Title: Evaluating an Online, Single-Session Intervention Targeting Self-Injurious Behavior in Adolescents

NCT: NCT04498143

Unique Protocol ID: University of Denver Institutional Review Board #1505797-3

Registered via Open Science Framework: 8-13-2020, https://osf.io/x5cd9

Evaluating an Online, Single-Session Intervention Targeting Self-Injurious Behavior in Adolescents

Data Collection Status: Have any data been collected for this study already? Note: 'Yes' is a discouraged answer for this pre-registration form.

No data has been collected already for this study.

What's the main question being asked or hypothesis being tested in this study?

To determine whether participants randomized to an online single session intervention (SSI) targeting non-suicidal self-injurious behavior ("Project SAVE") endorse:

1. Lower self-reported *likelihood* of future non-suicidal self injury immediately post-intervention (secondary outcome)
2. Lower self-reported *frequency* of suicidal ideation (in the past 3 months) at 3 month follow-up (secondary outcome)
3. Lower self-reported *frequency* of non-suicidal self-injury (in the past 3 months) at 3 month follow-up (primary outcome)

relative to those randomized to an active control condition (i.e. supportive therapy SSI; Schleider & Weisz, 2018) in a large sample of cisgender heterosexual & sexual and gender minority teenagers ages 13-16 years old.

Describe the key dependent variable(s) specifying how they will be measured.

Our primary and secondary dependent variables will be measured using the self-report version of the Self-Injurious Thoughts and Behaviors Interview-R (SITBI-R; Fox, Harris, Wang, Millner, Deming, & Nock, 2020). Items in this measure ask participants about their attitudes/beliefs/history concerning various self-injurious thoughts and behaviors, including suicidal ideation (SI) and non-suicidal self-injury (NSSI). While our participants will complete all questions from the SITBI-R, the present study evaluates items that measure past 3-month frequency of NSSI, past 3-month frequency of SI, likelihood of future NSSI (see below section).

https://shirleywang.rbind.io/papers/Fox_psychassess_2020.pdf

Primary Dependent Variable

NSSI Frequency in the Past 3 Months. Our primary outcome variable will compare past 3-month NSSI frequency at 3-month follow-up—for participants assigned to the Project SAVE SSI versus supportive therapy SSI (i.e. control condition). At pre-intervention and 3-month follow-up time points, participants will indicate "how many times they have purposely hurt themselves without wanting to die" in the past 3 months in an open-response text box. Notably, in the present study's survey flow, this question is displayed following display logic—such that

participants endorsing zero NSSI in the past 3-months (via a previous question) skip to the end of the question block and are not prompted to answer our main frequency outcome question. For these individuals, we will impute a past 3-month frequency value equal to 0 (as they have endorsed zero NSSI in the previous 3 months in the earlier question). More information about how we will handle very large numbers (i.e. outliers) for this outcome is reported below (see "Outliers and Exclusions" section; "Winsorizing Outliers" subsection).

Secondary Dependent Variables

Likelihood of Future NSSI. Likelihood of future NSSI will be measured immediately pre- and post-intervention. We will evaluate self-reported likelihood of future non-suicidal self-injury immediately post-intervention—comparing participants in the Project SAVE SSI versus supportive therapy SSI control conditions, and controlling for pre-intervention likelihood score. An item from the Self-Injurious Thoughts and Behaviors Interview-R (Fox et al., 2020) will assess participants' beliefs about the likelihood of future non-suicidal self injury on a 5-point Likert Scale (range = 0 to 4; 0 = "not at all"; 4 = "extremely").

SI Frequency in the Past 3 Months. We will compare past 3-month SI frequency at 3-month follow-up—for participants assigned to the Project SAVE SSI versus supportive therapy SSI (i.e. control condition). At pre-intervention and 3-month follow-up time points, participants will indicate "how many days did they have thoughts about killing themselves for more than a few minutes" in the past 3 months in an open-response text box. Notably, in the present study's survey flow, this question is displayed following display logic—such that participants endorsing zero history of SI in the past 3 months (via a previous question) skip to the end of the question block and are not prompted to answer our frequency outcome question. For these individuals, we will impute a past 3-month frequency value equal to 0 (as they have endorsed zero SI in the previous 3 months in the earlier question). More information about how we will handle very large numbers (i.e. outliers) for this outcome is reported below (see "Outliers and Exclusions" section; "Winsorizing Outliers" subsection).

All Other Variables

Demographics. Immediately pre-intervention, participants will be asked to report on the following demographic variables: age, sex assigned at birth, gender identity, race/ethnicity, sexual/romantic attraction and orientation, mental health treatment history, and zip code.

Perceived Socioeconomic and Social Status. Immediately pre-intervention, participants will be asked to rate their perceived socioeconomic and social status using the two items from the MacArthur Scale of Subjective Social Status-Youth Version (Goodman et al., 2001). Respondents indicate where they see themselves on a ladder with 10 rungs (range: 1 to 10 for both items, where 1 = families with most money/education/jobs and youth with highest respect/grades/social standing; 10 = families with least money/education/jobs and youth with lowest respect/grades/social standing).

https://doi.org/10.1542/peds.108.2.e31

Pubertal Development. Pubertal development will be measured pre-intervention via the the Pubertal Development Scale (Carskadon & Acebo, 1993). This scale asks participants to rate 5 items relating to their physical development milestones (e.g. “body hair growth”) on a 5-point Likert scale (1 = “not yet started”; 2 = “barely started”; 3 = “definitely started”; 4 = “seems complete”; 5 = “I don’t know/Not applicable”). Excluding “5” responses, higher scores generally indicate being further along in pubertal development.

https://www.jahonline.org/article/1054-139X(93)90004-9/pdf

Discrimination. The Expanded Everyday Discrimination Scale will be measured pre-intervention to gauge the levels of relatively minor, every-day, chronic discrimination experienced by participants. The original scale (Williams et al., 1997) asks participants, “In your day-to-day life, how often do any of the following things happen to you?” for 9 items (e.g. “you are treated with less courtesy than other people are”) on a 5-point Likert scale (1= “never”; 6 = “almost every day”). These items assess the participant’s observations about how others treat and act around them on a daily basis. The expanded version of this scale includes a 10th item, “You are followed around in stores.” Scores on the Expanded Everyday Discrimination Scale range from 10-60, with higher scores indicating higher levels of chronic discrimination experienced by participants. Respondents are also asked to identify what they believe to be the main reason(s) for these experiences (e.g. gender, race, age, etc.).

https://scholar.harvard.edu/files/davidrwilliams/files/measuring_discrimination_resource_june_2016.pdf

Self Hate. Self-hate will be assessed immediately pre-intervention, post-intervention, and follow-up using the Self Hate Scale (Turnell, Fassnacht, Batterham, Calear, & Kyrios, 2019), which is measured on a 7- point Likert scale. The Self Hate Scale asks participants to “Rate how true each of the following statements are for you right now, in this moment” for 7 items (e.g. “I hate myself”; 1 = “not at all true for me”; 7 = “very true for me”). Scores on the Self Hate Scale range from 7-49, with higher scores indicating higher levels of self hate.

https://www.sciencedirect.com/science/article/abs/pii/S0165032718313144

Disordered Eating. Disordered eating behaviors will be measured at pre-intervention and 3-month follow-up using The Dietary Restriction Screener (DRS-2; Haynos & Fruzetti, 2015). The DRS-2 is a 9-item measure evaluating restrictive eating, bingeing, and purging behaviors in participants. 6 items ask participants whether or not they have engaged in restrictive eating, bingeing, or purging behaviors in the past year or in the past 3 months (0 = no; 1 = yes). The other 3 items assess the frequency of these behaviors over the past 28 days.

https://link.springer.com/article/10.1007/s40519-014-0161-0

SSI Feedback. Acceptability and participant perception of each SSI will be assessed using the Program Feedback Scale (PFS). This scale was adapted for web-based SSIs (Schleider, Mullarkey, & Weisz, 2019) and is based on validated acceptability measures used for digital interventions. The PFS asks participants to "Please tell us how much you agree with each statement below" for 7 statements (e.g. "the activity was easy to use") using a 5-point Likert scale (1 = "really disagree"; 5= "really agree"). For these 7 items, scores can range from 7 to 35, with higher scores indicating greater acceptability of the SSI (how much participants enjoyed/ understood/ would recommend/ could easily use the SSI). The PFS also assesses qualitative participant feedback with 3 open-ended questions ("What did you like about the program?", "What would you change about the program?", and "Is there anything else you'd like to share with us about the program?"). In keeping with other web-based intervention research among youth, we will define a mean score of 3.5 or higher on the PFS (across all 7 items, 1-5 scale) as "acceptable". We will also report means and standard deviations for each individual item of the PFS (see https://psyarxiv.com/tdrpc/).

https://www.researchprotocols.org/2019/7/e13368/#box1

How many and which conditions will participants be assigned to? (optional)

Participants will be randomly assigned to 1 of 2 possible conditions, per a 1:1 allocation ratio: (1) an online, active control group program encouraging feelings disclosure (i.e. supportive therapy SSI), or (2) an online program targeting behaviors and thoughts relating to self-injury (i.e. Project "SAVE"—Stop Adolescent Violence Everywhere—SSI). Random assignment to condition will be conducted in a triple-masked manner. Therefore, participants will be masked to whether they received active treatment and investigators will be masked to which condition the participant is randomized to—as the randomization occurs automatically within the survey. Additionally, as the primary and secondary outcomes are self-report measures, we will not be utilizing an outcomes assessor (e.g. a clinical interviewer). Participants will be informed of their condition assignment following their completion of the full study protocol by debriefing form. Investigators will be informed of participant condition after data collection is complete (i.e. the data analysis phase).

Supportive Therapy ("Share Your Feelings") SSI (Schleider & Weisz, 2019; see https://osf.io/u4axs/) is a ~30-minute, self-administered, web-based program that uses components of supportive therapy to encourage feelings sharing. Like Project SAVE, this SSI takes approximately 30 minutes to complete and is self-administered online. The supportive therapy SSI encourages participants in the control group to identify and express their feelings by (1) explaining why sharing feelings is natural, important, and helpful and (2) including testimonials from teens who have shared their feelings with close others.

Project SAVE ("Stop Adolescent Violence Everywhere") SSI is a ~30-minute, self-administered, web-based program that uses components of cognitive behavior therapy and dialectical behavior therapy designed to decrease self-injurious behaviors in youth. The SSI's content is designed to maximize relevance for youth engaging in self-injurious behaviors by

directly addressing two commonly reported functions of self-injury in youth: using self-injury for self-punishment and/or for emotion regulation. The Project SAVE SSI has 4 general content sections: (1) explaining the science behind how changing your *actions* (i.e. decreasing self-injurious behaviors) can positively impact your *thoughts* and *emotions* over time; (2) providing scientific evidence and testimonials from other teens that have successfully decreased their self-injurious behaviors and noticed positive change as a result; (3) evidence-based tips for overcoming common obstacles to decreasing self-injurious behaviors in day to day life; and (4) offering an opportunity for youth to share their own thoughts and advice on what they have learned with other teenagers who are facing similar challenges. These content areas are reinforced by intervention design features intended to maximize persuasiveness and memorability, in line with other efficacious, web-based SSIs for adolescents.

Specify exactly which analyses you will conduct to examine the main question/hypothesis. (optional)

Testing for Differential Dropout

First, we will test for differential dropout by study condition. To do this, we will use Z-tests of differential proportions where we compare the proportion of people who drop out before completing the study (Y or N) as a function of treatment condition (0 = control group SSI, 1 = Project SAVE SSI). We will run two tests evaluating differential dropout: (1) measuring dropout prior to finishing our post-intervention questions, and (2) measuring dropout prior to finishing our 3-month follow-up questions. To be considered a study "completer" for the post-intervention questions, a participant must have made it through all post-intervention survey sections (i.e. all sections but "final block" section that presents information for study logistics). To be considered a study "completer" for the 3-month follow-up questions, a participant must have made it through all follow-up survey sections (i.e. all sections but "final block" section that presents information for study logistics).

If the *p* value is greater than .05 we will assume dropout was <u>not</u> dependent on condition assignment and can proceed with interpreting the effects of intervention assignment (i.e. proceed with the between-group analyses exactly as described below, starting in the section labeled "Hypothesis 1").

If the *p* value for this test is less than .05 we will conclude that dropout <u>was</u> dependent on condition assignment, and we will take the following steps to improve the interpretability of our results. Unequal dropout between treatment groups can introduce bias in trial results; however, this is not *always* the case (Bell, Kenward, Fairclough, & Horton, 2013; https://doi.org/10.1136/bmj.e8668). One major concern is that, if data are *not missing at random*, systematic differences may exist between "dropouts" and "completers" that cannot be corrected for, limiting the interpretability of our results. For example, if more control condition participants experienced a worsening of clinical symptoms and dropped out before completing follow-up measures, this worsening of symptoms would not be captured by completers' data (or by data imputed based on the available completers' data). Thus, should unequal dropout exist

between conditions in the present study, we will report the results of our main analyses below, highlighting (1) limitations to interpretation, and (2) ways in which this will inform future intervention development.

Further, if unequal dropout is present, we will run a sensitivity analysis to determine a range of possible effect sizes for our outcomes of interest, varying how much "dropout participants" may have changed on each outcome. To establish lower- and upper-end estimates of effect size, we will: (1) calculate residual change for each of our 3 outcomes, (2) identify the 25th and 75th percentile values for residual change in each outcome, (3) run analyses for each outcome where we impute the lower residual change value (25th percentile) for all dropouts/missing data, as well as the higher residual change value (75th percentile) for all dropouts/missing data. This approach explicitly assumes data are not missing at random, providing a more unbiased estimate of the overall treatment effect in the presence of unequal dropout.

Testing Hypothesis 1: *Do teenagers in the Project SAVE SSI condition report lower self-reported likelihood of future NSSI immediately post-intervention—relative to teenagers in the active control program?*

We will first test if the assumptions necessary to interpret a multiple linear regression are met. If not, we will follow the procedures below. We will document any of these changes when we report our results.

- If assumptions for interpreting a multiple linear regression *are not* met, we will perform a two-way ordinal regression (as this outcome will be Likert data), with pre-intervention likelihood of future NSSI and treatment condition entered as our independent variables, and post-intervention likelihood of future NSSI as the dependent variable. We will consider a p value of less than .05 for treatment condition (in favor of the Project SAVE SSI) as a significant effect of the Project SAVE treatment on likelihood of future NSSI.

- If the assumptions for interpreting a multiple linear regression *are* met, we will enter pre-intervention likelihood of future NSSI as a covariate and treatment condition as the predictor of post-intervention likelihood of future NSSI in a multiple linear regression. We will consider a p value of less than .05 for the treatment condition coefficient (in favor of the Project SAVE SSI) as a significant effect of the Project SAVE treatment on likelihood of future NSSI.

Testing Hypothesis 2: *Do teenagers in the Project SAVE SSI condition endorse lower frequency of SI in the past 3 months—relative to teenagers in the active control program?*

First, we create our secondary outcome variable (i.e. frequency of SI in the past 3 months) by imputing values equal to 0 for participants who endorsed zero SI in the previous 3 months in an earlier question. See "SI Frequency in the Past 3 Months" under the key dependent variables section for more details.

Next, we will determine whether a Poisson regression, a negative binomial regression, a zero-inflated Poisson regression, or a zero-inflated negative binomial regression is most appropriate for analyzing our past 3-month frequency outcome (count variable). To do this, we will run all four models, conducting model fit tests to compare AIC values, BIC values, and raw scores for each regression. If all three indices agree (i.e. select the same "best" model of the four possible), we will interpret the results of that regression. If they do not agree (i.e. select different "best" models from the four possible), we will report the results from each of the selected models.

Regardless of model type, we will enter pre-intervention past 3-month frequency of SI as a covariate, and treatment condition as a predictor, of past 3-month frequency of SI at follow-up in the chosen regression(s). We will consider a *p* value of less than .05 for the treatment condition coefficient (in favor of the Project SAVE SSI) as a significant effect of the Project SAVE treatment on frequency of SI in the past 3-months.

Testing Hypothesis 3: *Do teenagers in the Project SAVE SSI condition endorse lower frequency of NSSI in the past 3 months—relative to teenagers in the active control program?*

We will follow the same procedures for hypothesis 2 but replacing "SI frequency in the past 3 months" with "NSSI frequency in the past 3 months."

Correcting for Multiple Comparisons

We will use a false discovery rate approach to correct for multiple comparisons among our main confirmatory tests of hypotheses 1 - 3. As such, in all places where a significant effect is defined by a *p* value less than .05, this is referring to the post-correction *p* value (i.e. the *p* value after correction using the false discovery rate method).

Outliers and Exclusions. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations. (optional)

Type 1 Exclusions

Type 1 exclusions are exclusions that can be automated and do not require consensus among investigators. A research team member who will not be involved with conducting the primary study analyses will use type 1 exclusions to determine when participant recruitment is complete. Recruitment for the present study will end once 500 participants have been randomized *who pass all type 1 exclusions tests (i.e. are still eligible for inclusion after excluding for type 1 criteria*; see our section on sample size for more details).

Here, we will exclude participants that fail to meet the following study inclusion criteria:

- Must be 13 - 16 years old at the time of the baseline survey
- Must report comfort reading and writing in English

- Must endorse no learning disability, visual impairment, or other difficulty that makes it difficult to answer questions on a computer
- Must have access to a laptop or smartphone with internet access
- Must endorse purposefully hurting self without wanting to die within the past month
- Must endorse *either* "I do not like myself" or "I hate myself" on CDI-II item 7 in a screener survey

Here, we will exclude the following participants based on failure to meet the following quality check criteria:

- Participants who exit the study *prior* to condition randomization for our listed analyses (note: the present study will use an intention-to-treat approach, where every participant who was randomized to a study condition—and who is not excluded for one of the specific reasons outlined in this section—will be included our our analyses)

Type 2 Exclusions

Type 2 exclusions are exclusions that cannot be automated and require consensus among investigators. Participants excluded for type 2 reasons will be subtracted from the N = 500 total individuals who were randomized and met criteria for inclusion following type 1 exclusions.

Here, we will exclude the following participants based on failure to meet the following quality check criteria:

- Participants who respond with either copy/pasted responses from text earlier in the intervention (e.g. copy and pasting only text from a previous testimonial slide) to any of free response questions
- Participants demonstrating an obvious lack of English fluency in open response questions; these participants will be identified by consensus from all study team members
- Participants responding with random text in open response questions; these participants will be identified by consensus from all study team members
- Duplicate responses from the same individual in baseline or follow-up surveys (i.e. more than one response with an identical IP address). Where duplicate responses for the same individual are present in the same survey (e.g. more than 1 response for a single individual at follow-up), we will exclude the response that is less complete, retaining the more complete of the two responses. If both responses are 100% complete, we will retain the first of the two responses for that survey. Notably, if an individual completes the baseline survey more than once—*and happens to be randomized to both conditions*—we will exclude this individual's responses from our analysis altogether.
- We will also exclude for primary analyses (but may run sensitivity analyses including them) any participants who provide responses of 3 words or fewer to writing prompts that ask for at least 2 sentences or more.

These exclusions are based on previous single session intervention research conducted online

(for one example, see Schleider & Weisz, 2018: https://doi.org/10.1080/15374416.2017.1405353)

Imputing Data

After excluding individuals based on the above criteria, we will impute any missing data in our variables of interest (pre- and post-intervention likelihood of future NSSI; pre- and post-intervention desire for future NSSI; past 3-month frequency of NSSI at pre-intervention and follow-up). To do this, we will execute an expectation-maximization and bootstrapping algorithm via the Amelia II package in R (Honaker, King, & Blackwell, 2011; https://cran.r-project.org/web/packages/Amelia/Amelia.pdf). All variables from the corresponding statistical model (i.e. hypothesis tests 1 - 3) will be included in the algorithm used to create the imputed data for that test: (a) baseline values for the outcome—likelihood of NSSI, desire for NSSI, or past 3-month frequency of NSSI, (b) follow-up values for the outcome, and (c) condition assignment, as well as (d) other baseline variables that are likely associated with each outcome variable (for **past 3-month frequency of NSSI at follow-up**: lifetime history of NSSI (y/n), past-month frequency of NSSI, past-year frequency of NSSI, and lifetime frequency of NSSI at baseline; for **likelihood of NSSI post-intervention**: all of these same variables plus past 3-month frequency of NSSI at baseline; for **past 3-month frequency of SI at follow-up**: lifetime history of SI (y/n), past-month frequency of SI, past 3-month frequency of SI, past-year frequency of SI, and lifetime frequency of SI at baseline). Together, these variables (a - d) will be used to create the imputed datasets for each of our three hypotheses.

The number of imputed datasets we create will equal the percentage of missing data for each of our three outcome variables (post-intervention likelihood of NSSI, post-intervention desire for future NSSI, past 3-month frequency at follow-up), rounded up to the next integer value. For example, we will create 7 imputed datasets to test hypothesis 1 if 6.2% of the data for the post-intervention likelihood of NSSI outcome variable is missing. We will create 8 imputed datasets to test hypothesis 2 if 7.6% of the data for the post-intervention desire for future NSSI outcome variable is missing.

Lastly, we will run all of our confirmatory tests for hypothesis 1 - 3 as outlined in this pre-registration, within the original study data (i.e. removing/disregarding missing data in the original dataset and ultimately using non-imputed data). We will report the results in our supplementary materials for the present study. If main results/interpretations differ between analyses run within imputed vs. non-imputed data, we will report and discuss our paper's discussion.

Winsorizing Outliers

We will winsorize outliers in our two count variables of interest (past 3-month frequency of NSSI at pre-intervention and follow-up) before running tests for hypothesis 2 and 3. Outliers will be defined as values above the 95th percentile for the variable in question, calculated using the Winsorize() function in the DescTools package in R (Singnorell et al., 2020, https://cran.r-

project.org/web/packages/DescTools/DescTools.pdf); flagged outliers will be replaced with the 95th percentile value.

How many observations will be collected or what will determine the sample size? No need to justify decision, but be precise about exactly how the number will be determined. (optional)

The final sample size for *collected* data will be N = 500 randomized participants (250 per condition) who meet our type 1 inclusion criteria (i.e. who are not excluded by type 1 exclusion criteria outlined in the above section). The total number of participants who provide *usable* data will then be determined by the number of participants who meet our type 2 inclusion criteria (i.e. who are not excluded by type 2 exclusion criteria outlined above). Therefore, the present study will have a maximum of 500 participants, with a final sample size determined by the number of participants meeting our type 2 inclusion criteria specified above.

Other. Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?) (optional)

Anticipated Sample Demographics

The present study will recruit research participants via online social media advertisements. Based on samples recruited for other studies by members of this research team (using similar recruitment procedures; Fox et al., 2020, https://doi.org/10.1037/ccp0000486; Smith, Fox, Carter, Thoma, & Hooley, in press), we anticipate the following demographic representation among the present study's research participants:

65-75% White
5-15% Asian
5-15% Latinx
5-15% Black

The present investigators will recruit participants with the explicit goal of increasing racial diversity and representation among our research participants (see Roberts et al., 2020; https://doi.org/10.1177%2F1745691620927709). If we fail to meet the anticipated percentages for Asian, Latinx, or Black participants outlined above, we will specifically address lack of racial diversity as a potential limitation of generalizability within the text of the treatment results paper.

Additionally, we plan to recruit approximately 50% sexual or gender minority (SGM)-identifying participants via online advertisements that target LGBTQ+ interests/groups as well as advertisements targeted toward the general teen population.

Treatment Effects Among SSI Completers

An exploratory analysis will examine whether or not treatment assignment is a significant predictor of our primary outcome (i.e. NSSI frequency in the past 3 months at follow-up in SAVE vs. control conditions), *among participants who have completed an SSI*. To do this, we will run the same procedures as above for testing hypothesis 3—among the subsample of participants who completed a full online program (either Project SAVE SSI or the active control group SSI).

Treatment Effects Among SGM-Identifying Youth

An additional exploratory analysis of treatment effects within the specific subsample of SGM-identifying youth will be conducted in future research. Additionally, PFS total mean and item-level means will be reported specifically for these youth. These analyses will not be conducted nor reported in the present paper, and a more detailed pre-registration will be submitted at a later date.

Additional Exploratory Analyses

Additional secondary analyses will:

(1) employ predictive modeling techniques to predict which individuals benefited from, vs. did not benefit from, the Project SAVE intervention program.
(2) evaluate cross-sectional relationships between other variables collected within the SITBI-R and Self-Hate Scale (e.g. desire to stop future NSSI; likelihood of future NSSI; self-hatred).

These analyses will not be conducted nor reported in the present paper, and a more detailed pre-registration will be submitted at a later date.