

# **An educational parallel randomized controlled trial comparing training in gestalt psychotherapy with a response to narrative program vs. standard training and its effect on knowledge and clinical practice among healthcare professionals: A reproducible research protocol**

Raffaele Sperandeo  
Mauro Maldonato  
Bruno Melo  
Joao Ricardo Vissoci  
Paolo Valerio  
Ricardo Pietrobon

## **Abstract**

## **Introduction**

Gestalt psychotherapy is considered by experts as one of the most difficult forms for psychotherapy when it comes to the training of healthcare professional, primarily due to the complexity and depth of its concepts. Specifically, gestalt psychotherapy focuses on an existential or experiential form therapy, emphasizing concepts such as personal responsibility, the individual's experience in the moment, the relationship between therapist and client, the situation surrounding clients in their environmental and social contexts as well as the iterative, self-regulating adjustments clients make to their contexts. Although much of the knowledge and professional experience is gained while being coached in interacting with individual clients, gestalt therapy training could be substantially enhanced if more scalable, ancillary forms of training were to be available. Unfortunately, to our knowledge, no previous studies have been published evaluating alternative forms of professional training.

- lit review on Gestalt therapy training - phenomenological method, dialogical relationship, field-theoretical strategies, and experimental freedom
- lit review on Online learning for psychotherapy in general

The objective of this parallel randomized trial is therefore to compare standard training vs. an intervention involving {behavior} in educating college students regarding {course}, the study design, conductance and reporting following a reproducible research approach. We hypothesize that the intervention would result in decision making that is more fully supported by information gathered in

a more comprehensive manner through technological and socially means, actions that are more effective and adaptive and also more empathetic and connected to the needs of others.

## Methods

This trial was designed and reported in accordance with the recommendations provided by the [CONSORT statement](#)

## Ethics

This study was approved by the Ethics Committee at the University of Basilicata ([Università degli Studi della Basilicata](#)), informed consent having been waived since this was considered a research with minimal risk within an educational context. All participants were ensured that their participation was absolutely unrelated to their performance in their current disciplines and that they could drop out at any point if they so wanted.

## Trial Design

This is a parallel controlled randomized trial comparing (1) a narrative method where students react to videos with scenarios where fictitious patients present them with common challenges and where they have to respond using aspects of the phenomenological method, dialogical relationships, field-theoretical strategies, and of the experimental freedom, versus (2) a control with standard training through texts about the phenomenological method, dialogical relationships, field-theoretical strategies, and of the experimental freedom. The trial involved 106 participants who completed the trial, randomized on a 1:1 ratio to the intervention and control arms, with the intervention and respective follow-up lasting a total of four weeks. No changes to the initially methods design were implemented while the trial was in course.

## Participants

Our study involved 106 participants. Participants were undergraduate and graduate students at the University of Basilicata with an interest in Gestalt psychotherapy, located in the Basilicata region of Southern Italy. Demographic information is provided as a separate dataset in our reproducible research protocol to preserve participant confidentiality and prevent loss of confidentiality. For each participant we have collected information on gender, age, previous educational background, and baseline knowledge about psychotherapy.

## Interventions

The narrative training arm was constituted by a short course with weekly video lectures and exercises focused on Gestalt therapy. Specifically, a narrative method was where students had to react to videos with scenarios where fictitious patients present them with common challenges and where they have to respond using aspects of the phenomenological method, dialogical relationships, field-theoretical strategies, and of the experimental freedom. After watching the video, each participant responded to a series of multiple alternative questions where they were asked how they would react during a therapy session in relation to a series of situations that normally occur in daily practice. The videos and subsequent questions were devised to provide insights and develop skills that could later be used in a real scenario, similar to textbooks such as [Gestalt Therapy Verbatim](#), by Friedrich (Fritz) Perls, who developed Gestalt therapy with his wife Laura Perls.

The traditional training arm was constituted by a series of weekly texts discussing theoretical aspects of the phenomenological method, dialogical relationships, field-theoretical strategies, and of the experimental freedom. After reading each text, students were asked to respond to multiple alternative questions.

All educational material is made available within our Reproducible Research repositories, and delivered through the original [Open edX platform](#). Videos were provided in Italian. All remaining material was made available in Italian, although both Italian and English versions are provided within our Reproducible Research repositories (see subsequent section on Reproducible Research).

## Outcome variables

Our study had three main categories of outcome variables: Knowledge, User experience (UX) and impact on clinical practice, all described under the subsequent sections. Per CONSORT Statement guidelines, any eventual changes to outcome measures deviating from the initial plan and occurring throughout the trial were recorded in our versioning system ([Github](#)), thus ensuring full reproducibility. However, changes were avoided unless strictly necessary.

**Knowledge** Knowledge was measured through a group of questions (items) under four main domains, namely: phenomenological method, dialogical relationships, field-theoretical strategies, and of the experimental freedom. Given that there are no standardized scales to measure knowledge within each of these domain areas, we conducted a parallel validation using a combination of Classical Psychometric Theory and Item Response Theory.

An item (question) pool was initially created, covering all domains, ultimately constituting items sets with 30 items (questions) each, for a total of 120 items (questions).

A randomly selected subset of this item pool of 120 items was exposed to the participants in our trial, using an anchoring approach for linked scales where each set contains a constant group of items that is repeated in every set, thus ensuring that all sets were connected. Subsequently, a sequence of exploratory and confirmatory factor analyses, reliability evaluation within each individual construct, and validity triangulating across different items was used to achieve a psychometrically valid summed construct-specific and overall scores to serve as the variables representing the knowledge outcome of our trial. Details about each of these analyses are presented within the Data Analysis section.

Knowledge evaluation occurred at baseline and then at the end of weeks two and four. These assessments were automatically recorded within the edX Code platform under two databases ([MySQL](#) for participant information and [MongoDB](#) for exercise information), both with a common id.

**User eXperience** In this study we primarily measured User eXperience (UX) as participant satisfaction while participating in each of the two interventions. While the literature on the measurement of satisfaction within online education is vast, (see @banks2007reduction, @dibiase2005scaling and @summers2005comparison for a few examples) we have elected to use the scale developed by [Kuo et cols](#) since it is not only comprehensive but has also been shown to be psychometrically valid. The Sampson scale covers seven satisfaction sub-constructs, namely satisfaction with learner-learner interaction, learner-instructor interaction, learner-content interaction and overall satisfaction. Cronbach’s alpha reliability was found to be above 0.88 for all subscales, the Kuo scale also having been validated against student self-efficacy.

Satisfaction was assessed at the end of weeks two and four using a set of questionnaires provided within the Open edX platform. These assessments are automatically recorded within the platform.

**Impact on clinical practice** The last evaluated construct was the impact on clinical practice. Given that, to our knowledge, no previously validated scale has been devised to measure the impact of online learning on clinical practice, we have also devised a concomitant validation strategy.

Impact on clinical practice was measured through a group of questions (items) under one single domain (unidimensional scale). The item (question) focused on “How much has the knowledge you gained in this course has positively or negatively influenced your clinical practice, where 0 means no influence at all and 10 means an extremely influence.” . A five-point Likert alternative pattern then covered the spectrum of positive or negative influence.

This group of items was exposed to the participants in our trial. Similar to the knowledge domain, a sequence of exploratory and confirmatory factor analyses, reliability evaluation within each individual construct, and validity triangulating

across different items was used to achieve a psychometrically valid summed score serving as the outcome measuring impact on clinical practice for our trial. Further details on how each of these analyses was conducted are presented within the Data Analysis section.

Knowledge evaluation occurred at the end of weeks two and four. These assessments are also automatically recorded within the platform.

**Sample size** Given the absence of preliminary efficacy data, we calculated our sample size based on an assumption of a two-sample t-test for an index representing the worst case scenario for each of the three outcomes, with an effect size of 55%, a significance level and statistical power set at the traditional levels of 0.05 and 80%, respectively. Under these assumptions the estimated required sample size is of 52.9 per group, which we have rounded to 53, ultimately leading to a total sample size of 106 participants in total. Taking into account a worst case scenario dropout rate of approximately 20%, we therefore plan to enroll 140 participants.

**Interim analyses and stopping guidelines** Given the relatively short duration of our trial, we did not perform an interim analyses. In addition, given the low risk associated with the intervention, our guidelines did recommend that the trial be stopped in case of an unlikely loss of confidentiality that might compromise participants privacy, which has not occurred.

**Randomization: sequence generation** All randomization schedules were delivered through [Planout](#) within the edX Code platform. Briefly, Planout makes use of a pseudorandom number generator from the [Python language](#), which has been extensively tested. Randomization was blocked at 10 participants, thus decreasing the likelihood of a possible imbalance.

All randomization was delivered through the Open edX platform, thus ensuring that participants and investigators are blind to the allocation. Data analysts (RP and JRV) received a dataset from one of the computer scientists (BM) where the two randomization arms were designated by unspecific letters, thus ensuring that the analysis is also conducted in a blind fashion.

**Statistical methods** All analyses were conducted by the members in our team (RP, JRV). The data analysis started with an overall evaluation of individual variables to assess distributions for continuous variables (e.g., participant's age) in terms of their distributions, as well as the frequency/percentage for categorical variables (e.g., gender).

Missing values were handled differently if they happened to individual variables or full encounters, e.g., an entire evaluation for a week. Individual variables were

imputed starting by an exploratory visual analysis to better understand the potential underlying pattern behind missing values for each variable. This analysis was conducted using the [VIM package](#) within the [R language](#). Depending on the pattern of missing data, we might then choose one of several alternatives, ranging from not imputing when the percent missing might be negligible, imputing using predictive models from variables that are hypothesized to be associated with the missing values (e.g., age predicting residency year), or multiple imputation in cases where missing patterns might be happening at random (MAR or missing at random) ([@rubin1976inference](#)).

We evaluated for potential confounding of the intervention effect by assessing balance across baseline variables between the 3D and control arms. This evaluation occurred for the overall study. While imbalances were unlikely given that we are using a series of preventive measures (blinding, concealed allocation, blocking and stratification), eventual imbalances were evaluated on an individual basis and required adjusting in our modeling strategy (see below for details).

The psychometric analysis for parallel validation of our scales started with an Exploratory Factor Analysis to evaluate the underlying dimensional structure of our item pool. We then conducted a Confirmatory Factor Analysis using the hypothesized domains (see Outcome Variable section for extensive details). Once domains are established we used Cronbach’s alpha to measure internal reliability within each construct. Validity was measured by triangulation across different domains.

Given that all of our outcomes have multiple endpoints over time, we used a combination of mixed models and survival analysis to measure the efficacy associated with the 3D in comparison with the control arm. To formally test trends in each of our outcome variables over time (knowledge, User eXperience and impact on clinical practice) we developed mixed models that account for multiple measures as random a effect. For each time period we also assessed differences in scores. These bivariate analyses incorporated robust standard errors to account for the clustering effect. Survival models included Cox Proportional Hazards models measuring time until an information reached clinical practice. The hazards assumption were tested prior to analyses and eventual ties were resolved.

## **Reproducible research**

In order to follow a reproducible research protocol, we have written our final manuscript using Rmarkdown, which combines the R language with the [Markdown markup](#), ultimately allowing us to leave all calculations and data manipulation within the article text. All data sets and respective analytical scripts were provided in CSV (comma separated value) formats in open public repositories including [Figshare](#) and [Github](#). In addition, we have also stored the version of the software used in this analysis and deposited it within our Github

and Figshare repositories, ultimately decreasing the probability of reproducibility issues during subsequent attempts to reproduce our analyses.

## **Results**

## **Discussion**