

Supplementary Materials for  
**Randomized evaluation of a school-based, trauma-informed group  
intervention for young women in Chicago**

Monica P. Bhatt *et al.*

Corresponding author: Monica P. Bhatt, [mbhatt@uchicago.edu](mailto:mbhatt@uchicago.edu)

*Sci. Adv.* **9**, eabq2077 (2023)  
DOI: 10.1126/sciadv.abq2077

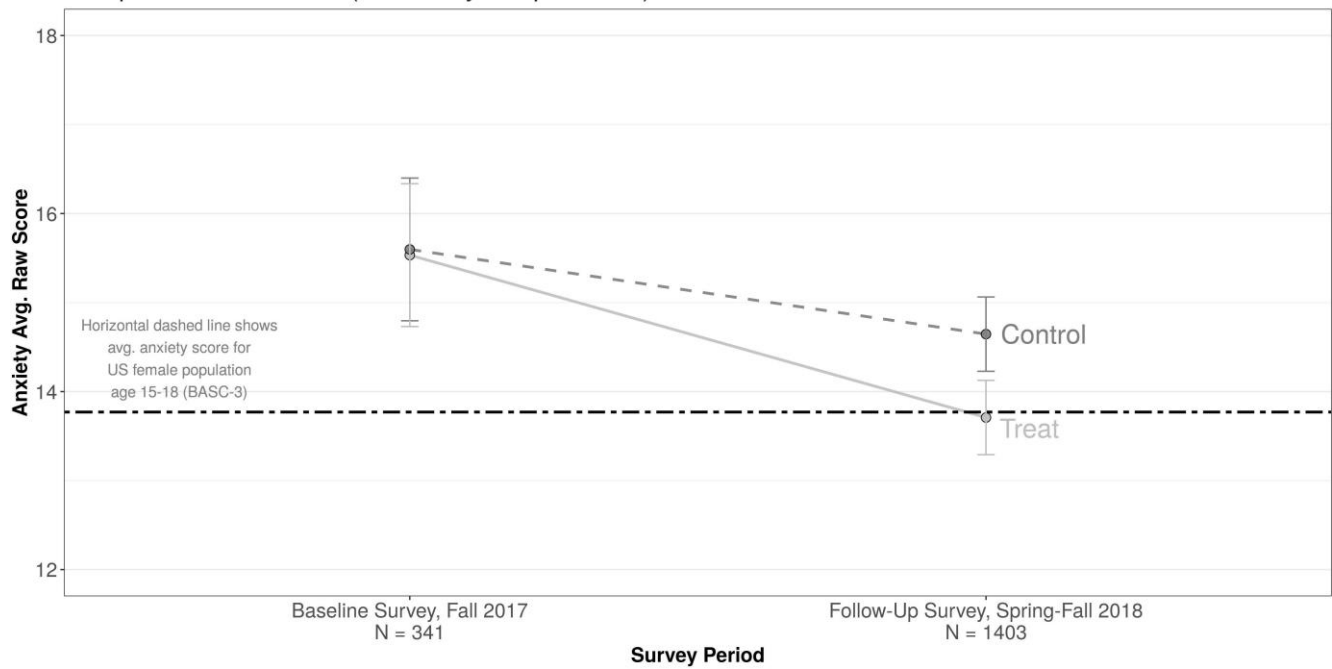
**The PDF file includes:**

Figs. S1 to S12  
Tables S1 to S21  
Legends for data files S1 to S3

**Other Supplementary Material for this manuscript includes the following:**

Data files S1 to S3

Sample: WOW Cohort I (All Survey Respondents)



**Fig. S1: Mental Health Trajectory for Anxiety (Effect of Being Offered WOW) - Unweighted**  
*Note: control means shown in the plot are regression adjusted, for both the baseline & follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level.*

Sample: WOW Cohort I (All Survey Respondents)

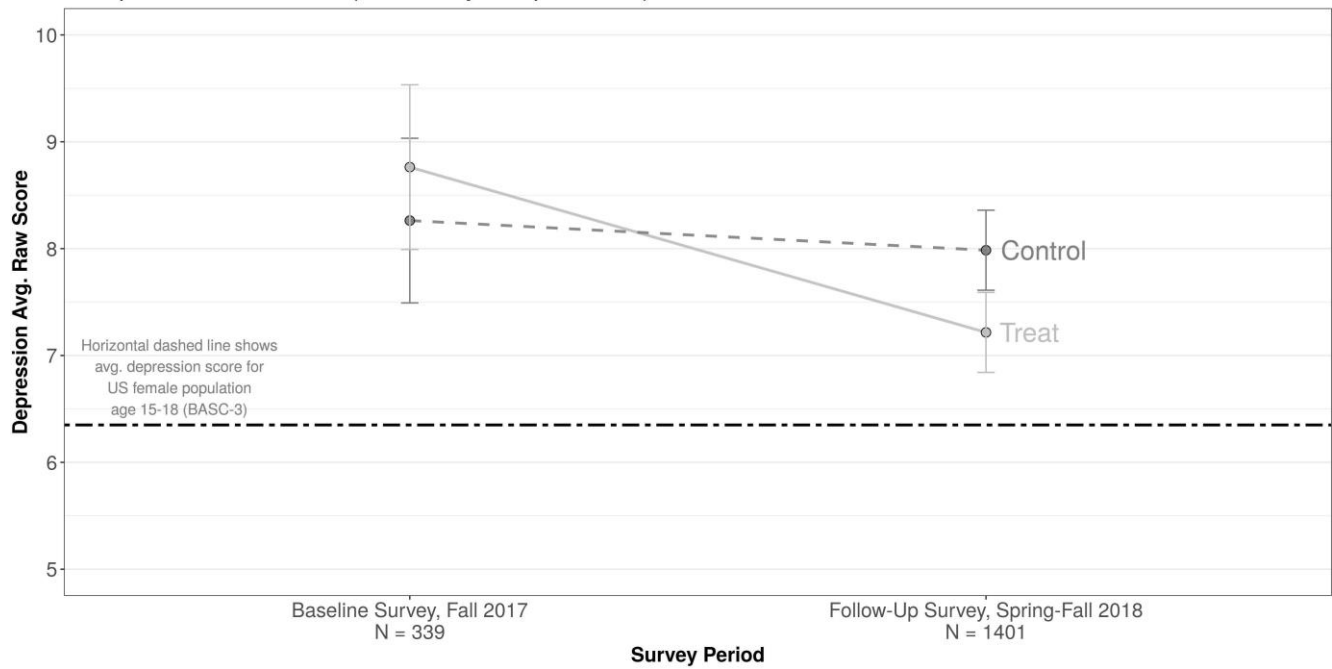
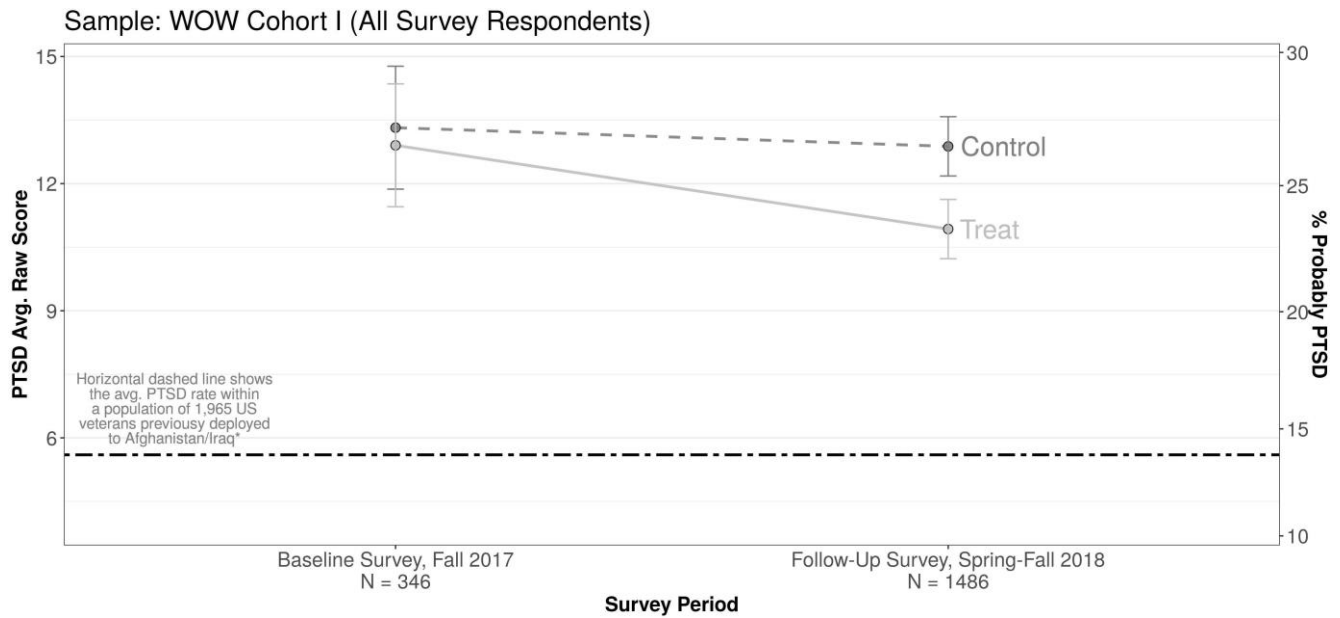
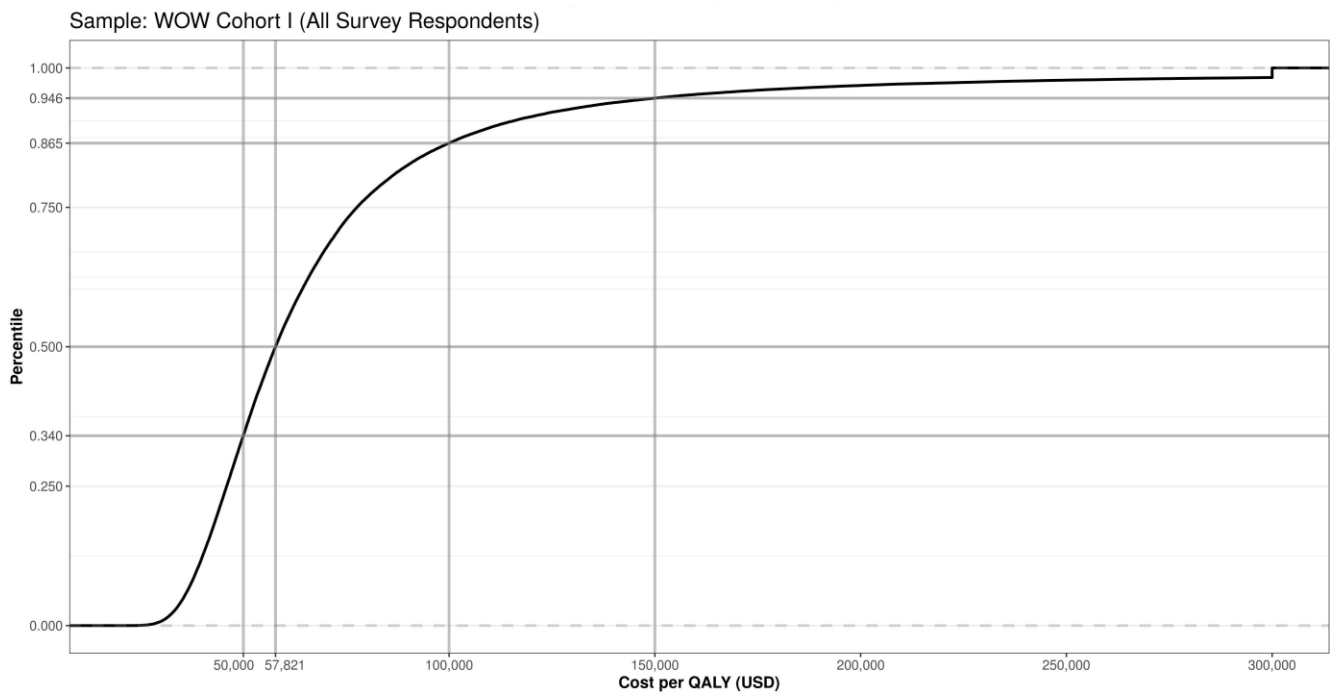


Fig. S2: Mental Health Trajectory for Depression (Effect of Being Offered WOW) - Unweighted  
*Note: control means shown in the plot are regression adjusted, for both the baseline & follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level.*



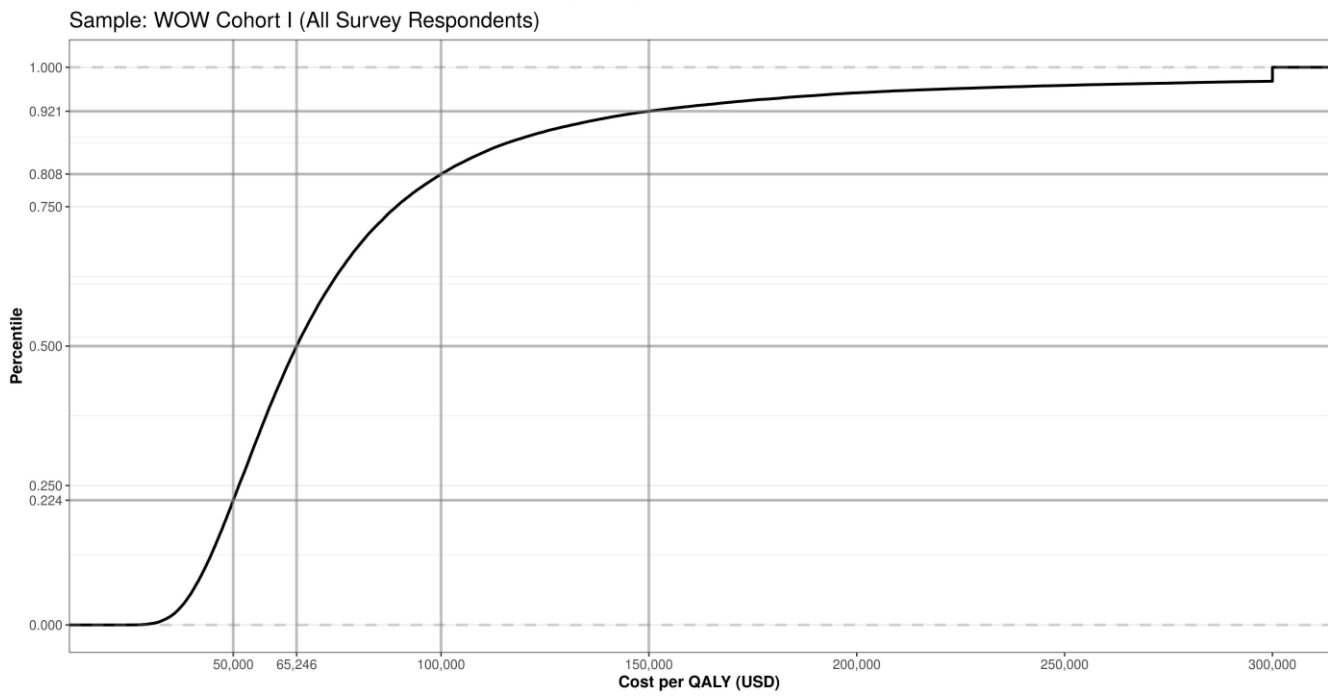
**Fig. S3: Mental Health Trajectory for PTSD (Effect of Being Offered WOW) - Unweighted**

*Note: control means shown in the plot are regression adjusted, for both the baseline & follow-up samples. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. The left hand y-axis indicates avg. PTSD raw scores (CATS); the right hand one shows the percentage with probably PTSD. The horizontal line showing the Probable PTSD incidence for the veterans' population should only be read using the right hand y-axis; the avg. PTSD raw scores (CATS) scale is only valid for the WOW sample (Source: Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery, Terri Tanielian & Lisa H. Jaycox, Center for Military Health Policy Research, 2008).*



**Fig. S4: Incremental Cost-effectiveness Ratio (ICER)**

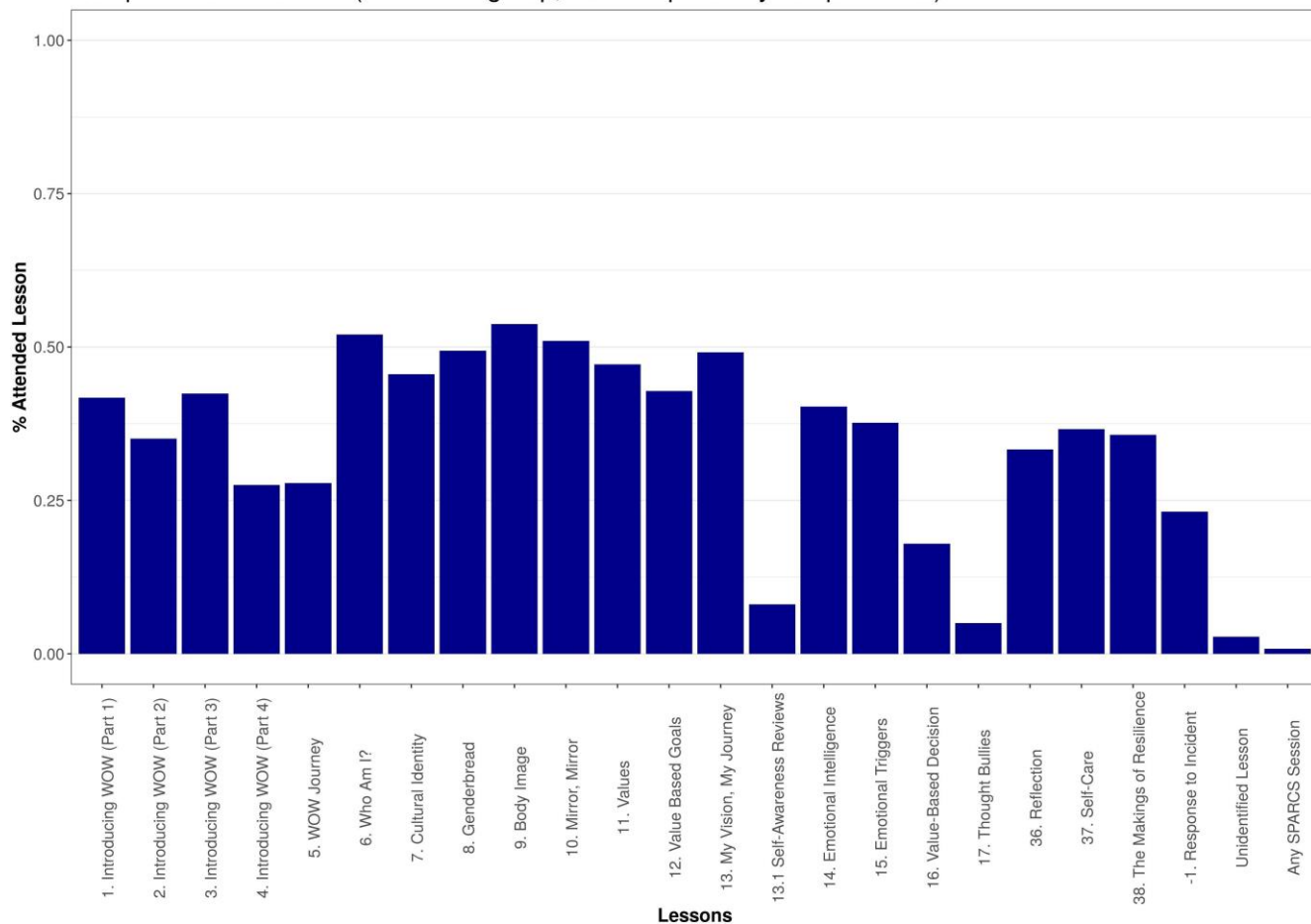
*Showing the distribution of cost per QALY obtained after a bootstrapping simulation of 100,000 iterations. The above distribution was estimated for the combined PTSD + depression model. Iterations that resulted on either negative costs or costs per QALY above the \$300,000 threshold were top-coded to \$300,000 (~1.72% of cases).*



**Fig. S5: Incremental Cost-effectiveness Ratio (ICER) - Unweighted**

*Showing the distribution of cost per QALY obtained after a bootstrapping simulation of 100,000 iterations. The above distribution was estimated for the combined PTSD + depression model. Iterations that resulted on either negative costs or costs per QALY above the \$300,000 threshold were top-coded to \$300,000 (~2.5% of cases).*

Sample: WOW Cohort I (Treatment group, Follow-up Survey Respondents)



**Fig. S6: Exposure to WOW Group Curriculum (Program Year 1)**

*Note: exposure to each lesson was calculated in the following way; if an individual attended at least one session for a given lesson, then we assigned them a value of '1', '0' otherwise (there can be several sessions per lesson). Exposure to each lesson was then calculated as the simple average across everyone in cohort I's treatment group that responded to the follow-up survey.*

Sample: WOW Cohort I (All Survey Respondents)

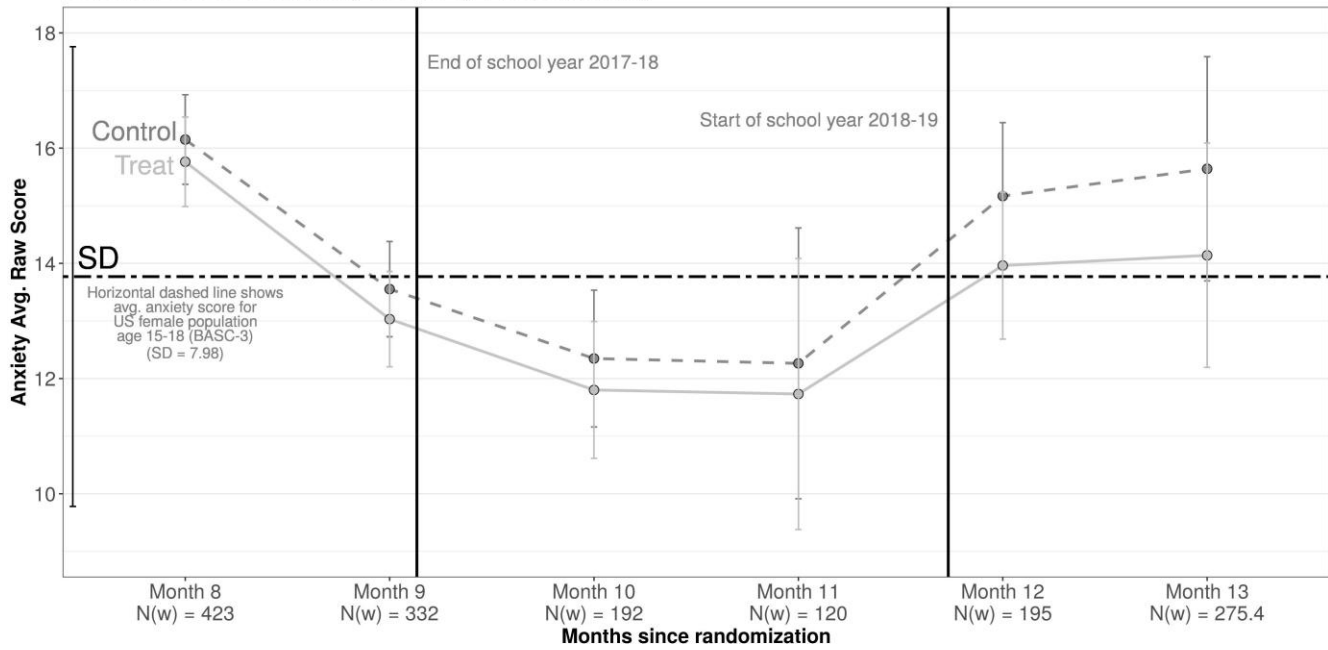


Fig. S7: Effect Duration Analysis for Anxiety (Effect of Being Offered WOW)

Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. Sample sizes shown on the x-axis indicate the number of participants surveyed on each month bucket. End of school year 2017-18 was June 18, 2018. Beginning of school year 2018-19 was September 4, 2018.



Sample: WOW Cohort I (All Survey Respondents)

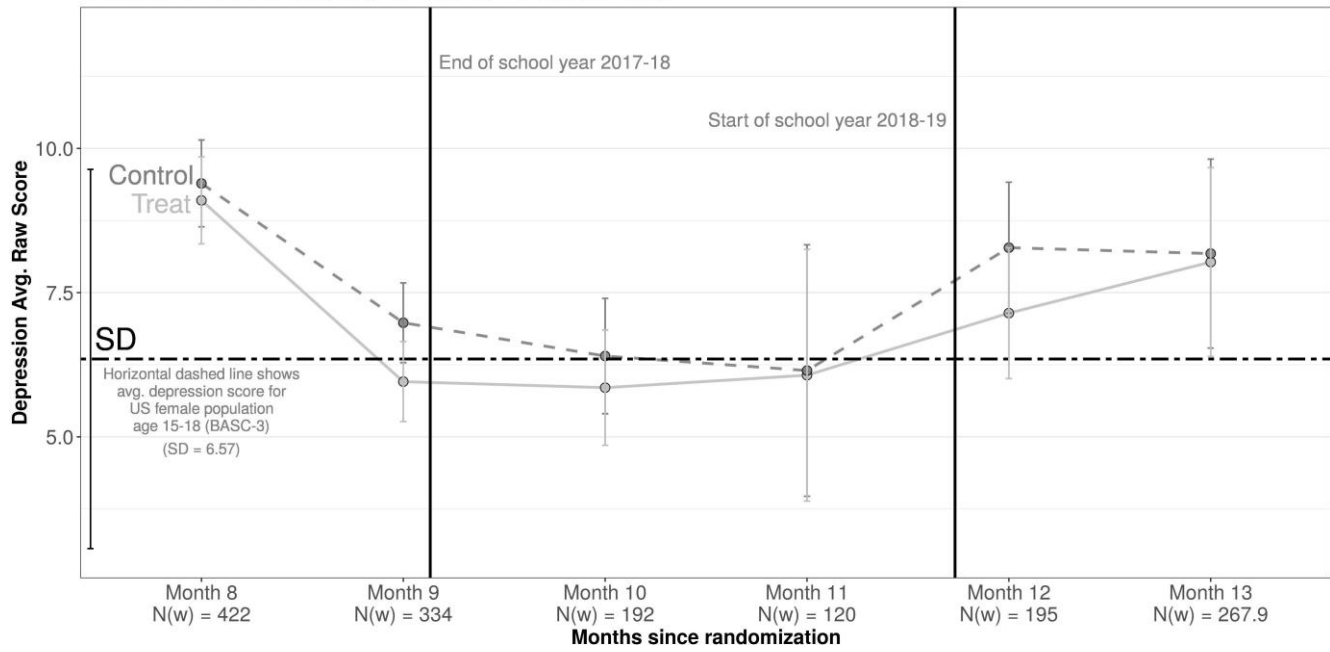


Fig. S8: Effect Duration Analysis for Depression (Effect of Being Offered WOW)

Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. Sample sizes shown on the x-axis indicate the number of participants surveyed on each month bucket. End of school year 2017-18 was June 18, 2018. Beginning of school year 2018-19 was September 4, 2018.

Sample: WOW Cohort I (All Survey Respondents)

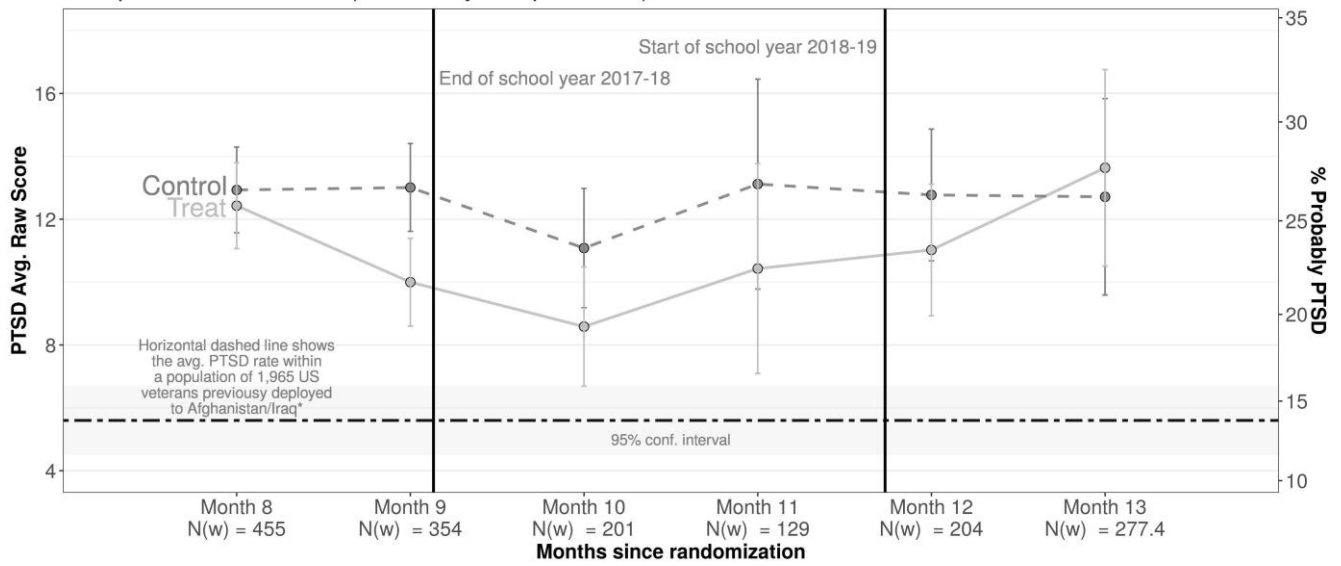
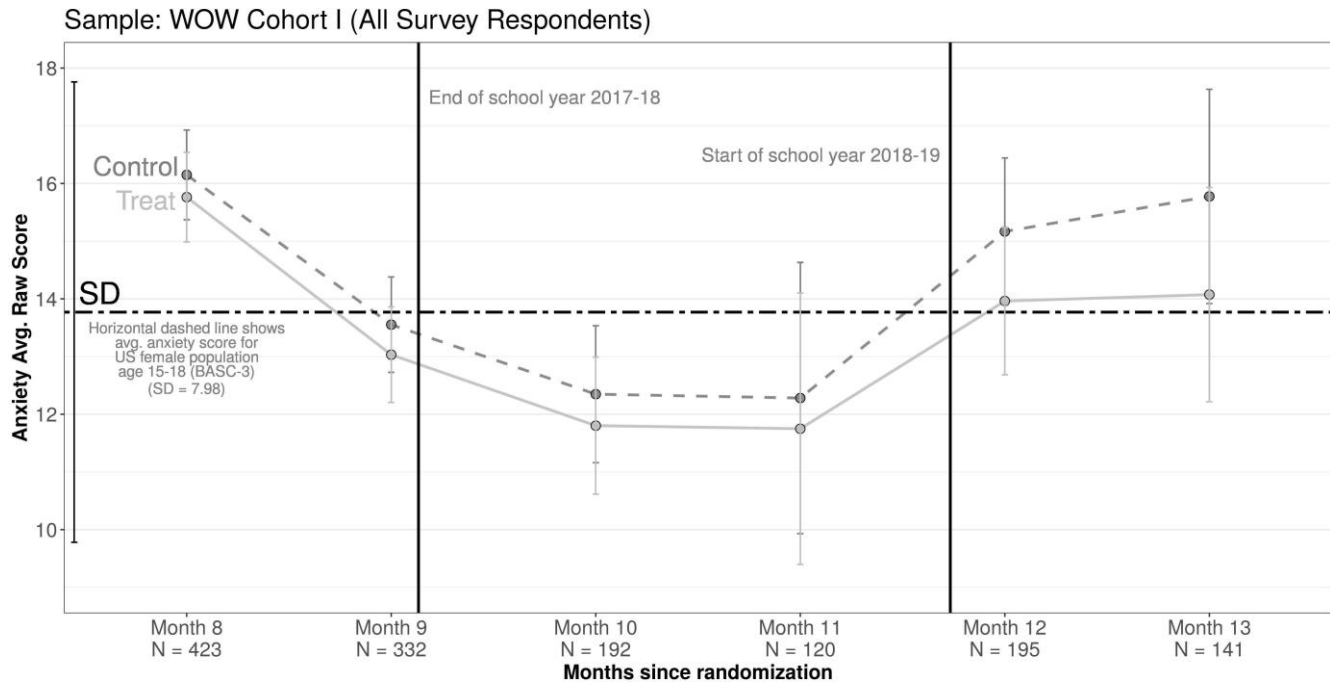


Fig. S9: Effect Duration Analysis for PTSD (Effect of Being Offered WOW)

Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. The left hand y-axis indicates avg. PTSD raw scores (CATS); the right hand one shows the percentage with probably PTSD. The horizontal line showing the Probable PTSD incidence for the veterans' population should only be read using the right hand y-axis; the avg. PTSD raw scores (CATS) scale is only valid for the WOW sample (Source: *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*, Terri Tanielian & Lisa H. Jaycox, Center for Military Health Policy Research, 2008).



**Fig. S10: Effect Duration Analysis for Anxiety (Effect of Being Offered WOW) - Unweighted**

*Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. Sample sizes shown on the x-axis indicate the number of participants surveyed on each month bucket. End of school year 2017-18 was June 18, 2018. Beginning of school year 2018-19 was September 4, 2018.*

Sample: WOW Cohort I (All Survey Respondents)

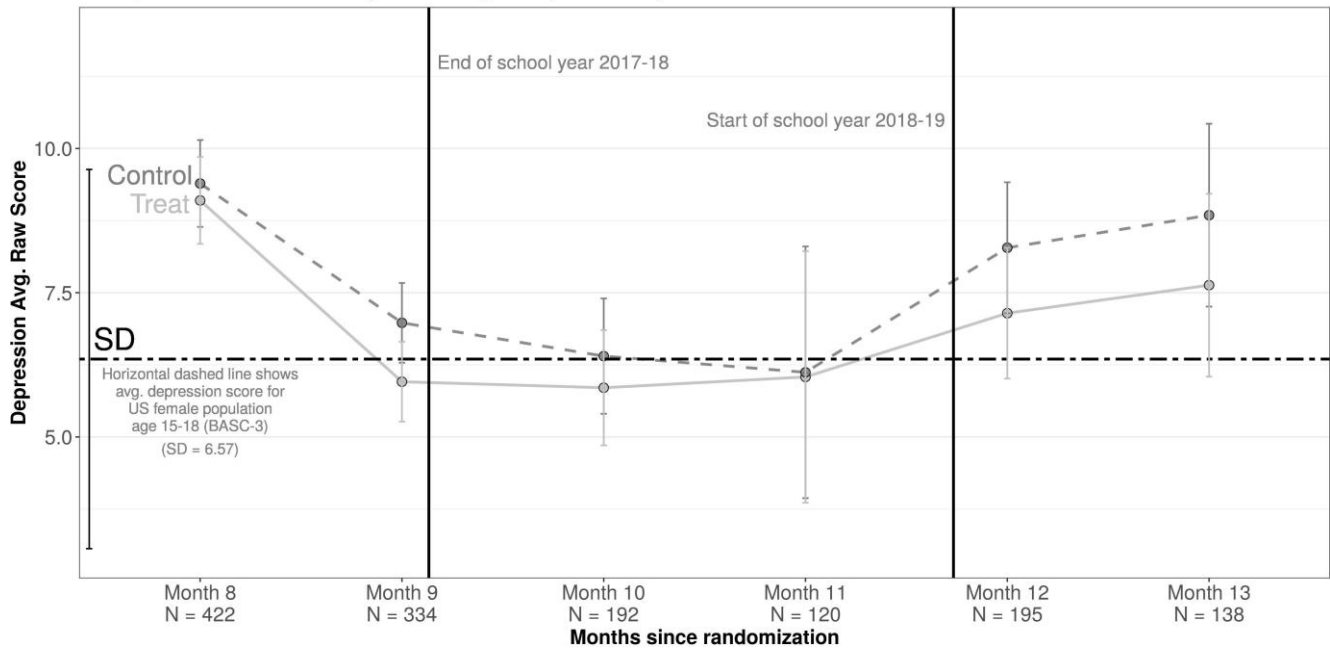


Fig. S11: Effect Duration Analysis for Depression (Effect of Being Offered WOW) - Unweighted  
*Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. Sample sizes shown on the x-axis indicate the number of participants surveyed on each month bucket. End of school year 2017-18 was June 18, 2018. Beginning of school year 2018-19 was September 4, 2018.*

Sample: WOW Cohort I (All Survey Respondents)

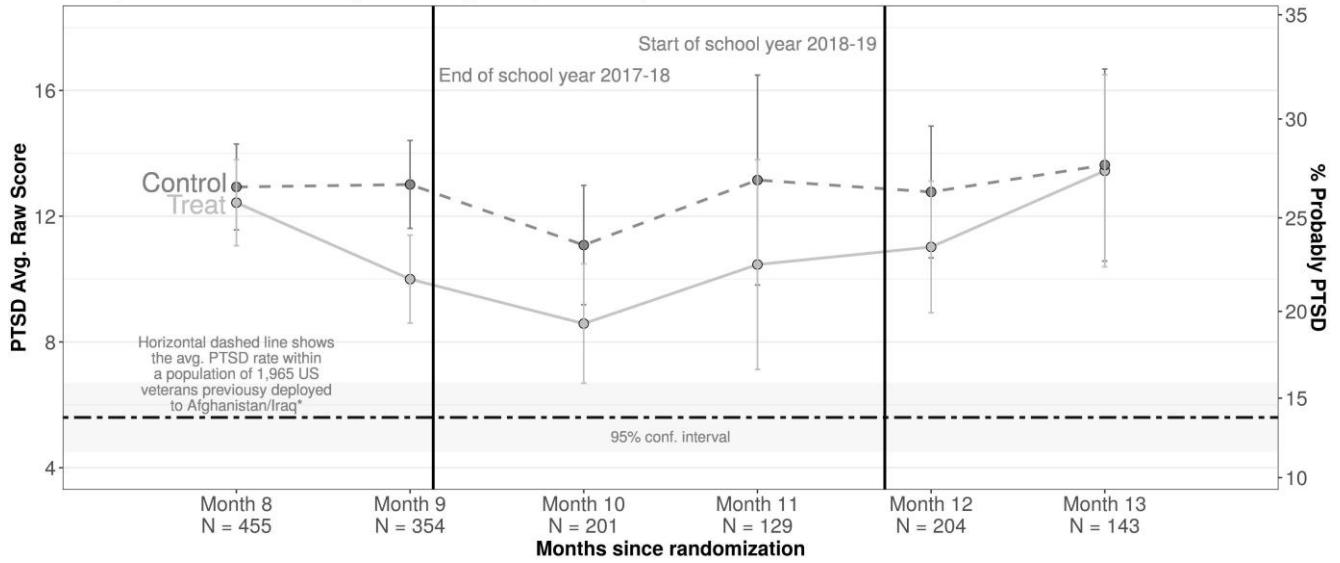


Fig. S12: Effect Duration Analysis for PTSD (Effect of Being Offered WOW) - Unweighted

Note: control means shown in the plot are regression adjusted. Treatment means are calculated as the adjusted control means plus the estimated effect of being offered WOW. Confidence intervals are at the 95% significance level. The left hand y-axis indicates avg. PTSD raw scores (CATS); the right hand one shows the percentage with probably PTSD. The horizontal line showing the Probable PTSD incidence for the veterans' population should only be read using the right hand y-axis; the avg. PTSD raw scores (CATS) scale is only valid for the WOW sample (Source: *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery*, Terri Tanielian & Lisa H. Jaycox, Center for Military Health Policy Research, 2008).

	Participants	Non-Participants	Total
Treatment	494 (65.1%)	265	759
Control	0	728	728
<b>Total</b>	<b>494</b>	<b>993</b>	<b>1,487</b>

Table S1: Treatment and Participation Status - Follow-up Survey Respondents  
*We consider that someone participated in the WOW program if they attended at least one counseling session during CPS school year 2017-18.*

Variable	Mean Resp. (N = 1487)	Mean Not Resp. (N = 516)	Dif.	p-value	Obs
Old for grade	.143	.128	.019	.269	2003
Age as of randomization	15.144	15.145	.007	.722	2003
Learning disability	.082	.099	-.016	.272	2003
Free/reduced lunch recipient	.948	.956	-.007	.508	2003
Black	.432	.326	.032*	.037	2003
Hispanic	.551	.633	-.011	.484	2003
Out-of-school suspensions	.073	.07	0	.994	2003
In-of-school suspensions	.114	.089	.031	.187	2003
Unexcused absences	8.877	9.989	-1.189*	.033	2003
Excused absences	4.031	4.65	-.401	.183	2003
Days present at school	163.472	160.68	2.479**	.01	2003
GPA	2.877	2.768	.117**	.002	2003
Disciplinary incidents	.299	.258	.039	.35	2003
Violent disciplinary incidents	.026	.025	-.002	.809	2003
Any disciplinary incidents?	.162	.149	.009	.611	2003
LEP indicator	.098	.104	.011	.479	2003
Primary Language Spanish	.463	.538	-.012	.532	2003

---

Joint F Test	Number of obs. =	2003
	F(17 , 1956) =	2.293
	Prob > F =	.002

---

**Table S2: Balance Table - Follow-up Survey Respondents vs. Non-respondents**

*Significance between the mean differences for survey respondents vs. non-respondents was estimated using a linear regression with block level fixed effects. All missing cases were imputed using block level averages. A joint F test was performed on demographic/academic covariates grouped together to test their joint significance.*

P-values estimated using heteroscedasticity-robust standard errors.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Variable	Treat	Control	Est	pvalue	Obs
<b>MENTAL HEALTH INDICATORS</b>					
Anxiety Raw Score	15.893	15.969	-.066	.938	329
Anxiety Per Score	59.25	58.565	.603	.839	329
Anxiety At Risk or Clinically Significant	.208	.267	-.064	.165	329
Depression Raw Score	8.754	8.553	.337	.677	326
Depression Per Score	61.395	60.535	1.363	.678	326
Depression At Risk or Clinically Significant	.257	.277	-.016	.745	326
PTSD Raw Score	12.641	13.38	-.372	.805	333
PTSD At Risk or Clinically Significant	.388	.362	.049	.364	333
<b>DEMOGRAPHIC INDICATORS</b>					
Old for grade	.153	.132	.017	.341	1487
Age as of randomization	15.163	15.124	.01	.61	1487
Learning disability	.091	.073	.015	.277	1487
Free/reduced lunch recipient	.947	.949	-.004	.735	1487
Black	.428	.435	-.005	.743	1487
Hispanic	.556	.545	.009	.605	1487
Out-of-school suspensions	.072	.073	0	1	1487
In-of-school suspensions	.116	.113	-.002	.948	1487
Unexcused absences	8.746	9.013	-.353	.495	1487
Excused absences	3.944	4.121	-.335	.196	1487
Days present at school	163.257	163.697	-.16	.855	1487
GPA	2.879	2.875	.015	.683	1487
Disciplinary incidents	.318	.279	.036	.438	1487
Violent disciplinary incidents	.025	.027	-.003	.726	1487
Any disciplinary incidents?	.173	.151	.02	.278	1487
LEP indicator	.108	.087	.017	.238	1487
Primary Language Spanish	.476	.451	.022	.246	1487
<hr/>					
Joint F Test I (Mental health & Demographic)	Number of obs. =	326			
	F(25 , 271) =	.745			
	Prob > F =	.809			
<hr/>					
Joint F Test II (Demographic only)	Number of obs. =	1487			
	F(17 , 1440) =	.722			
	Prob > F =	.782			

**Table S3: Balance Table - Follow-up Survey Respondents (Unweighted)**

*Significance between the mean differences for the treatment vs. control groups was estimated using a linear regression with block level fixed effects. All missing cases were imputed using block level averages. A joint F test was performed on demographic, academic & mental health covariates grouped together to test their joint significance. Obs(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.*

P-values estimated using heteroscedasticity-robust standard errors.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1



Sample	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N(w)	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
	<b>Anxiety Raw Score</b>					<b>Depression Raw Score</b>					<b>PTSD Raw Score</b>				
Baseline	341	15.98	-.064 (.821)	-.096 (1.178)	16.01	339	8.53	.5 (.789)	.752 (1.132)	8.5	346	13.77	-.416 (1.485)	-.62 (2.116)	14
Year One FU	1537.4	14.83	-.976* (.455)	-1.513* (.692)	15.48	1530.9	8.02	-.794+ (.41)	-1.227* (.625)	8.7	1620.4	12.93	-1.956** (.752)	-3.016** (1.14)	13.91
	<b>Anxiety Per Score</b>					<b>Depression Per Score</b>									
Baseline	341	58.68	.459 (2.902)	.691 (4.164)	58.04	339	60.27	2.299 (3.192)	3.461 (4.583)	59.47					
Year One FU	1537.4	54.89	-3.713* (1.656)	-5.755* (2.52)	57.37	1530.9	58.11	-4.041* (1.697)	-6.249* (2.584)	60.75					
	<b>Anxiety At Risk</b>					<b>Depression At Risk</b>					<b>PTSD At Risk</b>				
Baseline	341	.18	-.029 (.041)	-.044 (.059)	.21	339	.17	-.013 (.042)	-.019 (.06)	.16	346	.09	.053 (.033)	.079+ (.047)	.06
Year One FU	1537.4	.15	.006 (.021)	.01 (.032)	.16	1530.9	.13	-.009 (.019)	-.013 (.028)	.13	1620.4	.13	-.042* (.017)	-.065* (.026)	.17
	<b>Anxiety Clinically Significant</b>					<b>Depression Clinically Significant</b>					<b>PTSD Clinically Significant</b>				
Baseline	341	.08	-.033 (.026)	-.049 (.037)	.11	339	.1	.001 (.034)	.001 (.049)	.13	346	.29	-.009 (.049)	-.013 (.07)	.3
Year One FU	1537.4	.04	-.003 (.011)	-.005 (.017)	.05	1530.9	.09	-.019 (.017)	-.029 (.026)	.12	1620.4	.27	-.035 (.024)	-.055 (.036)	.28
	<b>Anxiety At Risk or Clinically Significant</b>					<b>Depression At Risk or Clinically Significant</b>					<b>PTSD At Risk or Clinically Significant</b>				
Baseline	341	.27	-.062 (.045)	-.094 (.065)	.32	339	.28	-.012 (.049)	-.018 (.071)	.29	346	.37	.044 (.053)	.066 (.076)	.36
Year One FU	1537.4	.2	.003 (.023)	.004 (.035)	.21	1530.9	.22	-.027 (.023)	-.042 (.035)	.25	1620.4	.39	-.078** (.026)	-.12** (.04)	.44

Table S4: Mental Health Primary Outcomes: Anxiety, Depression, and PTSD (including percentile scores when available)

Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property. N(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.

P-values estimated using heteroscedasticity-robust standard errors.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Sample	Anxiety Raw Score					Depression Raw Score					PTSD Raw Score				
	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	N	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
Baseline	341	15.98	-.064 (.821)	-.096 (1.178)	16.01	339	8.53	.5 (.789)	.752 (1.132)	8.5	346	13.77	-.416 (1.485)	-.62 (2.116)	14
Year One FU	1403	14.69	-.925* (.427)	-1.429* (.65)	15.59	1401	7.95	-.759* (.383)	-1.17* (.581)	8.72	1486	12.87	-1.952** (.715)	-2.999** (1.081)	13.89
Baseline	341	58.68	.459 (2.902)	.691 (4.164)	58.04	339	60.27	2.299 (3.192)	3.461 (4.583)	59.47					
Year One FU	1403	54.35	-3.424* (1.564)	-5.289* (2.383)	57.73	1401	57.65	-3.563* (1.631)	-5.494* (2.479)	60.71					
Baseline	341	.18	-.029 (.041)	-.044 (.059)	.21	339	.17	-.013 (.042)	-.019 (.06)	.16	346	.09	.053 (.033)	.079+ (.047)	.06
Year One FU	1403	.15	.011 (.02)	.017 (.03)	.15	1401	.13	-.004 (.018)	-.006 (.028)	.14	1486	.12	-.04* (.016)	-.061* (.024)	.16
Baseline	341	.08	-.033 (.026)	-.049 (.037)	.11	339	.1	.001 (.034)	.001 (.049)	.13	346	.29	-.009 (.049)	-.013 (.07)	.3
Year One FU	1403	.04	-.007 (.011)	-.012 (.016)	.05	1401	.09	-.021 (.015)	-.033 (.023)	.11	1486	.27	-.032 (.022)	-.049 (.034)	.27
Baseline	341	.27	-.062 (.045)	-.094 (.065)	.32	339	.28	-.012 (.049)	-.018 (.071)	.29	346	.37	.044 (.053)	.066 (.076)	.36
Year One FU	1403	.19	.004 (.021)	.006 (.032)	.2	1401	.22	-.025 (.022)	-.039 (.034)	.25	1486	.39	-.071** (.025)	-.11** (.037)	.43

Table S5: Mental Health Primary Outcomes: Anxiety, Depression, and PTSD (unweighted, including percentile scores when available)

Baseline covariates include: days present; number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the three arrest categories: violent, drug, and property. N(w) refers to the weighted sample size - i.e. the sum of the sample weights used in the statistical tests.

P-values estimated using heteroscedasticity-robust standard errors.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Outcome	Cost per QALY Gained presuming 365-day benefit	Days of posited benefit required to meet \$150,000/QALY threshold	Days of posited benefit required to meet \$100,000/QALY threshold
Depression as the only valued outcome	\$152,247/QALY	370	556
PTSD as the only valued outcome	\$76,137/QALY	185	278
Combined depression and PTSD	\$65,686/QALY	160	240

**Table S6: Cost per QALY Gained (unweighted)**

*Showing cost per QALY gained assuming a 365-day benefit. We also show the number of posited benefit days required to meet relevant cost-benefit thresholds.*

Lesson Number/Name	Avg. Number of Sessions	Median Lesson Date	% Groups that hosted each Lesson
1. Introducing WOW (Part 1)	1.54 (0.77)	2017-10-26	79.22%
2. Introducing WOW (Part 2)	1.22 (0.46)	2017-11-08	70.13%
3. Introducing WOW (Part 3)	1.48 (0.54)	2017-11-30	75.32%
4. Introducing WOW (Part 4)	1.19 (0.4)	2017-12-07	54.55%
5. WOW Journey	1.32 (0.62)	2017-12-16	51.95%
6. Who Am I?	1.62 (0.59)	2018-01-11	94.81%
7. Cultural Identity	1.21 (0.51)	2018-01-24	87.01%
8. Genderbread	1.51 (0.56)	2018-02-06	96.1%
9. Body Image	1.66 (0.5)	2018-02-20	98.7%
10. Mirror, Mirror	1.49 (0.5)	2018-03-06	100%
11. Values	1.09 (0.34)	2018-03-14	96.1%
12. Value Based Goals	1.06 (0.23)	2018-03-21	92.21%
-1. Response to Incident	1.27 (0.45)	2018-04-03	48.05%
13. My Vision, My Journey	1.34 (0.48)	2018-04-05	94.81%
14. Introduction to Emotional Intelligence	1.18 (0.42)	2018-04-18	87.01%
13.1 Self-Awareness Reviews	1 (0)	2018-04-21	12.99%
15. Emotional Triggers	1.46 (0.62)	2018-05-01	79.22%
17. Thought Bullies	1 (0)	2018-05-05	10.39%
16. Value-Based Decision	1.15 (0.44)	2018-05-08	44.16%
37. Self-Care	1.02 (0.13)	2018-05-16	80.52%
38. Bounce Back: The Makings of Resilience	1.08 (0.28)	2018-05-22	77.92%
36. Reflection	1.02 (0.13)	2018-05-30	71.43%

Table S7: Descriptives for each Group Lesson in the WOW Curriculum (SY 2017-18)

*Note: Standard deviation for the number of sessions across groups is shown in parenthesis.*

	N	Avg. Ses- sion/Lessons Attended	Max.	Min.	% Attended at least one Ses- sion/Lesson	Att. rate >= 50%	Att. rate >= 75%
Sessions	759	9.98	26	0	65.09%	52.7%	33.73%
Lessons	759	7.83	21	0	65.09%	54.94%	40.18%

Table S8: Attendance based on Groups Session/Lessons Provided (Program Year 1)

	Num. of participants with service records	Individual service records found	Avg. number of services provided	% of the treatment group that received services
Program Year 1	496	7170	14.46	65.35%
Summer 2018	144	409	2.84	18.97%

Table S9: Individual Services Provided by WOW During Program Year 1 vs Summer 2018

Variable	Mean T	Mean C	ITT	CCM	TOT	Obs	Obs (w)
<b>Anxiety Raw Score</b>							
Treat	13.69	14.691	-0.859+ (.444)	15.518	-1.361* (.678)	1403	1537.4
Time to Survey x Treat			-0.011 (.009)		-0.018 (.014)		
<b>Anxiety Per Score</b>							
Treat	50.671	54.355	-3.32* (1.612)	57.723	-5.28* (2.455)	1403	1537.4
Time to Survey x Treat			-0.048 (.032)		-0.078 (.048)		
<b>Anxiety At Risk or Clinically Significant</b>							
Treat	.194	.194	.009 (.022)	.2	.013 (.034)	1403	1537.4
Time to Survey x Treat			0 (0)		-0.001 (.001)		
<b>Depression Raw Score</b>							
Treat	7.217	7.95	-0.72+ (.403)	8.682	-1.134+ (.615)	1401	1530.9
Time to Survey x Treat			-0.008 (.008)		-0.014 (.013)		
<b>Depression Per Score</b>							
Treat	54.044	57.648	-3.842* (1.675)	61.297	-6.074* (2.558)	1401	1530.9
Time to Survey x Treat			-0.052 (.034)		-0.084+ (.051)		
<b>Depression At Risk or Clinically Significant</b>							
Treat	.201	.224	-0.023 (.023)	.246	-0.036 (.034)	1401	1530.9
Time to Survey x Treat			-0.001 (0)		-0.001 (.001)		
<b>PTSD Raw Score</b>							
Treat	10.931	12.867	-2.067** (.736)	14.095	-3.2** (1.122)	1486	1620.4
Time to Survey x Treat			-0.008 (.014)		-0.015 (.022)		
<b>PTSD At Risk or Clinically Significant</b>							
Treat	.318	.389	-0.081** (.025)	.449	-0.126** (.039)	1486	1620.4
Time to Survey x Treat			-0.001 (.001)		-0.001 (.001)		

Table S10: Effect Duration Estimates: Treatment Interacted with Time-to-Survey Indicator

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Variable	Mean T	Mean C	ITT	CCM	TOT	Obs
<b>Anxiety Raw Score</b>						
Treat	13.69	14.691	-0.897*	15.598	-1.441*	1403
			(.429)		(.662)	
Time to Survey x Treat			-.014		-.022+	
			(.008)		(.013)	
<b>Anxiety Per Score</b>						
Treat	50.671	54.355	-3.399*	57.919	-5.476*	1403
			(1.569)		(2.423)	
Time to Survey x Treat			-.057+		-.093*	
			(.03)		(.046)	
<b>Anxiety At Risk or Clinically Significant</b>						
Treat	.194	.194	.006	.206	.007	1403
			(.021)		(.033)	
Time to Survey x Treat			-.001		-.001	
			(0)		(.001)	
<b>Depression Raw Score</b>						
Treat	7.217	7.95	-.762*	8.765	-1.217*	1401
			(.382)		(.586)	
Time to Survey x Treat			-.011		-.018	
			(.008)		(.012)	
<b>Depression Per Score</b>						
Treat	54.044	57.648	-3.656*	61.063	-5.84*	1401
			(1.633)		(2.509)	
Time to Survey x Treat			-.052+		-.086+	
			(.031)		(.048)	
<b>Depression At Risk or Clinically Significant</b>						
Treat	.201	.224	-.026	.253	-.043	1401
			(.022)		(.034)	
Time to Survey x Treat			-.001+		-.001+	
			(0)		(.001)	
<b>PTSD Raw Score</b>						
Treat	10.931	12.867	-2.078**	14.137	-3.242**	1486
			(.717)		(1.097)	
Time to Survey x Treat			-.011		-.021	
			(.014)		(.022)	
<b>PTSD At Risk or Clinically Significant</b>						
Treat	.318	.389	-.077**	.445	-.122**	1486
			(.025)		(.038)	
Time to Survey x Treat			-.001		-.001	
			(0)		(.001)	

Table S11: Effect Duration Estimates, Treatment Interacted with Time-to-Survey Indicator - Unweighted

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1



Variable	Mean T	Mean C	ITT	CCM	TOT	Obs
<b>Anxiety Raw Score</b>						
Treat	13.632	14.536	-1.141*	15.93	-1.827*	1279
			(.464)		(.727)	
Time to Survey x Treat			-.02*		-.033*	
			(.01)		(.016)	
<b>Anxiety Per Score</b>						
Treat	50.435	53.89	-4.532**	59.651	-7.284**	1279
			(1.712)		(2.693)	
Time to Survey x Treat			-.091*		-.146**	
			(.036)		(.057)	
<b>Anxiety At Risk or Clinically Significant</b>						
Treat	.194	.187	.002	.208	.001	1279
			(.023)		(.035)	
Time to Survey x Treat			-.001		-.001	
			(.001)		(.001)	
<b>Depression Raw Score</b>						
Treat	7.164	7.808	-.875*	8.874	-1.393*	1280
			(.409)		(.633)	
Time to Survey x Treat			-.015		-.024+	
			(.009)		(.015)	
<b>Depression Per Score</b>						
Treat	53.913	57.148	-4.282*	61.936	-6.812*	1280
			(1.789)		(2.775)	
Time to Survey x Treat			-.07+		-.114+	
			(.038)		(.059)	
<b>Depression At Risk or Clinically Significant</b>						
Treat	.199	.216	-.03	.253	-.049	1280
			(.023)		(.036)	
Time to Survey x Treat			-.001+		-.001+	
			(.001)		(.001)	
<b>PTSD Raw Score</b>						
Treat	10.736	12.787	-2.183**	13.946	-3.393**	1361
			(.784)		(1.199)	
Time to Survey x Treat			-.018		-.031	
			(.017)		(.026)	
<b>PTSD At Risk or Clinically Significant</b>						
Treat	.31	.383	-.08**	.436	-.124**	1361
			(.026)		(.04)	
Time to Survey x Treat			-.001		-.001	
			(.001)		(.001)	

Table S12: Effect Duration Estimates, Treatment Interacted with Time-to-Survey Indicator (Phase 1 only) - Unweighted

Showing estimates for WOW participants surveyed during the first survey phase, only. That is, including everyone that responded to the follow-up survey before additional survey measures (e.g. probabilistic sampling) were implemented

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

Sample	N	Repeated any grade <sup>60</sup>			Drop-out <sup>61</sup>			Out-of-CPS transfer <sup>61</sup>					
		Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean	Control Mean	Intention to Treat	Effect of Participation (IV)	Control Complier Mean
Cohort I (year 2)	3,749	.07	-.012 (.008)	-.019 (.013)	.06	.03	-.003 (.005)	-.006 (.008)	.02	.05	-.004 (.008)	-.007 (.013)	.05
Cohort I (year 3)	3,749	.08	-.009 (.009)	-.015 (.014)	.08	.04	-.007 (.006)	-.012 (.01)	.02	.06	-.004 (.008)	-.006 (.013)	.05
Cohort I (year 3)	2,543	.85	.008 (.014)	.013 (.022)	.9								

Table S13: Academic Secondary Outcomes - Cohort I

Baseline covariates and randomization block fixed effects included in all models. Baseline covariates include: days; present number of in-school suspensions; number of out-of-school suspensions; number of each type of grade received (A, B, C, D, F); dummies for ages 14–15, 15–16, and 17+; and indicators for having a learning disability, having limited English proficiency, being in 9th or 10th grades, being old-for-grade, being Black, being Hispanic, and ever being arrested in each of the following arrest categories: violent, drug, and property. In this context, year 2(3) outcomes for Cohort I correspond to the period up to day20 academic year 2019-20(2020-21). (60, 61,62)

P-values estimated using heteroscedasticity-robust standard errors.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

	N	Mean C	ITT	TOT	CCM
Social Stress Raw Score	1419	7.738	-0.674 * (0.297)	-1.031 * (0.455)	8.339
Somatization Raw Score	1442	3.246	-0.112 (0.186)	-0.171 (0.282)	3.415
Sensation Seeking Raw Score	1412	10.11	-0.387 (0.262)	-0.594 (0.402)	10.583
Attention Problems Raw Score	1434	7.621	-0.106 (0.273)	-0.161 (0.418)	8.066
Atypicality Raw Score	1408	4.786	-0.176 (0.253)	-0.271 (0.389)	5.06
Hyperactivity Raw Score	1425	5.481	-0.029 (0.227)	-0.044 (0.350)	5.707
Locus of Control Raw Score	1411	5.135	-0.2 (0.233)	-0.308 (0.360)	5.562
Attitude to School Raw Score	1431	7.765	0.162 (0.237)	0.248 (0.363)	7.853
Attitude to Teachers Raw Score	1404	7.06	0.104 (0.244)	0.159 (0.374)	7.2
Sense of Inadequacy Raw Score	1418	9.91	-0.283 (0.327)	-0.433 (0.499)	10.256

Table S14: Clinical Scales

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$

	N	Mean C	ITT	TOT	CCM
Self Esteem Raw Score	1433	12.709	0.43 + (0.247)	0.658 + (0.379)	12.359
Self Reliance Raw Score	1429	17.62	0.104 (0.224)	0.158 (0.340)	17.602
Interpersonal Skills Raw Score	1419	18.422	0.198 (0.222)	0.303 (0.329)	18.232
Relationship with Parents Raw Score	1405	22.426	0.164 (0.372)	0.252 (0.570)	22.212

Table S15: Adaptive Scales

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$

	N	Mean C	ITT	TOT	CCM
Mania Raw Score	1405	9.133	-0.287 (0.313)	-0.44 (0.482)	9.533
Test Anxiety Raw Score	1432	9.207	0.01 (0.230)	0.016 (0.349)	9.446
Anger Control Raw Score	1409	8.051	-0.332 (0.290)	-0.512 (0.447)	8.508
Ego Strength Raw Score	1419	15.933	0.293 (0.224)	0.45 (0.343)	15.733

Table S16: Content Scales

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$

	N	Mean C	Mean T	ITT	ITT SE	TOT	TOT SE	CCMs
Expect to graduate from college	1467	0.459	0.422	-0.039	0.026	-0.059	0.040	0.502
Expect to attend a higher level of school after graduating from college	1467	0.362	0.410	0.052 *	0.025	0.08 *	0.039	0.326
Expect less than the above or doesn't know	1467	0.179	0.168	-0.013	0.019	-0.021	0.030	0.172

Table S17: How far in school do you expect to go?

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$

	N	Mean C	Mean T	ITT	ITT SE	TOT	TOT SE	CCMs
Participate or plan to participate in a club/program/activity	1467	0.818	0.875	0.06 **	0.019	0.091 **	0.028	0.838

**Table S18: Participation in programs/clubs/activities at school**

*Note: The full text of the questions was "What clubs, programs, or activities do you participate in or plan to participate in at school this year?" Respondents were able to select more than one option.*

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

	N	Mean C	Mean T	ITT	ITT SE	TOT	TOT SE	CCMs
In our family...								
...we really help and support each other	1472	1.554	1.620	0.066 *	0.029	0.101 *	0.044	1.521
... we spend a lot of time doing things together at home	1472	1.176	1.231	0.049	0.034	0.074	0.053	1.136
...we work hard at what we do in our home	1459	1.511	1.588	0.077 *	0.030	0.117 *	0.046	1.458
...there is a feeling of togetherness	1461	1.418	1.466	0.045	0.033	0.069	0.051	1.387
...we really get along well with each other	1461	1.359	1.409	0.049	0.031	0.075	0.047	1.338
... we can talk openly in our home	1465	1.347	1.343	-0.008	0.036	-0.012	0.054	1.331
...sometimes we tell each other about our personal problems	1467	1.275	1.276	-0.001	0.036	-0.002	0.054	1.272
...we begin discussions easily	1460	1.300	1.262	-0.03	0.036	-0.046	0.056	1.311
...we argue a lot (Reverse scaling)	1433	0.889	0.844	-0.032	0.037	-0.049	0.056	0.904
...we are really mad at each other (Reverse scaling)	1461	0.515	0.525	0.015	0.034	0.024	0.053	0.500
...we lose our tempers a lot (Reverse scaling)	1447	0.680	0.664	-0.005	0.037	-0.007	0.057	0.640
...we often put down each other (Reverse scaling)	1454	0.341	0.285	-0.054 +	0.029	-0.083 +	0.045	0.338
...we raise our voice when we are mad (Reverse scaling)	1469	1.090	1.009	-0.073 *	0.036	-0.112 *	0.055	1.118
My family members really support each other	1463	1.483	1.563	0.08 *	0.031	0.122 *	0.047	1.431
I am proud to be part of our family	1464	1.712	1.747	0.036	0.027	0.055	0.041	1.675
My family members are sometimes violent (Reverse scaling)	1457	0.242	0.207	-0.033	0.025	-0.051	0.039	0.259

Table S19: Questions about family life

Note: Respondents answered with "A lot," "Somewhat," and "Not at all."

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$



	N	Mean C	Mean T	ITT	ITT SE	TOT	TOT SE	CCMs
How well...								
...can you get teachers to help you when you get stuck on schoolwork?	1466	2.979	2.991	0.004	0.051	0.006	0.078	2.971
...can you express your opinions when other classmates disagree with you?	1464	2.720	2.746	0.028	0.055	0.042	0.084	2.712
...do you succeed in cheering yourself up when something upsetting has happened?	1466	2.384	2.489	0.096	0.063	0.146	0.096	2.328
...can you study when there are other interesting things to do?	1466	2.042	1.968	-0.086	0.060	-0.132	0.092	2.083
...do you succeed in calming down when you are very scared?	1460	2.320	2.285	-0.035	0.059	-0.054	0.090	2.318
...can you become friends with others?	1467	2.814	2.775	-0.051	0.058	-0.078	0.088	2.801
...can you study a chapter for a test?	1458	2.215	2.240	0.014	0.059	0.021	0.089	2.160
...can you have a chat with an unfamiliar person?	1458	2.049	2.016	-0.036	0.063	-0.055	0.097	2.049
...can you keep yourself from becoming nervous?	1461	1.951	1.987	0.034	0.059	0.052	0.089	1.901
...do you succeed in finishing all your homework every day?	1468	2.424	2.469	0.047	0.062	0.072	0.096	2.319
...can you work in harmony with your other classmates?	1455	2.748	2.776	0.03	0.053	0.046	0.080	2.702
...can you control your feelings?	1459	2.520	2.514	-0.007	0.061	-0.01	0.093	2.555
...can you pay attention during every class?	1458	2.695	2.704	0	0.051	0	0.078	2.667
...can you tell other children that they are doing something that you don't like?	1455	2.603	2.633	0.014	0.058	0.021	0.089	2.620
...can you give yourself a pep-talk when you feel low?	1461	2.418	2.439	0.026	0.064	0.039	0.097	2.410
...do you succeed in understanding all subjects in school?	1460	2.611	2.612	0.006	0.051	0.009	0.077	2.579
...can you tell a funny event to a group of children?	1457	2.563	2.546	-0.021	0.057	-0.031	0.088	2.626
...can you tell a friend that you don't feel well?	1457	2.824	2.931	0.109 +	0.058	0.167 +	0.089	2.761
...do you succeed in satisfying your parents with your schoolwork?	1458	2.794	2.741	-0.061	0.059	-0.093	0.089	2.799
...do you succeed in staying friends with others?	1456	2.764	2.861	0.102 +	0.056	0.156 +	0.086	2.654
...do you succeed in putting down unpleasant thoughts?	1441	2.229	2.327	0.088	0.062	0.134	0.094	2.139
...do you succeed in passing a test?	1451	2.567	2.561	0	0.050	-0.001	0.076	2.519
...do you succeed in preventing fights with others?	1448	2.596	2.731	0.137 *	0.063	0.209 *	0.097	2.441
...do you succeed in not worrying about things that might happen?	1458	2.057	2.088	0.014	0.063	0.022	0.095	2.046

**Table S20: Questions about behavior**

*Note: Respondents answered on a scale from 1-5 ranging from "Not at all" to "Very Well."*

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\* p<0.001, \*\* p<0.01, \* p<0.05, + p<0.1

	N	Mean C	Mean T	ITT	ITT SE	TOT	TOT SE	CCMs
I pay attention to how I feel	1470	2.788	2.773	0.008	0.055	0.012	0.085	2.763
I have no idea how I am feeling (Reverse scaling)	1461	1.338	1.320	-0.028	0.058	-0.043	0.089	1.354
I have difficulty making sense out of my feelings (Reverse scaling)	1451	1.552	1.512	-0.056	0.063	-0.085	0.096	1.627
I care about how I am feeling	1454	2.891	2.899	0.022	0.061	0.033	0.094	2.821
I am confused about how I feel (Reverse scaling)	1449	1.314	1.368	0.054	0.062	0.082	0.095	1.324
When I am upset...								
...I acknowledge my emotions	1448	2.303	2.445	0.166 *	0.066	0.254 *	0.101	2.211
...I become embarrassed for feeling that way (Reverse scaling)	1463	1.108	0.976	-0.141 *	0.064	-0.215 *	0.096	1.215
...I have difficulty getting work done (Reverse scaling)	1460	1.829	1.849	0.034	0.070	0.052	0.106	1.826
...I become out of control (Reverse scaling)	1453	0.850	0.869	0.022	0.061	0.033	0.093	0.824
...I believe that I will end up feeling very depressed (Reverse scaling)	1457	1.238	1.101	-0.147 *	0.072	-0.224 *	0.110	1.331
...I have difficulty focusing on other things (Reverse scaling)	1460	1.742	1.714	-0.02	0.068	-0.031	0.104	1.766
...I feel guilty for feeling that way (Reverse scaling)	1461	1.063	0.934	-0.144 *	0.064	-0.219 *	0.097	1.182
...I have difficulty concentrating (Reverse scaling)	1451	1.552	1.512	-0.056	0.063	-0.085	0.096	1.627
...I have difficulty controlling my behaviors (Reverse scaling)	1457	1.081	1.040	-0.045	0.066	-0.068	0.100	1.128
...I believe there is nothing I can do to make myself feel better (Reverse scaling)	1458	1.139	1.050	-0.092	0.064	-0.141	0.098	1.193
...I become irritated with myself for feeling that way (Reverse scaling)	1458	1.352	1.266	-0.099	0.068	-0.151	0.104	1.456
...I lose control of my behavior (Reverse scaling)	1455	0.822	0.785	-0.042	0.060	-0.065	0.091	0.871
...it takes me a long time to feel better (Reverse scaling)	1458	1.360	1.292	-0.07	0.066	-0.106	0.100	1.448

**Table S21: Questions about emotions**

*Note: Respondents answered with "Almost never," "Sometimes," "About half the time," "Most of the time," and "Almost always."*

Heteroscedasticity-robust standard errors shown in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.1$

**Data File S1: WOW Baseline Survey Data**

*Includes de-identified data showing answers to Behavior Assessment System for Children (BASC-3) questions that were used to create aggregated scores and indicators for Anxiety and Depression, collected at baseline. It also includes data showing answers to Children and Adolescent Trauma Screen (CATS) questions that were used to create aggregated scores/indicators for PTSD, collected at baseline. This file also includes data on alternative BASC-3 questions, used for secondary metrics (e.g., Self-esteem, self-reliance, etc.). The number of independent survey responses included in this file is 346.*

**Data File S2: WOW Follow-up Survey Data**

*Includes de-identified data showing answers to BASC-3 questions that were used to create aggregated scores and indicators for Anxiety and Depression, collected at the end of program year 1. It also includes data showing answers to CATS questions that were used to create aggregated scores/indicators for PTSD, collected at the end of program year 1. This file also includes data on alternative BASC-3 questions, used for secondary metrics (e.g., Self-esteem, self-reliance, etc.). The number of independent survey responses included in this file is 1487.*

**Data File S3: Survey Codebook**

*Shows the correspondence in between the variables included in the above datasets (e.g., q1\_1) and the questions that were asked, the scale they belong to, the answer's range, and the specific construct they belong to (if any).*