*H2020-SC1-DTH-2018-2020*

Type of Action: RIA Research and Innovation action

Topic: **Adaptive smart working and living environments supporting active and healthy ageing**

Grant Agreement no: 826266

Deliverable

# D6.4 Evaluation of CO-ADAPT Application

COADAPT

Start date of the project: December 1, 2018 Duration: 42 months

| Project funded by the European Commission within the Horizon 2020 programme for research, technological development and demonstration | | |
|---|---|---|
| Dissemination Level | | |
| PU | Public, fully open | **X** |
| CO | Confidential, restricted under conditions set out in Model Grant Agreement | ☐ |
| CL | Classified | ☐ |

## Notices

For information, please contact the project coordinator, Prof Giulio Jacucci, e-mail giulio.jacucci@helsinki.fi

This document is intended to fulfil the contractual obligations of the CO-ADAPT project, which has received funding from the European Union's Horizon 2020 Programme, concerning deliverable 6.4 described in contract 826266.

# Table of Revisions

| Version | Date | Description and reason | Author | Affected sections |
|---|---|---|---|---|
| **v0.1** | 02142022 | Table of Contents | T. Ciulli | All |
| **V0.2** | 02282022 | Ownership of each section and deadline | T. Ciulli | All |
| **V0.3** | 03012022 | Drafted Introduction and Psychological Model | T. Ciulli, G. Silvestri | 2, 3 |
| **V0.4** | 03032022 | Drafted Participatory design | T. Ciulli | 4 |
| **V0.5** | 03142022 | Drafted Participatory design, Experimental protocol, Methodology of intervention | G. Silvestri | 4, 5, 6 |
| **V0.6** | 03172022 | Drafted Experimental results, Data collection | G. Silvestri, T. Ciulli | 7, 8 |
| **V0.7** | 03212022 | Drafted The role of the APP and PHA | T. Ciulli | 6.1 |
| **V0.8** | 03232022 | Drafted APP and PHA data collection and User Engagement | G. Silvesti, C. Ferrante | 8.3 |
| **V0.9** | 03242022 | Drafted Conclusion and future outlook | T. Ciulli | 9 |
| **V1.0** | 03282022 | Final Revision | T. Ciulli, G. Silvestri | All |

## Partners

1 HELSINGIN YLIOPISTO (UH)

2 TYOTERVEYSLAITOS (FIOH)

3 INNOVATION SPRINT (INNO)

4 UNIVERSITA DEGLI STUDI DI TRENTO (UNITN)

5 UNIVERSITA DEGLI STUDI DI PADOVA (UNIPD)

6 IDEGO SRL (IDEGO)

7 BNP SRL (BNP)

8 AALTO KORKEAKOULUSAATIO SR (AALTO)

9 ETSIMO HEALTHCARE OY (ETSH)

10 ELECTROLUX ITALIA SPA (ELUX)

## Author(s)

- Simone Barbato (IDEGO)
- Tommaso Ciulli (IDEGO)
- Giorgia Silvestri (IDEGO)
- Lorenzo Di Natale (IDEGO)
- Carmelo Ferrante (UNITN)
- Giuseppe Riccardi (UNITN)

# List of Abbreviations

ACT - Acceptance and Commitment Therapy

ACT_TOT - Total Activities

A-DASS - Anxiety

ANX-HIGH - Anxiety high

ANX-LOW - Anxiety low

AU-BPNSFS - Autonomy

BPNSFS-24 - Basic Psychological Need Satisfaction and Frustration Scale

CA - Conversational Agent

CBT - Cognitive Behavioural Therapy

COM-BPNSFS - Competence

DASS-21 - Depression Anxiety Stress Scales short version

D-DASS - Depression

GSE - General Self Efficacy scale

G0 - Group zero

G1 - Group one

INT_TOT - Total Interaction

PHA - Personal Healthcare Agent

RE-BPNSFS - Relatedness

S-DASS - Stress

SDT - Self Determination Theory

SMT - Stress Management Therapy

USE_TOT - Total Usage

WHO - World Health Organisation Wellbeing

# List of Figures

# List of Tables

# Table of Contents

# 1 Executive Summary

In this document we report on the development and the evaluation of a mental health APP deployed in a real use case in Italy.

Some of the main objectives of task 6.4 discussed in this document are:

- 2 digit improvements on measures: self-confidence, motivation, compliance in behavioural change self-efficacy and wellbeing measures.
- Discuss and address ethical and deontological issues regarding the use of an APP.

In the document we report the organisational and operational steps taken to implement the APP, the psychological model, the co-participatory design with professionals, the experimental protocol, the data collected and the results. Moreover, we analyse the feedback collected from mental health professionals for the design and post-intervention as well as the feedback from end-users. Questionnaires were used to collect information on general psychological well-being and users' opinions. Overall, the APP evaluation and user experience were very good and useful and unexpected insights were collected.

In accordance with the above objectives, the final results showed a significant 2-digit improvement on the dimensions of general well-being, stress, anxiety, depression and sense of autonomy. For ethical issues, a study was carried out to build an intervention model to motivate users while maintaining their autonomy of choice. Furthermore, for deontological issues such as privacy, confidentiality and security, related to the use of APP, some colleagues from the deontological commissions of the regional psychologists' associations were involved and specific studies were conducted. Following this, a process of supervision and support by psychologists was foreseen.

# 2 Introduction

CO-ADAPT D6.4, describes what was obtained from evaluating a real-world application of what was developed in the CO-ADAPT project in a group of ageing working people. The objective of this task was to evaluate the impact of a methodology in which an APP, with a Conversational Agent inside, was used with a group of mental health professionals.

In this task, field experimentation of the CO-ADAPT mobile application was designed and implemented, which provides change program interventions through a conversational agent.

In Italy, the recruitment of at least 60 subjects was foreseen through a web-based platform for psychological counselling. At the end of the recruitment campaign, 95 people were recruited; of these, 76 were assigned to groups and only 56 went on to the next stages. The study focused on increasing subjects' awareness of change, motivation, self-confidence, and self-efficacy during counselling sessions. The study was conducted in Italian.

The experimental protocol was registered on ClinicalTrials.gov (ID: NCT05116553).

# 3 Psychological Model

In this task the purpose was to investigate the contribution of an APP within a real-world case, and more specifically, in the first demand analysis sessions that precede an eventual initiation of psychological treatment.

The focus was on reducing dropout rates and/or improving individuals' psychological well-being by working on motivation, self-confidence and self-efficacy.

In the field of psychology over time many theoretical models and also very different intervention approaches have emerged. So, in task 6.4 as a secondary objective we wanted to expand the possibility that this APP could be used by multiple professionals even with very different theoretical models.

One construct common to all approaches is motivation. Understood as a desire and drive to undertake a change and/or a psychological journey. It also seems to be associated with greater well-being and is central to behaviour change (Ryan, Lynch, Vansteenkiste & Deci, 2011).

Finally, the reasons why the people may need psychological counselling can be very different: to stop smoking, to support motivation in diet, exercise, for treatment of depression or anxiety, for mobbing at work, or to facilitate adaptation for a change in the workplace ect.

In any case, whatever the reason for the request for consultation or the psychological model used by the mental health professional, the first step is a psychological analysis that includes: motivation, goals, resources, any critical issues and psychological well-being assessment.

For all these reasons, the theoretical approach chosen for the treatment was that of Self Determination Theory (SDT), according to which wellness and motivation are based on the satisfaction of three fundamental psychological needs, such as autonomy, competence and relatedness. Autonomy concerns “a sense of initiative and ownership” with respect to the choices the individual makes; relatedness is about the experienced connection with others, and how they feel their relationships with them; competence is how the individual perceives the possibility of being able to succeed in their tasks and goals or decisions (Ryan & Deci, 2020). The non-achievement of these, especially if in a chronic way, represent a direct contributor to different kinds of psychopathology (Vansteenkiste & Ryan, 2013).

Relative to the methodology and APP tested in this real-world case, this included exercises from several approaches, such as Cognitive and Behavioural Theory (CBT, Beck, 1976) for the ABC technique (Ellis, 1962; Beck, 1976), time management exercises, and Acceptance and Commitment Therapy (ACT, Hayes, Strosahl & Wilson, 1999) for the defusion exercises.

The CBT model has its focus on dysfunctional thought and its aim is to change the mood by cognitive restructuring exercises to correct maladaptive beliefs. ABC technique is a tool that helps the individual to identify cognitive distortions through

the detailed description of an event, analysing and breaking down the behavioural and emotional dimension, with the feelings, intensity and bodily location of them (D5.3).

The ACT with the cognitive defusion leads to creation of a more psychological flexibility developing the ability to be in touch with the present moment, and proposes acceptance of that maladaptive thoughts. Defusion exercises help in this, modifying how an individual relates to negative thoughts.

# 4 Participatory design

The APP developed in Task 5.4 had been designed for data collection and to provide some guidance during an 8-session path to improve levels of anxiety, stress, and facilitate adaptation in the workplace.

Several CBT psychotherapists had been involved in the design of the APP, the control panel, and for the development of the materials (guides and exercises). Since task 6.4 the goal was to test the APP in a real case with numerous psychologists, not all psychotherapists or not all of the same theoretical model (for example psychodynamic or CBT), we conducted focus groups and interviews to collect feedback and opinions.

We divided this survey into three steps with the goal of analysing different parts:

A) Common goals of psychologists in initial interviews
B) Review APP and of the psychologists' Control Panel
C) Improved user experience

A) First of all, we have verified that independently of the theoretical model of intervention, all psychologists conduct a very similar analysis during the initial sessions: reasons for requesting consultation, goals, resources (social, cognitive, etc.), blocks to change, level of motivation to change behaviours, expectations of treatment success, and analysis of overall psychological well-being.

B) The second step was to review with the psychologists the APP and control panel designed for Task 5.4.

Most psychologists agreed that the elements present (the Conversational Agent questions, the exercises, the guides, etc.) were suitable to be used in their work without having to make significant corrections or changes. A few, however, suggested that they could add additional exercises useful for psychological analysis. After developing additional guides and exercises, feedback was collected and all were in agreement in the use of these new elements.

C) The third step, to improve the user experience according to the general objectives of task 6.4, was to redesign the APP with the intent to create a graphical timeline to follow the path of two weeks with steps activated by the psychologist (see paragraph 8.3).

A range of feedback was collected from psychologists and some users not psychologists at multiple stages to improve the user experience and make requested changes.

In almost all phases, focus groups were conducted online with at least 4-5 subjects. In some cases, single interviews were conducted.

# 5 Experimental Protocol

In this round of research the aim was to understand the influence of APP use during the initial phase of a psychological treatment; for this reason, only two groups were designed, defined as follows (Figure 1): the control group Group 0 (G0), in which were provided the two interviews and the use of the APP in a reduced form, only for filling out the psychological tests (pre and post), writing down why they had taken part in the project and making the appointment for the first meeting; the experimental group, Group 1 (G1), in which were provided the two interviews and the use of the complete app that included, in addition to the psychological tests, the motivation and the appointment, also the exercises, the guides and conversation with CA.

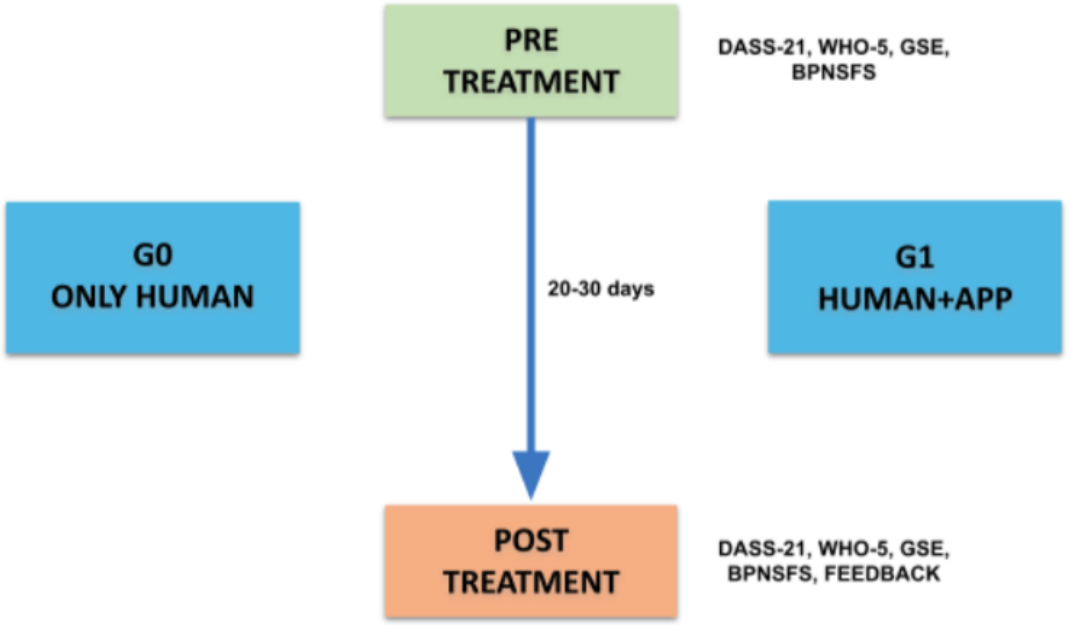


*Figure 1 - Experimental design and Test.*

Both groups were administered the same psychological tests: the first administration was scheduled at the beginning of the treatment, before the random division into groups, the second at the end of the second interview.

## 5.1 Subject samples and recruitment

A lead generation campaign began the recruitment process on October 15th; during this time contacts of interested people were collected and then contacted again once enrollment began on November 1st.

Due to the continued state of emergency for the COVID-19 pandemic, recruitment strategies have been adopted through the use of social media platform Facebook.

We chose Facebook because - while teens use more Instagram and Tik Tok, the over 55s represent the second most important audience for Facebook (see for example 19 Facebook Demographics to Inform Your Strategy in 2022 and Is Facebook for old people? Over-55s flock in as the young leave). In this way we could collect a sufficient number of people for creating four experimental groups, one of which was a control group.

The social strategy for recruitment was composed of two phases:

A) A first lead generation strategy aimed to attract people who were in our target audience, potentially interested or curious about the project. People were redirected from Facebook to a page on the co-adapt.it website where they could find more information about the project and leave their contact details to receive updates.

B) A second phase of the strategy, of retargeting, to engage people previously interested in the Lead phase.

Posts used and further details about the campaign are shown in table as follows (Table 1).

*Table 1 - Statistics from the social media campaigns*

| | N. | Period | General Goal | Target Insertion | Impression | Click | CTR |
|---|---|---|---|---|---|---|---|
|  | #1 | 15th Oct - 10th Nov | Lead generation & retargeting | Facebook | 61512 | 1184 | 1,92% |
|  | #2 | 15th Oct - 10th Nov | Lead generation & retargeting | Facebook | 190357 | 2212 | 1,16% |

From the Facebook social campaign 95 people were recruited; of these, 76 were assigned to the groups and only 56 went through the treatment, thus being considered for the final analyses.

The 56 users were equally distributed into the two groups, average age of 56,61 years (Mdn= 59), 30,4% male and the remaining 69,6% female.

Further information about the characteristics of the sample are shown in Table 2 below.

*Table 2 - Sample characteristics (N=56).*

| Characteristics | Values |
|---|---|
| **Age (years), mean (SD)** | 56.61 (13.14) |
| Group 0 | 57.04 (12.57) |
| Group 1 | 56.18 (13.91) |
| **Gender, n (%)** | |
| Male | 17 (30.4) |
| G0 | 8 (28.6) |
| G1 | 9 (32.1) |
| Female | 39 (69.6) |
| G0 | 20 (71.4) |
| G1 | 19 (67.9) |
| **Groups, n(%)** | |
| Group 0 | 28 (50) |
| Group 1 | 28 (50) |
| **Formal Education, n (%)** | |
| Secondary school | 2 (3.6) |
| High school | 20 (35.7) |
| Degree | 25 (44.6) |
| Master's degree or PhD[a] | 9 (16.1) |

a PhD: Doctor of Philosophy

### 5.1.1 Inclusion criteria

Included were those subjects (male and female) with good cognitive skills to understand project goals, duration and any difficulties.

### 5.1.2 Exclusion criteria

Exclusion criteria were the presence of severe depression, underlying psychiatric conditions, current use narcotics or other substances and neuropsychological mild impairments.

### 5.1.3 Pre and post-testing

Psychological tests used are described as follow:

1) *Well-being* (WHO-5)

To evaluate general well being the The World Health Organisation - Five Well-Being Index (WHO-5) was utilised (WHO, 1998). It's composed of 5 items on a 6-point Likert scale (from 0=never to 5=always) about how the individual felt in the past 2 weeks.

The results given by the test is a percentage from 0 to 100; higher scores indicate better well-being.

2) *Depression, Anxiety, Stress (DASS-21)*

With the *Depression Anxiety Stress Scales Short Version* (DASS-21; Henry & Crawford, 2005) were assessed the dimension of anxiety, depression and stress; the questionnaire has been used in its short form validated in Italian (Bottesi, Ghisi, Altoè, Conforti, Melli & Sica, 2015).

The questionnaire is composed of 21 items on a 5-point Likert scale (from 1=not at all to 5=Very much) divided by three factors, as mentioned before.

Each of the three constructs presents different cut-offs, classifying the individual according to 5 ranges of severity, from "normal" to "extremely severe".

In all of these subscales, higher scores indicate a higher level of discomfort.

3) *Self-Efficacy (GSE-10)*

To assess the levels of self-efficacy the General Self-Efficacy Scale was used (Schwarzer & Jerusalem, 1995) in its form validated in Italian (Sibilia, Schwarzer & Jerusalem, 1995).

This scale is composed of 10 items on a 4-point Likert scale (from 1=Not at all true to 4=Exactly true).

Higher scores in this scale indicate a higher level of self-efficacy.

4) *Motivation (BPNSFS-24)*

To assess the satisfaction of the three fundamental psychological needs, a factor that determines the well-being and motivation of the individual, according to SDT, the The Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS, Costa, Ingoglia, Inguglia, Liga, Lo Coco & Larcan, 2017) was adopted.

The test is composed of 24 items, 8 for each three (Autonomy, Competence and Relatedness) of psychological need; of these 8 items, 4 are about the satisfaction of that need, and the remaining 4 are for the dissatisfaction of the same.

The *autonomy* subscale aims to assess how much the individual feels "psychological free" in making them decisions, with the subscale of *competence* is meant to measure the feel of effectiveness and mastery the person experiences, and with the subscale of *relatedness* the aim is to assess their emotional sphere, how connected they feel to the people who are important to them.

There are many different scoring strategies for this test; in fact, each subscales of the specific need is also divided into satisfaction or dissatisfaction. No cut-offs are available; Higher scores are better in terms of motivation.

5) *Awareness*

To assess the subject awareness had been selected and translated 3 items from the Situational Self-Awareness scale (Govern & Marsch, 2001), to investigate the degree of awareness of the individual with respect to their own feelings, deep thoughts and what is happening around them.

The original scale is composed of 9 items divided per three main factors: public and private self-awareness and awareness of immediate surroundings; the items are phrased as declarative sentences on a 7-point Likert scale, from 1 (strongly disagree) to 7 (strongly agree).

The three items considered in the study are related to private and surrounding awareness. Higher scores indicate a greater awareness.

# 6 Methodology of intervention

At the end of the recruitment phase, all users were allowed to access the APP to complete the psychological tests (pre-test); then, they were asked to write the reason why they decided to participate in this project and to provide, again through the app, the availability to schedule the first interview with the psychologist.

Participants were divided randomly into the two groups, then proceeded with two different ways of treatment:

A) For what concerned the experimental group (G1), the path consisted of a first week of autonomous use of the APP.

After the first interview a second week of APP use, then finally the second interview and again tests at the end of the path, for a length of about two weeks. The development of the path is illustrated in Figure 2.


*Figure 2 - G1 path.*

B) The control group (G0) conducted the two interviews one week apart, then the initial tests were repeated as a post-test.

Relative to the experimental group (G1), exercises available on week one included relaxation exercises (7 videos) and therapeutic writing exercises (such as "letter to me," "as if," and “blessings"); on week 2, then after the interview, defusion exercises, mindfulness videos (3 videos), and two time management exercises, such as time boxes and pleasurable activities, were added.

## 6.1 The role of the APP and PHA

One of the objectives was to test an integrated methodology to the use of an APP with the aim of stimulating greater awareness, motivation and sense of self-efficacy to produce behaviour change and decrease the risk of dropout of counselling sessions.

To achieve these goals, experimental group 1 had the option of using an APP with more features than group 0 as described previously.

With these features, people were more engaged in using the APP, receiving information, gathering their thoughts, and exploring their resources and/or critical issues even outside of just the two sessions.

The hypothesis is that with the APP, people can achieve greater and faster emotional and cognitive insight, especially due to the Conversational Agent that aims to guide people in gathering events, thoughts, and emotions related to certain situations outside of the sessions with the psychologist (D5.3).

Finally, the information collected and the exercises completed could be viewed by the psychologist before the session. This helped psychologists save a lot of time in conducting their sessions and identifying with the person the best solution.

# 7 Experimental results

Three subjects were excluded from the analyses, due to their discontinuous course and the distance in time over which the tests were completed, then 28 subjects per group had been considered.

Initial analyses were conducted by performing parametric T-test to assess whether there were significant differences both between groups at T1 and T2, and within groups between times (Table 3).

Differences between groups, assessed by T-test for independent samples, had not been detected for any of the dimensions considered, neither in T1 nor in T2.

The only significant difference detected in G0 is the one related to the measures of WHO (t=-3.68 , *p*=.001) indicating an improvement of 20 percentage points, even better than that experienced by Group 1 (G0=20.6%, G1=17.4%). Improvement for WHO levels were significant for G1 (t=-3.08, *p*=.005) as well.

No other significant differences were found within Group 0.

For Group 1, many significant difference were detected:

Stress levels, for which lower values mean less discomfort, decreased by an average of about 26%, from an average T1=14,43 to T2= 10.57 (t=2.75, *p*=.011); in the same way, levels of anxiety shown a difference of 38,95%, indicating significant improvement (t=2.15, *p*=.041), as well as levels of depression, which fell by 43% (t=3.95, *p*=.001).

Significant differences were also detected for G1 in levels of Autonomy, measured by BPNSFS, with an increase of 10% (t=3.05, *p*=.005).

*Table 3 - T-test for paired samples, differences within groups between T1 and T2.*

| | | T1 Mean (DS) | T2 Mean (DS) | +% | t | *p* | D |
|---|---|---|---|---|---|---|---|
| **WHO** | G0* | 47,14 (18.71) | 56,86 (15.94) | +20,6%* | -3.68 | .001 | -0.70 |
| | G1* | 49,29 (21) | 57,86 (18.59) | +17,4%* | -3.08 | .005 | -0.58 |
| **S-dass** | G0 | 15,21 (8.42) | 13,21 (10.06) | -13,15% | 1.72 | .097 | 0.33 |
| | G1* | 14,43 (10.25) | 10,57 (7.48) | -26,75%* | 2.75 | .011 | 0.52 |
| **A-dass** | G0 | 4,79 (6.36) | 4.50 (6.36) | -6% | 0.30 | .770 | 0.06 |
| | G1* | 4,57 (5.60) | 2,79 (3.70) | -38,95%* | 2.15 | .041 | 0.41 |
| **D-dass** | G0 | 10,43 (9.31) | 8.00 (8.91) | -23,3% | 1.95 | .061 | 0.37 |
| | G1* | 13,29 (10.73) | 7,57 (7.08) | -43%* | 3.95 | .001 | 0.75 |
| **GSE** | G0 | 27,71 (5.73) | 28,61 (4.78) | +3,25% | -1.30 | .206 | -0.25 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Au-bpnsfs** | G1 | 28,07 (5.76) | 28,96 (4.98) | +3,17% | -1.22 | .232 | -0.23 |
| | G0 | 25,5 (5.57) | 26,89 (4.73) | +5,45% | -1.64 | .112 | -0.31 |
| | G1* | 24,64 (8.35) | 27,11 (6.82) | +10%* | -3.05 | .005 | -0.58 |
| **Re-bpnsfs** | G0 | 32,86 (5.16) | 32,96 (5.36) | +0,3% | -0.11 | .914 | -0.02 |
| | G1 | 32,96 (5.69) | 33,54 (5.70) | +1,76% | -0.61 | .545 | -0.12 |
| **Com-bpnsfs** | G0 | 29,32 (7.26) | 29,57 (6.14) | +0,85% | 0.29 | .771 | -0.06 |
| | G1 | 29,46 (7.41) | 30,14 (5.94) | +2,31% | -0.78 | .441 | -0.15 |
| **Awareness** | G0 | 17,54 (2.77) | 17,21 (2.47) | -1,88% | 0.70 | .488 | 0.13 |
| | G1 | 16,75 (2.99) | 17,32 (2.78) | +3,40% | -1.21 | .237 | -0.23 |

Further analysis was conducted to explore whether there was any type of interaction with the initial levels of distress experienced by individuals. To do this, the stress, anxiety, and depression scores, which originally presented 5 levels of severity, were reorganised in the following way: each of the variables considered was re-coded so that results belonging to the "Normal" range (the original lowest=1) had as their value 1, while all other ranges (from Mild to severe) had as their value 2, thus representing respectively the subgroup with low initial levels of anxiety and the subgroup with high initial levels of anxiety.

This division is best explained in Table 4.

Repeated measures ANOVA was performed, considering both group and the re-coded variables (anxiety, stress and depression) as factors between subjects, Time (T1 and T2) as within subjects factor, taking one at time each of the measures considered in the study as a dependent variable.

No interactions were found considering initial levels of stress or depression.

Some significant interactions were found considering levels of anxiety on depression, stress and competence.

More specifically:

Considering Initial levels of anxiety on the dependent variable of depression, which results are shown in table 4, regarding differences between times, results show significance for both G0 ($p$=.028, $\eta 2p$= 0.09) and G1 ($p$<.001, $\eta 2p$=0.30); more specifically, significance was maintained in G1 between T1 and T2 both for subgroups with low ($p$=.007, $\eta 2p$=0.13) and high anxiety ($p$<.001, $\eta 2p$=0.22), with a greater improvement for subgroups with high anxiety; differences are not maintained in G0 considering the subgroups of anxiety. This may indicate that the treatment has a greater impact on those subjects with higher anxiety levels.

Considering differences between groups, significance was revealed at T2 for subgroups with high anxiety (F=5.09, $p$=.028, $\eta 2p$=0.09).

*Table 4 - Differences within groups considering the initial levels of anxiety on dependent variable of depression.*

| | G0 | | | | | | G1 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **η2p** | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **η2p** |
| **D_dass** | 14.64 (2.10) | 11.03 (1.74) | 5.10 | (1,52) | .028 | 0.09 | 14.48 (1.99) | 7.33 (1.65) | 22.28 | (1,52) | <.001 | 0.30 |
| **Anx_ low** | 7.27 (1.94) | 5.73 (1.61) | 1.09 | (1,52) | .301 | 0.02 | 12.1 (1.99) | 7.81 (1.65) | 8.02 | (1,52) | .007 | 0.13 |
| **Anx_ high** | 22.00 (3.71) | 16.33 (3.08) | 4.01 | (1,52) | .051 | 0.07 | 16. 86 (3.44) | 6.86 (2.85) | 14.55 | (1,52) | <.001 | 0.22 |

Considering Initial levels of anxiety on the dependent variable of stress, which results are shown in Table 5, regarding differences between times, results showed significance for G1 ($p$=.002, $\eta^2_p$= 0.18) and more specifically for subgroups with high anxiety level ($p$=.008, $\eta^2_p$= 0.13). No significant differences between times were found for G0.

Considering differences between groups, significance was revealed for the subgroup with high anxiety at T2 (F=6.42, $p$=.014, $\eta^2_p$=0.11).

*Table 5 - Differences within groups considering the initial levels of anxiety on dependent variable of stress.*

| | G0 | | | | | | G1 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **η2p** | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **η2p** |
| **S_dass** | 18.77 (1.98) | 16.77 (1.89) | 1.63 | (1,52) | .208 | 0.03 | 15.81 (1.88) | 10.85 (1.79) | 11.11 | (1,52) | .002 | 0.18 |
| **Anx_ low** | 12.55 1.84) | 10.55 (1.75) | 1.90 | (1,52) | .174 | 0.04 | 13.05 (1.88) | 10.29 (1.79) | 3.46 | (1,52) | .069 | 0.06 |
| **Anx_ high** | 25.00 (3.52) | 23.00 (3.35) | 0.52 | (1,52) | .475 | 0.01 | 18.57 (3.26) | 11.43 (3.10) | 7.70 | (1,52) | .008 | 0.13 |

Considering Initial levels of anxiety on the dependent variable of competence, which results are shown in Table 6, regarding differences between times, results showed that levels of competence were significantly different for subgroups with high levels

of anxiety, more specifically in G1 ($p$=.025, $\eta^2_p$= 0.09). No significant differences between times were found for G0.

The effect on competence is higher in subgroups with higher levels of anxiety, especially for what concerns treatment of G1 (levels in G1 increase from T1=25.14 to T2=29); higher scores indicate a higher sense of competence.

Considering differences between groups, significance was revealed for the subgroup with high levels of anxiety at T2 (F=5.10, $p$=.028, $\eta^2_p$=0.09).

*Table 6 - Differences within groups considering the initial levels of anxiety on dependent variable of competence.*

| | G0 | | | | | | G1 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **$\eta^2_p$** | **T1 Mean (SE)** | **T2 Mean (SE)** | **F** | **df** | ***p*** | **$\eta^2_p$** |
| **Comp_ bpnsfs** | 26.17 (1.48) | 26.88 (1.25) | 0.48 | (1,52) | .490 | 0.01 | 28.02 (1.41) | 29.76 (1.19) | 3.27 | (1,52) | .076 | 0.06 |
| **Anx_ low** | 31.68. (1.37) | 31.59 (1.16) | 0.01 | (1,52) | .923 | 0.00 | 30.91 (1.41) | 30.52 (1.18) | 0.16 | (1,52) | .694 | 0.00 |
| **Anx_ high** | 20.67 (2.63) | 22.17 (2.22) | 0.70 | (1,52) | .408 | 0.01 | 25.14 (2.44) | 29.00 (2.01) | 5.37 | (1,52) | .025 | 0.09 |

# 8 Data Collection

In addition to the psychological data collected by the administration of psychological tests, further data were gathered by the administration of satisfaction questionnaires, both for psychotherapists and users, and the analysis of users' interactions with the APP.

## 8.1 Evaluation of Therapists feedback

A satisfaction questionnaire was designed to assess the evaluation of therapists about their experience.

The aim was to understand, on one hand, the specific experience of the professional with the use of the dashboard, the relation with the user and how much they felt comfortable with the proposed model of intervention; on the other hand, to understand if there were any differences dictated by having in charge a user with app or a user without app.

The questionnaire was built on the basis of the one used for the previous Task 5.4, then readapted.

At the beginning, there was a summary description about the goal of the task (6.4) just completed, to remind of the organisation of the groups, the usefulness of the APP specific for the two groups and the dashboard.

Some of the professionals only had users with APP or without APP, instead others had both of them. Then, the possible choices in the questionnaire at the beginning were: user with APP, user without APP, both.

At the end, a common section to investigate overall their perceptions on the proposed methodology, the use of tools such as the dashboard, to follow the users in the sessions, and the app.

The following tables describe the results found, divided by feedback related to the overall treatment (Table 7), feedback related to the use of the dashboard (Table 8) and feedback related to the relationship with the user (Table 9). All answers were provided on a scale of agreement from 1 to 5, where 1= not at all agree, 5= absolutely agree.

All these tables are divided in relation to which kind of users the professionals had, i.e. whether the app , no-app or both (as previously specified).

"*Only*" refers to those professionals who had only users with apps or only users without apps; with "*Both*" refers to those professionals who had both app and non-app users.

More specifically, results in Table 7 show the rates of satisfaction experienced by the professionals, how much they find the treatment useful to better understand users' reason, and if they would like to implement this service in their own practice.

*Table 7 - Satisfaction professional Treatment*

| | **Treatment** | | | |
|---|---|---|---|---|
| | **Only** | | **Both** | |
| | **G0** | **G1** | **G0** | **G1** |
| **Satisfaction** | 3 | 3,6 | 4,1 | 3,9 |
| **Utility** | 3,4 | 3,2 | 4,2 | 4,1 |
| **Implementation** | 3,2 | 3,8 | 4,2 | 4,1 |

In Table 8 are shown results related to the use of the dashboard by professionals; this Table is divided into two parts: the upper (a) shows results to the general item, valid both for users with the app and without it, about how easy the professional found it to work with the dashboard.

The lower part (b) of the table is about those features only available to those who had users with the APP, investigating how useful they found the chance to view written ABCs and exercises done by the users, both for professionals only with a user with app and professionals who had both kinds of users.

*Table 8 - Satisfaction professional Dashboard.*

| | **Dashboard** | | | |
|---|---|---|---|---|
| **a)** | **Only** | | **Both** | |
| | **G0** | **G1** | **G0** | **G1** |
| **Easy** | 3,6 | 3,6 | 4,2 | 4,2 |

| **b)** | **G1** | **G1** |
|---|---|---|
| **Utility - ABCs** | 3,1 | 4,2 |
| **Utility - Exercises** | 3 | 4,3 |

Section (b) shows that professionals who had both kinds of users (No APP and with APP) found the chance to see what users wrote in their ABCs and the exercises more useful than therapists who only had the user with the APP.

Table 9 shows results related to the relationship developed with the users; this table is divided into two parts as well.

The upper part (a) shows results related to how professionals felt the alliance and the relationship with the users, how much they felt the users were engaged in the treatment and the privacy they felt for their users.

In the lower part (b) are shown results related to the contribution of the APP in increasing: the understanding the professional was able to develop toward the user, the alliance developed between the two parts, and the engagement in the treatment shown by the user.

*Table 9 - Satisfaction with the relationship developed by professionals with users.*

| | Users' Relation | | | |
|---|---|---|---|---|
| **a)** | **Only** | | **Both** | |
| | **G0** | **G1** | **G0** | **G1** |
| **Alliance** | 3 | 3,7 | 4 | 4,1 |
| **Relation** | 3 | 3,8 | 4,4 | 4,3 |
| **Engagement** | 2,6 | 2,9 | 3,9 | 3,9 |
| **Privacy** | 4 | 4,8 | 4,8 | 4,9 |

| **b)** | **G1** | **G1** |
|---|---|---|
| **Better understanding** | 2,8 | 4,1 |
| **Better Alliance** | 2,1 | 3,4 |
| **Better Engagement** | 2,7 | 3,6 |

Section (b) shows that those professionals who had both kinds of users (No APP and with APP) had a perception of the APP's impact on the relationship with the users bigger than of those professionals who only had users with the APP.

From these results we can deduce that APP combined with CA does not negatively influence the relationship or alliance between the user and the professional.

## 8.2 Evaluation of Users feedback

At the end of treatment, after completing the psychological tests, a satisfaction questionnaire was proposed to all users to investigate their satisfaction about the overall treatment, the usefulness they have found and the ease of using the application.

Two different types of questionnaires were proposed, depending on the group they belonged to.

The Table 10 shows the results of those aspects common to both groups, such as satisfaction with the proposed treatment, satisfaction with respect to the feedback returned at the end of each compilation (pre and post), whether they would recommend a treatment like this to their loved ones, and how much they felt their privacy was respected; all answers were provided on a scale of agreement from 1 to 5, where 1= not at all agree, 5= absolutely agree.

*Table 10 - Common questions about the treatment.*

| | **Satisfaction** | |
|---|---|---|
| | **G0** | **G1** |
| **Treatment** | 4.38 | 4.34 |
| **Feedback** | 4.15 | 4.10 |
| **Recommend** | 4.42 | 4.31 |
| **Privacy** | 4.88 | 4.76 |

Table 11 shows the results with respect to the use of the app.

Since G0 also had partial use of the APP, the top part of the table shows the results with respect to the use of the APP in its shared features, such as testing and the possibility of scheduling the appointment with the psychologist.

The bottom part, on the other hand, shows the results with respect to G1's use of the APP in its full version. More specifically, in the table are shown results with respect to the usefulness of the proposed content, such as guides, exercises and videos, and the ease of use of the same; all answers were provided on a scale of agreement from 1 to 5, where 1= not at all agree, 5= absolutely agree.

*Table 11 - APP usage, common and specific questions.*

| **APP Usage (both G0 and G1)** | | |
|---|---|---|
| | **G0** | **G1** |
| **Ease** | 4.35 | 4.66 |
| **APP Usage (specific G1)** | | |
| **Utility** | 4.14 | |
| **Ease** | 4.28 | |

In Table 12 are shown the results related to the experience with the Conversation Agent.

More specifically, if they found it easy to talk to the CA and if it was helpful; with "effort" it is meant if the requests were excessive, with "recommended" if they would recommend it to their loved ones and finally with "personal usage" it was investigated if, in case the APP was available on the stores (iOs, Android) they would download it for personal use; all answers were provided on a scale of agreement from 1 to 5, where 1= not at all agree, 5= absolutely agree.

*Table 12 - CA usage.*

| | **CA usage (specific G1)** |
|---|---|
| **Ease** | 4.03 |
| **Utility** | 3.31 |
| **Effort** | 2.31 |
| **Recommend** | 3.72 |
| **Personal Usage** | 3.45 |

In the last table (Table 13) are shown results of the amount of people who decided to continue the treatment with the professional at the end of 2 sessions included in the project.

*Table 13 - People who continued after the end of 2-session treatment.*

| | **Total n** | **Continued n(%)** |
|---|---|---|
| **G1** | 28 | 6 (21.43) |
| **G0** | 28 | 5 (17.86) |

## 8.3 APP and PHA data collection and User Engagement

To perform the trial, a modified version of the APP realised in WP4 and WP5 have been created. This modified version maintained the original functionalities described in D5.4 adding some more exercises and features customised for this experimentation.

Under these new modifications a specific therapy could be assigned to each user. The therapy is constituted of many steps to be taken in order to achieve the prescribed

objective. These steps include questionnaires, exercises, audio and video contents, feedbacks and interviews with psychologists.

For this reason the APP has been created to start with a dashboard showing the assigned path with the performed and pending steps.

Every step in the path is automatically unlocked when some condition is reached: the previous step has been completed or a defined number of days has passed or the interview with the psychologist has occurred.


*Figure 3 - W6.4 Users APP Dashboard example.*

During the path, the activities carried out by each user within the application were recorded, such as opening pages, guides, exercises done and permanence on them.

These data were then analysed and organised in order to create indices of use, as described below.

Initially, a mapping of the activities carried out and the participant's interactions with the App was produced. Specifically:

***Exercises***

Under this tag, all the exercises completed belonging to the different categories were considered; each type of exercise had different parameters so that it could be defined as done.

- *Therapeutic writing exercises*: these kinds of exercises provide written output, consistent with the demands of the exercise itself. So, the recorded written response was sufficient to determine the exercise was completed.

- *Mindfulness and relaxation*: these kinds of exercises were represented by videos with different length. So, the parameter for defining that the exercise had been performed was dwell time, at least 3 minutes.

- *Defusion exercises*: in this category there were short exercises in which the user was invited to reflect and repeat in mind in different ways a thought that is unpleasant for him, all without having to write anything down. So, the parameter for defining that the exercise had been performed was dwell time, at least 40 seconds.

***Openings***

Under this tag, all the opened exercises, but which do not meet the hypothesised criteria (e.g. shorter than identified dwell times) to be considered as done.

***ABCs***

The ABCs exercises were the ones proposed by the conversational agent and consisted of a series of questions that accompanied the user in describing a given event.

So, under this, all the ABCs written by the subject were considered.

***Guides***

Under this tag, all openings of each guide were considered, even if opened more than once in time.

***Reason***

One of the first steps in the app usage was to write down the reason why the subject decided to take part in the project.

This was then repeated at the end of the first interview, so we have

- R#1: for the written reason before the 1st interview.

- R#2: for the written reason after the 1st interview.

After the data collection, three general use indices were created.

*Int_Tot: Total Interaction*

This index identifies in numerical form the amount of general interactions.

It is an index that corresponds to the sum of the number of exercises done + number of openings + number of guides + number of ABCs + number of reasons written.

Ind_tot = openings + exercises + guides + ABCs + Reasons

*Use_tot: Total Usage*

This index identifies in numerical form the amount of use of the App, which excludes simple interactions (i.e. Openings).

It is an index that corresponds to the sum of the number of exercises done + number guides + number of ABCs + number of reasons written.

Use_tot: exercises + guides + ABCs + Reasons

*Act_tot: Total activities*

This index identifies in numerical form the amount of exercises in a strict sense done. In Total Activities have been reported those interactions that with more certainty we can say have been completed (as they present a better traceability of use).

It is an index that corresponds to the sum of the number of exercises done + number of ABCs + number of reasons written.

Act_tot: Exercises + ABCs + Reasons

All these indexes (both partial and total) were then divided by the two weeks, taking into account personal specifics with respect to the date of the first session, that determined the transition from the first to the second week of use.

Important to remember that in Week 2 more exercises were added; in the first week, in fact, only 10 exercises were available, while in the second week 11 exercises were added, bringing the total number of exercises to 21.

For this reason, with regard to the exercises of the second week, a further division was made to analyse how many old and how many new exercises had been performed.

Statistical analyses were conducted to understand if there were any differences between the usage in Week 1 over Week 2.

Table 14 shows the mean of the different kinds of interaction comparing Week 1 and Week 2.

For all measures significant increases were found except for Guides.

Measures for which significant differences are found are indicated with an asterisk (*).

*Table 14 - Mean of interactions with the APP*

| | Week #1 | Week #2 |
|---|---|---|
| **Exercises*** | 0.69 | 3.93 |
| **Openings*** | 1.25 | 7.67 |
| **Guides** | 3.46 | 4.29 |
| **ABCs*** | 0.17 | 1.38 |
| **Int_tot*** | 5.82 | 20 |
| **Uso_tot*** | 4.45 | 12.13 |
| **Att_tot*** | 1.82 | 6.77 |

In table 15 are shown two different kinds of information related to the exercises done in both Week 1 and Week 2.

In the upper part of the table are shown the specific means for the exercises of the week, differentiating between the exercises those that were already available in the first week (old#2) and those that were new (new#2).

In the lower part of the table, on the other hand, the percentages of exercises, divided by Week 1, New of Week 2 and Old of Week 2; each percentage is relative to the totality of exercises carried out in the weeks of treatment.

*Table 15 - Means and percentages of exercises done divided per week of treatment.*

| | **Exercises per week** | |
|---|---|---|
| | **n** | **Mean** |
| **Week#1** | 14 | 0.69 |
| **Week#2** | 110 | 3.93 |
| **Old #2** | 29 | 1.04 |
| **New #2** | 81 | 2.89 |
| | **All Exercises** | |
| | n | % |
| **Old #1** | 14 | 11.29 |
| **Old #2** | 29 | 23.39 |
| **New #2** | 81 | 65.32 |
| **Total** | 124 | 100 |

Statistical analyses conducted on exercises completed in Week 2 show a significant difference between the use of old and new kinds of exercises (Old= 1.04; New= 2.89 *p*=.003).

# 9 Ethical issues

The ethical issues addressed by the research team concerned two main problems:

- The use of an ethical psychological model of behaviour change.
- The use of an APP for a clinical health service (deontology, professional secrecy, etc.).

For the first point, since one of the objectives was to decrease the risk of dropout that can occur in requests for psychological help, we identified strategies that would motivate users to remain in the pathway.

Obviously, it was necessary and fundamental to understand the issue of autonomy of user choice and also to analyse the ethical issues in guiding another person's choices.

The key point, therefore, is autonomy understood as "*the capacity to think, decide, and act on the basis of such thought and decision freely*" (Gillon, 1985).

A parallel analysis was carried out on the Deontological Code of Italian Psychologists which incorporates the obligation for professionals to protect the autonomy in the choices of their users and does not discriminate on the basis of religion, ethnicity, nationality, social background, socio-economic status, gender, sexual orientation, disability (Art. 3 and 4, https://www.psy.it/codice-deontologico-degli-psicologi-italiani).

Thus, the task was to provide the user with all the necessary information and also to strategically guide him/her in making positive choices for his/her psychological well-being without interfering with autonomy.

Many approaches to change psychology have been studied and the most ethical model chosen is the Self-Determination Theory model, which includes and explores the concept of autonomy in choice as a fundamental factor in preserving motivation for change and general well-being (Ryan, Lynch, Vansteenkiste & Deci, 2011).

For the second point, as already clarified and addressed in D 5.3, for issues concerning the development of APPs with regard to privacy, confidentiality and security, we explored and followed the recommendations provided by Bakker et al., (2016), and the ones by the Division of Digital Psychiatry of the American Psychiatric Association (https://www.digitalpsych.org/).

Furthermore, the indications of the National Council of Italian Psychologists concerning deontological issues for IT tools used in the field of mental health were considered (https://www.psy.it/wp-content/uploads/2015/04/Atti-Tipici_DEF_interno-LR-1.pdf).

Constant monitoring by other psychologists of both the activity of the psychologists involved in the interviews and the APP was planned, and contacts for possible emergency requests were distributed.

We can report that there were no emergency and/or dangerous situations for the users. Some problems have emerged with regard to some users who for organisational reasons and/or personal issues could not continue to the end.

# 10 Conclusion and future outlook

This report described what was developed and verified in task 6.4 and specifically the use of what was developed in task 5.3 and appropriately modified applied to a real case in Italy.

Much importance was given to the involvement of mental health professionals and possible users in almost all phases of design. To test the initial hypotheses of this task, an experimental RCT design was constructed with two groups.

The data collected verified at least in part that using an APP with a CA significantly increases people's sense of autonomy by 10% in just two weeks (a variable considered to be the basis of motivation and well-being). In addition, it is interesting how with such a short path, subjects report a significant decrease in measures of stress, anxiety, and depression. A particularly interesting finding was also collected, namely that for those people who describe themselves as anxious, the use of an APP with CA brings a significant improvement in lowering the levels of depression, stress and sense of autonomy compared to the group that did not have extended use of the APP.

This could indicate that some kinds of individuals (e.g. those with a medium/high levels of anxiety) might find the use of this type of tools more useful or effective than others.

Overall, users reported high levels of satisfaction and a sense of respect for their privacy, even though the information entered was monitored by professionals. It appears how users value the use of a digital tool (APP and CA) as a companion to a human professional.

On the other hand, professionals have reported that being able to access the user's information entered during everyday life is very important to improve the psychological assessment, decrease the time of the sessions or make it more efficient and provide more appropriate guidance to users.

Finally, despite the fact that few users actually continued the sessions, it appears that many of the subjects in the experimental group (APP and CA), reached high levels of well-being and clarity in their motivations such that they did not feel the need to continue but to eventually remain in contact with professionals.

These results encourage us to reevaluate the initial hypotheses and consider these tools as important means of screening and prevention.

# 11 References

Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health. 2016 Mar 1;3(1):e7. doi: 10.2196/mental.4984. PMID: 26932350; PMCID: PMC4795320.

Beck AT (1976). Cognitive Therapy and the Emotional Disorders. International Universities Press.

Bottesi, G., Ghisi, M., Altoè, G., Conforti, E., Melli, G., & Sica, C. (2015). The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples. Comprehensive psychiatry, 60, 170-181.

Costa, S., Ingoglia, S., Inguglia, C., Liga, F., Lo Coco, A., & Larcan, R. (2017): Psychometric evaluation of the Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS) in Italy. Measurement and Evaluation in Counseling and Development, 51, 193-206.

Ellis A. (1962). Reason and emotion in psychotherapy. New York: Lyle Stuart.

Gillon, R. (1985). Autonomy and the principle of respect for autonomy. British Medical Journal, 290, 1806-1808.

Govern, J. M., & Marsch, L. A. (2001). Development and validation of the situational self-awareness scale. Consciousness and cognition, 10(3), 366-378.

Hayes SC, Strosahl KD, & Wilson KG (1999). Acceptance and Commitment Therapy. An experiential approach to behaviour change. New York: Guilford.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. British Journal of Clinical Psychology, 44(2), 227-239.

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary educational psychology, 61, 101860.

Ryan, R. M., Lynch, M. F., Vansteenkiste, M., & Deci, E. L. (2011). Motivation and Autonomy in Counseling, Psychotherapy, and Behavior Change: A Look at Theory and Practice 1ψ7. The Counseling Psychologist, 39(2), 193–260. https://doi.org/10.1177/0011000009359313.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35- 37). Windsor, England: NFER-NELSON.

Sibilia, L., Schwarzer, R., & Jerusalem, M. (1995). Italian adaptation of the general self-efficacy scale. Resource document. Ralf Schwarzer website.

Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: basic psychological need satisfaction and need frustration as a unifying principle. Journal of psychotherapy integration, 23(3), 263.

WHO. (1998). Wellbeing Measures in Primary Health Care/The Depcare Project. WHO Regional Office for Europe: Copenhagen.