Exploring the Relationship between the *It's On Us* Sexual Assault Prevention Campaign and Students' Active Bystander Practices

Winn Decker, Tiffany J. Davis (Affiliation), Sang Lu, Xiaoping Zhang,
Jon Stallings (Faculty Advisor)

Introduction

As a sexual assault prevention and bystander intervention program, the It's On Us campaign offers the tools and resources necessary to engage students, particularly men, to have difficult conversations about sexual assault. Research on the It's On Us campaign will help provide guidance to initiative leaders regarding next steps of the campaign, as well as help student affairs leaders and institutions determine if they should allocate funds and time toward the campaign. Furthermore, as this is a campaign supported by the White House, research outcomes will help sustain the campaign and drive policy decisions with data and evidence. As a new take on sexual assault prevention education, research is necessary to see if the *It's On Us* Campaign empowers students to be active bystanders. The purpose of this study is to examine the effectiveness of the *It's On Us* Campaign in engaging students in active bystander practices. Bystander intentions, bystander confidence and bystander actions were all studied and used to answer these questions:

(1) Make inference from demographics of students that participate in *It's On Us* programs. What is the relationship between institutional differences in type and size and bystander intentions and behavior? Is there a difference between gender or sexual orientation on bystander intentions and behavior?

(2) Does the *It's On Us* Campaign effectively engage students in positive bystander actions? Are students more active bystanders because of the *It's On Us* Campaign? What is the relationship between students' involvement in the *It's On Us* campaign and measure of bystander intention and behaviors?

(3) From a subjective perspective, does the *It's On Us* Campaign increase students' confidence in acting as a bystander? From the students' perceptions, how are the popularity and effectiveness of different types of programs in this Campaign?

Data

All data is collected during a two-month data collection period during spring 2017 via a self-administered, untimed questionnaire. We select the information of interest in the survey to create a dataset that includes the following compositions:

- Demographic information: class standing, race/ethnicity, sexual orientation, gender, and student involvement. The demographic questions represent the variables that are grouped (i.e., race/ethnicity and student involvement) for the data analysis and interpretation of participants' active bystander practices scores.
- Involvement information: frequency, time, and duration of participations in *It's On Us* programming and associated activities are recorded.
- Perceptions of these different programs' effectiveness: knowledge, skills, and ability to be active bystanders.
- Bystander practice: measured by a behavioral test composed of 38 questions to examine the participant's intention and actions in the past (before joining the campaign) and now (after joining the campaign). An example of the practice question is: Use 1-not at all likely to 5-very likely to gauge the likelihood of this practice -- Checked-in with a friend who looked drunk when she went to a room with someone else at a party.

We selected students in college and the ones who finished the total questionnaire (a total of 1835 responses) to conduct data analysis for this research. There are a total of 68 survey questions being studied and currently 1032 students being recorded. For each record, missing data for a specific survey question will be only excluded in that indicator but will not interfere with the analysis for other indicators. Notice that for gender information, there used to be 7 levels in the original survey question. However, with a majority of population falling into 'Woman or Female or Feminine' and 'Man or Male or Masculine' category and less than 4% falling into other levels, we decided to only consider the difference between the 'Woman or Female or Feminine' and 'Man or Male or Masculine' for our current research.

Methodology

1. The definition of bystander practice score

Before we investigate the factors that would affect the bystander practice, a criteria to define 'good' or 'bad' for bystander practice should be determined. In the survey, bystander practice is measured by a behavioral test composed of 38 questions to examine the participant's intention and actions in the past (before joining the campaign) and now (after joining the campaign) with 5 levels (from 1:not at all likely to 5: very likely) . The levels in each question can also be considered a score that can represent how strong are the participant's intentions or behavior. Hence, it's reasonable to define an overall score to describe the bystander practice by adding up the 'score' for each sub-question.

However, the 38 survey questions examined different aspects of bystanders' practice. Based on previous research (McMahon et al., 2014) and some new modifications, we divide this questionnaire for bystanders' practice into four types of questions: (1) high-risk situations, (2) intervention opportunities before, during, or after an assault (3) proactive opportunities. (4) low-risk situations.

Based on the four aspects, we calculate the participant's' current practice score for each of them and make further analysis separately:

High-risk situation score = Sum of the scores corresponding to the bystander's practice under high-risk situations (composed of 3 survey questions, so it ranges from 3-15).

Intervention score = Sum of the scores corresponding to the bystander's practice concerning intervention opportunities before, during, or after an assault (composed of 7 survey questions, so it ranges from 7-35).

Proactive score = Sum of the scores corresponding to the bystander's practice concerning proactive opportunities (composed of 4 survey questions, so it ranges from 4-20).

Low-risk situation score = Sum of the scores corresponding to the bystander's practice under low-risk situations (composed of 4 survey questions, so it ranges from 4-20).

2. Statistical Tests for the relationship between demographic information and bystander practice

After the determination of how the practice is measured by four aspects, we look into the distributions of participants' scores for the four aspects and find that, for each of the four aspect, the distribution is obviously left skewed. For example, in high-risk situation aspect (Figure 1), most of the participants tend to score themselves 4 or 5 instead of 1 to 3 for each survey question so the scores are not normally distributed. Due to this characteristic of the distribution, t-test and analysis of variance (ANOVA) which have the assumption of normality are not appropriate in our case.

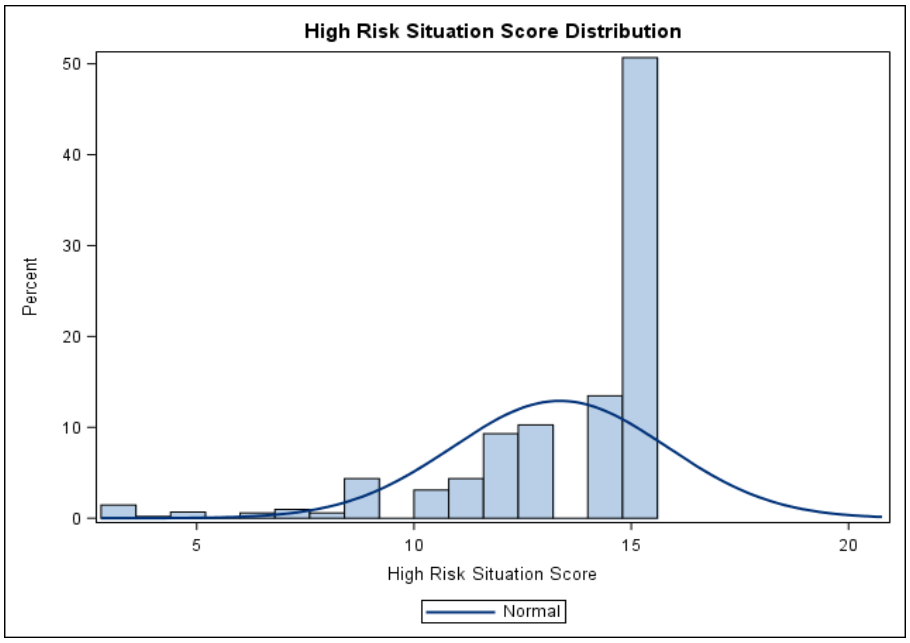

Figure 1. High-risk situation score distribution

Since our primary interest is to test the association between demographic indicators and participants' practice, we group the data according to demographic indicators, calculate the proportion of people who have the average score of at least 4 for each group and test for equality of the proportions using the Pearson chi-square test. This is also equivalent to the Z test for comparing two independent proportions.

For example, when testing the gender difference in high-risk situation score, we calculate the proportion of people who have the average scores of at least 4 (which leads to the overall score at least 12) for both male and female group.

The data can be arranged in a 2×2 table:

Table 1. Proportion test for gender difference in high-risk situation score

	Proportion of score at least 12	Proportion of score less than 12
Male	60.11%	39.89%
Female	78.17%	21.83%

It is assumed that each man in the sample of male had the same probability of responding score of 12 or more. The same is assumed for the sample of female, though of course, the two probabilities may differ. We test the equality of two proportions obtained from independent samples using the Pearson chi-square test. The null hypothesis can be denoted by:

H_0: The proportion of people having score at least 12 are equal in male and female.

Decision rule is to reject H_0 if the calculated p-value < .05.

The same test is implemented for all the demographic indicators.

3. Logistic regression for the relationship between students' involvement in It's On Us campaign and bystander practice

In order to investigate if It's On Us campaign has effectively engaged students in active bystander practice, a regression model is built to describe the relationship between students' involvement and bystander practice. Binary logistic regression estimates the probability π that a characteristic is present given the values of explanatory variables, in this case variables that indicates the previous practice, involvement activity and demographics.

The logistic regression model has the following assumptions:

1. The data $Y_1, Y_2, ..., Y_n$ are independently distributed, i.e., participants are independent.
2. Distribution of Yi is Bernoulli(n_i, π_i), i.e., binary logistic regression model assumes binomial distribution of the response.
3. Assume linear relationship between the logit of the response and the explanatory variables; $\text{logit}(\pi_i) = \beta_0 + \beta X_i$

In our study, the sample size is sufficiently large with each observation independent to meet the assumptions.

Let Y be a binary response variable,

$Y_i = 1$ if the average practice score ≥ 4 in observation (person) i;

$Y_i = 0$ if the average practice score < 4 in observation (person) i;

Let $X = (X_1, X_2, ..., X_5)$ be a set of explanatory variables. x_i is the observed value of the explanatory variables for observation i.

$\pi = \Pr(Y = 1|X = x)$ = probability of "having the average practice score ≥ 4 for the specific aspect"

Model 1:

$$\text{logit}(\pi_i)=\log(\pi_i/\ 1-\pi_i)= \beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\beta_3 X_{i3}+\beta_4 X_{i4}+\beta_5 X_{i5}$$

where, X_1= a set of categorical variables of previous score with '1' representing the response 'Yes/likely' and '0' representing the response 'No/not likely' for each aspect of practice. Eg. for high-risk situation score, X_1 is composed of Q17_2, Q17_4, Q17_5, with each of them has the level 1 or 0 indicating whether the question were answered 'Yes/likely' in the test about practice in the past .

X_2 = categorical variable of the involved time, which has 4 levels

X_3 = total counts of programs attended

X_4 = categorical variable of whether the participant has pledged

X_5 = a set of significant demographic indicators derived from statistical tests

Model 2:

$$\text{logit}(\pi_i)=\log(\pi_i/\ 1-\pi_i)= \beta_0+\beta_2 X_{i2}+\beta_3 X_{i3}+\beta_4 X_{i4}+\beta_5 X_{i5}$$

Compared to model 1, model 2 drops out the previous score indicators.

4. Evaluation on the popularity, effectiveness and controversy of different programs in It's On Us campaign

In order to verify whether the *It's On Us* Campaign has increased students' confidence in acting as a bystander from a subjective perspective, a set of descriptive statistics is drawn from the data to investigate the popularity, effectiveness and controversy of different programs in *It's On Us* campaign.

For the popularity of a program, we use the total counts of students who attended this program as the measurement since the participant size in program is the most appropriate statistics to describe the popularity. The total counts were calculated from the involvement information.
For the effectiveness of a program, we collected students' perception (measured by a score from 1-5) towards each program they attended and use a mean score to represent the effectiveness of this program.
Mean score = Sum of the scores for this program/ Sum of the students who attended this program,
Look further into the scores, we can get a standard deviation of students' perception to a specific program and this standard deviation can be interpreted as the controversy of this program, as a large standard deviation means a large variation in students' perception towards it.
There are three ways through which people can know or participate in *It's On Us* Campaign. People may attend a specific program or experience marketing or involve in the broad program. From the comparison of programs falling into each of the three categories and the internal comparison of the programs in one category, we can advised on the design of program in the future for *It's On Us* campaign. For the comparison of effectiveness between the three categories of programs, we calculate the proportion of people who scored at least 4 for each category and test for equality of the proportions using the Pearson chi-square test.

Results

1. Statistical Tests for the relationship between demographic information and bystander practice
In the results, a table of significant factors that would affect bystander practice score is listed below. The small p-value for the Pearson chi-square test ($p<.05$) indicates that the null hypothesis of equal proportions can be rejected and that the proportions are unequal.

Table 2. Significant demographic indicators for four aspects of bystander practice score

Aspect of bystander practice	Significant demographic indicators
High-risk situation	Gender, Race
Intervention opportunities	NA
Proactive opportunities	Gender, School type

Low-risk situation	Class status (marginal significant with p-value=0.07)

From the test results, we can conclude that, in the four aspects of practice, high-risk situation score is significantly affected by gender and race; proactive opportunities score is significantly affected by gender and school type; Low-risk situation score is strongly affected by class status with p-value=0.07; While we didn't find any demographic indicators significantly affect intervention opportunities.

Specifically, for high-risk situation score, female group has 18.06% higher of people who have the average score at least 4 than male group (Table 1). While for proactive opportunities, female group has 27.34% higher proportion of 'good' response (Table 3).

Table 3. Proportion test for gender difference in proactive opportunities score

	Proportion of score at least 16	Proportion of score less than 16
Male	30.05%	69.95%
Female	57.39%	42.61%

School type only has the significant effect on proactive opportunities score with public school has 8.35% higher proportion of students with higher score (Table 4).

Table 4. Proportion test for school type difference in proactive opportunities score

	Proportion of score at least 16	Proportion of score less than 16
Private	46.63%	53.37%
Public	54.98%	45.02%

For race, we can conclude that there might be some clue showing its effect on practice score since the majority (more than 70%) of participants are white and there are not enough sample size to represent the performance of other race. However, when we simply divide the population into white group and non-white group, the white group have a slightly better performance while the difference appears to be not significant (Table 5).

Table 5. Proportion test for race difference in high-risk situation score

	Proportion of score at least 12	Proportion of score less than 12
White	75.22%	24.78%
Non-white	73.26%	26.74%

2. Logistic regression for the relationship between students' involvement in It's On Us campaign and bystander practice

For model 1 (previous score included), most of the significant indicators are previous scores (Table 6). When previous scores are not included, gender, total number of programs attended and involved time are significant factors that would affect practice score (Table 7).

Table 6. Significant indicators in model 1 for four aspects of bystander practice score

Aspect of bystander practice	Significant indicators in logistic regression model
High-risk situation	Previous score (Q17_2)
Intervention opportunities	NA
Proactive opportunities	Gender, previous score (Q17_16, Q17_17,Q17_19)
Low-risk situation	Previous score (Q17_1,Q17_9)

Table 7. Significant indicators in model 2 for four aspects of bystander practice score

Aspect of bystander practice	Significant indicators in logistic regression model
High-risk situation	Gender, total number of programs attended
Intervention opportunities	Total number of programs attended
Proactive opportunities	Gender, involved time, total number of programs attended
Low-risk situation	NA

The comparison of the two models demonstrates that, the student's current practice is mainly dependent on previous practice. However, when previous score is excluded, total number of programs attended is another strong indicator of current practice score.

3. Evaluation on the popularity, effectiveness and controversy of different programs in It's On Us campaign

For each category, properties are shown below in the tables for each program.

Table 8. Properties of the programs in 'Specific' Category

	Total counts	Mean Score	Standard Deviation
It's On Us pledge drive/tabling	1099	3.103	1.161
Round Table Discussion	144	3.972	0.953
It's On Us Freeze	48	3.542	1.010
Dorm Storm	30	2.933	1.388
Open Mic Night	106	3.330	1.177
It's On Us Rally	253	3.534	1.143
Student Panel on Sexual Assault	332	3.819	0.979
Administrative Panel on Sexual Assault	159	3.730	1.071
Student and Administrative Panel on Sexual Assault	206	3.786	1.047
The Hunting Ground film screening	656	3.899	1.092
Film screening (other than the Hunting Ground)	135	3.526	1.119
It's On Us PSA (Public Service Announcement)	487	3.522	1.069
It's On Us Twitter Rally/Town Hall	309	3.333	1.207

Table 9. Properties of the programs in 'Marketing' Category

	Total Counts	Mean Score	Standard Deviation
Vice President Joe Biden's Speech at the Grammy Awards 2016	1232	3.642	1.191
Lady Gaga's performance at the Grammy Awards 2016	1447	3.572	1.266
Vice President Biden's Open Letter to a Courageous Young Woman	1055	3.755	1.179
Celebrity Endorsement/ Interviews for the It's On Us Campaign	925	3.360	1.215
It's On Us PSA during Sweet/Vicious	414	3.208	1.249
It's On Us Commercials during NCAA sporting events	1854	3.025	1.252
It's On Us Emails	1554	3.219	1.183
It's On Us Facebook	886	3.284	1.208
It's On Us Twitter	1273	3.641	1.142
It's On Us Official Website	322	3.584	1.187
Interactions with an It's On Us Student Advisory Committee Member	114	3.175	1.384
Media articles about the campaign	591	3.504	1.154

Table 10. Properties of the programs in ‘Broad’ Category

	Total Counts	Mean Score	Standard Deviation

Round Table Discussion	288	4.024	1.067
Open Mic Night	149	3.443	1.249
Student Panel on Sexual Assault	528	3.905	1.056
Administrative Panel on Sexual Assault	287	3.864	1.106
The Hunting Ground Film Screening	607	4.031	1.022
Film Screening	155	3.735	1.039
Online Webinar/Panel	305	3.630	1.047
School Made PSA	571	3.508	1.128

For the comparison of general category of programs, the proportion test results are shown in table 11.

Table 11. Proportion test for program difference in students' evaluation score

	Number (proportion) of score ≥4	Number (proportion) of score <4
'Specific' programs	2313(33.2%)	4648(66.8%)
'Marketing' programs	5799(33.2%)	11667(66.8%)
'Broad' programs	2271(35.5%)	4135(64.5%)

We can see that the three categories of programs have very close proportion of people who gave 'good' evaluation with score ≥4. It is noticeable that student tend to consider only about ⅓ of the programs they attended as "effective". Also, from the Pearson Chi-square test result, 'broad' programs have significantly better performance than the other two types of programs.

Discussions

1. Conclusions and result interpretation

From the results, there are several conclusions of interest in this study:

- Among all the demographic indicators, gender is the most important factor that affect the practice score. Female tend to have better performance in practice test. This might because female is the main population of victim, who would be more likely to learn how to protect themselves and other female victims.
- The school type (public school and private school) has significant effect on the student's proactive opportunities. Students in public school tend to have better performance, which might due to the difference in school management or education on sexual assault prevention. This might also give us a clue that education on campus is important for students' behavior, creating an environment in which sexual assault is unacceptable should be more extensively advocated by schools.
- Since the overwhelming majority of participants were white (70%), this shows the need for It's On Us to further expand its efforts to reach underrepresented racial populations. The origins of the sexual assault prevention education are rooted in the white feminist movement, and other groups may have been slower to adopt active bystander education. By creating programming that incorporates students from multiple racial groups, it may support a community approach to bystander intervention.
- When 'involved time', 'total number of programs attended' and 'whether the student have taken the pledge' are used in the predictive model, 'total number of programs attended' tend to be the 'strongest' indicator that can be used as the measurement of involvement activity. Specifically, the total number of programs attended can influence the student's behaviors or intentions in high-risk situation, intervention opportunities and proactive opportunities. This can partly indicate the effectiveness of programs of It's On Us campaign. This shows that the simple mandatory training for incoming first-year students is not enough, as this usually only happens during orientation. Instead, schools should focus on a holistic, year-long approach that encourages students to attend multiple educational programs throughout the year.
- When consider the two models together, we find that previous score is the main factor that would affect the student's practice score. Specifically, for high-risk situation, the

score of one question can be the only significant indicator in the several previous practice score. This may indicate that some specific questions in the questionnaire can highly represent students' bystander behavior and intentions but the other questions might be not that representative.

- From the comparison of specific programs in three categories, we find the programs with relatively higher effectiveness (high average score) include 'Round Table Discussion', 'Student Panel on Sexual Assault' , 'The Hunting Ground Film Screening', etc. These programs with high effectiveness has a common feature: they all highly involve students themselves on discussion or communication. Students tend to be more engaged during this kind of programs and they can express and exchange their ideas or experience through this kind of programs. This result can be a reference for the future design of program in It's On Us campaign: We might consider to increase or advocate on more programs which allow students to take their initiative or they can highly engaged themselves into the program. The peer education and students' own discussion might be very efficient for education. Media presences was a large component in the early years of the It's On Us campaign, as to help spread the idea of the campaign. Now that there is more name-recognition, It's On Us should work to develop more programs that integrate students into the conversation, while also supporting the media presence to grow to new communities. However, Public Service Announcements and other short videos do not appear to be sufficient alone.
- Student involvement in educational activities appears more effective than those that focus on an administrative approach. These top down educational programs from administration, while good intentioned, may create a disconnect between students and the material. By including students in the programming and training, students appear to engage more with the materials.
- From the overall comparison of three categories of programs. The programs with average score higher than four is around ⅓, which is much lower than our expectation. The design of programs might be not desirable enough currently. We might more closely hear the

feedback of each program from participants and pay more attention to the effectiveness of programs.

2. Limitations and future directions

There are several limitations that should be noted when interpreting the results:

1. Generalization issues: First, the questionnaire was previously designed for a longitudinal study of students' bystander practice overtime and the two parts of practice, practice in the past and practice currently, were measured at least two times. But in our study, due to restriction of time and resources, students' practice in the past and current practice were measured at the same time in a same questionnaire. It increases the difficulty of detecting the change of practice score from now to the past. We can only based the measurement of previous behavior on the student's memory and that can certainly bring inaccuracy to the results. Second, the questionnaire were previously based on a sample that was composed of incoming students on 1 particular campus; therefore, these findings are limited to this sample. Geographic differences or variations in campus culture might cause problems on generalization to all college campuses, and so replication of the analyses with additional samples is needed to see if the factor structure remains. In our case, sample population is mainly composed of students that have involved in It's On Us campaign, most of them are active bystander with higher practice score than normal students and this result might not be able to reflect an average level of college students' practice.
2. Questionnaire design limitations: Due to the restriction of the form of questionnaire, the students' responses are all subjective. The self-evaluation scores might be not accurately reflect their real intentions or behaviors. Additionally, the questions on the questionnaire the are not meant to be exhaustive lists of bystander attitudes or behaviors. Besides, the measurement scale of previous score and current score is not consistent. For previous score, the responses all include two levels (Yes or No) while for current score, the responses include 5 levels. This also increase the difficulty in measuring change of practice score. Further testing on these indicators is needed to confirm the reliability of current questionnaire and future investigation of a more representative questionnaire with consistent measuring scale is expected.

3. Regression model choices: In our study, a logistic regression with binary outcome is used to find the possible significant factors that affecting bystander practice score, which means we have projected the multi-level measure to 2 levels and the criteria (average score higher or lower than 4) is defined 'manually'. This method can probably lose some information of the data and make it harder to detect the significant factors. Also, from the correlation test results, there is a significant positive correlation between previous score and total number of programs attended, which means students with higher score are proved to be more actively involved in these programs. As we mentioned, the majority of sample population in our study is students with high practice score, thus, most of them have very small space of improvement in practice score so it's even harder to detect the individual effect of involvement on the practice score. Model 2 can be a reference of the campaign's effectiveness but a improved model is needed in future study.

References

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

It's On Us. (2015). Retrieved November 1, 2015 from http://itsonus.org/#tools

McMahon, S., Allen, C. T., Postmus, J. L., McMahon, S. M., Peterson, N. A., & Lowe Hoffman, M. (2014). Measuring bystander attitudes and behavior to prevent sexual violence.

McMahon, S., Peterson, N. A., Winter, S. C., Palmer, J. E., Postmus, J. L., & Koenick, R. A. (2015). Predicting bystander behavior to prevent sexual assault on college campuses: The role of self-efficacy and intent. *American Journal of Community Psychology, 56*(1-2), 46-56. doi:10.1007/s10464-015-9740-0

The White House. (2014). The first report of the White House task force to protect students from sexual assault. Retrieved May 9, 2014 from http://www.whitehouse.gov/sites/default/files/docs/report_0.pdf.

Appendix

Table 12. Demographic Questions in the Survey Being Studied

Question Index	Demographic information	Number of Options
Q1	Class Status	8 levels
Q2	College Type (private/public)	2 levels
Q3	College Type (2-year/4-year)	2 levels
Q4	College Scale (size)	5 levels
Q5	Major	Text
Q6	Housing	3 levels
Q7	Working hours	5 levels
Q8	Activities/Clubs	Text
Q9	Leadership	Text
Q32	Age	5 levels
Q33	Race	8 levels
Q34	Gender	7 levels
Q35	Sexuality	10 levels

Table 13. Involvement Questions in the Survey Being Studied

Question Index	Involvement information	Number of Options
Q47	Popularity (heard of)	2 levels
Q54	Ways to know about it	Text
Q48	Pledge (involvement)	2 levels

Q10	Involved time	4 levels
Q52	Present on campus	2 levels
Q53	Supporting methods	4 levels
Q12	Programs attended-specific	14 levels
Q13	Marketing experience	11 levels
Q11	Broad Programs attended	10 levels

Table 14. Perceptions of Different Programs in the Survey Being Studied
(All questions in this part are measured by score 1-5, with 1: not useful at all to 5: very useful)

	Specific Programs- See Q12	Marketing-See Q13	Broad programs-See Q11
Awareness	Q26	Q36	Q21
Knowledge	Q27	Q37	Q24
Skills	Q28	Q38	Q25

Table 15. Bystander Intention and Bystander Actions in the Survey Being Studied
(All questions in this part are measured by score 1-5, with 1: not at all likely to 5: very likely)

Question Index	Time	Number of Options
Q17	In the past	19 sub-questions, 2 levels for each sub-question
Q22	Now	19 sub-questions, 5 levels for each sub-question