

Project SOLVE: A School-based Trial of a Universal Single-Session Digital Problem-Solving Intervention for Adolescent Mental Health

Olivia Fitzpatrick, Jessica Schleider, & John Weisz

Hypothesis, Aims, and Analytic Plan
February 20, 2021

Study Aims and Hypotheses

The goals of this study include:

- 1.) Examining whether proximal clinically-relevant outcomes (e.g., hopefulness; hopelessness) are improved immediately following completion of Project SOLVE.
- 2.) Assessing perceived acceptability and usefulness of Project SOLVE immediately following completion.
- 3.) Examining whether trajectories of internalizing symptoms (anxiety and depression) across the baseline to 3-month assessment period are improved following completion of Project SOLVE.

Hypotheses associated with each of these primary aims include:

- 1.) Proximal clinically-relevant outcomes—hopefulness, as measured by the State Hope Scale “Pathways” subscale, and hopelessness, as measured by the [Beck Hopelessness Scale–4 item version](#)—will improve from immediately pre to immediately post Project SOLVE.
- 2.) Adolescents will perceive Project SOLVE as acceptable and useful, indexed by mean ratings of at least 3.0 out of 5.0 across the seven feedback items.
- 3.) Trajectories of anxiety and depression symptoms, as measured by the [Behavior and Feelings Survey](#), will improve across the baseline to 3-month assessment period.

Sample Size Justification

According to *a priori* power analyses conducted in G*Power 3.1, the estimated sample size of 1,000 adolescents will provide sufficient power ($1-\beta$; 80%) for us to detect overall mean group differences, of small (.2; $N=788$), medium (.5; $N=128$), and large (.8; $N=52$) effect sizes, using two-tailed tests with α set at .05.

Analytic Plan

All analyses will be conducted in R. To examine immediate intervention effects, we will conduct a series of linear regression models using the *lm* package. For example, we will test intervention condition as a predictor of post-intervention Beck Hopelessness Scale (BHS) scores, with pre-intervention BHS and Behavior and Feelings Survey (BFS) scores input as covariates to control for baseline levels of hope and mental health symptoms. Additional potential covariates (e.g., demographic variables) will be identified by conducting model fit comparison analyses, which yield Akaike Information Criterion (AIC) metrics that inform model selection (Sakamoto, et al., 1986).

To examine longer-term intervention effects, we will conduct a series of linear mixed-effects regression models using the *lmer* package. For example, we will test intervention condition as a predictor of trajectories of BFS scores across the follow-up assessment period, with pre-intervention

BFS scores input as covariates. We will determine the optimal number of levels in these multilevel models—e.g., two levels (measurement occasions nested within participants) vs. three levels (measurement occasions nested within participants, nested within school sites)—by conducting model fit comparison analyses (Sakamoto, et al., 1986).

We will conduct the proposed analyses with the entire sample. We will also conduct separate analyses for the subsample of students with elevated internalizing symptoms—operationalized as 1 standard deviation above the mean, consistent with well-established standardized instruments of youth mental health (e.g., Youth Self Report; Achenbach & Rescorla, 2001)—to assess the potentially different effects of Project SOLVE as a universal versus indicated prevention program.

Effect Size Reporting

We will calculate both within-subject and between-subject effect sizes for intervention effects.

Calculating within-subject effect sizes for intervention effects for continuous outcomes (reflecting pre- to post-intervention change). Using the MOTE R Package, we will compute and report d_{average} ($d.\text{dep.t.avg}$), including the d effect size estimate and lower/upper 95% confidence intervals, and d_z ($d.\text{dep.t.diff}$), including the d effect size estimate, lower/upper 95% confidence intervals, t -value, and p -value. We will also report the Common Language (CL) effect size (see package documentation for details).

Calculating between-subject effect sizes for change in continuous outcomes (change in an outcome from pre- to post-intervention in an intervention group vs. a control group). We will compute and report Cohen's d based on residualized change scores (computed by regressing the post-intervention outcome variable onto the pre-intervention outcome variable) in the intervention versus the comparison group. Residualized change scores reflect the magnitude of change in an outcome of interest from baseline to a given follow-up point (more information here). Using the means and standard deviations from the residualized change scores (computed separately for each intervention condition), we will use the MOTE R Package to compute and report d_s ($d.\text{int.t}$). Reported values will include the d effect size estimate, lower/upper 95% confidence intervals, t statistic, and p -value. We will also report the Common Language (CL) effect size (see package documentation for details).

Calculating between-subject effect sizes for post-intervention proportions across groups (e.g., percent of participants with clinically-elevated symptoms at follow-up in treatment versus control groups). Using the MOTE R Package, we will compute and report d ($d.\text{prop}$), including the d effect size estimate and lower/upper 95% confidence intervals. We will also report the Common Language (CL) effect size (see package documentation for details).

Exploratory Aims

Goals that do not have associated hypotheses and are exploratory in nature include:

- 1.) Qualitatively assessing adolescent problem-solving skills following Project SOLVE via thematic analysis of text-response data collected during end-of-session problem-solving activity (see Thematic Analysis Plan for details).
- 2.) Examining whether trajectories of hopefulness and hopelessness improve across the baseline to 3-month assessment period and across the baseline to 12-month assessment period.

- 3.) Examining whether trajectories of internalizing symptoms (anxiety and depression) improve across the baseline to 12-month assessment period.
- 4.) Examining whether trajectories of externalizing symptoms (misbehavior) improve across the baseline to 3-month assessment period and the baseline to 12-month assessment period.
- 5.) Assessing whether academic record data provided by our school partners (e.g., grades; attendance) improve across the baseline to 3-month assessment period and the baseline 12-month assessment period.
- 6.) Exploring important qualitative and consumer-related data (e.g., the kinds of problems that adolescents identify during the problem-solving activity in Project SOLVE; which adolescents select which “valued others” vignettes).
- 7.) Examining whether perceived control, as measured via a single item (“Right now, I feel like things are out of my control; 0 [not at all] to 10 [a lot] scale), improves immediately pre-intervention to post-intervention.
- 8.) Exploring moderation tests of various baseline variables as moderators of the effectiveness of and adolescent engagement with Project SOLVE.

Thematic Analysis Plan

We will conduct thematic analyses using the guidelines proposed by Braun and Clarke (2006). Specifics steps that we will take to iteratively build a coding manual are outlined below. The number of coders (anticipated 2-3 coders) and the percentage of responses coded at each step depend on the total number of responses received from participants.

Qualitative Iterative Coding Steps:

- 1) Lead coder will randomly select and distribute 10% of qualitative responses to coders.
- 2) Coders independently read through 10% of responses and jot down response patterns they observe. This involves assigning a category/description to each response.
- 3) Coders meet and discuss response patterns ascertained; start developing code list and preliminary definitions/descriptions of codes.
- 4) Lead coder will develop a full codebook (including codes, definitions, and examples of coded text) and randomly select subset of responses and distribute selected responses and codebook to coders.
- 5) Coders use preliminary codebook to independently code new subset of responses.
- 6) All coders meet and discuss coding of subset of responses and reconcile discrepancies through discussion; discuss updates to codebook.

**Steps 5 and 6 may be repeated if discrepancies are too extensive and further calibrating is deemed necessary

- 7) Lead coder updates the codebook, if needed, and sends updated codebook to all coders
- 9) All coders read through updated codebook and share feedback with all coders
- 10) Lead coder finalizes codebook and send final codebook to all coders

Coding for interrater reliability begins:

- 11) Lead coder randomly selects a subset of 20% of the total individual responses (that have not been previously read/coded) and distribute to coders to code independently.
- 12) Coders code all responses on Excel and send final coded material to lead coder. Coding occurs directly in a format that will facilitate quantitative analyses: Each code will be a variable and coders will assign a 1 (if the response endorses that code) or 0 (if the response does not endorse that code)

- 13) A team member calculates interrater reliability for all coders on each code. Acceptable interrater reliability is *kappa* value $> .60$. If acceptable interrater reliability is not established (sometimes happens for one coder on a certain code), lead coder or other coder may meet with this person who is not reliable and discuss and practice.
- 14) Once interrater reliability is established, the lead coder distributes the remaining uncoded responses evenly among coders to code independently.
- 15) Coders independently code their subset of responses.
- 16) Lead coder (or any other coder) assembles a master document with all coded responses.
- 17) Two coders will check accuracy of master document with all coded responses.

The study team will keep a record of meeting discussions, especially when resolving discrepancies in coding, (e.g meeting minutes), at each step. At the end of study completion, the study team will make the coding manual and the qualitative data publicly available.

Missing Data

We will impute any missing data using the expectation-maximization and bootstrapping algorithm implemented with Amelia II in R after running a sensitivity analysis to ensure that data is not missing at random. These imputed datasets allow for more conservative intent-to-treat analyses than listwise deletion or last-observation carried forward. We plan to impute as many datasets as there are percent of missing data for an outcome – rounding up to the next highest percentage (e.g., if 2.4% of data is missing on an outcome, we will create 3 imputed datasets). This process will allow us to retain high power even in the presence of missing data. Cohen's *d* effect sizes and 95% confidence intervals for analyses will be calculated using *t*-values for the treatment effects obtained from the analyses with the MOTE package in R.

False Discovery Rate

The false discovery rate (FDR) will be applied to identify potential false-positive results. *Q*-values will be computed for *p*-values from each model using an online calculator applying Benjamini and Hochberg's approach (1995; www.sdmproject.com/utilities/?show=FDR). Results from tests described above will be considered significant if FDR corrected $q < 0.05$.

Methodological Details Not Mentioned Elsewhere

Intervention and Follow-Up Assessment Completion Period

Intervention and follow-up assessment completion period may vary within and across sites, in response to the fluctuating needs and contexts of school-based partners. Thus, the average number of days elapsed across the completion periods for each component will be reported for both sites.