

Original Paper

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

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

¹Interdisciplinary Center for Research in Health and Behavioral Sciences (CIICSAC), Universidad Adventista del Plata (UAP)., National Scientific and Technical Research Council (CONICET)., Libertador San Martín, Entre Ríos, Argentina

²X2AI inc., San Francisco, CA, United States

³Department of Psychology, Palo Alto University, Palo Alto, CA, United States

* all authors contributed equally

Corresponding Author:

Maria Carolina Klos, PhD candidate, MA

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

25 de Mayo 99

Libertador San Martín, Entre Ríos, 3103

Argentina

Phone: 54 3435064263

Email: mcarolinaklos@gmail.com

Abstract

Background: Artificial intelligence–based chatbots are emerging as instruments of psychological intervention; however, no relevant studies have been reported in Latin America.

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

Methods: This was a pilot randomized controlled trial. The experimental condition used Tess for 8 weeks, and the control condition was assigned to a psychoeducation book on depression. Comparisons were conducted using Mann-Whitney *U* and Wilcoxon tests for depressive symptoms, and independent and paired sample *t* tests to analyze anxiety symptoms.

Results: The initial sample consisted of 181 Argentinian college students (158, 87.2% female) aged 18 to 33. Data at week 8 were provided by 39 out of the 99 (39%) participants in the experimental condition and 34 out of the 82 (41%) in the control group. On an average, 472 (SD 249.52) messages were exchanged, with 116 (SD 73.87) of the messages sent from the users 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 differences between the experimental and control groups were found from the baseline to week 8 for depressive and anxiety symptoms. However, significant intragroup differences demonstrated that the experimental group showed a significant decrease in anxiety symptoms; no such differences were observed for the control group. Further, no significant intragroup differences were found for depressive symptoms.

Conclusions: The students spent a considerable amount of time exchanging messages with Tess and positive feedback was associated with a higher number 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.

(JMIR Form Res 2021;5(8):e20678) doi: [10.2196/20678](https://doi.org/10.2196/20678)

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 have highlighted the importance of optimizing health care services for individuals who are not receiving any form of psychological care [2]. Furthermore, the epidemiological data collected in Argentina emphasizes the need for strategies that prevent delays to treatment access [1]. Behavioral intervention technologies (BITs) are a novel and effective delivery format that can expand the mental health services offered and facilitate early access to those in need [3]. Chatbots are examples 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 analyzing depression and anxiety have been conducted in Argentina.

Chatbots developed using artificial intelligence (AI) are emerging in the field of psychology [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 depressive symptoms in students during a follow-up after 2 weeks. Fulmer et al [7] reported a reduction in depressive and anxiety symptoms in college students using 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 decreased depressive and anxiety symptoms in college students, these studies were performed in the United States. To the best of our knowledge, there are no studies on chatbots used for addressing mental health disorders in Spanish-speaking populations. Other examples of chatbots with empirical support are Manage Your Life Online (MYLO) 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 prediabetes treatment [10]; and Wysa, a chatbot that uses cognitive behavioral therapy, behavioral reinforcement, and mindfulness techniques to support patients with depression [11]. Research studies 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 for strategies to reduce the frustration arising from feelings of incomprehension by users when the chatbot does not respond accurately.

The present study aims to assess the viability and acceptability of psychological interventions delivered through Tess to college students in Argentina. The objectives of this study were as follows: (1) identify participant flow from recruitment to follow-up; (2) understand aspects related to the usage patterns of Tess, such as the number of messages sent and exchanged; (3) examine the relationship between the feedback provided by the participants and the number of messages exchanged with Tess; and (4) compare the outcomes on depression and anxiety

between and within groups among the college students who completed the study. Although the focus of this research was not the effectiveness of the chatbot, comparisons were made between the experimental and control groups to obtain preliminary data for future randomized controlled trials given the importance of obtaining preliminary information about the viability and acceptability of Tess as a means of psychological intervention for 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

The participants were college students in Entre Ríos, Argentina. The inclusion criteria were as follows: being a resident of Argentina, 18 years or older, and a college student, as well as providing informed consent. Recruitment was conducted through presentations in different university courses. Participants who provided consent were assigned to the experimental or control group by simple randomization conducted through a Python algorithm.

Intervention

Experimental Group

The experimental group utilized Tess, an AI-based chatbot that delivers brief text conversations as comprehensive support for mental health. Tess sends reminders, psychoeducational content, and emotional support responses based on what the users express. Tess combines words and emojis in the messages for providing a more user-friendly experience. Tess responds with prescribed statements to replicate empathetic answers that are appropriate for the emotion or concern expressed by the participants. For example, a participant expressing 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 (eg, "yes, thank you") to an intervention based on cognitive behavioral therapy (CBT) and negatively (eg, "no, not really") to emotion-focused therapy, Tess would then offer more CBT-based interventions. For users who answered in a negative or neutral manner, Tess would offer alternative interventions.

In 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 United States [7]. During the 8-week intervention for this test, Tess initiated contact asking about the emotions and moods of the participants once a day during the initial weeks and every other day in the following weeks. All the conversations with Tess occurred through Facebook messenger.

Control Group

An electronic psychoeducation book focusing on affective symptoms was provided to the participants in the control group [16]. The provided evidence-based information and resources helped 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 in the experimental and control groups were analyzed. The perceived feedback of the participants was collected after each conversation with Tess through the following question: “Was our conversation helpful?” The answers from the users were coded as positive, negative, or ambivalent and assigned values of 1, 2, and 3, respectively. For instance, if a user responded saying “yes, thank you,” then that response was coded as positive.

Measures

The **Patient Health Questionnaire-9 (PHQ-9)** [17] is a self-reporting questionnaire comprising 9 items that evaluate the frequency and severity of depressive symptoms during the last 2 weeks. Each of the 9 items is based on the *Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV)* criteria, which are scored from 0 (not at all) to 3 (nearly every day). The PHQ-9 is one of the most used measures to assess depressive symptoms and has been validated in Argentina with adequate psychometric properties (Cronbach $\alpha=.87$) [18]. The first 2 items are considered screening criteria (PHQ-2); if these are scored with 0 or 1, then an absence of symptoms is assumed. Scores ranging from 5 to 9 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 2 weeks. Items are based on the diagnostic criteria of the *DSM-IV* and scored from 0 (not at all) to 3 (nearly every day). Rodríguez de Behrends and Brenlla [20] reported an adequate reliability level (Cronbach $\alpha=.74$) for 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), National Registry of Health Research (RENIS, reference number: CE000237), and Ministerial Resolution of the Ministry of Health of the Province of Entre Ríos (reference number: 3999). This resolution is recorded in ACT 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 selection (electronic signature) on a closed form.

Data were collected through Tess. All personally identifiable information was eliminated in the transcriptions downloaded from Tess. The downloaded data were processed and stored 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 8 weeks and both groups were granted free access to Tess for a year. If a participant expressed suicidal ideations, Tess was programmed to provide the National Line of Suicide Prevention numbers, the crisis text line, and 911, and encourage seeking professional help.

Data Analysis

The data collected was entered and analyzed using the Statistical Package for the Social Sciences (Version 20.0; IBM Corporation) [21]. The number of messages exchanged was considered to assess the feasibility and acceptability of Tess. Additionally, the participants' qualitative feedback was analyzed by two researchers (CK and ME) 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 owing 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 the baseline characteristics between samples, a ***t* test** was performed for independent samples to compare anxiety levels. The **Mann-Whitney *U* statistic** (respecting the ordinal nature of the variables; namely, if the first 2 items were scored 0 or 1, the system did not ask the subsequent items) was used to compare depressive symptoms.

To evaluate the effects between conditions, a ***t* test** was performed for the independent samples to assess the anxiety symptoms and the **Mann-Whitney *U* statistic** was used to compare the mean ranges of depression. To assess the longitudinal effects from the baseline to week 8 within conditions, a ***t* test** was performed for related samples assess the anxiety symptoms and the **Wilcoxon test** was performed to compare the mean ranges of depression. To complement the significance test, the effect sizes in the intragroup and intergroup tests were calculated. For the *t* tests, the effect size was calculated using Cohen *d*; measures between 0.2 and 0.3 were labeled “small effect,” around 0.5 as “moderate effect,” and above 0.8 as “large effect” [23]. For the Mann-Whitney *U* and Wilcoxon tests, the *r* formula was calculated based on the *z* scores. The measures between 0.1 and 0.3 were labeled “small effect,” between 0.3 and 0.5 as “moderate effect,” and above 0.5 as large effect” [23,24].

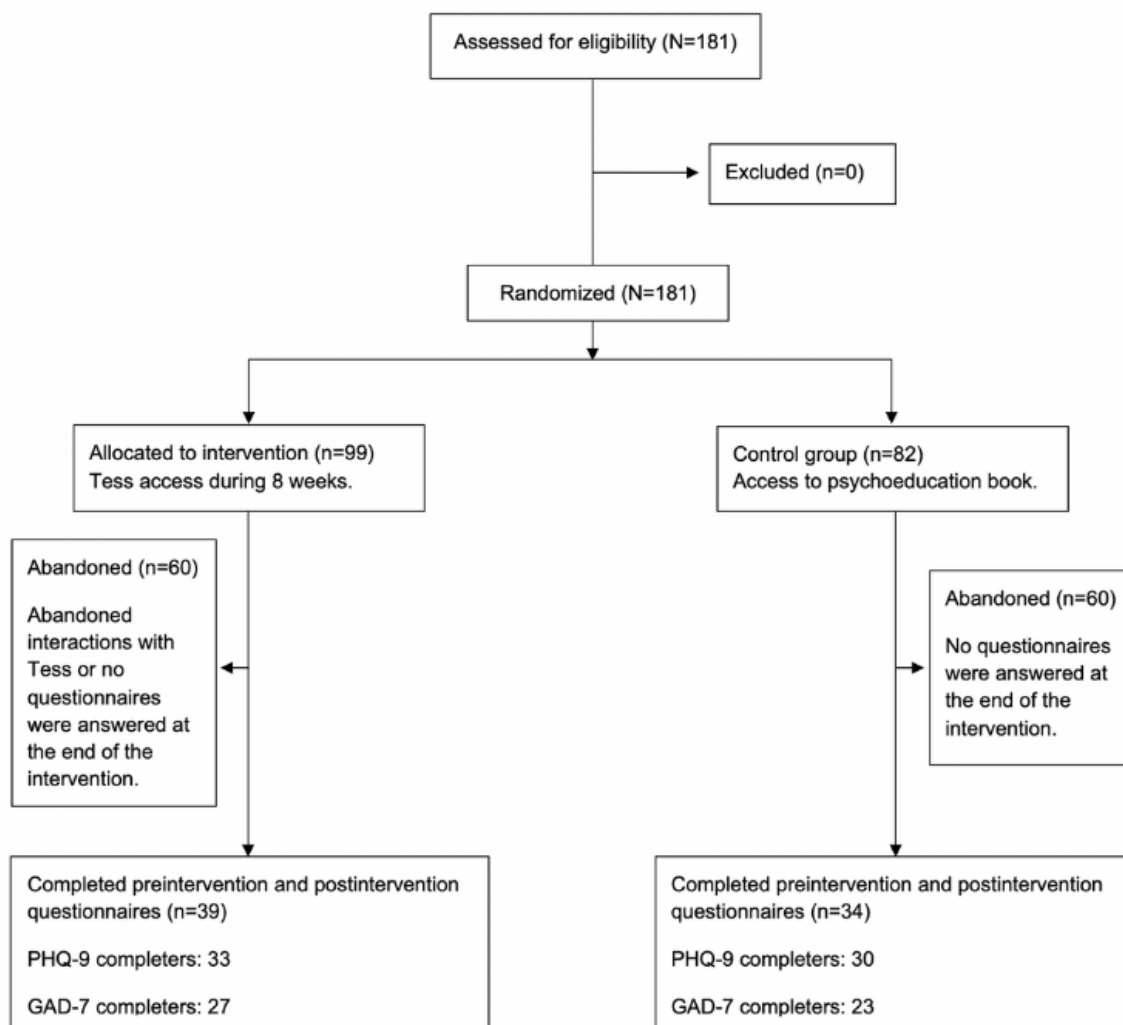
Results

Initial Observations

The initial sample consisted of 181 college students in Argentina, aged 18 to 33, with 158 (87.2%) identifying as female. Among the 181 students, 99 (55%) were randomized to the experimental condition and 82 (45%) to the control condition. Data at week 8 were provided by 39 out of the 99 (39%) participants in the experimental condition and 34 out of the 82 (41%) in the control group. Regarding data on the depressive symptoms, 33 (33%) participants in the experimental

condition and 30 (37%) in the control condition provided data at week 8. Regarding data on anxiety symptoms, 27 (27%) participants in the experimental condition and 23 (28%) in the control condition provided data at week 8 (see Figure 1).

Figure 1. CONSORT flow diagram. GAD-7: Generalized Anxiety Disorder Scale-7, PHQ-9: Patient Health Questionnaire-9.



Messages Exchanged

Regarding the participants' engagement with Tess (39/99), after 8 weeks, there was an average of 472 exchanged messages (SD 249.52), where the minimum interaction level involved 162 messages and the maximum involved 1290. More specifically, an average of 116 (SD 73.87) of the exchanged messages were sent from the user to Tess.

Feedback

Feedback from most participants (25/39) at week 8 was coded as positive (ie, *Yes, you really understand me, Tess. Thanks for talking to me. My anxiety has decreased and I can confidently go outside again.*). A minor number of participants (7/39) provided ambivalent (ie, *Not much, but it's ok, I am capable*)

or negative (ie, *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 Scheffé test showed statistically significant differences between those participants providing positive feedback and those providing negative feedback ($P=.04$); nevertheless, a higher number of messages exchanged with Tess was associated with positive feedback. No differences were observed between the participants providing ambivalent and positive feedback or ambivalent and negative feedback (See Table 1).

Table 1. Number of interactions per user feedback.

Variable	Positive, mean (SD)	Ambivalent, mean (SD)	Negative, mean (SD)	Statistical values	
				F	P value
Interactions	551.24 (52.54)	374.43 (73.11)	287.43 (23.35)	4.37	.02

Potential Impact of Tess on Indicators of Depression and Anxiety

Baseline Characteristics

There were no statistically significant differences at the baseline in the anxiety ($t_{48}=.16$; $P=.9$) and depression scores ($U=451.50$; $P=.5$) between the experimental and the control groups (See Table 2).

Table 2. Comparison of the average values and ranges for the anxiety and depression variables at the baseline and week 8 between the experimental and control groups.

Variable	Experimental group (n=39)	Control group (n=34)	Statistics	P value
Anxiety at baseline, mean (SD)	15.59 (5.30) ^a	15.35 (5.75) ^b	$t=0.16$.90
Anxiety at week 8, mean (SD)	13.04 (7.12) ^a	16.26 (5.79) ^b	$t=1.74$.09
Depression at baseline, middle range	33.32 ^c	30.55 ^d	$U=451.50$.50
Depression at week 8, middle range	30.58 ^c	33.57 ^d	$U=448.00$.48

^an=27.

^bn=23.

^cn=33.

^dn=30.

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 8 (See Table 2). Regarding the effect sizes, the mean scores for anxiety in the experimental group were lower than for the control group after 8 weeks and the effect size of the intervention was moderate ($d=.5$; 95% CI [-6.96 to .51]). For depressive symptoms, the experimental group reported a lower mean score than the control

group and the effect size of the intervention was nonexistent ($r=.09$).

Within-Group Differences

Within the experimental condition, a statistically significant decrease in the symptoms was observed from the baseline to week 8 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 either in the experimental condition ($Z=1.76$; $P=.08$) or in the control condition ($Z=0.00$; $P>.99$) (See Table 3).

Table 3. Comparison of the average values and ranges within groups for anxiety and depression variables from the baseline to week 8.

Variable and condition	Baseline	Week 8	Statistics	P value
Anxiety, mean (SD)				
Experimental	15.59 (5.30)	13.04 (7.12)	$t=2.15$.04
Control	15.35 (5.75)	16.26 (5.79)	$t=1.00$.33
Depression, middle range				
Experimental	8.83	7.14	$Z=1.76$.08
Control	6.50	6.50	$Z=0.00$.99

Discussion

Important Findings

The use of chatbots (ie, 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 the best of our knowledge, there are no studies on the use of chatbots for mental health in Latin America. This trial was intended to evaluate an AI-based chatbot (Tess) in a sample comprising Argentinian college students. The specific objectives were as follows: (1) understand the participant flow from recruitment to follow-up; (2) report aspects related to the usage patterns of Tess, such as the number of

messages sent and exchanged; (3) examine participant feedback; and (4) compare the preliminary measures of depression and anxiety.

Regarding the usage patterns of Tess, there are three findings that support a satisfactory level of engagement. First, a considerable number of participants in the experimental (39/99, 39%) and control (34/82, 41%) conditions remained in the study throughout the 8-week study period. The completion rates found in the current study are better than that observed in most unpaid and unsupported Internet-based interventions for depression and anxiety, where 90% of the users withdraw after the first two sessions [25]. Furthermore, in studies using mobile apps, the follow-up completion rates were comparable (53%); the mean percentage of complete “adherers” was 36% for depression

and 41% for anxiety [26]. When compared to a chatbot study for US college students, a lower attrition rate was reported (31% and 9% in the control and experimental conditions, respectively); however, this study compensated participants and the follow-up was at 2 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 number of messages sent from the user to Tess was 116 (SD 73.87). A previous study on the usage patterns of the depression modules of Tess showed a much lower average number (17.57) of messages sent to Tess by adult users [27]. It is possible that college, younger, and Latinx students are more willing to engage in conversations with chatbots than older populations in the United States. Two previous studies involving college students in the United States did not report the number of messages sent by the user to Tess [6,7]. Regarding the messages exchanged, Fulmer et al [7] reported a comparable number of total messages exchanged during a period of 4 weeks (M 286; SD 104.6), whereas the total number of messages exchanged in the current study was during a period of 8 weeks.

Third, feedback provided by those in the experimental condition was mostly positive (eg, *Yes, you really understand me, Tess. Thanks for talking to me.*). Among the participants offering negative feedback, there was a predominant dissatisfaction regarding the accuracy of some interventions (eg, *Sometimes I ask you something and you don't specifically respond to what I asked.*). Feedback is a key component for AI-based chatbots as it allows systems to tailor the dialogues to the user. Interestingly, the positive and negative feedbacks 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 could lead to better customization the intervention messages. This finding is relevant as it supports the need to collect user feedback for achieving optimal levels of customization and increasing engagement that could lead to higher intervention doses.

Regarding the impact of Tess on anxiety and depressive 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 8 weeks of intervention and a near-significant trend ($P=0.07$) for depressive symptoms. Analyzing the effect sizes showed that Tess had a moderate effect on anxiety and no effect on depression in the experimental group. These outcomes were unexpected given that previous studies using Tess [7] and another conversational agent called Woebot [6] reported significant reductions in anxiety and depressive symptoms; both studies used a similar control group (a psychoeducation book). Moreover, in the current study, depression was measured using PHQ-9 as a categorical and ordinal variable, whereas Fulmer 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, although the findings of the current study were not statistically significant, the direction of the change observed for anxiety and depression was as expected; therefore, it is possible that low-intensity interventions delivered via chatbots may require a higher dose to yield a between-group effect when delivered to Argentinian students. Third, Tess provides many conversations based on different theoretical approaches, and this may have resulted in less therapeutic power. However, Fulmer et al [7] observed significant effect using similar conversations. Fourth, it is possible that during 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 conducted with intervention completers; therefore, future studies with larger samples (including completer and intent-to-treat analyses) are needed. Second, only college students from a specific region in Argentina were included in this study, and the socioeconomic aspects of the sample were not assessed; thus, the inclusion of a more diversified sample is suggested. Third, there was a high dropout rate throughout the 8-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 the limited capacity of most digital interventions to capture the attention and motivation of users. 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, as most participants who remained until completion of the study were female, male participation was scarce. Fifth, the control group had access to a psychoeducation book, and there was no information on whether they read it. As chatbot research is in its initial stages, further studies could benefit from offering waitlists rather than self-help books. Although offering a waitlist could present an ethical dilemma, this would mitigate the potential effects of not having an intervention if short-term studies are conducted.

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

Conclusions

Students spent a considerable amount of time exchanging messages with Tess and positive feedback was associated with higher numbers of messages exchanges. The initial results showed preliminary evidence regarding the effectiveness of Tess in addressing anxiety symptoms, but there was no

significant effect on 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 health care access, developing affordable strategies such as chatbots may become effective tools to address these needs. AI-based chatbots have the ability to reach higher levels of customization and may thus be of service to educational and mental health care centers aiming to deliver interventions to

targeted users that are accessible at any time without geographical restrictions. Additionally, chatbots may be used as standalone resources for those who have no access to treatment or as a complement to traditional treatments. Although 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.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (v 1.6.1).

[PDF File (Adobe PDF File), 1170 KB-Multimedia Appendix 1]

References

1. Cía AH, Stagnaro JC, Aguilar Gaxiola S, Vommaro H, Loera G, Medina-Mora ME, et al. Lifetime prevalence and age-of-onset of mental disorders in adults from the Argentinean study of mental health epidemiology. *Soc Psychiatry Psychiatr Epidemiol* 2018 Feb;53(4):341-350. [doi: [10.1007/s00127-018-1492-3](https://doi.org/10.1007/s00127-018-1492-3)]
2. Framework for the implementation of the regional strategy on mental health 2011. Pan American Health Organization. Washington, D.C: PAHO; 2011. URL: <https://iris.paho.org/bitstream/handle/10665.2/3535/PAHO%20framework%20implem%20reg%20strat%20on%20mh%202011.pdf?sequence=1&isAllowed=y> [accessed 2019-01-20]
3. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry* 2013 Jul;35(4):332-338 [FREE Full text] [doi: [10.1016/j.genhosppsych.2013.03.008](https://doi.org/10.1016/j.genhosppsych.2013.03.008)] [Medline: [23664503](https://pubmed.ncbi.nlm.nih.gov/23664503/)]
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 Aug;1-13. [doi: [10.1159/000501812](https://doi.org/10.1159/000501812)]
5. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. *JMIR Ment Health* 2019 Oct;6(10):e14166 [FREE Full text] [doi: [10.2196/14166](https://doi.org/10.2196/14166)] [Medline: [31628789](https://pubmed.ncbi.nlm.nih.gov/31628789/)]
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 Ment Health* 2017 Jun;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](https://doi.org/10.2196/mental.7785)] [Medline: [28588005](https://pubmed.ncbi.nlm.nih.gov/28588005/)]
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 Ment Health* 2018 Dec;5(4):e64 [FREE Full text] [doi: [10.2196/mental.9782](https://doi.org/10.2196/mental.9782)] [Medline: [30545815](https://pubmed.ncbi.nlm.nih.gov/30545815/)]
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. *Behav Cogn Psychother* 2014 Nov;42(6):731-746. [doi: [10.1017/S135246581300060X](https://doi.org/10.1017/S135246581300060X)] [Medline: [23899405](https://pubmed.ncbi.nlm.nih.gov/23899405/)]
9. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interv* 2017 Dec;10:39-46 [FREE Full text] [doi: [10.1016/j.invent.2017.10.002](https://doi.org/10.1016/j.invent.2017.10.002)] [Medline: [30135751](https://pubmed.ncbi.nlm.nih.gov/30135751/)]
10. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Transl Behav Med* 2019 May;9(3):440-447. [doi: [10.1093/tbm/ibz043](https://doi.org/10.1093/tbm/ibz043)] [Medline: [31094445](https://pubmed.ncbi.nlm.nih.gov/31094445/)]
11. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth* 2018 Nov;6(11):e12106 [FREE Full text] [doi: [10.2196/12106](https://doi.org/10.2196/12106)] [Medline: [30470676](https://pubmed.ncbi.nlm.nih.gov/30470676/)]
12. Beck J. 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? *Behav Cognit Psychother* 2009 Jun;23(04):325. [doi: [10.1017/S135246580001643X](https://doi.org/10.1017/S135246580001643X)]
16. Depression in College Students. National Institute of Mental Health. 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 JBW. The PHQ-9. *J Gen Intern Med* 2001 Sep;16(9):606-613. [doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)]
18. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, et al. Validation and calibration of the patient health questionnaire (PHQ-9) in Argentina. *BMC Psychiatry* 2019 Sep 18;19(1):291-310 [FREE Full text] [doi: [10.1186/s12888-019-2262-9](https://doi.org/10.1186/s12888-019-2262-9)] [Medline: [31533674](https://pubmed.ncbi.nlm.nih.gov/31533674/)]
19. Newman MG, Zuellig AR, Kachin KE, Constantino MJ, Przeworski A, Erickson T, et al. Preliminary reliability and validity of the generalized anxiety disorder questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy* 2002 Mar;33(2):215-233. [doi: [10.1016/S0005-7894\(02\)80026-0](https://doi.org/10.1016/S0005-7894(02)80026-0)]
20. Rodríguez de Behrends M. 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. 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. 2015 Nov. URL: http://www.psi.uba.ar/investigaciones/eventos_cientificos/vii_congreso_memorias/6_tecn_proces_eval.pdf [accessed 2019-01-20]
21. IBM SPSS Statistics for Windows, Version 22. 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. *Front Psychol* 2018 Apr;9:556. [doi: [10.3389/fpsyg.2018.00556](https://doi.org/10.3389/fpsyg.2018.00556)]
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
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;1(21):19-25 [FREE Full text]
25. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. *PLoS One* 2013 Jul;8(7):e62873 [FREE Full text] [doi: [10.1371/journal.pone.0062873](https://doi.org/10.1371/journal.pone.0062873)] [Medline: [23843932](https://pubmed.ncbi.nlm.nih.gov/23843932/)]
26. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. *J Consult Clin Psychol* 2020 Jan;88(1):1-13. [doi: [10.1037/ccp0000459](https://doi.org/10.1037/ccp0000459)] [Medline: [31697093](https://pubmed.ncbi.nlm.nih.gov/31697093/)]
27. Dosovitsky G, Pineda B, Jacobson NC, Chang C, Escoredo M, Bunge EL. Artificial intelligence chatbot for depression: descriptive study of usage. *JMIR Form Res* 2020 Nov;4(11):e17065 [FREE Full text] [doi: [10.2196/17065](https://doi.org/10.2196/17065)] [Medline: [33185563](https://pubmed.ncbi.nlm.nih.gov/33185563/)]
28. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]

Abbreviations

AI: artificial intelligence
BITs: behavioral intervention technologies
CBT: cognitive behavioral therapy
DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders-IV*
GAD-7: Generalized Anxiety Disorder Scale-7
MYLO: Manage Your Life Online
PAHO: Pan American Health Organization
PHQ-9: Patient Health Questionnaire-9

Edited by G Eysenbach; submitted 07.07.20; peer-reviewed by E Bendig, M Rahmadiana; comments to author 17.08.20; revised version received 11.11.20; accepted 29.05.21; published 12.08.21

Please cite as:

Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL
 Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial
JMIR Form Res 2021;5(8):e20678
 URL: <https://formative.jmir.org/2021/8/e20678>
 doi: [10.2196/20678](https://doi.org/10.2196/20678)
 PMID: [34092548](https://pubmed.ncbi.nlm.nih.gov/34092548/)

©Maria Carolina Klos, Milagros Escoredo, Angela Joerin, Viviana Noemí Lemos, Michiel Rauws, Eduardo L Bunge. Originally published in JMIR Formative Research (<https://formative.jmir.org/>), 12.08.2021. This is an open-access article distributed under

the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.