poor to good [13]. Roth [3] notes that “although groups of raters working together can achieve high concordance on the CTS-R, the reliability of ratings by independent judges is uncertain". Second, resource limitations create substantial barriers to effective competence development. This includes demands on clinical supervisors, limited access for trainees to diverse clinical presentations, reliance on time-demanding assessment methods, and resource constraints affecting feedback quality [8,14]. Third, there are also ethical considerations around developing competence ‘on the job,’ with a need to ensure patient safety and well-being while fostering the

trainee's professional development. The British Psychological Society's Code of Ethics and Conduct (2021) emphasizes that practitioners must work within their competence level while maintaining responsibility for client welfare [15]. This creates inherent tension where trainees must gain experience with real clients while ensuring no harm occurs through inexperience. While this is managed through informed consent, supervision, adherence to professional codes of conduct, and a constant focus on the trainee's competence and ethical decision-making [15,16], it requires a high level of supervisory support, which can be resource-intensive for training programs managing increasing cohort sizes. The expansion of the psychological professions workforce places additional demands on supervision infrastructure, potentially affecting the quality and availability of supervisory support [1,14]. These constraints highlight the need for supplementary training approaches that can provide additional practice opportunities while maintaining the essential human oversight required for ethical competence development.

Traditional Role-Play Methods in Mental Health Training

Role-play exercises are a cornerstone of competence-based training, offering structured opportunities to practice and receive feedback on specific therapeutic skills [17,18]. While role plays provide valuable opportunities for skill development, they face substantial limitations in terms of ecological validity, scalability, and consistency [19,20]. Within psychological practitioner training, standardized role-plays using trained actors offer structured learning, though validity concerns persist [10]. Student to student role plays face practical challenges such as inconsistencies in role play quality due to student anxiety when role-playing with peers, authenticity limitations as students may deliberately facilitate their peers' success rather than presenting realistic clinical challenges, ethical concerns regarding self-disclosure, and logistical difficulties across large student cohorts [17,18,21–23]. Research has identified four authenticity barriers: students' insufficient knowledge of psychiatric symptoms, reluctance to portray distressing emotions, artificial dynamics from pre-existing relationships, and inadequate environmental settings [17,24]. Interestingly, studies show higher-performing students often experience greater anxiety during role-plays than lower-performing peers—a "competence-anxiety paradox" linked to skilled trainees' greater awareness of therapeutic complexity [18]. Teacher demonstration models, where instructors model therapeutic techniques with volunteer "clients," provide excellent teaching opportunities but limited hands-on practice for students [18].

These limitations contribute to the "theory-practice gap" where classroom knowledge fails to translate effectively to clinical application [20,25].

Technological Innovations in Mental Health Training

Traditional role-play limitations have driven advancements in simulation technology for clinical skills training. Innovations included decision tree-based interactive videos that improved the standardization of training experiences and assessment procedures [26,27]. This approach has reduced psychological strain on actors playing patient roles, as they no longer needed to maintain complex character portrayals for extended periods, and improved scalability by eliminating the need for extensive simulant training that required 10

hours for students or 3-4 hours for clinicians [27]. While research shows skills can improve through these methods, there is variability in the degree to which students find interactive client systems engaging and useful [26,28]. Machine learning and language models have advanced patient simulations through contextual understanding and human-like text generation, which can increase realism comparable to human patients [29], across diverse scenarios, though responses sometimes appeared overly idealized [30]. While these approaches facilitate practice, they do not provide competence-based feedback to further skill development, which is essential for meeting professional training standards and ensuring therapists develop the specific competences required for NHS service delivery.

Rationale for Current Study

Approaches to operationalize competence development through frameworks and scales, need to build on learning models such as experiential learning [31], deliberate practice methodologies [32], and technology-enhanced training [7,33]. Therefore, an AI-driven simulation platform that incorporates competence-based feedback mechanisms offers a promising solution to address the identified limitations in traditional training approaches whilst maintaining alignment with established professional frameworks and educational principles. Such a platform promotes experiential learning through active experimentation within a safe environment combined with competence-based reflection to facilitate improvement, whilst providing structured learning opportunities that support deliberate practice methodologies.

Despite promising advances in AI-based virtual patient simulations, important questions remain about their feasibility, acceptability, and educational effectiveness in real-world training environments. Most studies to date have relied on evaluations over short exposure periods [26,28,30,34]. While structured Large Language Model (LLM) approaches show promise in creating more realistic patient simulations, methodological limitations make it difficult to determine whether these innovations translate to measurable improvements in clinical competence. Additionally, the practical implementation of these technologies within existing training curricula remains underexplored, with limited evidence on how trainees integrate such tools into their broader learning experiences.

Unlike previous research focusing primarily on perceived fidelity or effectiveness [28–30], our mixed-methods approach examines usage patterns, integration with existing curricula, and specific mechanisms through which such simulations might enhance learning. Collecting both quantitative engagement metrics and qualitative feedback, can provide preliminary evidence of the practical value of AI simulations in psychological therapist training. Furthermore, this study explores how AI-driven simulations might complement traditional role-play methods, potentially addressing some of the limitations identified in previous research.

Study Aims

This pilot study aimed to evaluate the educational potential of an AI-driven simulation approach for psychological therapist training. Specifically, the study aimed:

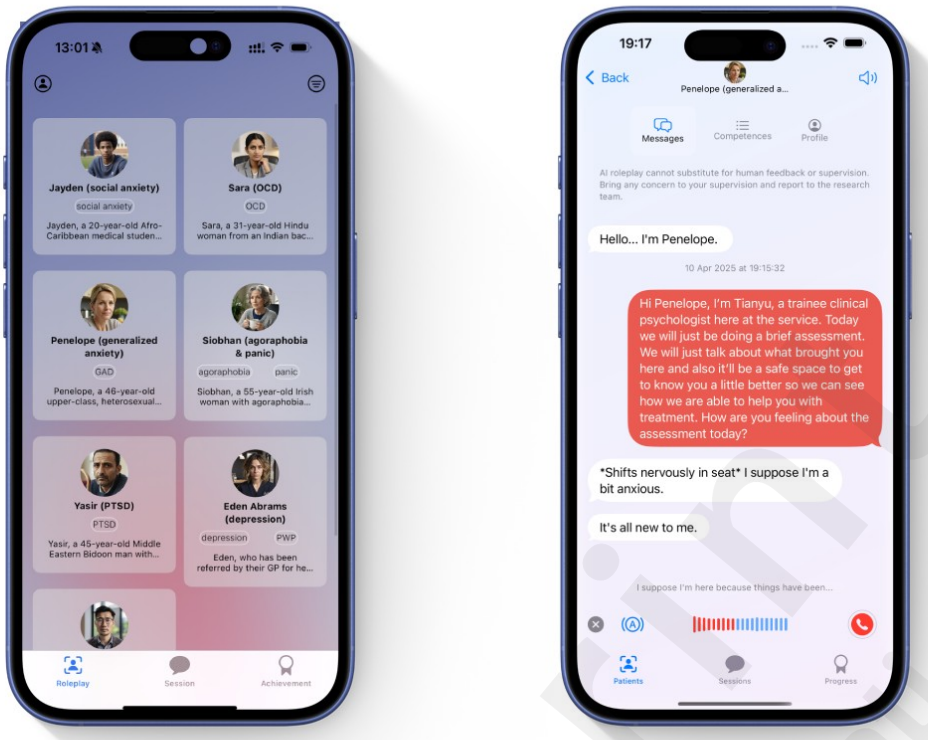
1. To assess the acceptability of simulation-based learning as a supplemental tool within established training programs for psychological therapists.
2. To explore trainees' perceptions of the educational value of simulated therapeutic interactions compared to traditional training methods.
3. To examine patterns of voluntary engagement with simulation-based practice among trainees from different professional training backgrounds.
4. To gather preliminary evidence regarding self-reported competence development following simulation-based practice experiences.

Methods

The CBT Trainer Platform

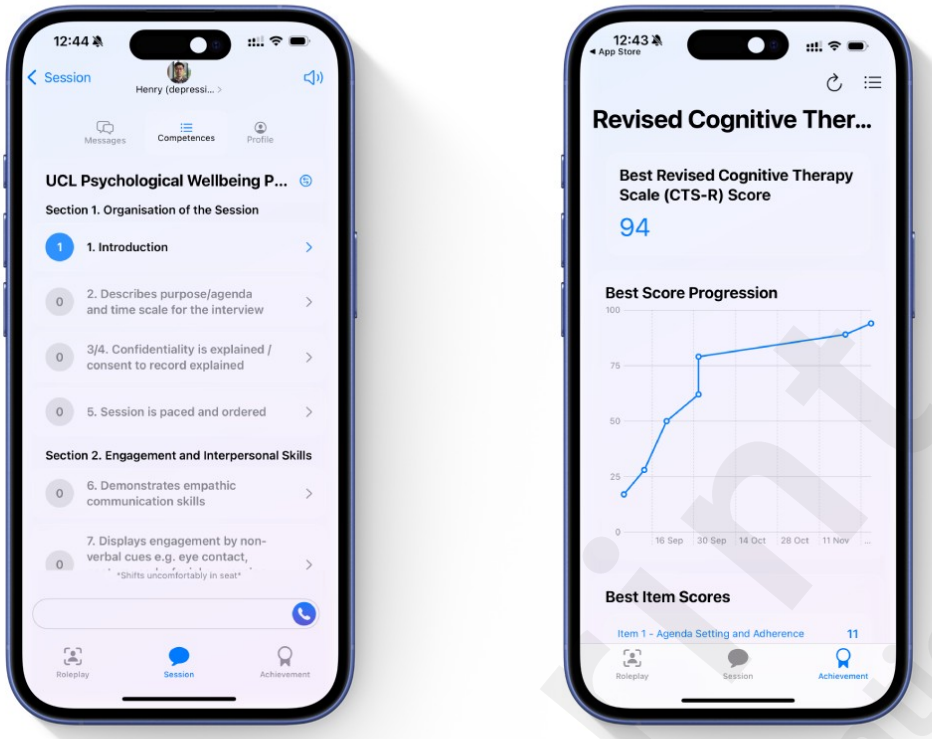
The CBT Trainer platform [35] is a AI-based virtual patient platform designed for mental health training. The application presented participants with interactive scenarios replicating therapeutic sessions for common mental health conditions including depression, generalized anxiety disorder, social anxiety disorder, post-traumatic stress disorder, and agoraphobia. The platform primarily enabled voice- and text-based interaction with virtual patients exhibiting realistic clinical presentations. Users began by selecting from seven virtual patients representing diverse demographics and clinical presentations. During voice or text-based interactions, the system continuously evaluated responses against competence frameworks, providing real-time feedback visible in the interface.

The CBT Trainer assessed trainees against a competence scale—either the Cognitive Therapy Scale-Revised (CTS-R) [4] or the UCL Psychological Wellbeing Practitioner (PWP) Assessment Competence Scale [36]. CBT Trainer indicates whether specific competences were demonstrated or required further development. After practice sessions, participants received structured feedback, highlighting competences met, and also those not, with specific guidance to support their development of key therapeutic competences. The platform was designed to complement, not replace, traditional training methods by providing additional opportunities for independent skill practice. The CBT Trainer application's content and features were kept unchanged throughout the data collection period to ensure consistency across all participant interactions. Figure 1 and Figure 2 illustrates the application's user interface diagrams, including the patient selection screen displaying diverse clinical presentations, the interactive roleplay screen and the competence feedback screens.



[Insert Figure 1 here]

Figure 1. CBT Trainer Patient Selection and Role-play Interfaces. (Left) Patient selection interface shows patient profiles representing different demographics, ages, ethnicities, and clinical presentations including depression, social anxiety, generalized anxiety, agoraphobia & panic, and PTSD. (Right) Interactive simulation interface demonstrating voice-based therapeutic interaction with virtual patient Penelope, showing the conversational flow with real-time competence tracking visible in the interface background. The system emphasizes ethical considerations by clearly stating that "AI roleplay cannot substitute for human feedback or supervision" and reminding trainees to bring concerns to their supervision.



[Insert Figure 2 here]

Figure 2. Competence Assessment and Progress Tracking Interfaces. (Left) UCL Psychological Wellbeing Practitioner (PWP) Assessment Competence Scale interface showing real-time competence evaluation across core domains including "Organization of the Session" and "Engagement and Interpersonal Skills." Each competence item displays scoring indicators that update as trainees demonstrate skills during interactions. (Right) Cognitive Therapy Scale-Revised (CTS-R) progress tracking screen displaying longitudinal competence development and visual progression chart showing improvement over time from 16 September to 11 November, alongside individual item scores for specific competences such as "Agenda Setting and Adherence."

Study Design

This pilot study employed a two-stage mixed-methods approach to evaluate an AI-based virtual patient platform for training cognitive behavioral therapists. This included questionnaires, interviews, and digital engagement data to assess acceptability and educational impact.

Participants

Participants were recruited from psychological practitioner training programs at University College London during the 2024-25 academic year, targeting two specific programmes: the Doctorate in Clinical Psychology and the Low-intensity Psychological Well-being Practitioner training. Eligible participants were required to be

currently enrolled in one of these programs and have access to an iOS device with internet connectivity capable of running the CBT Trainer application.

Procedure

The study was conducted in two stages:

Stage 1 involved four participants in one-hour in-person usability testing sessions, including interviews about training experiences, and supervised interaction with the virtual patient application, concluding with a usability assessment.

Stage 2 expanded the study to include 101 participants who completed an initial online screening. After completing online screening, informed consent and demographic questionnaires, eligible participants downloaded the application to their personal device. Participants were encouraged to use the platform for at least 60 minutes during the study period to gain meaningful experience, though they were explicitly informed that there was no mandatory usage requirement and they could engage with the application as much or as little as they preferred. Two reminder emails were sent to all participants during weeks 1 and 3 of the 4-week study period, providing tips on integrating the platform with their training programme activities. The stage concluded with a comprehensive online survey.

All 101 participants completing the screening survey were entered into a draw for one of three £45 vouchers, but no incentives were provided for app download or engagement to preserve ecological validity of usage patterns.

The research version of CBT Trainer application was distributed through Apple's TestFlight beta testing platform, which required participants to complete multiple steps including receiving and accepting email invitations, downloading the TestFlight app, and then installing CBT Trainer.

Measures

Survey Instruments

System Usability Scale (SUS): A standardized 10-item questionnaire measuring platform usability, with scores ranging from 0-100 [37]. Scores above 68 are considered above average, and scores above 80 are considered excellent [38].

Post-Study Questionnaire: A custom questionnaire designed to capture participants' experiences with the CBT Trainer platform. The questionnaire comprised four main sections: (1) three subjective learning outcome measures assessing helpfulness, skill improvement, and perceived realism using 0-100 rating scales; (2) a checklist assessment of specific competence development across therapeutic domains; (3) comparative evaluation questions rating CBT Trainer against traditional role-play methods; and (4) open-ended qualitative questions exploring learning impact and user experience.

Engagement Metrics

Platform usage was tracked through several dimensions: role-play time (time spent in active interaction with simulated patients), session quantity (total number of role-play sessions created per user), and interaction depth (average exchanges between trainees and virtual patients per session, where each exchange comprised one trainee input paired with one AI patient response).

Data Analysis

Descriptive statistics were computed for all continuous variables (e.g., SUS scores, engagement metrics, self-rated learning outcomes) and are presented as means (M), standard deviations (SD), medians, and ranges. Categorical data (e.g., demographic variables, competence development checklist items, comparative advantages) are presented as frequencies (n) and percentages (%). Missing data were addressed by using complete case analysis, resulting in varying sample sizes across different analyses, which are explicitly reported alongside each result (e.g., N=33 for SUS scores, N=31 for competence ratings). Qualitative responses from open-ended questionnaire questions were analyzed using content analysis to identify key themes in participant feedback [39].

Outcomes

The research team established pre-specified outcomes prior to study commencement, following implementation study guidelines [40]:

Primary Outcomes:

1. Platform engagement: Target of >50% of participants using the application for ≥10 minutes, or >25% using it for ≥30 minutes
2. Perceived educational value: Target of >70% of participants reporting positive impact on training

Secondary Outcomes:

1. Recruitment and retention: Target of ≥22 participants across both study phases
2. System usability: Target SUS score >68

Ethical Considerations

The study received approval from the UCL Research Ethics Committee (21883/006). The study protocol was pre-registered with the Open Science Framework (https://osf.io/mskb7). Participants were informed about the experimental nature of the study, and support pathways were established for technical issues or concerns. Data privacy and security were prioritized, with personal identifiers kept separate from research data. Participants were explicitly informed that the AI-based system was intended as a supplement to, not a replacement for, existing clinical supervision and training.

Results

Phase 1: Initial Usability Assessment

Phase 1 involved four doctoral clinical psychology students (all women, aged 25-34 years). Findings informed platform refinements, including enhanced navigation and modification of virtual patient communication patterns to better reflect authentic clinical progression.

Phase 2: Extended Implementation

Participant Characteristics

Figure 3 illustrates participant flow and attrition throughout the study phases. Of 101 invited, 92 (91.1%) were eligible based on iOS device compatibility. Attrition occurred primarily at app download, where 25 eligible participants (27.2%) did not proceed likely due to TestFlight app download complexity and technical barriers, and at survey completion, where 28 active users (47.5%) did not complete follow-up assessments, likely mistaking the end-of-study survey as optional educational monitoring rather than mandatory research data collection. Technical difficulties with older iOS devices prevented 8 participants (8.7% of eligible) from engaging despite successful installation. Overall, 59 participants (64.1% of eligible) successfully engaged with the platform, with 31-33 completing comprehensive survey assessments for secondary analyses.

[Insert Figure 3 here]

Figure 3. Participant Flow Diagram

Of the 59 engaging participants, 84.7% (n=50) are from the Low Intensity Cognitive Behavioral Interventions programme and 15.3% (n=9) are from the Doctorate in Clinical Psychology programme. Participants were predominantly women (83.1%, n=49) with a mean age of 28.37 years (SD=7.25, range 22-62). The ethnic distribution included White (57.6%, n=34), Black/African/Caribbean/Black British (15.3%, n=9), Asian/Asian British (11.9%, n=7), Other ethnic groups (8.5%, n=5), and Mixed/Multiple ethnic groups (6.8%, n=4). The majority of participants were at the early stages of their training, with a mean enrollment time of 1.61 months in their respective programs (SD=2.65, median=1.0, range 0-18 months). Regarding familiarity with competence assessment frameworks, 39.0% (n=23) reported being somewhat familiar, 25.4% (n=15) were neutral, 16.9% (n=10) were somewhat unfamiliar, 16.9% (n=10) were very unfamiliar or had never heard of

them, and 1.7% (n=1) were very familiar. Exploratory analyses revealed no substantial differences in engagement patterns, demographics, or outcomes between the two training programmes, supporting combined analysis of the sample.

Platform Engagement

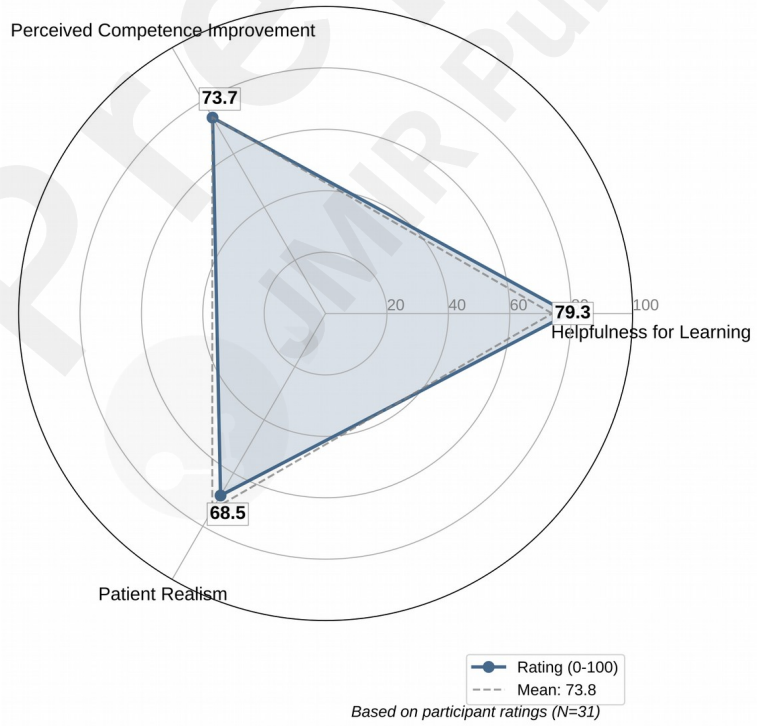
Of the 59 engaging participants , they spent an average of 95.24 minutes (SD=134.58, median=45.34) engaging in roleplays. The data was heavily right-skewed (most values clustered at the low end with a long tail of high values. While most participants engaged for shorter periods, a smaller number engaged for very long periods (several hours). Users created an average of 4.15 role-play sessions each (SD=3.55, range 1-13), with each session containing an average of 49.69 interactions.

System Usability

The CBT Trainer platform achieved high usability ratings with a mean SUS score of 82.2 (SD=12.93, median=82.5; N=33), placing it in the "excellent" category according to established benchmarks [38].

Self-Reported Competence Development

Participants reported high agreement that "The simulated patient interactions were helpful for my learning" (M=79.35, median=80.0, N=31) and "My clinical skills has improved after using CBT Trainer" (M=73.67, median=81.0, N=30), as illustrated in Figure 3. The statement "The simulated patients were realistic to real patients" received moderately strong agreement (M=68.45, median=70.0, N=31).

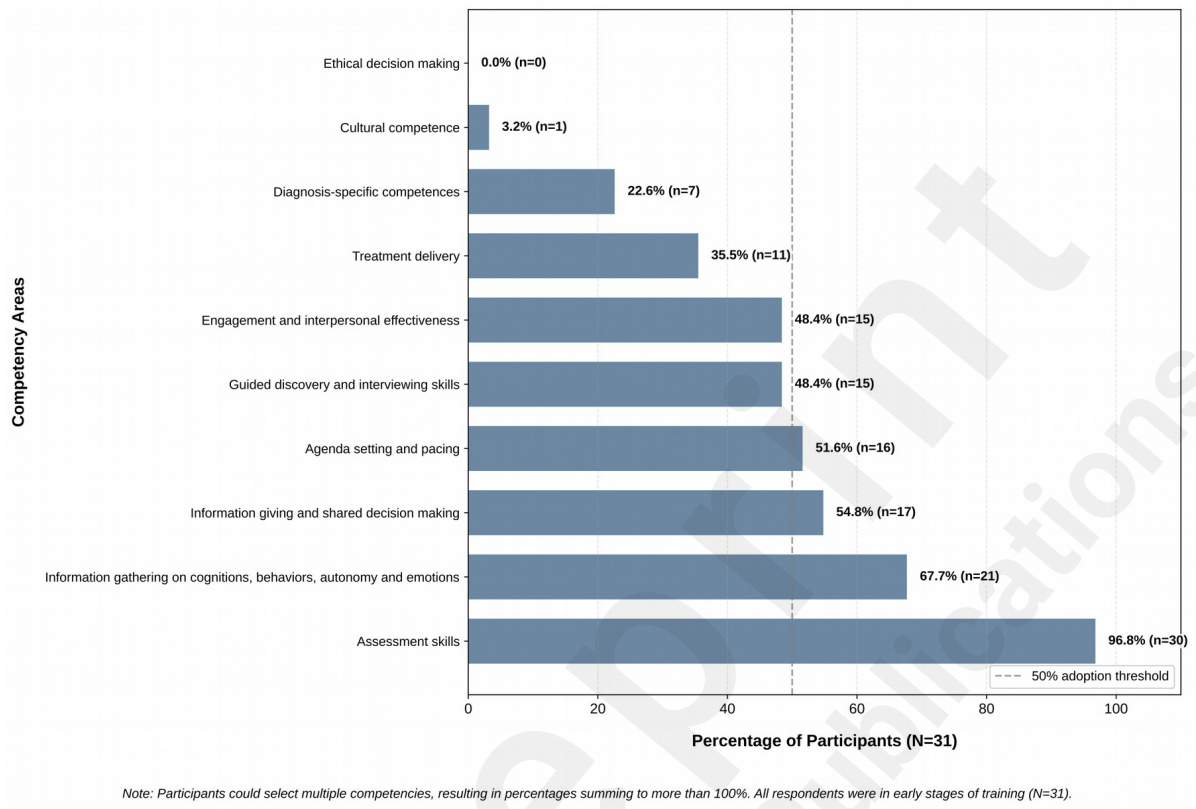


[Insert Figure 4 here]

Figure 4: **Educational Impact Ratings Reported by CBT Trainer Users.**

The distribution of self-reported competence development varied considerably across the assessed domains

(Figure 5). Some competence areas showed limited improvement, including ethical decision making (0.0%, n=0) and cultural competence (3.2%, n=1). The most frequently reported competence improvements (Figure 5) were in assessment skills (96.8%, n=30), followed by information gathering on cognitions, behaviors, autonomy and emotions (67.7%, n=21), and information giving and shared decision making (54.8%, n=17).



[Insert Figure 5 here]

Figure 5: **Competence Development Areas Reported by CBT Trainer Users.**

Qualitative Feedback on Learning Impact

Content analysis of qualitative feedback gathered through open-ended questions in the post-study questionnaire revealed several key themes. The most frequently cited benefit was specific skill development (71%, n=15), followed by feedback and competence assessment (38%, n=8), and accessibility for independent practice (29%, n=6).

Development of specific clinical competences was frequently mentioned, with participants reporting improvements in information gathering: *"This really helped me to improve my information gathering skills using the 4Ws and ABCs"* (Participant 15). Risk assessment skills were specifically highlighted: *"CBT Trainer helped me to practice risk assessment and safety planning"* (Participant 1) and finding ways to *"work my risk questions in a way that was conversational"* (Participant 4).

Session structure and pacing emerged as another improvement area: *"It helped with my pacing and knowing when to move on"* (Participant 4), while another appreciated being able to *"look at my timings afterwards to work out how much time I spent in each section"* (Participant 15). Additional participants mentioned the

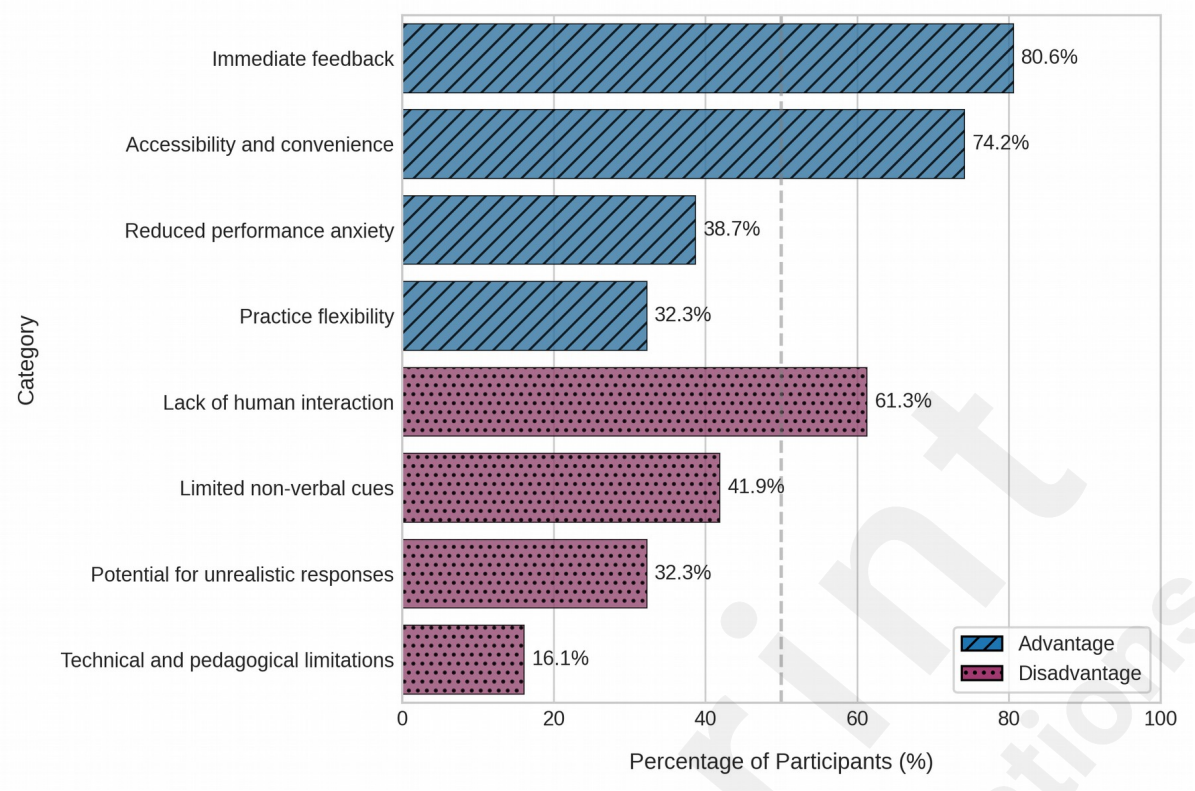
platform *"helped me to funnel and this positively impacted my clinical skills"* (Participant 6) and *"Identified areas that needed improvement, such as funneling and gathering information"* (Participant 9).

Participants valued the safe, judgment-free learning environment: *"practice and experiment with new techniques and skills during training without fear of judgement and with different client presentations"* (Participant 2). Accessibility and convenience were emphasized: *"I found it difficult to continuously practice with my friends and family... Being able to use the platform meant that I could practice an entire assessment repeatedly"* (Participant 3). Another noted that *"knowing that I could practice anytime of the day made me feel so much more calm about my exam as I didn't have to rely on other people to practice with"* (Participant 16).

Preparation for real-world clinical encounters was also noted: *"preparation for 'slow start' on first-year placements allowing channeling of nerves and planning for first sessions"* (Participant 10) and helping participants *"become more confident with my script and questioning style but also how to adjust quickly to what the patient said"* (Participant 16). Additional benefits included improved communication skills, with one participant noting *"Normally I stutter when talking to patients and this helped me to think before I speak"* (Participant 12).

Comparison to traditional role play

Participants were asked to evaluate "How does CBT Trainer compare to traditional role-play exercises with peers?" and rated CBT Trainer against traditional classroom role-play with a mean score of 5.90 (SD=1.94, median=6.0) on a 10-point scale favoring CBT Trainer, indicating moderate comparability. Figure 6 show participants' perceptions of the comparative advantages and disadvantages of CBT Trainer relative to traditional role-play methods.



[Insert Figure 6 here]

Figure 6: Perceived Advantages and Disadvantages of CBT Trainer Compared to Traditional Training Methods by CBT Trainer Users.

Primary and Secondary Outcomes

All pre-specified implementation targets were achieved. For platform engagement, 79.66% (n=47) of 59 active users engaged with the platform for ≥10 minutes, and 59.32% (n=35) used it for ≥30 minutes, both exceeding target thresholds. Perceived educational value similarly exceeded the target of >70% reporting positive impact with mean ratings of 77.58/100 for educational impact and 96.2% reporting improvement in assessment skills. Recruitment and retention substantially exceeded targets with 101 initial participants and 59 active users. The mean SUS score of 82.2 was well above the target threshold of >68, placing the platform in the "excellent" usability category.

Discussion

This pilot study evaluated an AI-driven virtual patient platform for training CBT practitioners against competencies. The findings demonstrate the potential of this technological approach to augment traditional training methods for CBT skills. We discuss these results in relation to existing literature, consider their implications for psychological training, and discuss limitations and future research directions.

Acceptability and Implementation

The study provides evidence for the acceptability and feasibility of implementing AI-based virtual patient simulations in psychological therapist training programs. Participants actively engaged in role-play

interactions with uptake exceeding our pre-specified outcome target. This level of engagement suggests that trainees recognize potential value in supplementing traditional training with CBT trainer and the simulations approximated realistic therapeutic encounters in terms of communication quantity and depth. The platform successfully demonstrates engagement and excellent usability, strongly supporting the platform's feasibility for implementation in clinical training contexts. [38]. These acceptability and feasibility findings align with previous research for AI-based simulated patients among healthcare students [26,29]. Our study extends these findings to psychological therapist trainees specifically, demonstrating that this population similarly values interactive simulations as learning tools.

Educational Impact

Participants reported substantial improvement in assessment skills and information gathering, suggesting that the platform effectively supports development of fundamental therapeutic competences. These improvements in core clinical skills supported by strong ratings for educational value and perceived clinical skill improvement mirror a similar study that observed significant gains in reflection skills among users of their patient-like conversational agent [28].

The self-reported benefits in areas such as risk assessment, therapeutic pacing, and interpersonal effectiveness suggest that virtual patient interactions can facilitate development across multiple competence domains. Structured AI simulations can enhance targeted clinical skills [41]. However, the varying degrees of improvement across different competence areas indicate that certain therapeutic skills may be more amenable to technological simulation than others.

CBT trainer’s capacity to facilitate immediate reflection on therapeutic interactions supports experiential learning [31], particularly enhancing the critical reflective observation and abstract conceptualization phases that bridge concrete experience and active experimentation. This immediacy addresses a fundamental requirement of deliberate practice [32], where skill acquisition depends on timely opportunities for reflection and refinement following performance attempts.

Participants consistently highlighted the benefit of practicing in a judgment-free environment. Traditional training models face interconnected challenges: performance anxiety in role-plays [18,22], evaluation apprehension inhibiting reflection [42] and delayed feedback cycles preventing systematic improvement of specific therapeutic microskills [7,43]. Participants' emphasis on the judgment-free practice environment suggests the platform addresses these challenges simultaneously, creating psychological safety that enables the deliberate, reflective skill development that conventional approaches struggle to achieve.

Comparative Advantages and Challenges

The balanced perspective offered by participants when comparing AI-based practice to traditional role-play methods highlights the complementary nature of these approaches. When asked about the platform's strengths, participants most frequently identified immediate feedback and convenience as significant educational

benefits. Many users also valued the reduced performance anxiety that virtual practice offered compared to peer-based role-plays. These advantages suggest that AI-based simulations can fill specific pedagogical gaps in existing training programs [18,34].

Participants also recognized certain limitations of the technology, particularly regarding interpersonal elements. The recognized lack of non-verbal cues and potential for unrealistic responses echo concerns raised regarding the authenticity of AI-based clinical simulations [30]. These limitations reinforce that current AI technology serves as a complement to, rather than replacement for, conventional training models incorporating interpersonal engagement and supervision.

The pattern of responses regarding comparative advantages aligns with the "pedagogical affordances" framework [44], which suggests that different educational technologies offer unique benefits for specific learning objectives. The participants' nuanced assessment of when and how the platform added value to their training reflects a sophisticated understanding of how technological tools fit within their broader educational experience.

Limitations and Future Directions

Our findings should be interpreted with several limitations in mind. Technical limitations restricting participation to iOS users may have introduced selection bias. A substantial proportion of eligible participants who initially expressed interest did not progress to downloading or using the application. This pattern of attrition suggests the importance of understanding barriers to technology adoption in clinical training contexts, whether technical, motivational, or practical in nature. Future research should systematically investigate factors influencing technology adoption among clinical trainees, including surveys of non-users to identify specific barriers and preferences.

Methodologically, our mixed-methods approach provided rich experiential data, but the absence of a control group prevents causal inferences about the platform's impact on skill development. The study relied on self-reported skill improvement rather than objective competence assessment, weakening conclusions about educational efficacy. The one-month evaluation period offers limited insight into long-term engagement patterns. Future evaluations should incorporate standardized competence evaluations, longitudinal designs across training stages, cross-platform accessibility, and randomized controlled trials. Promising research directions include investigating simulation of diverse clinical presentations and therapeutic modalities, optimal curricular integration, and the relationship between simulation fidelity and learning outcomes.

Conclusions

This pilot study demonstrated that an AI-based virtual patient simulation shows promise as a supplementary training tool for psychological therapists. The platform achieved high usability ratings with participants reporting significant skill development, particularly in assessment competences, suggesting effective targeting

of specific training needs in cognitive behavioral therapy education. Key strengths included accessibility, immediate feedback, and psychological safety—allowing practice without the anxiety that may be associated with peer role-plays, consistent with previous simulation-based learning research. The findings support further work to implement AI-based virtual patient simulation in psychological therapist training by providing accessible, on-demand practice opportunities with feedback supporting competence development, enhancing traditional approaches in today's technology-enabled educational landscape.

Acknowledgements

We thank Dr. Rachel Newman and Dr. Joshua Buckman for their support during participant recruitment, and thank all participants for their participation.

This research received no external financial or non-financial support from any funding agency, commercial or not-for-profit sectors.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

TTZ is the founder and director of Soy Educational Technologies Limited that developed the CBT Trainer platform evaluated in this study. No other investigators report any conflicts of interest.

Abbreviations

CBT - Cognitive Behavioral Therapy

AI - Artificial Intelligence

SUS - System Usability Scale

CTS-R - Cognitive Therapy Scale-Revised

PWP - Psychological Wellbeing Practitioner

NHS - National Health Service

References

1. NHS England. NHS England » NHS Long Term Workforce Plan. 2023. Available from: https://www.england.nhs.uk/long-read/nhs-long-term-workforce-plan-2/ [accessed May 23, 2025]

2. Roth AD, Pilling S. Using an evidence-based methodology to identify the competences required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. Behavioural and Cognitive Psychotherapy United Kingdom: Cambridge University Press; 2008;36(2):129–147. doi: 10.1017/S1352465808004141

3. Roth AD. A New Scale for the Assessment of Competences in Cognitive and Behavioural Therapy. Behav Cogn Psychother 2016 Sept;44(5):620–624. doi: 10.1017/S1352465816000011

4. Blackburn I-M, James IA, Milne DL, Baker C, Standart S, Garland A, Reichelt FK. Revised Cognitive Therapy Scale. 2001. doi: 10.1037/t65807-000

5. Young J, Beck A. Cognitive Therapy Scale. 1980. doi: 10.1037/t00834-000

6. Reichelt FK, James IA, Blackburn I-M. Impact of training on rating competence in cognitive therapy. Journal of Behavior Therapy and Experimental Psychiatry 2003 June 1;34(2):87–99. doi: 10.1016/S0005-7916(03)00022-3

7. Bennett-Levy J. Therapist skills: A cognitive model of their acquisition and refinement. Behavioural and Cognitive Psychotherapy United Kingdom: Cambridge University Press; 2006;34(1):57–78. doi: 10.1017/S1352465805002420

8. Muse K, Kennerley H, McManus F. The why, what, when, who and how of assessing CBT competence to support lifelong learning. the Cognitive Behaviour Therapist 2022 Jan;15:e57. doi: 10.1017/S1754470X22000502

9. Muse K, McManus F. A systematic review of methods for assessing competence in cognitive–behavioural therapy. Clinical Psychology Review 2013 Apr 1;33(3):484–499. doi: 10.1016/j.cpr.2013.01.010

10. Liness S, Beale S, Lea S, Byrne S, Hirsch CR, Clark DM. Evaluating CBT Clinical Competence with Standardised Role Plays and Patient Therapy Sessions. Cogn Ther Res 2019 Dec 1;43(6):959–970. doi: 10.1007/s10608-019-10024-z

11. Sharpless BA, Barber JP. The Examination for Professional Practice in Psychology (EPPP) in the era of evidence-based practice. Professional Psychology: Research and Practice US: American Psychological Association; 2009;40(4):333–340. doi: 10.1037/a0013983

12. Marriott BR, Cho E, Tugendrajch SK, Kliethermes MD, McMillen JC, Proctor EK, Hawley KM. Role-Play Assessment of Therapist Adherence and Skill in Implementation of Trauma-Focused Cognitive-Behavioral Therapy. Adm Policy Ment Health 2022 May;49(3):374–384. doi: 10.1007/s10488-021-01169-9

13. Loades ME, Armstrong P. The challenge of training supervisors to use direct assessments of clinical competence in CBT consistently: a systematic review and exploratory training study. The Cognitive Behaviour Therapist 2016/10/14 ed Cambridge University Press; 2016;9:e27.

doi: 10.1017/S1754470X15000288

14. Johnson J, Corker C, O'connor DB. Burnout in psychological therapists: A cross-sectional study investigating the role of supervisory relationship quality. Clinical Psychologist Taylor & Francis; 2020 Nov 1; doi: 10.1111/cp.12206

15. British Psychological Society. Code of Ethics and Conduct. 2021. p. bpsrep.2021.inf94. doi: 10.53841/bpsrep.2021.inf94ISBN:978-1-85433-804-4

16. British Psychological Society. Supervision guidance for psychologists. 2024. p. bpsrep.2024.rep178. doi: 10.53841/bpsrep.2024.rep178ISBN:978-1-85433-904-1

17. Caltabiano M, Errington E, Sorin R, Nickson A. The potential of role-play in undergraduate psychology training / Marie Caltabiano … [et al.]. Asian Journal of University Education (AJUE) Penerbit UiTM (UiTM Press); 2018;14(1):1–14.

18. Melluish S, Crossley J, Tweed A. An Evaluation of the Use of Simulated Patient Role-Plays in the Teaching and Assessment of Clinical Consultation Skills in Clinical Psychologists' Training. Psychology Learning & Teaching SAGE Publications; 2007 Sept 1;6(2):104–113. doi: 10.2304/plat.2007.6.2.104

19. Doolen J, Giddings M, Johnson M, Nathan GG de, Badia LO. An Evaluation of Mental Health Simulation with Standardized Patients. International Journal of Nursing Education Scholarship De Gruyter; 2014 Jan 1;11(1):55–62. doi: 10.1515/ijnes-2013-0075

20. Himmelbauer M, Seitz T, Seidman C, Löffler-Stastka H. Standardized patients in psychiatry – the best way to learn clinical skills? BMC Medical Education 2018 Apr 6;18(1):72. doi: 10.1186/s12909-018-1184-4

21. Badger LW, Macneil G. Rationale for Utilizing Standardized Clients in the Training and Evaluation of Social Work Students. Journal of Teaching in Social Work Routledge; 1998 Jan 26;16(1–2):203–218. doi: 10.1300/J067v16n01_13

22. Fall KA, Levitov JE. Using Actors in Experiential Group Counseling Leadership Training. The Journal for Specialists in Group Work Routledge; 2002 June 1;27(2):122–135. doi: 10.1177/0193392202027002002

23. Pilnick A, Trusson D, Beeke S, O'Brien R, Goldberg S, Harwood RH. Using conversation analysis to inform role play and simulated interaction in communications skills training for healthcare professionals: identifying avenues for further development through a scoping review. BMC Med Educ 2018 Dec;18(1):267. doi: 10.1186/s12909-018-1381-1

24. Ruiz Rodríguez J, Bados López A, Fusté Escolano A, García-Grau E, Saldaña García C, Balaguer Fort G, Lluch T, Arcos Pros M. Peer counselling versus role-playing: Two training methods of therapeutic skills in clinical psychology. Psicothema 2018 Feb 1;1(38):21–26. doi: 10.7334/psicothema2016.286

25. Englar RE. Using a Standardized Client Encounter to Practice Death Notification after the Unexpected Death of a Feline Patient Following Routine Ovariohysterectomy. Journal of Veterinary Medical Education University of Toronto Press; 2019 Nov;46(4):489–505. doi: 10.3138/jvme.0817-111r1

26. Prescott J, Ogilvie L, Hanley T. Student therapists' experiences of learning using a machine client: A proof-of-concept exploration of an emotionally responsive interactive client (ERIC). Counselling and Psychotherapy Research 2024;24(2):524–531. doi: 10.1002/capr.12685

27. Zalewski B, Guziak M, Walkiewicz M. Developing Simulated and Virtual Patients in Psychological Assessment – Method, Insights and Recommendations. Perspect Med Educ 2023;12(1):455–461. PMID:37901884

28. Tanana MJ, Soma CS, Srikumar V, Atkins DC, Imel ZE. Development and Evaluation of ClientBot: Patient-Like Conversational Agent to Train Basic Counseling Skills. Journal of Medical Internet Research 2019 July 15;21(7):e12529. doi: 10.2196/12529

29. De Mattei L, Morato MQ, Sidhu V, Gautam N, Mendonca CT, Tsai A, Hammer M, Creighton-Wong L, Azzam A. Are Artificial Intelligence Virtual Simulated Patients (AI-VSP) a Valid Teaching Modality for Health Professional Students? Clinical Simulation in Nursing 2024 July 1;92:101536. doi: 10.1016/j.ecns.2024.101536

30. Maurya RK. A Qualitative Content Analysis of ChatGPT's Client Simulation Role Play for Practicing Counseling Skills. OSF; 2023. doi: 10.31234/osf.io/vwuxh

31. Kolb D. Experiential Learning: Experience As The Source Of Learning And Development. Englewood Cliffs, NJ: Prentice Hall.; 1984. ISBN:978-0-13-295261-3

32. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychological Review US: American Psychological Association; 1993;100(3):363–406. doi: 10.1037/0033-295X.100.3.363

33. Rakovshik SG, McManus F. Establishing evidence-based training in cognitive behavioral therapy: A review of current empirical findings and theoretical guidance. Clinical Psychology Review 2010 July 1;30(5):496–516. doi: 10.1016/j.cpr.2010.03.004

34. Rogers SL, Hollett R, Li YR, Speelman P. An Evaluation of Virtual Reality Role-Play Experiences for Helping-Profession Courses. Teaching of Psychology SAGE Publications Inc; 2022 Jan 1;49(1):78–84. doi: 10.1177/0098628320983231

35. CBT Trainer. CBT Trainer - AI-Powered Psychotherapy Training | Virtual Patient Roleplay. 2025. Available from: https://www.soymh.com

36. NHS England. National Curriculum for Psychological Wellbeing Practitioner (PWP) Programmes. 2022. Available from: https://www.ucl.ac.uk/brain-sciences/sites/brain_sciences/files/pwp_review_-_final_report.pdf [accessed Sept 20, 2025]

37. Brooke J. SUS - A quick and dirty usability scale. Usability evaluation in industry Taylor & Francis.; 1996. p. 189–194.

38. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction Taylor & Francis; 2008 July 29;24(6):574–594. doi: 10.1080/10447310802205776

39. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res

2005 Nov;15(9):1277–1288. PMID:16204405

40. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, Rycroft-Malone J, Meissner P, Murray E, Patel A, Sheikh A, Taylor SJC. Standards for Reporting Implementation Studies (StaRI) Statement. BMJ 2017 Mar 6;i6795. doi: 10.1136/bmj.i6795

41. Wang R, Milani S, Chiu JC, Zhi J, Eack SM, Labrum T, Murphy SM, Jones N, Hardy K, Shen H, Fang F, Chen ZZ. PATIENT-Ψ: Using Large Language Models to Simulate Patients for Training Mental Health Professionals. arXiv; 2024. doi: 10.48550/arXiv.2405.19660

42. Rapisarda C, Jencius M, McGlothlin J. Master's students' experiences in a multicultural counseling role-play. International Journal for the Advancement of Counselling Germany: Springer; 2011;33(4):361–375. doi: 10.1007/s10447-011-9139-z

43. Chow DL, Miller SD, Seidel JA, Kane RT, Thornton JA, Andrews WP. The role of deliberate practice in the development of highly effective psychotherapists. Psychotherapy US: Educational Publishing Foundation; 2015;52(3):337–345. doi: 10.1037/pst0000015

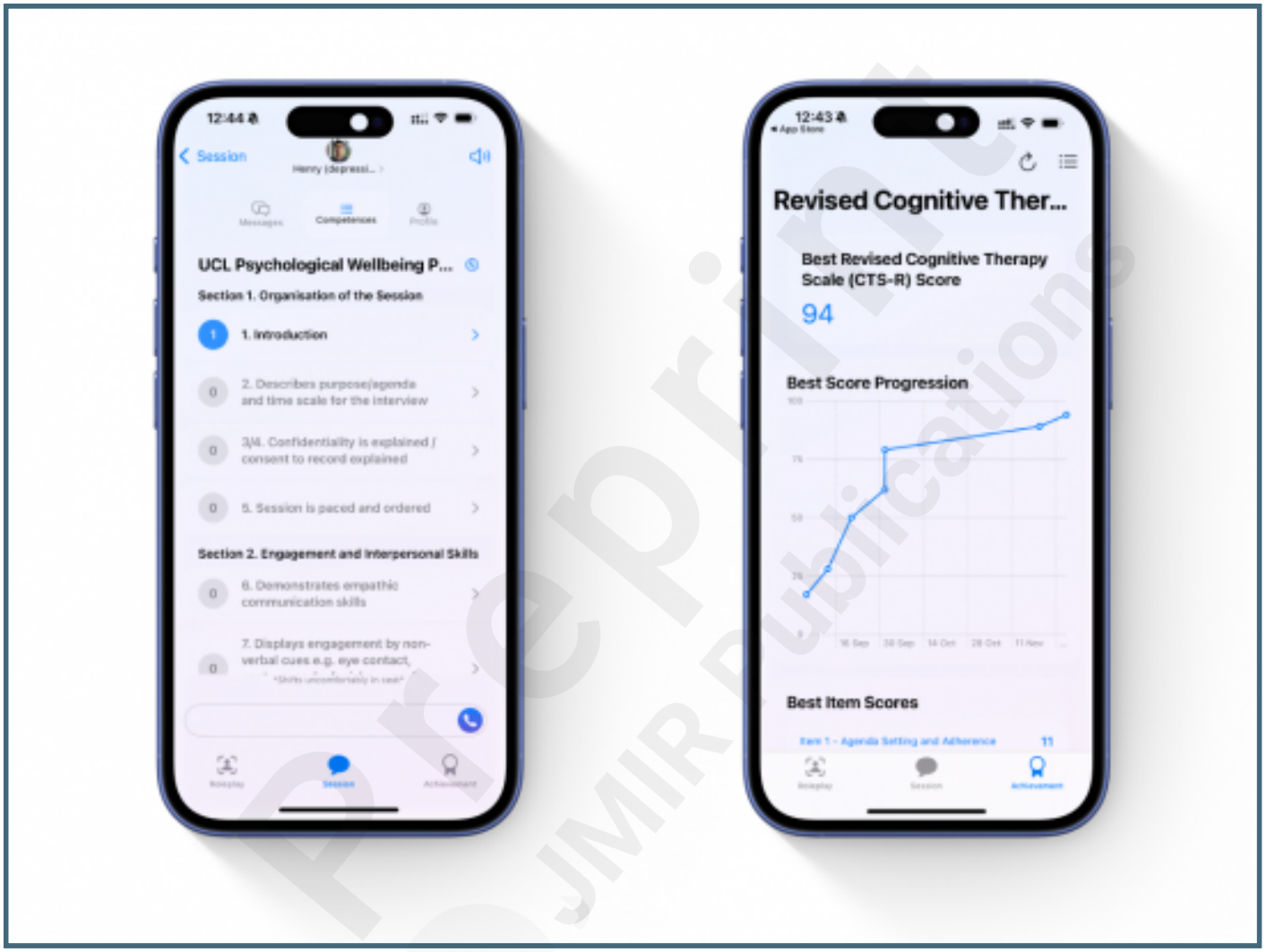
44. Laurillard D. Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies. 2nd ed. London: Routledge; 2013. doi: 10.4324/9781315012940ISBN:978-1-315-01294-0

Supplementary Files

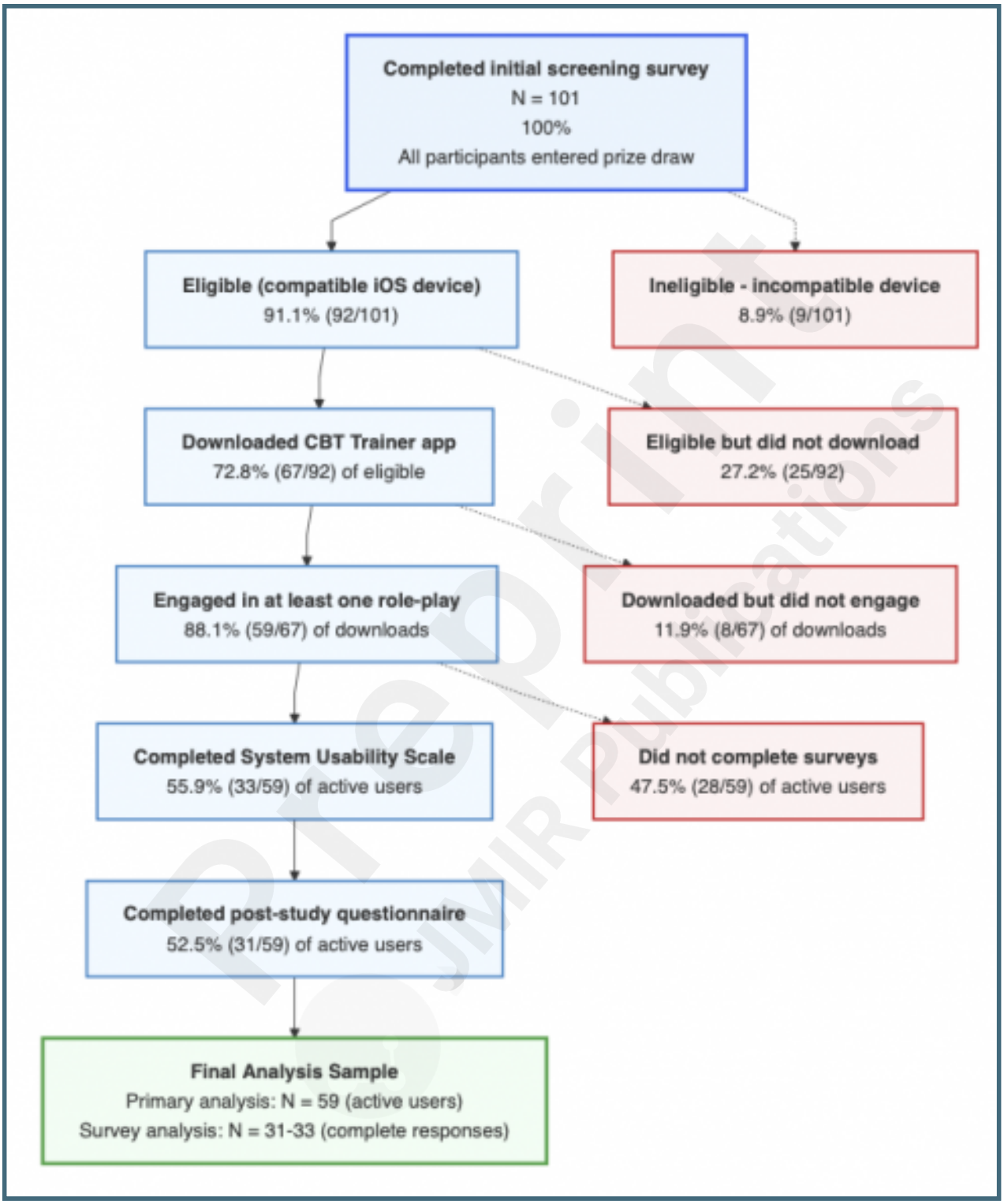
Figures

CBT Trainer Patient Selection and Role-play Interfaces. (Left) Patient selection interface shows patient profiles representing different demographics, ages, ethnicities, and clinical presentations including depression, social anxiety, generalized anxiety, agoraphobia & panic, and PTSD. (Right) Interactive simulation interface demonstrating voice-based therapeutic interaction with virtual patient Penelope, showing the conversational flow with real-time competence tracking visible in the interface background. The system emphasizes ethical considerations by clearly stating that "AI roleplay cannot substitute for human feedback or supervision" and reminding trainees to bring concerns to their supervision.

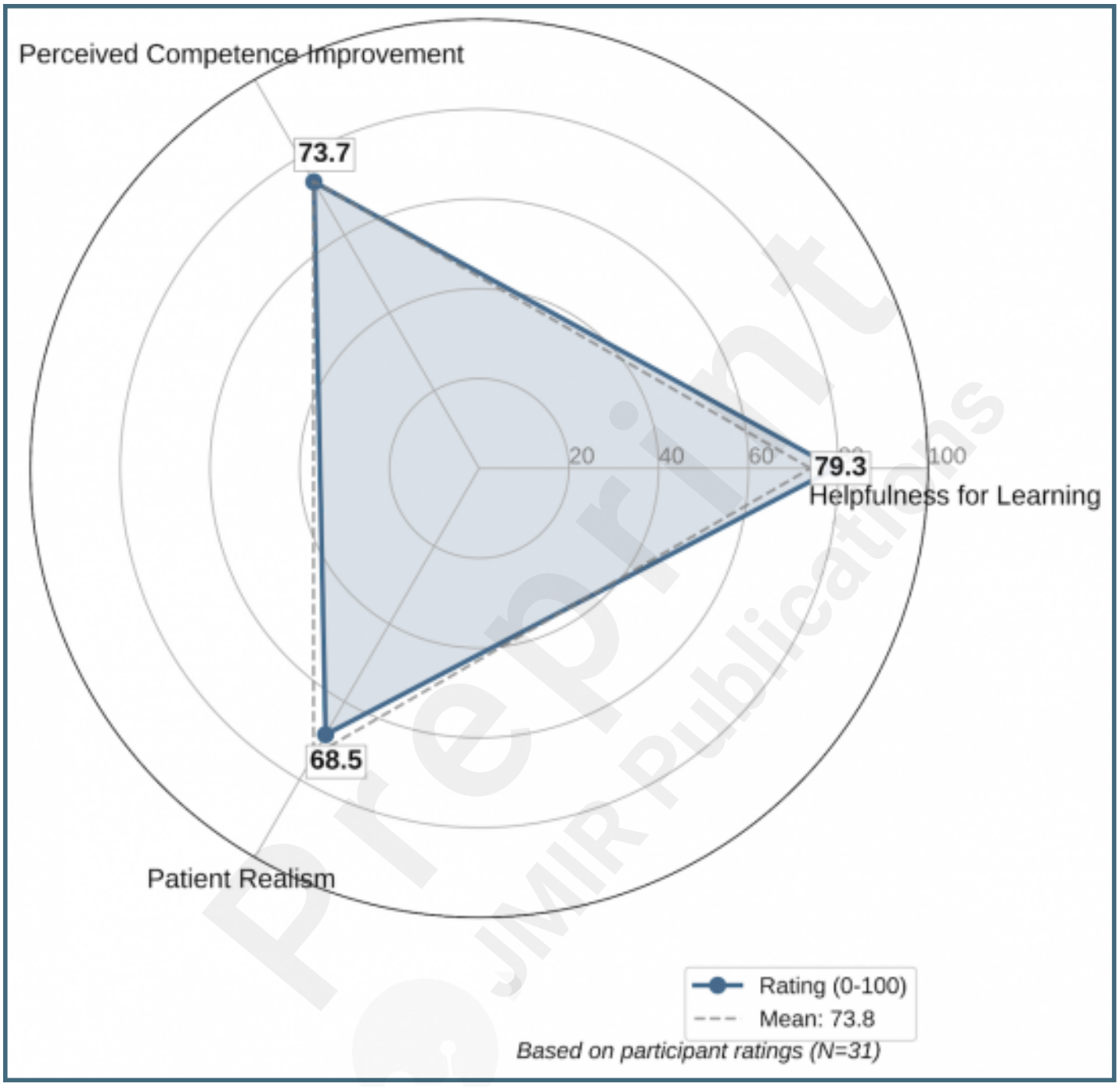
Competence Assessment and Progress Tracking Interfaces. (Left) UCL Psychological Wellbeing Practitioner (PWP) Assessment Competence Scale interface showing real-time competence evaluation across core domains including "Organization of the Session" and "Engagement and Interpersonal Skills." Each competence item displays scoring indicators that update as trainees demonstrate skills during interactions. (Right) Cognitive Therapy Scale-Revised (CTS-R) progress tracking screen displaying longitudinal competence development and visual progression chart showing improvement over time from 16 September to 11 November, alongside individual item scores for specific competences such as "Agenda Setting and Adherence".



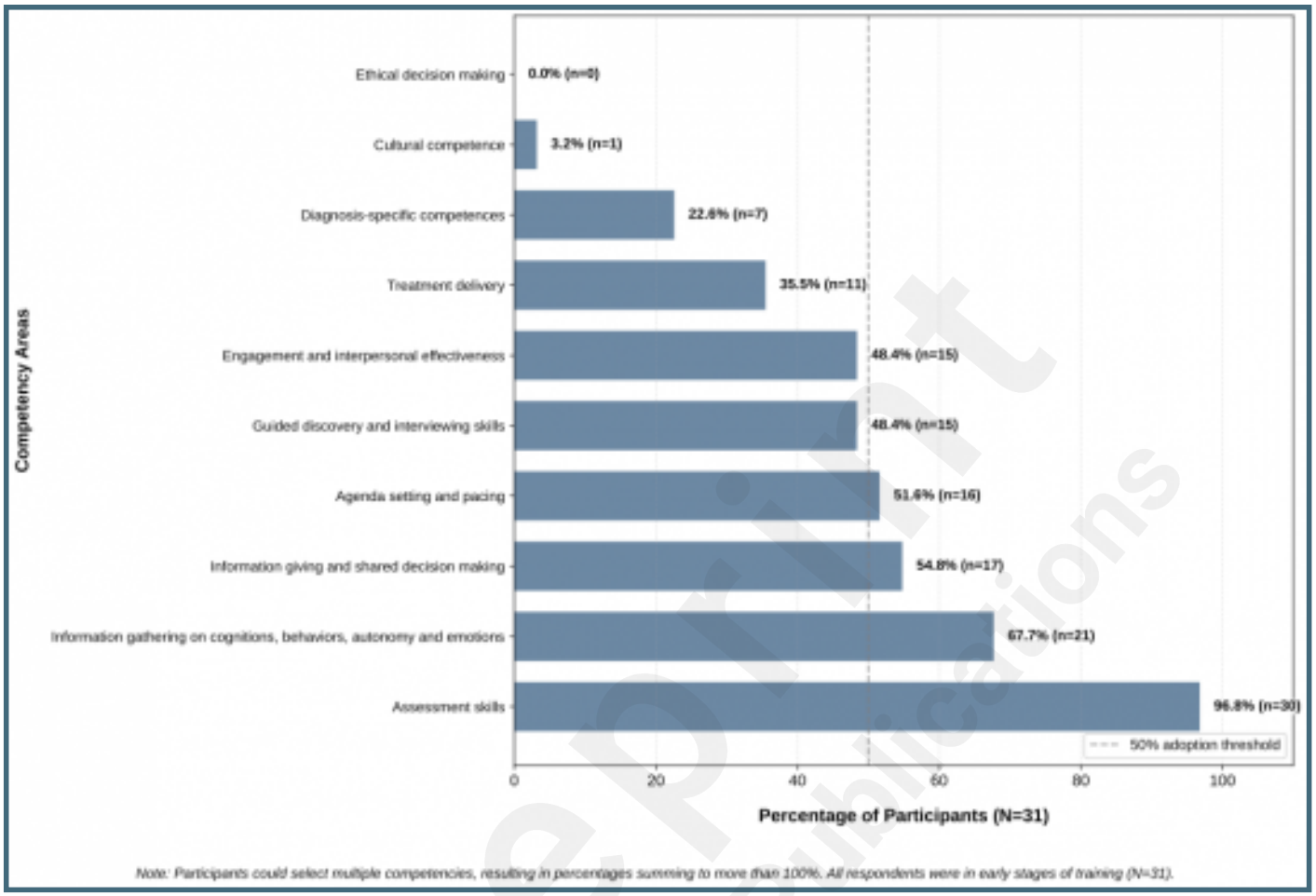
Participant Flow Diagram.



Educational Impact Ratings Reported by CBT Trainer Users.



Competence Development Areas Reported by CBT Trainer Users.



Perceived Advantages and Disadvantages of CBT Trainer Compared to Traditional Training Methods by CBT Trainer Users.