

SUPPLEMENTARY INFORMATION

In this document, we further describe the analyzed skills and include more details on the results of our statistical analyses.

S.1 Modules and Skills

Supplementary Table 1 includes details on the modules and skills examined.

Module	Skill	Subskill
Mindfulness	Observe ^b	Breathing Sounds Visuals Everyday Life
	Describe ^b	Visuals Thoughts Expressions Everyday Life
	Participate ^b	Counting Jumping Laugh Club Saying Words Walking Everyday Life
	NonJudgementally ^b	Observe Judgments Rephrase Judgmental Statements
	One Mind ^b	–
Emotion Regulation	Check the Facts ^a /Brief CtF	– –
	Opposite Action ^a /Brief OA	– –
	Problem Solve ^a /Brief PS	– –
Distress Tolerance	Distract ^b	Comparisons Emotions Pushing Away Sensations Thoughts
	Self-Soothe ^a	Hearing Smell Taste Touch Vision
	TIP ^b	Intense Exercise Paced Breathing Progressive Muscle Relaxation Temperature

^aSkills with both pre- and post-ratings (with which we could calculate skill improvement).

^bSkills with post-ratings only.

Supplementary Table 1: Modules, skills, and subskills examined in our analyses.

Finding	Estimate	p Value
DT Distract via Pushing Away vs DT Distract via Comparisons	-1.20	0.0227
DT Distract via Sensations vs DT Distract via Comparisons	-1.43	0.0171
M Participate via Saying Words vs DT Distract via Comparisons	-1.04	0.0104
M Non-Judgementally Observe Judgments vs DT Distract via Pushing Away	0.992	0.0307
M Non-Judgementally Rephrase Judgmental Statements vs DT Distract via Pushing Away	1.25	0.0000617
M Non-Judgementally Observe Judgments vs DT Distract via Sensations	1.23	0.0302
M Non-Judgementally Rephrase Judgmental Statements vs Distract via Sensations	1.49	0.000351
M Non-Judgementally Rephrase Judgmental Statements vs Distract via Thoughts	0.901	0.0472
M Participate via Saying Words vs M Non-Judgementally Observe Judgments	-0.836	0.00468
M Focus On One Thing at a Time vs M Non-Judgementally Rephrase Judgmental Statements	-0.710	0.00307
M Describe Visuals vs M Non-Judgementally Rephrase Judgmental Statements	-0.749	0.00234
M Observe Breathing vs M Non-Judgementally Rephrase Judgmental Statements	-0.819	0.00000986
M Observe Sounds vs M Non-Judgementally Rephrase Judgmental Statements	-0.828	0.0000127
M Participate via Saying Words vs M Non-Judgementally Rephrase Judgmental Statements	-1.10	0.000000218

Supplementary Table 2: Skill effectiveness: examining pairwise differences in skills for post-ratings. Positive deltas indicate the first skill results in worse post-ratings than the second skill. M=Mindfulness, ER=Emotion Regulation, DT=Distress Tolerance.

S.2 Overall Skill Use

In our Tukey test examining overall skill improvement between subskills, the *Emotion Regulation* skill of *Problem Solve* resulted in more improvement than the *Distress Tolerance* skills of *Self-Soothe via Hearing* (by 0.64 points; $t=-3.252$, $p<0.05$), *Self-Soothe via Touch* (by 0.90 points; $t=-4.388$, $p<0.01$), and *Self-Soothe via Vision* (by 0.63 points; $t=-3.361$, $p<0.05$). The *Emotion Regulation* skill *Check the Facts* also resulted in more improvement than the *Distress Tolerance* skill *Self-Soothe via Touch* (by 0.79 points; $t=-4.108$, $p<0.01$).

See Supplementary Table 2 for results of our Tukey test examining overall post-skill rating between subskills.

S.3 Skill vs. Scale Improvement

For the PHQ-9 model, our Tukey test examining *age bucket* revealed that being under 25 ($z=3.414$, $p<0.01$) and 25-35 ($z=3.167$, $p<0.01$) years old correlates with more improvement than 35-45 years old. Being under 25 ($z=2.698$, $p<0.05$) and 25-35 ($z=2.869$, $p<0.05$) years old also correlates with more improvement than being greater than 45 years old. Our examination of *family bucket* revealed that having zero family members living close correlates with greater improvement than having 1 or 2 ($z=3.618$, $p<0.01$), 3 or 4 ($z=3.021$, $p<0.05$), or more than 4 ($z=-2.781$, $p<0.05$). Finally, the Tukey results for *favorite module* revealed that preferring the *Addiction* module correlates in more improvement than the *Distress Tolerance* module ($z=3.638$, $p<0.01$), the *Emotion Regulation* module ($z=3.391$, $p<0.01$), and the *Mindfulness* module ($z=3.564$, $p<0.01$).

For the OASIS model, we found that being 25-35 correlates with greater improvement than being 35-45 ($z=2.939$, $p<0.05$) or older than 45 ($z=3.278$, $p<0.01$).

Consistent with our other analyses, these models use the buckets defined in Section 3.2 for the variables of *age*, *number of family members close*, and *education*. If these variables are instead treated as continuous, the resulting models yield no significant effects. This discrepancy may be due to a nonlinear relationship to those variables or due to our relatively limited dataset.

S.4 Models

Supplementary Table 3 presents a list of skill use characteristic features used in predictive modeling. Supplementary Table 4 presents the classifier performance across the four different learning algorithms and various feature sets.

Feature	Description	Type
Skill ID	Unique skill identifier	Categorical
Day of week	Day of week for skill use	Categorical
Time of day	Time of day for skill use grouped by morning (6-12), afternoon (12-18), evening (18-24), and night (0-6).	Categorical
Consecutive use of any skill	Boolean indicating whether or not another skill is used within 15 minutes prior to the current skill use	Categorical
Consecutive use of the same skill	Boolean indicating whether or not the same skill is used within 15 minutes prior to the current skill use	Categorical
Cumulative average skill improvement	Cumulative average of skill improvements since the study intake	Numerical
Last improvement	Skill improvement of the last used skill (0 for the first skill use)	Numerical
Pre-rating	Pre-rating of emotional intensity or distress	Numerical

Supplementary Table 3: Skill use characteristic features used for training predictive models.

Feature set	DT	LR	GB	RF
All	0.721 / 0.628	0.643 / 0.665	0.705 / 0.699	0.659 / 0.632
All but skill	0.62 / 0.457	0.636 / 0.616	0.64 / 0.565	0.663 / 0.648
All but pre-rating	0.721 / 0.628	0.663 / 0.69	0.698 / 0.708	0.667 / 0.657
All but mood disorder	0.721 / 0.628	0.655 / 0.685	0.709 / 0.703	0.64 / 0.621
All but anxiety disorder	0.721 / 0.628	0.64 / 0.666	0.698 / 0.707	0.671 / 0.609
All but education	0.721 / 0.628	0.643 / 0.671	0.671 / 0.703	0.698 / 0.677
All but favorite module	0.721 / 0.628	0.651 / 0.664	0.713 / 0.703	0.655 / 0.672
All but least favorite module	0.721 / 0.628	0.647 / 0.659	0.694 / 0.705	0.647 / 0.662

Supplementary Table 4: Predictive model performances (denoted by “test accuracy / test AUC”) across four different learning algorithms—decision tree (DT), logistic regression (LR), gradient boosting (GB), and random forest (RF)—and various feature sets.