

Artificial Intelligence Chatbot for Anxiety and Depression in University Students: a Pilot Randomized Controlled Trial

Maria Carolina Klos, Milagros Escoredó, Angie Joerin, Viviana Noemí Lemos, Michiel Rauws, Eduardo L. Bunge

Submitted to: JMIR Formative Research
on: July 07, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	26
Figures	27
Figure 1.....	28



Artificial Intelligence Chatbot for Anxiety and Depression in University Students: a Pilot Randomized Controlled Trial

Maria Carolina Klos^{1*} PhD candidate, MA; Milagros Escoredo^{2*} MA; Angie Joerin^{2*} MA; Viviana Noemí Lemos^{3*} PhD; Michiel Rauws^{2*} MBA; Eduardo L. Bunge^{4*} PhD

¹Interdisciplinary Center of Mathematical and Experimental Psychology Research (CIIPME) Interdisciplinary Center for Research in Health and Behavioral Sciences (CIICSAC). Adventista del Plata University National Scientific and Technical Research Council (CONICET) Libertador San Martín, Entre Ríos AR

²X2AI inc. San Francisco US

³Interdisciplinary Center of Mathematical and Experimental Psychology Research (CIIPME) Interdisciplinary Center for Research in Health and Behavioral Sciences (CIICSAC). Adventista del Plata University National Scientific and Technical Research Council (CONICET) Libertador San Martín AR

⁴Palo Alto University Palo Alto US

* these authors contributed equally

Corresponding Author:

Maria Carolina Klos PhD candidate, MA

Interdisciplinary Center of Mathematical and Experimental Psychology Research (CIIPME)

Interdisciplinary Center for Research in Health and Behavioral Sciences (CIICSAC). Adventista del Plata University National Scientific and Technical Research Council (CONICET)

25 de Mayo 99

Libertador San Martín, Entre Ríos

AR

Abstract

Background: The use of artificial intelligence based chatbots as an instrument of psychological intervention is emerging, however no studies have been reported in Latin America.

Objective: This study aims to evaluate usage patterns and whether the use of a chatbot is effective for relieving depression and anxiety symptoms compared to a control group utilizing a psychoeducation book in Argentina.

Methods: This was a randomized controlled trial study utilizing the chatbot Tess throughout eight weeks. The initial sample consisted of 181 Argentinian college students ages 18 to 33, 87.2% female. Of those, 33 participants in the experimental condition and 30 in the control condition provided data on depressive symptoms at week eight, and 27 participants in the experimental condition and 23 in the control condition provided data on anxiety symptoms at week eight. Between and within group comparisons were analysed using Mann-Whitney U and Wilcoxon tests for depression symptoms, and Independent and Paired Samples t Tests to analyze anxiety symptoms.

Results: There was no significant intergroup differences between the experimental group and the control group for depression and anxiety symptoms from baseline to week eight ($P > .05$). However, there were significant intragroup differences, where the experimental group showed a significant decrease in anxiety symptoms ($P = .04$) and no differences were observed for the control group ($P = .33$). No significant differences were found for depressive symptoms within the groups ($P > .05$). The effect size of the intervention was moderate for anxiety ($d = .50$) and small for depression ($r = .09$). In regards to participants engagement after eight weeks, there was an average of 472 exchanged messages ($M = 472.15$; $SD = 249.52$) and a higher number of messages exchanged with Tess was associated with positive feedback ($F(2,36) = 4.37$; $P = .02$).

Conclusions: Students engaged a considerable amount of time exchanging messages with Tess and positive feedback was associated with higher numbers of messages exchanged. The initial results show promising evidence for the use of Tess for anxiety symptoms and a lower effect on depressive symptoms in Argentinian college students. Research on chatbots is still in its initial stages and further research is needed.

(JMIR Preprints 07/07/2020:20678)

DOI: <https://doi.org/10.2196/preprints.20678>

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 visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#)

Original Manuscript

Preprint
JMIR Publications

Original paper**Artificial Intelligence Chatbot for Anxiety and Depression in University Students: a Pilot Randomized Controlled Trial****Abstract**

Background: The use of artificial intelligence based chatbots as an instrument of psychological intervention is emerging, however no studies have been reported in Latin America.

Objective: The objective of the present study was to evaluate the viability, acceptability and potential impact of Tess, a chatbot, on symptoms of depression and anxiety in university students.

Methodology: This was a pilot randomized controlled trial. The experimental condition used Tess for eight weeks and the control condition was assigned to a psychoeducation book on depression. Comparisons were conducted using Mann-Whitney *U* and Wilcoxon tests for depression symptoms, and Independent and Paired Samples *t* Tests to analyze anxiety symptoms.

Results: The initial sample consisted of 181 Argentinian college students (87.2% female) ages 18 to 33. A total of 39 (39%) participants in the experimental condition and 34 (41%) in the control group, provided data at week eight. There was an average of 472 ($SD=249.52$) messages exchanged and an average of 116 ($SD=73.87$) of the messages were sent from the user in response to Tess. A higher number of messages exchanged with Tess was associated with positive feedback ($F_{2,36}=4.37$; $P=.02$). No significant intergroup differences from baseline to week eight between the experimental group and the control group were found for depression and anxiety symptoms. However, significant intragroup differences demonstrated that the experimental group showed a significant decrease in anxiety symptoms; no differences were observed for the control group. No significant differences were found for depressive symptoms within the groups.

Conclusions: Students engaged a considerable amount of time exchanging messages with Tess and positive feedback was associated with higher numbers of messages exchanged. The initial results show promising evidence for the usability and acceptability of Tess in the Argentinian population. Research on chatbots is still in its initial stages and further research is needed.

Keywords: Artificial intelligence; chatbots; conversational agents; mental health; anxiety; depression; college students

Introduction

The most prevalent disorders in Argentina are anxiety (16.4%) and mood (12.3%) disorders. The average age for the onset of these conditions is 20 years [1]. The Pan American Health Organization (PAHO) and the Argentinian Ministry of Health highlighted the importance of optimizing healthcare services for individuals who are not receiving any kind of psychological care [2]. Furthermore, the epidemiological data collected in Argentina emphasizes the need for strategies that prevent a delay to treatment access [1]. Behavioral Intervention Technologies (BITs) are a novel and effective delivery format that can expand mental health services offered and facilitate early access to those in need [3]. Chatbots are an example of BITs that represent an opportunity for addressing delays associated with access to treatment for depression and anxiety [4]. However, no studies on the use of chatbots for depression and anxiety have been conducted in Argentina.

The use of chatbots utilizing artificial intelligence (AI) in the field of psychology is emerging [5]. Currently, there are two chatbots that have addressed anxiety and depressive symptoms - Woebot [6] and Tess [7]. Woebot is a chatbot based on the cognitive-behavioral approach with evidence for the reduction of anxiety and depression symptoms in students at two weeks follow-up. Fulmer et al., [7] reported a reduction in depressive and anxiety symptoms in college students utilizing Tess, a chatbot that provides support and psychoeducation through an integrative approach. Although the research completed by Fulmer et al. [7] and Fitzpatrick, et al. [6] reported a decrease in depression and anxiety symptoms on college students, these studies were performed in the USA and to our knowledge there are no studies on chatbot for mental health for Spanish-speaking populations. Other examples of chatbots with empirical support are: MYLO (Manage Your Life Online) that focuses on problem solving [8]; Shim, for well-being based on the cognitive-behavioral approach and elements of positive psychology [9]; Tess for pediatric obesity and pre-diabetes treatment [10]; and Wysa, a

chatbot that uses cognitive behavioral therapy, behavioral reinforcement and mindfulness techniques to support patients with depression [11]. Research on chatbots for mental health have several limitations such as small sample sizes and short term follow ups [6,7]. Additionally, current chatbots for mental health promotion present several problems such as the lack of recognition of the emotional tone of users, crisis identification and management, as well as the need of strategies to reduce the frustration arising from feelings of incomprehension by the user when the chatbot does not respond accurately.

The present study aims to assess the viability and acceptability of a psychological intervention delivered through Tess to college students in Argentina. The objectives of this study were to: i) identify participants' flow from recruitment to follow-up; ii) understand aspects related to the usage patterns of Tess, such as the number of messages sent and exchanged; iii) examine the relation between feedback expressed by participants and the number of messages exchanged with Tess; iv) compare between and within groups outcome on depression and anxiety, among college students who completed the study. Although the focus of this research is not the effectiveness of the chatbot, comparisons were made between experimental and control groups in order to have preliminary data for future randomized controlled trials. Given the importance of having preliminary information about the viability and acceptability of Tess as a means of psychological intervention with college students in Argentina.

Methods

Trial design

This was a pilot randomized controlled parallel-group trial. The experimental group had

access to Tess for 8 weeks and the control group to a psychoeducation electronic book.

Participants

Participants were college students in Entre Ríos, Argentina. The inclusion criteria included: being a resident of Argentina, 18 years or older, a college student, and providing informed consent. The recruitment was conducted through presentations in different university courses. Participants that provided consent were assigned to experimental or control condition by simple randomization conducted through an algorithm in python.

Intervention

Experimental group: The experimental group utilized Tess, an AI-based chatbot that provides conversations through text, to deliver brief conversations as comprehensive support for mental health. Tess sends reminders, psychoeducational content and emotional support responses based on what the users expose. Tess combines words and emojis in the messages for a more user-friendly experience. Tess responds with prescribed statements in order to replicate empathetic answers that are appropriate for the emotion or concern introduced by the participants. For example, a participant indicating anxious feelings would be offered a relaxation strategy. The conversations offered by Tess were based on the Cognitive Behavioral model [12], Emotion-Focused Therapy [13], Solution-Focused Brief Therapy [14], and Motivational Interviewing [15]. Such conversations were developed by mental health experts. After each conversation, Tess asked, “Was our conversation helpful?”, if a user responded positively (e.g., “yes, thank you”) to a CBT-based intervention and negatively (e.g. “no, not really”) to an Emotion-Focused Therapy, Tess would then offer more CBT-based interventions. For users that answered in a negative or neutral manner, Tess would offer alternative interventions.

For the present study, customized conversations for university students in Argentina were elaborated, revised and tested within the framework of a previous study developed in the US [7]. In the eight-week intervention of this test, Tess initiated contact asking about their emotions and moods with the participants once a day during the first weeks and every other day the following weeks. All conversations with Tess were done through Facebook messenger.

Control Group: An electronic psychoeducation book focused on affective symptoms was provided to participants in the control group [16]. The provides evidence-based information and resources to help students identify and seek treatment for depressive symptoms.

Engagement and feedback

Engagement was measured using the number of messages exchanged with Tess. In addition, the dropout rates were analyzed, both in the experimental group and in the control group. Participant's perceived feedback was collected after each conversation with Tess, through the following question: "Was our conversation helpful?" Answers from users were codified as positive, negative or ambivalent and assigned a value of one, two, or three, respectively. For instance, if a user responds "yes, thank you" that response was coded as positive.

Measures

The Patient Health Questionnaire – 9 (PHQ-9) [17] is a self-reporting questionnaire comprising nine items that evaluate the frequency and severity of depression symptoms during the last two weeks. Each one of the 9 items are based on the DSM-IV criteria which are scored from zero (not at all) to three (nearly every day). It is one of the most used measures to assess depression symptoms and has been validated in Argentina, with adequate psychometric properties ($\alpha = .87$) [18].

The two first items are considered as screening criteria (PHQ-2); if these are answered with zero or one, then an absence of symptoms is assumed. Scores ranging from five to nine are interpreted as mild, from 10 to 14 as moderate, from 15 to 20 as moderately severe and over 20 as severe.

The Generalized Anxiety Disorder Scale (GAD-7) [19] is a 7-item self-reporting scale that evaluates the frequency and severity of anxious thoughts and behaviors during the last two weeks. Items are based on the diagnostic criteria of the DSM-IV and scored from zero (not at all) to three (nearly every day). Rodríguez de Behrends and Brenlla [20] reported an adequate reliability level ($\alpha = .74$) in the Argentinian population.

Ethical aspects

The study was approved by the Research Ethics Committee of the Faculty of Health Sciences (FCS) of the Universidad Adventista del Plata (UAP), with No CE000237 of the National Registry of Health Research (RENIS), and No 3999 of Ministerial Resolution of the Ministry of Health of the Province of Entre Ríos. This resolution is recorded in ACT No. 1-2019 of the registration of this committee. Participants expressed their consent in a form according to the personal data protection law (Argentine National Law 25.326) through checkbox (e-signature) on a closed form.

Data collection was collected through Tess. All personally identifiable information was eliminated from transcriptions downloaded from Tess. Processing and storage of downloaded data were made using secure servers and were compliant with the Health Insurance Portability and Accountability Act. Upon completion of the study, the control group obtained access to Tess for eight weeks and both groups were granted free access to Tess for a year. In the event that a participants expressed suicidal ideation, Tess was programmed to provide the National Line of Suicide Prevention numbers, the crisis text line, 911, and to encourage seeking professional help.

Data Analysis

The data collected was entered and analyzed using 20.0.0 *Statistical Package for the Social Sciences* [21]. The number of messages exchanged was considered in order to assess the feasibility and acceptability of Tess. Additionally, participants' qualitative feedback was analyzed by two researchers (C.K. and M.E.) and coded into three categories: positive, negative and ambivalent. A data analysis protocol was carried out. The treatment of missing data through multiple imputation or plausibility analysis techniques was not possible due to the high percentage of participants who dropped out of the intervention [22].

A one-factor Analysis of Variance (ANOVA) was applied to determine if feedback (positive, negative or ambivalent) impacted the number of interactions that users had with Tess. To examine baseline characteristics between samples, a *t*-test for independent samples was performed to compare anxiety levels and a Mann-Whitney *U* (respecting the ordinal nature of the variable - i.e., if the first two items scored 0 or 1 the system did not ask the subsequent items) test was used to compare depressive symptoms.

To evaluate the effects between conditions a *t*-test for independent samples was performed to assess the anxiety symptoms and the Mann-Whitney *U* test was used to compare the mean ranges of depression. To assess longitudinal effects from baseline to week eight within conditions, a *t*-tests for related samples was performed to assess the anxiety symptoms and the Wilcoxon test was used to compare the mean ranges of depression. In order to complement the significance test, the effect size in the intragroup and intergroup tests were calculated. For *t*-tests, the effect size was calculated with Cohen *d*, measures between .20 and .30 are considered “small effect”, around .50 “moderate effect” and above .80 “large effect” [23]. For Mann-Whitney *U* and Wilcoxon tests, the formula *r* was

calculated based on Z scores, measures between .10 - .30 are considered “small effect”, between .30 - .50 “moderate effect” and above .50 “large effect” [23].

Results

The initial sample consisted of 181 college students in Argentina, ages 18 to 33, with 87.2% identifying as female. Of those, 99 were randomized to the experimental condition and 82 to the control condition. A total of 39 (39%) participants in the experimental condition and 34 (41%) in the control group, provided data at week eight. Regarding data on the depressive symptoms, 33 participants in the experimental condition and 30 in the control condition provided data at week eight. Regarding data on anxiety symptoms 27 participants in the experimental condition and 23 in the control condition provided data at week eight (See figure 1).

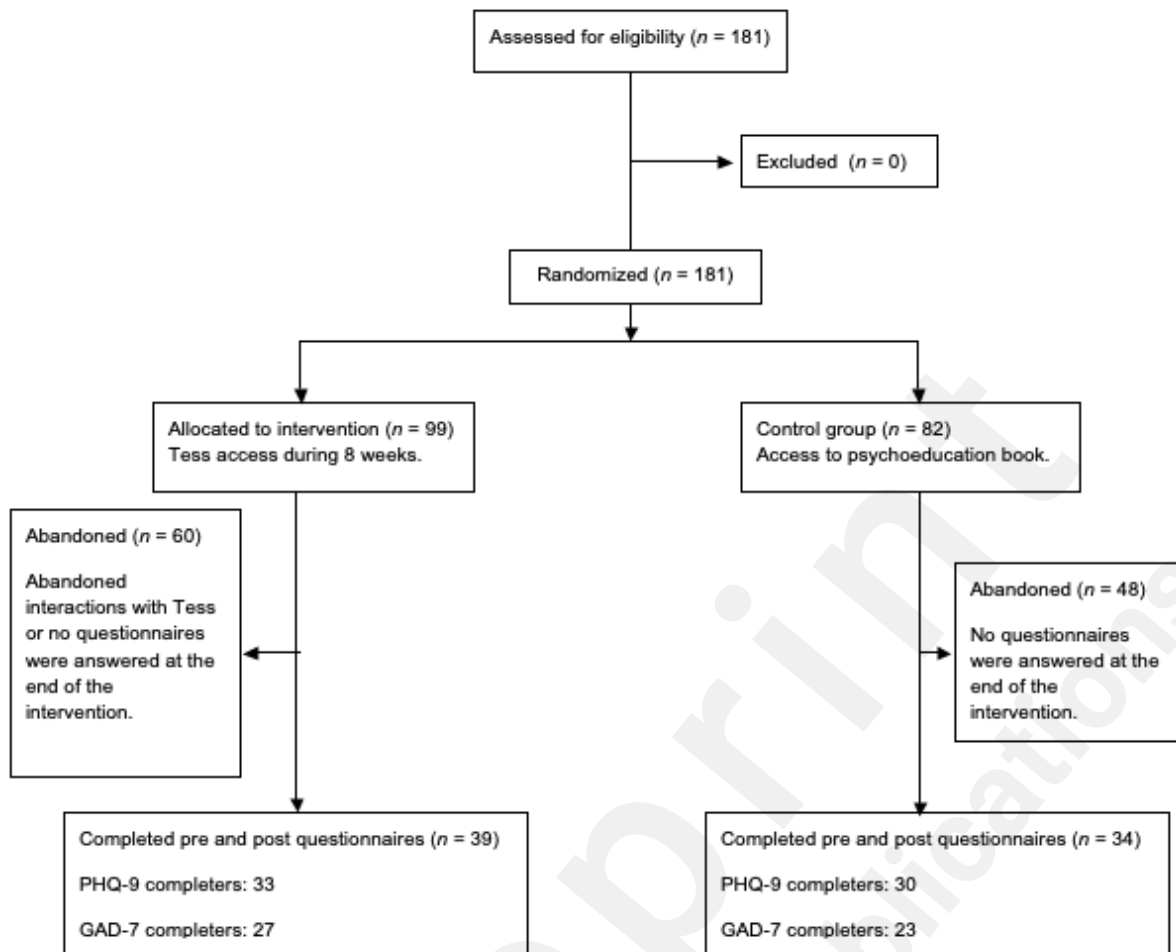


Figure 1. CONSORT *diagram*

Messages exchanged

In regards to participants' engagement with Tess (n=39), after eight weeks, there was an average of 472 exchanged messages ($SD = 249.52$), where the minimum interaction was 162 messages and the maximum, 1290. More specifically an average of 116 ($SD = 73.87$) of the messages were sent from the user to Tess.

Feedback

Feedback from participants at week eight was coded as positive for most participants ($n = 25$) (i.e. “Yes, you really understand me, Tess. Thanks for talking to me. My anxiety has decreased and I can confidently go outside again”). A minor portion provided ambivalent feedback ($n = 7$ - i.e. “Not much, but it’s ok, I am capable”) or negative ($n = 7$ - i.e. “Sometimes I ask you something and you don’t specifically respond to what I asked”). A one-factor ANOVA was applied to determine if feedback (positive, negative or ambivalent) impacted the number of interactions that users had with Tess. Results showed that feedback from users was associated with the number of messages exchanged with Tess ($F_{2,36} = 4.37$; $P = .02$). Post hoc contrasts resulting from the Scheffe Test showed statistically significant differences between those that reported a positive feedback and those who reported a negative feedback ($P = .038$) whereas a higher number of messages exchanged with Tess was associated with positive feedback. No differences were observed between those that reported ambivalent and positive feedback or ambivalent and negative feedback (See Table 3).

Table 1
Number of interactions per user feedback

Variable	Positive		Ambivalent		Negative		Statistical Values	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>P</i>

Interactions	551.24	52.54	374.43	73.11	287.43	23.35	4.37	.020
---------------------	--------	-------	--------	-------	--------	-------	------	-------------

Potential impact of Tess on indicators of depression and anxiety

Baseline Characteristics

There were no statistically significant differences at baseline in anxiety ($t_{48}=.16$; $P = .88$) and depression scores ($U = 451.50$; $P = .52$) between the experimental and the control group (See Table 1).

Between Group Differences

No statistically significant differences were observed between the experimental and the control groups in the average scores for anxiety ($t_{48}= 1.74$; $P = .09$) or in the average ranges for depression ($U = 448.00$; $P = .48$) at week eight (See Table 1). Regarding the effect sizes, the mean scores for anxiety in the experimental group were lower than for the control group after eight weeks and the effect size of the intervention was moderate [$d = .50$, 95% IC (-6.96 - .510)]. For depressive symptoms, the experimental group reported a lower mean score than the control group and the effect size of the intervention was non-existent ($r = .09$).

Table 2

Comparison of the average values and ranges between groups for the anxiety and depression variables at baseline and week eight between the experimental and control groups.

Variable	Experimental Group		Control Group		Statistical Values	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>P</i>
Anxiety-baseline	15.59 ^a	5.30	15.35 ^b	5.75	.16	.876
Anxiety-week eight	13.04 ^a	7.12	16.26 ^b	5.79	1.74	.089
Mann-Whitney U test	MR		MR		U	P
Depression-baseline	33.32 ^c		30.55 ^d		451.50	.522
Depression-week eight	30.58 ^c		33.57 ^d		448.00	.477

Note: MR: middle range. ^an = 27; ^bn= 23; ^cn= 33; ^dn= 30

Within Group Differences

Within the experimental condition, a statistically significant decrease of symptoms was observed from baseline to week eight for the anxiety scores ($t_{26} = 2.15$; $P = .04$); the control condition did not demonstrate any significant changes ($t_{22} = 1.00$; $P = .33$). Regarding depressive symptoms no significant differences were found for depression in the experimental condition ($Z = 1.76$; $P = .08$) or the control condition ($Z = .00$; $P > .99$) (See Table 2).

Table 3

Comparison of the average values and ranges within groups for the anxiety and depression variables from baseline to week eight.

Variable	Condition	Baseline		Week eight		Statistical Values	
<i>t test</i>		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>P</i>
Anxiety	Experimental	15.59	5.30	13.04	7.12	2.15	.041
	Control	15.35	5.75	16.26	5.79	1.00	.327
<i>Wilcoxon</i>		MR		MR		Z	P

<i>test</i>					
Depression	Experimental	8.83	7.14	1.76	.078
	Control	6.50	6.50	.000	.99

Note: MR: middle range.

Discussion

The use of chatbots (i.e., conversational agents) to address mental health conditions may contribute to the treatment of large populations and attend to the needs of those who do not have access to treatment. To our knowledge there are no studies on the utilization of chatbots for mental health in Latin America. This trial was intended to evaluate an AI-based chatbot (Tess) in a sample of Argentinian college students. The specific objectives were: i) understand the participants flow from recruitment to follow-up; ii) report aspects related to the usage patterns of Tess, such as the number of messages sent and exchanged; iii) examined the participant's feedback; iv) compare preliminary measures of depression and anxiety.

Regarding usage patterns of Tess, there are three findings that support a satisfactory level of engagement. First, a considerable number of participants in the experimental (39%) and control (41%) conditions remained in the study throughout the eight-week study period. The completion rates found in the current study are better than what is observed in most unpaid and unsupported internet interventions for depression and anxiety where 90% of users withdraw after the first two sessions [25]. Furthermore, in studies using smartphone applications the follow-up completion rates were comparable to (53%); and the mean percentage of complete "adherers" for depression was 36% and for anxiety 41% [26]. When compared to a U.S. chatbot study for college students, a lower attrition rate was reported (31% and 9% in the control and experimental condition, respectively); however, this study compensated participants and follow-up was at two weeks, making it difficult to compare the outcomes [6].

Second, participants in the experimental condition had exchanged a considerable number of

messages with Tess ($M = 472$; $SD = 249.52$), and the mean messages sent from the user to Tess was 116 ($SD = 73.87$). A previous study on usage patterns of the depression modules of Tess showed a much lower average number (17.57) of messages sent to Tess by adults users [27]. It is possible that college, younger, and Latinx students are more willing to engage in conversation with chatbots than older populations in the U.S. Two previous studies with college students in the U.S. did not report the number of messages sent by the user to Tess [6,7]. Regarding messages exchanged, Fulmer et al. [7], reported a comparable number of total messages exchanged during a period of four weeks ($M = 286$; $SD = 104.6$), while the total number of messages exchanged in the current study was during a period of eight weeks.

Third, feedback provided by those in the experimental condition was mostly positive (e.g., *“Yes, you really understand me, Tess. Thanks for talking to me.”*). Among those that sent negative feedback there was a predominant dissatisfaction regarding the accuracy of some interventions (e.g., *“Sometimes I ask you something and you don’t specifically respond to what I asked.”*). Feedback is a key component for AI chatbots as it allows the systems to tailor the dialogues to the user. Interestingly, the positive and negative feedback were associated with the number of messages exchanged. Users who reported higher satisfaction had the highest number of exchanged messages, it is possible that providing positive feedback yielded better customization for the participants of the intervention’s messages exchanged. This finding is relevant as it supports the need to collect user feedback in order to achieve optimal levels of customization and increase engagement that could lead to higher intervention doses.

Regarding the impact of Tess on anxiety and depression symptoms, no statistically significant differences were found between groups. Interestingly, when comparing within group scores, the experimental group showed a significant decrease in anxiety symptoms after eight weeks of the intervention and a near-significant trend ($P = 0.07$) for depressive symptoms. When analyzing the effect sizes, in the experimental group the Tess had a moderate effect for anxiety and no effect for

depression. These outcomes were unexpected given that previous studies utilizing Tess [7] and utilizing another conversational agent called Woebot [6] reported significant reductions in symptoms of anxiety and depression, both studies used a similar control group (a psychoeducation book). Of note, in the current study depression was measured using PHQ-9 as categorical and ordinal variables while Fulmmer, et al. [7] and Fitzpatrick, et al. [6] used it as a continuous measure.

The lack of between-group differences could be explained by several factors. First, the current study was underpowered. Second, while the findings of the current study were not statistically significant, the direction of the change observed for anxiety and depression were in the expected direction, therefore, it is possible that low intensity interventions delivered via chatbot may require a higher dose in order to yield a between group effect when delivered to Argentinian students. Third, Tess provides many conversations from different theoretical approaches and this may have results in a less therapeutic power. However, Fulmer et al. [7] found significant effect using similar conversations. Fourth, it is possible that in the adaptation of the dialogues from English to Spanish the quality of the intervention may have been reduced.

Limitations and Future Directions

This pilot study has several limitations. First, the current analysis was done with intervention completers, future studies with larger samples (and including analysis of both completers and intent-to-treat) are needed. Second, only college students from a specific region in Argentina were included in this study and the socio-economic aspects of the sample were not assessed, the inclusion of a more diversified sample is suggested. Third, there was a high dropout rate throughout the eight-week period. This is congruent with the findings reported by most studies that use technology-based intervention (see “The Law of Attrition,”) [28]. A high dropout rate may be due to a limited capacity of most digital interventions to capture the users attention and motivation. Additionally, high dropout

rates in studies with digital interventions were linked to the fact that as access is easy, a lower level of commitment is required from the user to enroll in the study compared to traditional face-to-face interventions. Fourth, the majority of participants who remained until completion of the study were female, so male participation was scarce. Fifth, the control group had access to a psychoeducation book and there is no information on whether they read it or not. Since chatbot research is in its initial stages, further studies could benefit from comparing to waitlist groups rather than self-help books. While offering a waitlist could represent an ethical dilemma, if studies are short term this would mitigate the potential effects of not having an intervention.

Future Chatbot studies may benefit from designing chatbots with more conversation from a specific therapeutic approach, rather than utilizing few conversations from several approaches. Additionally, analyzing the impact of chatbots as adjuncts to face-to-face psychotherapy, and comparing these treatments combined vs. face-to-face psychotherapy alone would yield important insight to the advancement of research on chatbots for mental health. Finally, a simple randomization was used; future studies may consider utilizing unequal randomization (2:1) so that more participants enter to the experimental group or a stratified randomization procedure so that participants with similar characteristics can be assigned evenly to the experimental and control groups.

Conclusions

Students engaged a considerable amount of time exchanging messages with Tess and positive feedback was associated with higher numbers of messages exchanged. The initial results show preliminary evidence for the use of Tess for anxiety symptoms but no effect at impacting depressive

symptoms in Argentinian college students. Given the high prevalence of anxiety and depression in Argentinian college students [1], and the need to expand mental healthcare access, developing affordable strategies such as chatbots may become an effective tool to address these needs. AI-based chatbots have the ability to reach higher-levels of customization and, thus, may be of service to education and mental healthcare centers who aim to deliver interventions that are accessible at any time and without geographical restrictions for targeted users. Additionally, chatbots may be used as stand alone resources for those who have no access to treatment and/or as a complement to traditional treatments. While the initial evidence on the efficacy of chatbots is promising, research on chatbots is still in its initial stages and presents several limitations. Thus, more robust evidence is needed.

References

1. Cía AH, Stagnaro JC, Gaxiola SA, et al. Lifetime prevalence and age-of-onset of mental disorders in adults from the Argentinian Study of Mental Health Epidemiology. *Social psychiatry and psychiatric epidemiology*. 2018; 53(4):341-350. doi: 10.1007/s00127-018-1492-3
2. Pan American Health Organization. Framework for the Implementation of the Regional Strategy on Mental Health 2011. <https://iris.paho.org/bitstream/handle/10665.2/3535/PAHO%20framework%20implem%20reg%20strat%20on%20mh%202011.pdf>. Accessed January

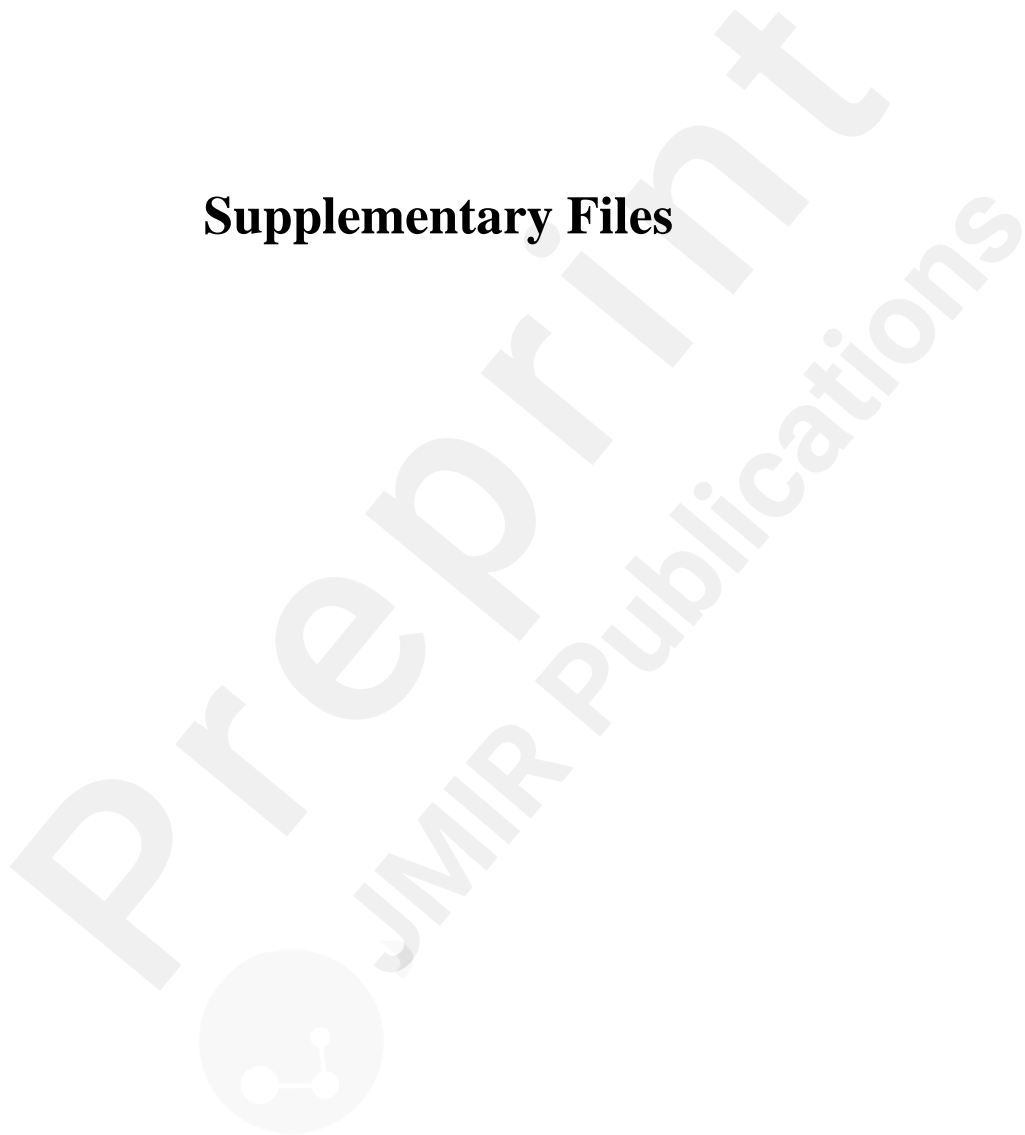
- 20, 2019.
3. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *General hospital psychiatry*. 2013; 35(4):332-338. doi: 10.1016/j.genhosppsych.2013.03.008
 4. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health—a scoping review. *Verhaltenstherapie*. 2019: 1-13. doi:10.1159/000501812
 5. Gaffney H, Mansell W, Tai S. Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Mental Health*. 2019; 6(10): e14166. doi: 10.2196/14166
 6. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*. 2017; 4(2): e19. doi: 10.2196/mental.7785
 7. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. *JMIR Mental Health*. 2018; 5(4): e64. doi:10.2196/mental.9782
 8. Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behavioral and cognitive psychotherapy*. 2014; 42(6): 731-746. doi: 10.1017/s135246581300060x
 9. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet interventions*. 2017; 10: 39-46. doi: 10.1016/j.invent.2017.10.002
 10. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Translational behavioral medicine*. 2019; 9(3): 440-447. doi: 10.1093/tbm/ibz043
 11. Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11): e12106. doi: doi:10.2196/12106
 12. Beck JS, Beck AT. *Cognitive Behavior Therapy: Basics and Beyond, Second Edition*. New York City: The Guilford Press; 2011.
 13. Greenberg L. *Emotion-focused Therapy: Coaching Clients to Work Through Their Feelings*. Washington, DC: American Psychological Association; 2002.
 14. Pichot T, Dolan Y. *Solution-Focused Brief Therapy: Its Effective Use in Agency Settings*. Haworth marriage and the family. New York: Routledge; 2003.
 15. Rollnick S, Miller WR. What is Motivational Interviewing? *Behavioural and Cognitive Psychotherapy*. 2009; 23(04):325. doi:10.1017/s135246580001643x
 16. National Institute of Mental Health. *Depression in College Students*. URL: https://www.nimh.nih.gov/health/publications/espanol/la-depresion-en-la-adolescencia/19-mh-8089s_157885.pdf [accessed 2018-11-10]
 17. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*. 2001; 16(9):606-613. doi: 10.1046/j.1525-

1497.2001.016009606x

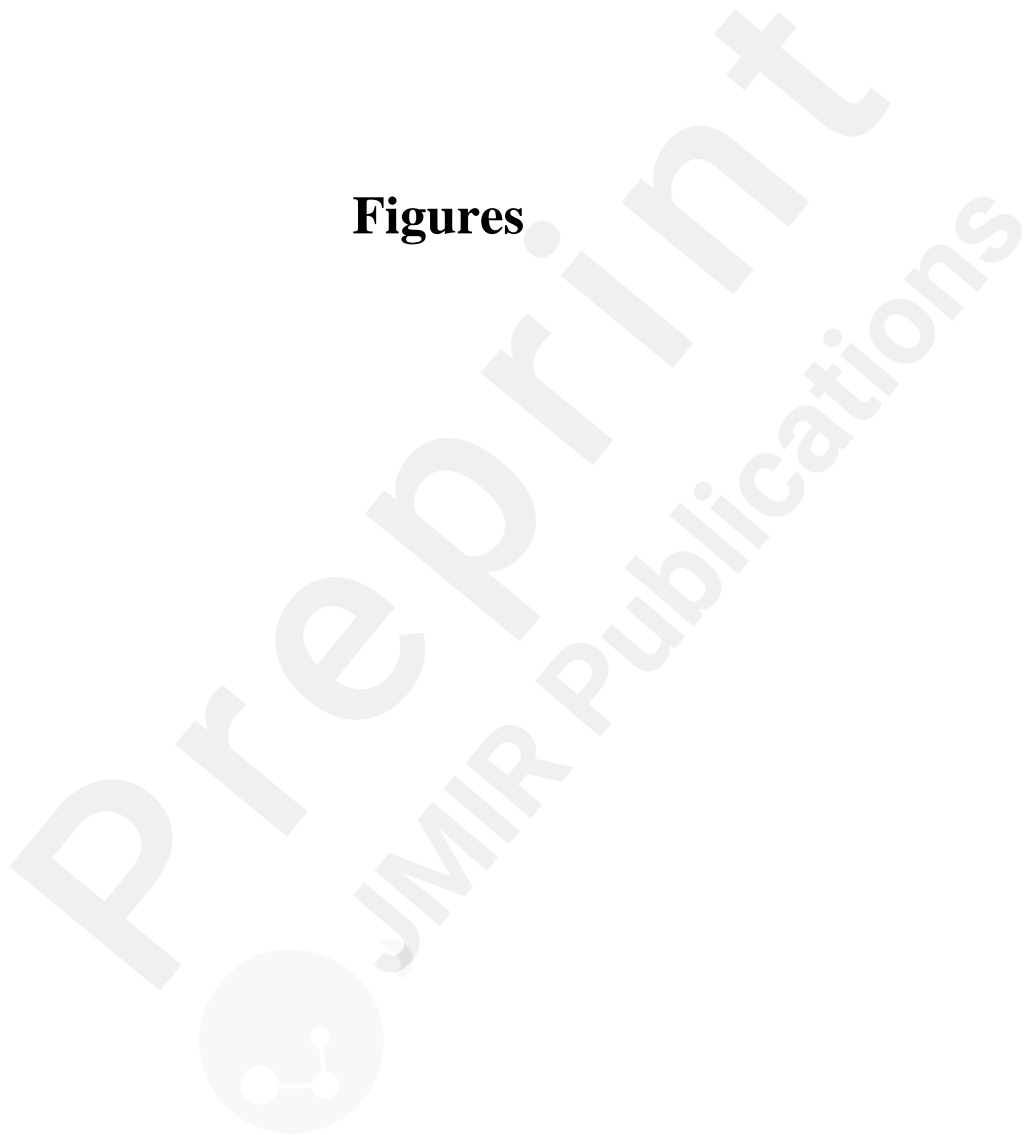
18. Urtasun M, Daray FM, Teti GL, et al.. Validation and calibration of the patient Health Questionnaire (PHQ9) in Argentina. *BMC psychiatry*. 2019; 19(1):1-10. doi: 10.1186/s12888-019-2262-9
19. Newman MG, Zuelling AR, Kachin KE, Constantino MJ, Cashman L. The reliability and validity of the GAD-Q-IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy*. 2002; 33(2):215-233. doi:10.1016/s0005-7894(02)80026-0
20. Rodríguez de Behrends M, Brenlla, ME (2015, november). Estudio preliminar de las propiedades psicométricas del Cuestionario de Trastorno de Ansiedad Generalizada-IV (GAD-Q-IV-Generalized Anxiety Disorder Questionnaire-IV) en población de Buenos Aires. Presentation in VII Congreso Internacional de Investigación y Práctica Profesional en Psicología XXII Jornadas de Investigación XI Encuentro de Investigadores en Psicología del MERCOSUR. Facultad de Psicología-Universidad de Buenos Aires. Retrieved from http://www.psi.uba.ar/investigaciones/eventos_cientificos/vii_congreso_memorias/6_tecn_proces_eval.pdf
21. IBM Corp. Released. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp; 2013.
22. Fernández-García MP, Vallejo-Seco G, Livácic-Rojas P, Tuero-Herrero, E. The (Ir) Responsibility of (Under) Estimating Missing Data. *Frontiers in Psychology*. 2018; 9. doi:10.3389/fpsyg.2018.00556
23. Cohen J. *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates; 1988. ISBN: 0805822232
24. Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*. 2014; 21(1):19-25.
25. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, Rapee RM, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. *PloS one*. 2013;8(7):e62873. doi:10.1371/journal.pone.0062873
26. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of consulting and clinical psychology*. 2020; 88(1):1–13. doi:10.1037/ccp0000459
27. Dosovitsky G, Pineda B, Chang C, Jacobson N, Bunge E. Understanding Artificial Intelligence Chatbot Usage for Depression (Preprint). doi: 10.2196/preprints.17065
28. Eysenbach G. The law of attrition. *Journal of medical Internet research*. 2005; 7(1):e11. doi: 10.2196/jmir.7.1. e11

Preprint
JMIR Publications

Supplementary Files



Figures



CONSORT diagram.

