Construct Validity of the General Outcomes of Bystander Action Scale: When Bystander Consequences Predict Future Bystander Behavior

Alison Krauss

Southern Methodist University, alkrauss@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_psychology_etds

Part of the Clinical Psychology Commons

Recommended Citation


This Dissertation is brought to you for free and open access by the Psychology at SMU Scholar. It has been accepted for inclusion in Psychology Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
CONSTRUCT VALIDITY OF THE GENERAL OUTCOMES
OF BYSTANDER ACTION SCALE: WHEN BYSTANDER
CONSEQUENCES PREDICT FUTURE
BYSTANDER BEHAVIOR

Approved by:

__________________________
Ernest N. Jouriles, Ph.D.
Professor of Psychology

__________________________
Renee McDonald, Ph.D.
Senior Associate Dean
Professor of Psychology

__________________________
David Rosenfield, Ph.D.
Professor of Psychology

__________________________
Victoria L. Banyard, Ph.D.
Professor of Social Work
CONSTRUCT VALIDITY OF THE GENERAL OUTCOMES
OF BYSTANDER ACTION SCALE: WHEN BYSTANDER
CONSEQUENCES PREDICT FUTURE
BYSTANDER BEHAVIOR

A Dissertation Presented to the Graduate Faculty of the
Dedman College
Southern Methodist University
in
Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy
with a
Major in Clinical Psychology
by
Alison Krauss
B.S., Psychology, Northern Illinois University
Masters of Arts, Southern Methodist University

August 4, 2021
Sexual and physical relationship violence are prevalent public health concerns on college campuses. Bystander training programs have recently been introduced as a means to prevent or intervene upon such violence, and generally lead to increases in prosocial bystander behavior. However, little is known about the consequences of engaging in prosocial bystander behavior. The General Outcomes of Bystander Action Scale (GOBAS) has been proposed to examine such consequences, and preliminary evidence points to some validity of this scale. The current research aims to add upon this research by replicating the factor structure of the GOBAS and providing evidence of criterion validity. In Study 1, we utilized a sample of 730 undergraduate students to perform a confirmatory factor analysis on the GOBAS. Although we were able to replicate the original factor structure, we could only do so when items on this measure were dichotomized. In Study 2, we utilized a subsample of our original sample ($N = 495$) to examine the criterion validity of the GOBAS. Results broadly confirm that the GOBAS predicts future bystander behavior, with important difference in this relation emerging as a function of moral courage. These findings point to the construct and criterion validity of the GOBAS and highlight the need to include these consequences in the study of prosocial bystander behavior.
# TABLE OF CONTENTS

LIST OF FIGURES .......................................................................................................................... VI

LIST OF TABLES ........................................................................................................................... VII

INTRODUCTION ............................................................................................................................. 8

   Establishing Validity ..................................................................................................................... 12

   Correlates of Consequences ....................................................................................................... 14

   Current Research ......................................................................................................................... 17

STUDY 1 ......................................................................................................................................... 18

METHOD ....................................................................................................................................... 18

   Participants .................................................................................................................................. 18

   Procedures ................................................................................................................................... 19

   Measures ..................................................................................................................................... 19

      Bystander Consequences .......................................................................................................... 19

   Data Analytic Plan and Sample Size Justification ........................................................................ 20

RESULTS ........................................................................................................................................ 20

   Descriptive Statistics ............................................................................................................... 20

   Confirmatory Factor Analysis .................................................................................................... 21

STUDY 2 ......................................................................................................................................... 26

METHOD ....................................................................................................................................... 27

   Participants .................................................................................................................................. 27
GENERAL OUTCOMES OF BYSTANDER ACTION SCALE

Procedures ........................................................................................................................................ 28

Measures ........................................................................................................................................ 28

Bystander Consequences .................................................................................................................. 28

Bystander Behavior ............................................................................................................................ 29

Moral Courage .................................................................................................................................. 29

Bystander Efficacy .............................................................................................................................. 30

Participant sex ..................................................................................................................................... 31

Data Analytic Plan and Sample Size Justification ............................................................................. 31

RESULTS ........................................................................................................................................... 32

Descriptive Statistics ........................................................................................................................ 32

Associations Between Positive Consequences, Negative Consequences, and Bystander Behavior ........................................................................................................................................ 33

Exploratory Analyses ......................................................................................................................... 35

GENERAL DISCUSSION ..................................................................................................................... 36

Limitations ........................................................................................................................................... 41

Future Research ................................................................................................................................. 44

Conclusion ........................................................................................................................................... 46

APPENDIX A: SUPPLEMENTARY RESULTS ..................................................................................... 47

Exploratory Analyses ........................................................................................................................ 49

REFERENCES ...................................................................................................................................... 50
LIST OF FIGURES

Figure 1 Two-factor model of the General Outcomes of Bystander Action Scale………………68

Figure 2 Visual representation of the regression coefficient between consequences and bystander behavior for individuals low in moral courage………………69

Figure 3 Visual representation of the regression coefficient between consequences and bystander behavior for individuals with average levels of moral courage………………………………………………………………………………………………70

Figure 4 Visual representation of the regression coefficient between consequences and bystander behavior for individuals high in moral courage………………71
LIST OF TABLES

Table 1 Descriptive statistics for each item of the General Outcomes of Bystander Action Scale .................................................................60

Table 2 Descriptive statistics for each item of the General Outcomes of Bystander Action Scale by sex ...............................................................61

Table 3 Sample size per response option by item on the General Outcomes of Bystander Action Scale ...............................................................62

Table 4 Kendall’s tau correlation coefficient between items on the General Outcomes of Bystander Action Scale ...............................................................63

Table 5 Maximum likelihood estimates with robust standard errors of factor loadings for enhanced model of the General Outcomes of Bystander Action Scale .............64

Table 6 Diagonally weighted least squares estimators with robust standard errors of factor loadings for the General Outcomes of Bystander Action Scale ..............65

Table 7 Means, standard deviation, and Kendall’s tau correlations of study variables to assess criterion validity .............................................................66

Table 8 Means, standard deviation, and Kendall’s tau correlations of study variables to assess criterion validity when utilizing the original item scoring method of the GOBAS .............................................................67
Introduction

Sexual and physical relationship violence on college campuses is a prevalent and harmful public health concern. A recent review of research on the prevalence of sexual victimization suggests that up to 25% of women and 18% of men report an unwanted sexual encounter while in college (Gialopsos, 2017). Additionally, around 26% of women and 30% of men report physical relationship violence victimization since beginning college (Milletich et al., 2010). Both sexual and physical relationship violence victimization results in a number of adverse outcomes, including depression, anxiety, somatization, and poor academic performance (Amar & Gennaro, 2005; Jordan et al., 2014). Sexual victimization is associated with additional concerns related to unprotected sex, such as sexually transmitted infections and unplanned pregnancy (Black, 2011).

In light of the high prevalence and adverse outcomes of sexual and physical relationship violence, many universities are implementing bystander training programs as one strategy to prevent such violence on college campuses. These programs target students who may be witnesses, or bystanders, to potential violence, and encourage students to act in ways to prevent violence and assist victims of violence. These acts, referred to as bystander behaviors, include checking on a friend and their partner after hearing sounds of yelling and fighting coming from their room, or stopping a friend from taking an intoxicated person to their room. Dozens of empirical studies have documented the efficacy of bystander training programs in increasing prosocial bystander behavior among college students. Indeed, recent meta-analytic reviews indicate that students exposed to bystander training programs report more bystander behavior
compared to those in control conditions (Jouriles et al., 2018; Kettrey & Marx, 2019; Katz & Moore, 2013). Although researchers have yet to examine whether such increases in bystander behavior result in reduced campus rates of violence, these initial findings are promising. In fact, the American College Health Association now recommends all universities implement bystander training programs in an effort to reduce sexual and relationship violence (American College Health Association, 2016).

As bystander training programs increase in popularity among universities, researchers are beginning to turn their attention to determinants of bystander behavior. For example, theory on bystander behavior to prevent sexual violence highlights the importance of risk awareness, accepting responsibility to intervene to prevent violence, and efficacy in one’s ability to successfully intervene (Banyard, 2011; Burn, 2009). Empirical findings largely support this theory, with efficacy to intervene emerging as one of the most consistent predictors of bystander behavior (Banyard & Moynihan, 2011). Others have documented the effect of peer norms around bystander behavior and feelings of connectedness at one’s campus as correlates of bystander behavior (Jouriles et al., 2020; Murphy Austin et al., 2016).

A variable that is likely to be a determinant of bystander behavior, but one that has received very little research attention to date, is the consequences or outcomes individuals experience after engaging in acts to prevent sexual and physical relationship violence, and to assist victims of such violence. Learning theory suggests that behavior is influenced by its consequences (Skinner, 1958). Very basically, experiencing positive events (i.e., reinforcement) after performing a behavior is thought to increase the occurrence of that behavior in the future, while experiencing negative events (i.e., punishment) in response to behavior is theorized to decrease the occurrence of that behavior in the future. Although researchers who are interested in
predicting bystander behavior have yet to fully examine how consequences of bystander behavior influence the occurrence of bystander behavior, learning theory suggests that it likely plays an important role.

In recent years, researchers have noted that a substantial number of students experience adverse outcomes after engaging in bystander behavior. For instance, Taylor et al. (2019) revealed that victims of violence report 15% to 23% of helpful bystanders are physically hurt or threatened when attempting to intervene. Similarly, Krauss and colleagues (2017) found that 16% to 20% of students who reported engaging in bystander behavior experienced a negative consequence, such as being physically hurt, threatened, harassed, or faced with legal problems. Such outcomes can clearly be harmful to students who are attempting to help others by engaging in bystander behavior. Such outcomes may also decrease the likelihood of these students engaging in bystander behavior in the future.

Understanding outcomes of bystander behavior has clear practical implications. For example, if bystander behavior can lead to severe negative consequences, such as being physically hurt, advocates of bystander training programs should carefully consider whether promoting bystander behavior inadvertently puts students at risk of harm. Although previous research on consequences to bystander behavior does not directly assess whether strategies in bystander training programs increase the risk for severe negative consequences, developers of such programs might consider what content best promotes bystander behavior while ensuring the safety of bystanders. For instance, if intervening in a heated argument between a couple is likely to lead to a bystander becoming physically hurt, bystander training programs should suggest alternative methods of intervention that reduce this risk, such as alerting a resident adviser (RA) or campus police of the situation. Relatedly, if bystander training programs promote behavior
that is likely to result in severe negative consequences, these programs may lose credibility among students and become less effective. For instance, if students feel a program is promoting behavior that may put them in danger of becoming physically hurt, they may view the program as less credible and ignore other useful strategies in the program.

Furthermore, the learning theory principles outlined previously have implications for researchers interested in promoting bystander behavior and examining predictors of such behavior. Specifically, bystander training programs could attempt to utilize natural positive reinforcements to increase bystander behavior among students. This could be achieved by promoting specific behaviors that are likely to result in positive consequences or encouraging students to provide praise to their friends who perform bystander behaviors. Similarly, programs should avoid encouraging behavior that likely leads to natural negative consequences in an effort to increases the likelihood of behaviors continuing beyond the program. Such strategies may enhance the effects of bystander training programs on bystander behavior.

In response to the need for a better understanding of the consequences and outcomes of bystander behavior, Banyard and colleagues (2019) developed a compendium of measures to assess consequences of bystander behavior. One such measure, referred to as the General Outcomes of Bystander Action Scale, assesses outcomes to engaging in any bystander behavior during the past 6 months. The measure includes three items assessing positive consequences (e.g., “People praised me for what I did”), and five items assessing negative consequences (e.g. “I was harassed because of what I did”). The development of this measure was guided by principles of learning theory that emphasize the importance of reinforcement (positive consequences) and punishment (negative consequences) in response to an individual’s bystander behavior.
The initial psychometric evaluation of this measure included an exploratory factor analysis which evidenced a two-factor solution: one factor reflecting positive consequences and one factor reflecting negative consequences. Additionally, a confirmatory factor analysis in an independent sample replicated this two-factor structure. Both samples used for the exploratory factor analysis and confirmatory factor analysis included a combination of psychology subject pool participants from a medium-sized New England public university and Amazon Mechanical Turk participants. Both the positive and negative consequences subscales evidenced adequate internal reliability in the study, with $\alpha = .85$ for positive consequences and $\alpha = .95$ for negative consequences. Both subscales were also related to respondent-specific measures of consequences (i.e., consequences from the victim or from the perpetrator of the event). Banyard et al. (2019) found that students reported an average of $3.60 (SD = 2.40)$ positive consequences and $1.49 (SD = 2.19)$ negative consequences over a 6-month reference period, with at least 20% of participants reporting each type of consequence. Male participants reported significantly more positive consequences ($M = 4.00, SD = 2.41$) and negative consequences ($M = 2.09, SD = 2.43$) compared to female participants ($M = 3.30, SD = 2.36$ for positive consequences; $M = 1.04, SD = 1.87$ for negative consequences). Investigation of this measure has not extended beyond this initial examination. The current study aims to provide additional validity for the General Outcomes of Bystander Action Scale by replicating its factor structure and examining theoretical correlates of consequence of bystander behavior.

**Establishing Validity**

The accuracy of a measure is determined by several forms of validity of the measure, including construct and criterion validity. Construct validity refers to the extent to which a scale measures the underlying theoretical construct in question (Cronbach & Meehl, 1955). Construct
validity is often determined through psychometric evaluation of a measurement model (i.e., through exploratory and confirmatory factor analysis). Such analyses lend insight into whether items on a measure are assessing a single construct and how much unique variance each item assesses compared to other items. In addition, exploratory factor analysis can illustrate if a measure is comprised of multiple related dimensions (in the case of the General Outcomes of Bystander Action Scale, these dimensions would include positive and negative consequences). Evaluation of a measurement model through these means is an iterative process that requires replication in multiple samples to ensure that factor structures are not dependent on sample-specific characteristics. In this case, the measurement model of the General Outcomes of Bystander Action Scale has been examined in a sample of psychology subject pool participants and community participants, but has not been replicated in a broader sample of college students. This particular population is important to consider given the popularity of bystander training programs in college.

Construct validity is also thought to subsume criterion validity, or the extent to which a measure is related to variables theorized to be associated with the construct. There are several types of criterion validity, including concurrent validity and predictive validity. Concurrent validity involves whether a measure is related to hypothesized correlates at the same time point (or concurrently). Predictive validity includes how well a measure predicts theoretical correlates across time. Banyard and colleagues’ (2019) research that introduced the General Outcomes of Bystander Action Scale included initial evidence of construct validity via factor analysis. However, criterion validity of this measure has yet to be examined.
Correlates of Consequences

As stated previously, learning theory suggests that experiencing positive events contingent upon behavior increases the likelihood of the behavior in the future, while experiencing negative events contingent upon the behavior decreases the likelihood of the behavior in the future (Skinner, 1958). Consistent with these learning theory principles, a large body of research points to peer influence as a key factor in prosocial behavior (for a review, see Brown et al., 2008). That is, peers can influence an individual’s behavior by providing approval or positive feedback contingent upon that behavior, as well as by providing disapproval or negative feedback contingent upon that behavior.

In one such study illustrating this learning theory principle in a peer context, adolescents were told they would participate in an online game assessing group decision making with three other group members (van Hoorn et al., 2014). Each participant was given tokens amounting to a small amount of money and were given the opportunity after each round to either donate some tokens to the group (conceptualized as the prosocial behavior) or retain all of their tokens. Adolescents were also told a group of 10 peers acted as spectators and provided feedback on their choices during the game. Each participant was randomly assigned to one of three spectator feedback conditions: approval for donating tokens, approval for keeping tokens, and no feedback. Adolescents who received peer approval for prosocial behavior (donating tokens) demonstrated a significant increase in prosocial behavior as the game progressed. Additionally, prosocial behavior decreased when this supportive feedback was no longer present. Adolescents who received approval for keeping tokens demonstrated a significant decrease in prosocial behavior during the game.
Recent work suggests prosocial bystander behavior among college students may be susceptible to similar peer influence. In one cross-sectional study, experiencing negative reactions from victims or other bystanders was associated with less self-reported willingness to engage in bystander behavior again, while positive reactions were correlated with greater willingness to engage in future bystander behavior (Moschella & Banyard, 2018). One might therefore expect that bystander consequences would predict future bystander behavior, such that positive consequences would result in more future bystander behavior and negative consequences would result in less future bystander behavior when presented with an opportunity to engage in such behavior.

Previous work suggests bystander behavior is often met with both positive and negative consequences. For instance, one study of college students found that 42% of students reported experiencing at least one negative consequence in response to bystander behavior, while around 36% reported at least one positive consequence (Moschella et al., 2018). Of students who reported at least one positive consequence, just under half (48%) reported a co-occurring negative consequence. Thus, according to the findings of this study, it is not uncommon for bystanders to experience both negative and positive consequences in response to bystander behavior.

Very little is known about the effects of concurrent positive and negative consequences on subsequent behavior. Within the field of behavioral learning, Farley and Fantino (1978) propose a model of behavior in which the propensity to engage in behavior is predicted by the absolute reinforcement value, a single composite score that takes into account the amount of reinforcement and punishment received. While this theory suggests that reinforcement and punishment act in equal and opposing ways, others have speculated that reinforcement and
punishment have unique effects on behavior that cannot be modeled as equal, opposing forces (Kubanek et al., 2015). There is scant research in the field of prosocial behavior on the predictive effects of co-occurring positive and negative consequences on later behavior. Thus, it appears important to model the co-occurrence of positive and negative consequences, although their interactive effects remain unknown.

Furthermore, the effect of positive or negative consequences of prosocial bystander behavior likely have different effects on different types of people. Specifically, some traits may predispose individuals to engage in bystander behavior to prevent sexual assault and relationship violence, regardless of the consequences they may experience (Banyard, 2015). One such trait is moral courage, or one’s willingness to engage in prosocial behavior despite high social costs (Osswald et al., 2010). In other words, by definition, individuals who engage in acts of moral courage are unlikely to be influenced by negative or positive consequences; rather, these acts are motivated by a pursuit of upholding moral values rather than avoiding negative consequences or gaining approval from others (Osswald et al., 2010). For example, individuals high in moral courage are theorized to be willing to defend a victim of sexual harassment or interrupt a situation involving racial discrimination even when faced with severe negative consequences, such as being threatened or attacked by the perpetrator (Galdi et al., 2017).

It is therefore likely that individuals high in moral courage are not influenced as much by consequences, compared to individuals low in moral courage. Indeed, research on prosocial behavior in related fields supports this assertion (Hannah et al., 2011; Poteat & Vecheo, 2016). In one such study, participants were randomly assigned to a condition that was theorized to either make them more or less likely to intervene in an instance of workplace sexual harassment (Galdi et al., 2017). Moral courage was positively related to intervention behavior, regardless of
condition, suggesting that this trait may strongly influence individuals to intervene in such situations independent of other encouragement to do so (or not to do so). Thus, individuals high in moral courage seem to engage in prosocial behavior not because of external encouragement but because such behavior aligns with their moral character.

Although some have incorporated moral courage into the theory of bystander behavior to prevent sexual assault on college campuses (Banyard, 2015), there are few, if any, empirical examinations of how this trait influences bystander intervention behavior. Despite this, research from other fields suggests a moderating effect of moral courage on the relation between consequences of bystander behavior and future bystander behavior. Specifically, individuals high in moral courage may be likely to intervene to prevent sexual assault regardless of their experience of positive and negative consequences, weakening the relation between consequences and future bystander behavior. For such individuals, intervening to prevent sexual assault may be in line with their moral character. In this case, negative consequences may be unlikely to deter this behavior, and positive consequences may not demonstrate reinforcement of this behavior. On the other hand, individuals low in moral courage may be susceptible to the effect of consequences on future behavior. For these individuals who may not consider intervening to prevent sexual assault in line with their values, negative consequences might act to discourage future bystander behavior while positive consequences may be likely to promote bystander behavior.

**Current Research**

The current research examines the construct validity of the General Outcomes of Bystander Action Scale in two studies. In Study 1, we aimed to replicate the factor structure of the General Outcomes of Bystander Action Scale in a sample of first year students. To date, there
are no investigations of the factor structure of this measure in an entirely college student sample. In Study 2, we used a subsample of this group to examine the criterion validity of this scale as it relates to future bystander behavior. In Study 1, we hypothesized that (1) a two-factor structure reflecting the positive consequences and negative consequences subscales would demonstrate adequate model fit.

**Study 1**

**Method**

**Participants**

First year undergraduate students (N = 730) were recruited during the fall semester of 2016 from three universities: two private universities in the South and Midwest and one public university in the Northeast. Based on Banyard and colleagues’ (2019) original procedure, we limited our analytic sample to reflect students who could have experienced a consequence during the study period. Specifically, individuals can only experience a consequence of bystander behavior if they have engaged in bystander behavior. Therefore, we excluded participants who reported no bystander behavior on the Bystander Behavior (for Friends) Scale and thus no opportunity to experience a consequence of bystander behavior at the baseline assessment (scale described below). This resulted in a sample size of n = 495.

Within this analytic sample (n = 495), the majority of students were female (65.9%), non-Hispanic (83.4%), and White (71.1%). Students also identified as Asian (8.9%), Black or African American (7.7%), “More than one race” (6.5%), American Indian/Alaska Native (0.6%), and Native Hawaiian or Other Pacific Islander (0.6%). Twenty students reported their race as
“Unknown or Not Reported” and three students did not report a race. The average age of students was 18.15 (SD = 0.71).

Participants included in our analytic sample differed from those not included on sex ($\chi^2 = 4.56, p = .03$) and race (coded 0 = white, 1 = non-white; $\chi^2 = 4.13, p = .014$). Specifically, inclusion in the analytic sample was associated with a higher proportion of male students and a higher proportion of white students. Participants in the analytic sample did not differ from excluded participants in age ($t(726) = -0.01, p = .99$) or ethnicity ($\chi^2 = 0.001, p = .97$).

**Procedures**

Data were collected as part of a larger study on the effectiveness of a bystander training program. The institutional review board at the first author’s institution approved all procedures. Students were recruited within the first two months of the fall semester of their first year; eligible participants included any first-year student 18 years or older. Participants completed the baseline assessment in a monitored research lab, dormitory, or university common space. Students completed their baseline assessment an average of 26.08 days after the start of the semester ($SD = 8.93$). Participants received course credit or a gift card for participation; students who did not wish to participate were offered an alternative assignment for course credit.

**Measures**

**Bystander Consequences**

Students rated their experience of positive and negative consequences at baseline using the General Outcomes of Bystander Action Scale (GOBAS; Banyard et al., 2019). Students reported how often they experienced each of the three positive consequences items and five negative consequences items during the past 2 months on a response scale ranging from 0 to 8+.
times (full scale is presented in Table 1). Items within the positive consequences subscale and negative consequences subscale are summed to reflect a composite positive consequences score and a composite negative consequences score; higher scores reflect more consequences in the past 2 months. To evaluate internal consistency, we utilized the greatest lower bond (GLB) coefficient as it is robust to many assumptions required for Cronbach’s alpha (Trizano-Hermosilla & Alvarado, 2016). GLB in the analytic sample was .87 for the positive consequences subscale and .83 for the negative consequences subscale.

Data Analytic Plan and Sample Size Justification

To evaluate our hypothesis, we conducted a confirmatory factor analysis (CFA) using maximum likelihood estimation with robust standard errors through R lavaan (Rosseel, 2012), a statistical package developed for structural equation modeling. We used several model fit indices to evaluate the fit of the 2-factor structure: root mean square error of approximation (RMSEA; adequate fit defined as ≤ .08), comparative fit index (CFI; adequate fit defined as ≥ .90), Tucker-Lewis index (TLI; adequate fit defined as ≥ .90), and root mean square residual (SRMR; adequate fit defined as ≤ .08) (Hu & Bentler, 1998). Pervious research suggests that with 2 factors and 8 indicator variables, a sample size of 160 results in adequate power (.80) to conduct a CFA (Wolf et al., 2013).

Results

Descriptive Statistics

Table 1 displays means, standard deviations, and ranges of each item on the GOBAS. We also report prevalence rates by item, or the percentage of participants who endorsed at least one
consequence. Across all items, 31.9% \((n = 158)\) of students reported at least one consequence during the past 2 months. Table 2 lists these descriptive statistics separately by males and females. Of note, both means and prevalence rates of consequences appear greater among male students compared to female students. We also examined whether students at each university differed in the number of consequences they experienced. Results of a one-way ANOVA revealed no difference across university for positive consequences, \(F(2, 492) = 0.48, p = .62\), or negative consequences, \(F(2, 492) = 0.10, p = .91\).

Table 3 reports sample sizes and percentage of participants reporting each response item of each item on this scale. As is evident from these tables, the majority of students reported “0” on each item, with positive consequence items endorsed more often than negative consequence items. The most frequently endorsed item was “People said positive things about me to others,” while the least frequently endorsed items were “I was physically hurt because of what I did” and “I got in trouble as a result of my action (e.g., charged with underage drinking).” With the exception of one participant, all students who reported a negative consequence also reported at least one positive consequence. On the other hand, of students who reported a positive consequence, 25.5% \((n = 40)\) reported at least one negative consequence. Table 4 lists correlations between each item. As expected, items within the same subscale generally displayed higher correlations than items across subscales, with some exceptions.

**Confirmatory Factor Analysis**

We first attempted to replicate Banyard and colleague’s (2019) factor structure of the GOBAS (Figure 1) through CFA. We specified that items 1 through 3 load onto a positive consequences latent factor, while items 4 through 8 load onto a negative consequences latent factor. We refer to this original model as the “base model.”
The base model evidenced poor model fit, RMSEA = .10 (90% confidence interval [CI] [.04, .15]), CFI = .90, TLI = .86, SRMR = .09 (Hu & Bentler, 1999). Modification indices suggested that we model the correlations between items 6 and 7, and items 7 and 8.

Theoretically, these items seem to capture similar content (i.e., significant negative consequences that are likely to involve a substantial amount of time or result in legal difficulties). We therefore reran the CFA modeling these correlations (we refer to this model as the “enhanced model”). The enhanced model evidenced poor-to-adequate fit, RMSEA = .08 (90% CI [.05, .12]), CFI = .94, TLI = .89, SRMR = .09. Specifically, this model evidenced adequate fit according to the CFI and evidenced poor fit by the RMSEA, TLI, and SRMR. Although modification indices suggested adding some parameters to improve model fit, these suggestions did not align with the theoretical conceptualization of bystander consequences (e.g., loading item 8 onto the positive consequences latent factor). We therefore retained this model as the model that evidenced the best model fit while maintaining the theoretical integrity of the construct. Table 5 presents unstandardized and standardized factor loadings for this model.

As the enhanced model evidenced poor fit across some model fit indices, we explored other strategies for improving model fit. Namely, descriptive statistics suggest that the items on this measure may not operate on a continuous scale ranging from 0 to 8. This seems to be especially true for the negative consequences items. We therefore reran the CFA using three different modified response scales that may more accurately capture how the measure operates. These modified response scales included: one where the highest option was 6+ consequences, one where the highest option was 3+ consequences, and one with a dichotomous response scale where 0 represents never experiencing that item and 1 represent experiencing at least one instance of that consequence.
GENERAL OUTCOMES OF BYSTANDER ACTION SCALE

We first reran the base model with the GOBAS recoded as 0 to 6+ consequences. This model again demonstrated poor model fit, RMSEA = .10 (90% CI [.05, .15]), CFI = .90, TLI = .85, SRMR = .10. We then reran the enhanced model with the recoded items; again, this model evidenced poor-to-adequate fit, RMSEA = .09 (90% CI [.06, .12]), CFI = .93, TLI = .88, SRMR = .09. Thus, this recoded response scale did not substantially improve model fit.

Next, we reran the model with the response scale recoded as 0 to 3+ consequences. Again, the base model evidenced poor fit, RMSEA = .11 (90% CI [.06, .15]), CFI = .88, TLI = .83, SRMR = .09, while the enhanced model suggested poor-to-adequate fit, RMSEA = .09 (90% CI [.06, .12]), CFI = .93, TLI = .88, SRMR = .08.

Finally, we reran the model with a dichotomous response scale using diagonally weighted least squares estimators with robust standard errors. Contrary to previous models, the base model with this revised response scale evidenced excellent model fit, RMSEA = .03 (90% CI [.05, .06]), CFI = .99, TLI = .99, SRMR = .09. Table 6 presents unstandardized and standardized factor loadings of this model. Thus, scoring each item dichotomously on the GOBAS resulted in the best model fit, while three different continuous scoring methods result in poor-to-adequate fit.

Discussion

Study 1 aimed to replicate the two-factor structure of the General Outcomes of Bystander Action Scale. When utilizing the original scoring method, and when modeling the correlation between items 6 and 7 and items 7 and 8, this measure evidenced poor-to-adequate model fit across several model fit indices. Thus, although Banyard et al. (2019) achieved adequate model fit with this factor structure, the current results suggest our data are not accurately modeled in this way. When using a dichotomous item scoring method, model fit was improved such that model fit indices demonstrated excellent model fit. Though we did not find support for
acceptable model fit using Banyard et al.’s (2019) original scoring method, we did replicate the 2-factor structure of the General Outcomes of Bystander Action Scale with a dichotomous scoring method. Altogether, these results suggest two types of related consequences (positive and negative consequences) that are best modeled dichotomously.

These findings suggest that items on the General Outcomes of Bystander Action Scale operate in a dichotomous fashion, rather than on a continuous scale. Conceptually, this suggests that it is less important to consider the number of times an individual experiences a consequence, and more important to consider whether or not they experienced the consequence. This conclusion is in line with other research suggesting that single or infrequent reinforcement and punishment can influence behavior (Bowman-Perrott et al., 2016). Thus, our results point to the importance of assessing whether or not a consequence has occurred during a given time frame rather than the number of times it occurred.

The magnitude of our standardized factor loadings of the final CFA solution also largely mirror those of Banyard et al. (2019); that is, factor loadings range from around .70 to .95. However, some differences emerged in the pattern of these loadings between our factor solution and Banyard et al.’s. Specifically, while Banyard et al. found that item 3 loaded strongest on the positive consequences subscale, we found item 1 loaded with the highest factor loading. Similarly, Banyard et al. found that item 4 loaded strongest on the negative consequences subscale, while we found that both items 4 and 5 evidenced the highest factor loading. However, these rank orders are driven by seemingly small differences in factor loadings (e.g., Banyard et al. reported item 4’s factor loading as .94, whereas we found item 5 evidenced a factor loading of .95). Thus, these differences in the magnitude of factor loadings between the two studies may not be meaningful.
Interestingly, item 6 evidenced a somewhat low factor loading (.69) in the current study, although the factor loading reported by Banyard et al. (2019) was .91. This factor loading may be due, in part, to the low prevalence rate (0.6%) of this item in the current study, especially among female students (0.2%). Variability of this item is likely limited due to the low endorsement rate, which may have skewed the factor loading in the current study. Thus, further replication of the factor structure is warranted.

There are a number of reasons why our findings regarding the optimal scoring method diverge from Banyard et al.’s (2019). For one, our findings may be due to the sample used in the current study, and are thus in need of replication in other college samples. In addition, the time frame used in the current study (past 2 months) is shorter than that used by Banyard and colleagues (past 6 months). This time frame may have inadvertently restricted the variability in the number of consequences students could experience in the study period, which may have influenced our results. For instance, it may be that students do not have much opportunity to experience multiple consequences in a 2-month time period, which may limit our ability to test a factor structure of a frequency scale rather than a dichotomous scale. However, a 2-month time frame is commonly used in research on bystander behavior (Banyard et al., 2014; Bennet et al., 2014; Jouriles, McDonald, et al., 2016); thus, our results may be representative of common research designs in this field.

Overall, the results of Study 1 suggest that consequences of bystander behavior are best modeled as two separate dimensions (positive consequences and negative consequences) of dichotomously coded items. We were able to partially replicate the findings of Banyard et al. (2019), providing additional evidence of the construct validity of this measure. However, aspects of construct validity, such as criterion validity, for this measure remain unknown.
Study 2

After replicating the factor structure of the GOBAS in Study 1, we aimed to extend this work in Study 2 by evaluating the criterion validity of this measure with theorized correlates of consequences, namely future bystander behavior. More specifically, as reviewed earlier, learning theory suggests that experiencing positive consequences as a result of bystander behavior increases the likelihood of bystander behavior in the future, while experiencing negative consequences as a result of bystander behavior decreases the likelihood of the behavior in the future. As moral courage appears important in considering the influence of consequences on future bystander behavior, we also examined the moderating effect of moral courage on the relation between positive and negative consequences and bystander behavior.

In Study 2, we hypothesized that for individuals low in moral courage, (1) positive consequences at baseline would predict more bystander behavior at a 2-month follow up, (2) while negative consequences at baseline would predict less bystander behavior at a 2-month follow up. We also hypothesized that for individuals high in moral courage, (3) neither negative consequences nor positive consequences would predict future bystander behavior. While we modeled the interactive effects of positive consequences and negative consequences to account for their co-occurrence, we made no a priori hypothesis about their interactive effect on bystander behavior. We also included two consistent predictors of bystander behavior (efficacy to engage in bystander behavior and sex) as control variables in this model to determine if positive and negative consequences are predictive of bystander behavior above and beyond other established determinants.
Method

Participants

Participants in Study 2 were drawn from the sample \((n = 495)\) we used in Study 1. Students who participated in Study 1 also were asked to participate in a 2-month follow-up which included an assessment of bystander behavior. However, students cannot engage in bystander behavior without an opportunity to engage in bystander behavior. That is, an individual must witness a potentially violent situation (e.g., seeing someone take an intoxicated person up to their room) in order to engage in bystander behavior to prevent violence (e.g., stopping that person and checking in with them). We therefore further limited our sample in Study 2 to students who both participated in the 2-month follow up \((n = 457)\) and reported at least one opportunity to engage in bystander behavior at the 2-month follow up on the Bystander Behavior (for Friends) Scale. This resulted in an analytic sample of \(n = 357\).

Within this analytic sample, the majority of students were female (66.1%) and non-Hispanic (86.0%). The majority also identified as White (73.4%), followed by Asian (8.7%), Black or African American (7.6%), “More than one race” (5.6%), American Indian/Alaska Native (0.8%), and Native Hawaiian or Other Pacific Islander (0.8%). Eight participants indicated their race as “Unknown or Not Reported”, and three participants did not indicate a race. The average age of participants was 18.15 \((SD = 0.80)\).

Students in our analytic sample for Study 2 differed from those excluded from the Study 1 sample in terms of ethnicity \(\chi^2 = 6.07, p = .02\). Specifically, students in the analytic sample were more likely to be non-Hispanic than excluded students. Participants in the analytic sample
did not differ from excluded participants on sex ($\chi^2 = 0.04, p = .92$), race ($\chi^2 = 3.78, p = .06$), or age ($t(492) = -0.02, p = .98$).

**Procedures**

Data for Study 2 included the same baseline assessment described in Study 1. In addition to completing this assessment, students were invited to return to the lab approximately 2 months later for a follow up assessment ($M = 56.96$ days after baseline, $SD = 9.05$ days). Of the 495 students who completed the baseline assessment, 457 returned for the 2-month follow up (92.3% retention).

**Measures**

**Bystander Consequences**

Students rated their experience of positive and negative consequences at baseline using the General Outcomes of Bystander Action Scale (GOBAS; Banyard et al., 2019), which was described in Study 1. Study 1 suggested a dichotomous scoring method best fit the data in the current research. We therefore recoded the item response scales such that 0 represents the absence of that item and 1 represents experiencing that item at least once in the past 2-months. Items were then summed within each subscale to form a total positive consequences and negative consequences score. This item scoring method results in a slightly different interpretation of each subscale. Specifically, scores represent the number of different types of consequences experienced in the past 2 months within subscales, rather than the total number of consequences in the past 2 months. GLB using the dichotomized scoring method was .87 for the positive consequences subscale and .72 for the negative consequences subscale in the current sample.
Bystander Behavior

At baseline and the 2-month follow up, participants completed the 19-item Risky Situations subscale of the Bystander Behavior (For Friends) Scale (BBS; Banyard et al., 2014), a self-report measure of bystander behavior. Students reported their engagement in each of the 19-items over the past 2 months. We modified the response format such that students indicated whether they had an opportunity to intervene, and, if so, whether they intervened with prosocial bystander behavior. For example, the item “I saw a friend and their partner in a heated argument” included the following response options: (1) “Yes, and I asked if everything was ok” (opportunity and bystander behavior); (2) “Yes, but I didn’t ask if everything was ok, but there were some good reasons why I didn’t” (opportunity and no bystander behavior); (3) “No” (no opportunity). For the purpose of the current study, we recoded responses to reflect the original scoring method (described in Banyard et al., 2014). Specifically, items were scored such that opportunity and behavior = 1, and opportunity and no behavior or no opportunity = 0. All 19 items were summed to reflect the total number of different situations in which a student engaged in bystander behavior over the past 2 months; possible scores ranged from 0 to 19. GLB in the current sample was .86 at baseline and .89 at the 2-month follow up. The risky situations subscale of the BBS is associated with theoretical correlates of bystander behavior, such as responsibility and behavioral intentions to intervene (Banyard et al., 2014).

Moral Courage

At baseline, students reported their level of moral courage on a modified version of the Woodard-Pury Moral Courage Scale (Woodard & Pury, 2007). The original measure included 23-items assessing one’s agreement and fear in performing several prosocial acts likely to result
GENERAL OUTCOMES OF BYSTANDER ACTION SCALE

in negative consequences, such as “I would confront a parent abusing his or her child in public.”
We revised the measure for the current study to reflect situations specific to intervening to
prevent sexual violence; due to the specific nature of these situations, this also reduced the length
of the measure to 7 items. Sample items include “I would do something to protect a friend from a
potential unwanted sexual experience, even if it brought intense public criticism,” and “I would
prevent a friend from having sex with an intoxicated person, even if I thought my peers didn’t
want me to.” Participants reported their agreement to each item on a 5-point scale ranging from
strongly disagree (0) to strongly agree (4); possible scores ranged from 0 to 28. Students also
rated the how much fear they would experience enacting each item. However, we only used
scores on the agreement response scale as the original measure was validated on this response
scale alone. Items were summed to form a total score such that higher scores indicate greater
agreement. GLB in the analytic sample was .92. The Woodard-Pury Moral Courage Scale is
related to psychological wellbeing (Keller, 2016), and similar scales of moral courage are related
to prosocial behavior (Hannah et al, 2011; Poteat & Vecho, 2016).

**Bystander Efficacy**

Participants completed a shortened version of the 14-item Bystander Efficacy Scale
(Banyard et al., 2007) at the baseline assessment. Students rated how confident they felt in
preforming 5 specific bystander behaviors on a scale from 0 (can’t do) to 100 (very certain).
Sample items include “Criticize a friend who tells me they had sex with someone who is passed
out or didn’t give consent,” and “Do something to help a very drunk person who is being brought
upstairs to a bedroom by a group of people at a party.” Items were summed such that higher
scores indicate greater confidence in preforming bystander behaviors. GLB in the analytic
sample was .83. Previous research demonstrates this 5-item version of the Bystander Efficacy
Scale correlates with self-reported bystander behavior (Jouriles, Rosenfield, et al., 2016; Jouriles et al., 2019).

**Participant sex**

Participants reported their sex on a single item: “What is your gender?” Response options included “Male” and “Female.”

**Data Analytic Plan and Sample Size Justification**

To evaluate our hypotheses, we utilized Poisson regression with an estimated scale factor predicting bystander behavior at the 2-month follow up. Poisson regression is an appropriate alternative to linear regression when the dependent variable represents a count distribution, such as a count of bystander behavior (Famoye & Singh, 2006). We utilized an estimated scale factor as our data were overdispersed; this technique appropriately adjusts for overdispersion and mitigates concerns related to the use of Poisson regression with overdispersed data (Breslow, 1990). We included positive consequences, negative consequences, moral courage, sex (0 = male, 1 = female), bystander efficacy, and bystander behavior at baseline as predictor variables in this model. As data for the current study was collected as part of a larger study evaluating a bystander intervention across three universities, we also included intervention condition (0 = control condition, 1 = intervention condition) and dummy codes for university as control variables. To model the co-occurrence of positive and negative consequences, we included an interaction between positive and negative consequences predicting bystander behavior. As we also hypothesized that the effect of such consequences on bystander behavior depends on levels of moral courage, we included a three-way interaction of positive consequences, negative consequences, and moral courage in our model, as well as each necessary component of this
interaction. Thus, we included 13 total predictor variables in our model: positive consequences, negative consequences, moral courage, sex, bystander efficacy, baseline bystander behavior, condition, two dummy codes for university, the interaction between positive and negative consequences, the interaction between positive consequences and moral courage, the interaction between negative consequences and moral courage, and the three-way interaction between positive consequences, negative consequences, and moral courage.

As stated previously, we used a dichotomous item scoring method in Study 2 analyses given the results of Study 1. We also attempted to replicate these analyses with the original scoring method, as we felt this was the most conservative test of criterion validity. Results with the dichotomous scoring method are reported in text, while results with the original scoring method are reported in Appendix A.

A post-hoc power analysis using G*Power (Faul et al., 2007) indicated that with alpha set at .05, 13 predictor variables, and a sample size of 357, power exceeded 0.99 to detect a small effect in a Poisson regression ($OR = 1.3$).

Results

Descriptive Statistics

Table 7 displays means, standard deviations, and Kendall’s tau correlation of variables used in tests of criterion validity. Positive consequences and negative consequences were positively correlated, suggesting that students who experience a variety of positive consequences tend to experience a variety of negative consequences. Both positive and negative consequences were positively correlated with bystander behavior at the baseline assessment and 2-month follow up. These positive relations at the bivariate level are unsurprising, given that positive and
negative consequences are a function of bystander behavior. Positive and negative consequences were not correlated with any of the remaining study variables.

Moral courage, sex, and bystander efficacy were positively associated with bystander behavior at the 2-month follow up (but not at the baseline assessment). In addition, bystander behavior at the baseline assessment was associated with bystander behavior at the 2-month follow up. A paired-samples t-test indicated no mean difference between bystander behavior at the baseline assessment and bystander behavior at the 2-month follow up, $t(356) = 0.26, p = .79$.

Similar to Study 1, 33.9% ($n = 121$) of students reported at least one consequence in Study 2. All students who reported a negative consequence reported at least one positive consequence in Study 2. Of students who reported a positive consequence, 24.8% ($n = 30$) reported a negative consequence.

**Associations Between Positive Consequences, Negative Consequences, and Bystander Behavior**

We first evaluated whether the effect of positive and negative consequences on bystander behavior differed by levels of moral courage by examining the three-way interaction between these variables. The three-way interaction between negative consequences, positive consequences, and moral courage was significant, $b = -0.10, SE = 0.05, p < .05$, suggesting that the influence of one type of consequence on the relation between the other type and bystander behavior depends on level of moral courage. To further investigate this finding, we reran analyses with moral courage centered at the following values: one standard deviation below the mean (what we refer to as low moral courage), at the mean (average moral courage), and one standard deviation above the mean (high moral courage).
At low levels of moral courage, there was a significant interaction between positive and negative consequences, $b = 0.68$, $SE = 0.34$, $p < .05$. Regardless of the number of different co-occurring positive consequences, negative consequences were unrelated to bystander behavior: $b = -1.73$, $SE = 1.03$, $p = .09$ in the absence of positive consequences; $b = -1.24$, $SE = 0.74$, $p = .11$ at mean numbers of different positive consequences; and $b = -0.48$, $SE = 0.41$, $p = .24$ at high numbers of different positive consequences. Positive consequences were also unrelated to bystander behavior in the absence of negative consequences, $b = 0.06$, $SE = 0.09$, $p = .51$, and at mean numbers of different negative consequences, $b = 0.16$, $SE = 0.10$, $p = .12$. However, at high numbers of different negative consequences, positive consequences predicted greater bystander behavior at the 2-month follow up, $b = 0.49$, $SE = 0.24$, $p = .03$. Broadly, these results suggest that for individuals low in moral courage, positive consequences predicted greater bystander behavior at high numbers of different co-occurring negative consequences, while negative consequences did not affect bystander behavior. Figure 2 displays a visual representation of regression coefficients between consequences and bystander behavior.

When moral courage was centered at its mean, there was a not significant two-way interaction between positive and negative consequences, $b = 0.24$, $SE = 0.14$, $p = .09$, suggesting that the effect of one type of consequences on bystander behavior was not affected by numbers of the other type of consequence. Regardless of number of different positive consequences, negative consequences were unrelated to bystander behavior: $b = -0.39$, $SE = 0.40$, $p = .33$. Alternatively, positive consequences predicted greater bystander behavior at the 2-month follow up, regardless of number of different negative consequences, $b = 0.10$, $SE = 0.05$, $p = .04$. Thus, these findings suggest that for individuals with average levels of moral courage, positive
consequences appear influential on bystander behavior while negative consequences do not (Figure 3).

Finally, at high levels of moral courage, there was not a two-way interaction between negative consequences and positive consequences, $b = -0.15, SE = 0.13, p = .24$, suggesting the effect of one type of consequence on bystander behavior does not depend on the other type of consequence. Negative consequences were related to more bystander behavior at the 2-month follow up, $b = 0.78, SE = 0.35, p = .03$, as were positive consequences, $b = 0.14, SE = 0.06, p = .02$. The general pattern of results indicate that for individuals high in moral courage, both positive and negative consequences are associated with increased bystander behavior (Figure 4).

As expected, bystander behavior at the baseline assessment predicted bystander behavior at the 2-month follow up, $b = .08, SE = .01, p < .001$. Surprisingly, sex, $b = 0.13, SE = 0.11, p = .22$, and bystander efficacy, $b =0.001, SE = 0.001, p = .08$, were unrelated to bystander behavior at the 2-month follow up. Condition was also not significantly associated with bystander behavior at the 2-month follow up, $b = -0.08, SE = 0.10, p = .43$. University was a significant predictor of bystander behavior, such that students at the South and Midwest university reported greater bystander behavior at the 2-month follow up than students at the Northeast university: $b = -0.24, SE = 0.12, p = .04$ for South vs. Northeast; $b = -0.27, SE = 0.12, p = .03$ for Midwest vs. Northeast. Students at the South and Midwest universities did not differ in bystander behavior, $b = -0.02, SE = 0.11, p = .88$.

**Exploratory Analyses**

In addition to our primary analyses, we also examined the association between moral courage and future bystander behavior. As the interaction between positive consequences,
negative consequences, and moral courage was significant, we explored the relation between moral courage and bystander behavior at different levels of these consequences.

In the absence of positive or negative consequences, moral courage was unrelated to bystander behavior, $b = -0.001$, $SE = 0.01$, $p = .97$. Similarly, in the absence of negative consequences, moral courage was unrelated to bystander behavior at mean numbers of different positive consequences, $b = 0.01$, $SE = 0.01$, $p = .63$, and high numbers of different positive consequences, $b = 0.02$, $SE = 0.02$, $p = .45$. Similarly, in the absence of positive consequences, moral courage was unrelated to future bystander behavior at mean numbers of different negative consequences, $b = 0.04$, $SE = 0.02$, $p = .08$. However, in the absence of positive consequences and high numbers of different negative consequences, moral courage was related to greater bystander behavior, $b = 0.19$, $SE = 0.09$, $p = .04$. This relation remained at co-occurring levels of negative and positive consequences: $b = 0.04$, $SE = 0.02$, $p = .04$ for mean numbers of different positive and negative consequences; $b = 0.09$, $SE = 0.04$, $p = .02$ for high numbers of different positive and negative consequences. Taken together, these findings suggest that moral courage is associated with bystander behavior at the 2-month follow up, but only in the presence of negative consequences.

General Discussion

In the current research, we examined the construct validity of the General Outcomes of Bystander Action Scale, a measure of consequences that has only undergone initial psychometric evaluation. In Study 1, we aimed to replicate the factor structure of this measure. Our results indicated that the General Outcomes of Bystander Action Scale evidences a two-factor structure reflecting positive and negative consequences as distinct, but related, dimensions. This factor solution evidenced poor-to-adequate fit when utilizing the original item scoring method and
excellent fit when items were scored dichotomously. Additionally, our descriptive statistics largely mirror those reported by Banyard and colleagues (2019); positive consequences appear more common than negative consequences, and male students seem to report more consequences than female students.

Study 2 aimed to provide further evidence of construct validity by examining the association between positive and negative consequences and bystander behavior over a 2-month period. As expected, our results broadly indicate that positive consequences are related to more future bystander behavior for individuals with low-to-average levels of moral courage. However, negative consequences were unrelated to bystander behavior among these individuals. For individuals high in moral courage, both positive and negative consequences relate to greater bystander behavior at the 2-month follow up. Taken together, these two studies provide further evidence of construct validity for the positive consequences subscale of the General Outcomes of Bystander Action Scale.

These findings may suggest that the negative consequences subscale of the General Outcomes of Bystander Action Scale does not represent a valid assessment of negative consequences to engaging in bystander behavior. However, there are some other possible interpretations of our findings. Research on reducing antisocial behavior notes that punishment is not always effective in influencing future behavior (Atkins et al., 2002; Chen & Raine, 2017). It may be that bystander behavior is not sensitive to the effects of punishment, and our lack of significant relations between negative consequences and future bystander behavior reflect this phenomenon rather than a concern with the measurement of consequences. However, research on the effects of punishment of prosocial behavior is sparse. Thus, this interpretation should be considered with caution.
Our results regarding the negative consequences subscale may also be attributable to the low prevalence rate of negative consequences in the current study rather than inaccuracy of the measure. Specifically, previous research has found that between 15% and 23% of students report a negative consequence (Banyard et al., 2019; Krauss et al., 2017; Moschella et al., 2018; Taylor et al., 2019), while we found that only 8.3% reported a negative consequence. Low prevalence rates represent possibly low variability of the negative consequences subscale in our model, which can make it difficult to detect significant relations. Therefore, future studies should continue to examine the criterion validity of the negative consequences subscale using different research designs to confirm whether this subscale represents a valid measure of negative consequences.

The lower prevalence rates of the current study may be due to a number of factors, including the specific sample used and the timing of the current study. Specifically, we utilized a sample of first year students assessed within the first two months of their first semester in college, whereas others have used non-first year college samples or study designs implemented later in participants’ college careers. Because consequences rely on another person initiating the act, it may be that negative consequences are less common among early first-year students as they have not yet formed many peer relationships in college. It is also possible a 2-month time period does not allow for much opportunity to engage in bystander behavior, and therefore to experience a consequences of such behavior. Although such low prevalence rates of negative consequences can be promising, these findings should be interpreted with caution as previous work suggest they may be higher. Researchers should continue to monitor and examine the rate of negative consequences across different samples.
Regarding Study 2, our results generally support our first two hypotheses in regards to positive consequences: for individuals with low-to-average levels of moral courage, positive consequences are related to greater bystander behavior at the 2-month follow up. These relations did appear to depend on levels of co-occurring consequences. Specifically, relations between one type of consequences and bystander behavior depended on levels of the other type of consequence. For instance, for individuals low in moral courage, positive consequences are related to greater bystander behavior, but only for students with a high number of different co-occurring negative consequences. One interpretation of these results is that positive consequences seem especially important when negative consequences are also present. High numbers of positive consequences appear to mitigate the effect of negative consequences on bystander behavior, while positive consequences in the absence of negative consequences have little to no effect on future bystander behavior. This is generally in line with previous theory documenting the co-occurrence of these consequences (Moschella et al., 2018), and speaks to the need for further empirical work in understanding how these consequences interact to affect bystander behavior.

Contrary to our third hypothesis, positive and negative consequences were associated with bystander behavior for individuals high in moral courage. In fact, both types of consequences were related to greater future bystander behavior. These results might be interpreted to suggest students high in moral courage are reinforced by both positive and negative consequences to engage in bystander behavior in the future. Although not initially predicted, such an interpretation is in line with some previous theory. For instance, research in other fields finds that any recognition of behavior, even that meant to discourage it, can be reinforcing (Brenner & Fox, 1998; Helbing et al., 2010). It may be that for individuals high in
moral courage, any recognition of their bystander behavior acts to reinforce this behavior in the future. Relatedly, students high in moral courage may view negative consequences as a “badge of honor” associated with engaging in behavior that they highly value. Alternatively, students high in moral courage may engage in many bystander behaviors, and thus have ample opportunity to experience a consequence. Since both positive and negative consequences are a function of bystander behavior, the positive association between both types of consequences and bystander behavior may simply reflect opportunity to experience a consequence. Including baseline bystander behavior may have controlled for some of this effect, but only partially. Although it is clear consequences of bystander behavior influence future behavior for individuals high in moral courage, researchers should continue to investigate the underlying cause of these associations.

The findings of the current study not only provide evidence for the validity of the General Outcomes of Bystander Action Scale (specifically the positive consequences subscale), but also suggest important implications for prevention work. Specifically, such findings help to further inform theories of bystander behavior, which largely focus on situational aspects of intervening (e.g., risk awareness in the current situation, barriers specific to the current situation; Burn, 2009). Our results suggest that positive consequences of bystander behavior experienced prior to a situation affect whether students intervene in that situation. In addition to furthering theory of bystander behavior, efforts to increase bystander behavior may be aided by utilizing consequences of bystander behavior. For instances, developers of bystander training programs may want to consider included encouragement to provide positive reinforcement to friends who engage in bystander behavior as these positive consequences may potentiate the effects of bystander training programs.
Although the current research was not designed to investigate the role of moral courage in understanding bystander behavior, our results also highlight the importance of this understudied variable. Specifically, our exploratory analyses demonstrated that moral courage was related to future bystander behavior, but only when negative consequences were present. This supports theoretical and empirical work on moral courage that highlights its importance as a predictor of prosocial behavior when individuals are likely to experience negative consequences (for review, see Osswald et al., 2010). Our results align with theory on the relation between moral courage and bystander behavior (Banyard, 2015) and point to the need for further research of this understudied variable.

Limitations

The current research is not without limitations. Although data for the current two studies was drawn from three universities, these universities were not selected to be representative of all colleges in the US. It is therefore unclear if our results replicate to samples of other college students. Moreover, we utilized a sample of incoming first year students as this time period is developmentally important for understanding and encouraging bystander behavior (Bowman, 2010; McMahon et al., 2015). However, it remains unknown if our results replicate to older college students who may have less exposure to situations that pose high risk for sexual assault (Cranney, 2015).

Relatedly, both analytic samples used in the current research were majority White and non-Hispanic, limiting the extent to which our results can be generalized to more diverse student populations. In fact, there may be reason to believe that different ethnic and racial groups experience varying amounts of consequences of bystander behavior compared to White students. For one, previous research suggests that African American students report more bystander
behaviors than White students (Brown et al., 2014), which may result in a greater number of consequences. In addition, students of minority race and ethnicity are more often victims of crime during college than White students (Rayburg et al., 2003; Coulter et al., 2017), and may receive more negative consequences in response to bystander behavior than White students. Future research should examine differences in consequences of bystander behavior among racial and ethnic minority students.

An important aspect of learning theory not fully addressed in this study includes an individual’s learning history. That is, an individual’s history of engaging in bystander behavior and experiencing consequences of this behavior prior to the study period likely impacts their behavior during the study period. Previous research notes sexual violence is prevalent before college (Anderson et al., 2019), and adolescents engage in bystander behavior to prevent such violence (Sargent et al., 2017). Participants’ reports of bystander behavior and consequences of such behavior in the current study are likely influenced by their previous experience intervening to prevent sexual assault. Furthermore, this previous history could conceivably affect the relations between consequences and future bystander behavior investigated in the current study. For instance, an individual who has experienced several positive consequences after intervening in bystander behavior in the past may not be deterred by a single negative consequence in the current study, while an individual who has already experienced several negative consequences may be greatly deterred by even one additional negative consequence. Although we could not control for participants' learning history in the current study, the time frame of the current study allowed us to examine students’ learning history beginning at the start of their college careers. Therefore, the current study examines an important time period for the development of bystander behavior in college.
The assessment of bystander behavior in the current study has some important limitations. Recent work suggests that measures of bystander behavior that rely on a yes/no response scale are problematic for a number of reasons (for review, see Hoxmeier et al., 2017 and McMahon et al., 2017). Researchers have provided a number of different assessment strategies that may more accurately capture bystander behavior among college students (Hoxmeier et al., 2017; McMahon et al., 2014; McMahon et al., 2017; Jouriles, Kleinsasser, et al., 2016; Yule & Grych, 2017). Although we utilized a slightly modified version of the BBS to account for the previously documented limitations of this measure, we rescored items such that they reflected a yes/no response scale. We chose to do so to allow us to interpret our results within the context of previous work; that is, the majority of work on identifying determinants of bystander behavior uses a yes/no scoring method. With initial evidence from this research suggesting that consequences predict bystander behavior scored in this way, future research should extend these findings to explore how consequences affect future bystander behavior scored in alternative ways.

Although a strength of this study is the use of longitudinal data to establish temporal precedence, one limitation of this research is the short time frame used. We demonstrated that consequences of bystander behavior predict future bystander behavior over a 2-month period, but our results do not address how bystander behavior is affected over a longer period of time (such as an entire year). We assessed students during their first semester of their first year of college; as stated earlier, this time is developmentally important in constructing and modifying prosocial responses to sexual violence. However, such a time frame does not allow us to investigate how consequences of bystander behavior affected future behavior across a longer period of time. Future research should attempt to extend the current study to examine changes in bystander
GENERAL OUTCOMES OF BYSTANDER ACTION SCALE

behavior and consequences across a broader study period, such as throughout the first year of
college.

Future Research

On the whole, the current research provides additional evidence for the construct validity
of the General Outcomes of Bystander Action Scale, and specifically the positive consequences
subscale of this measure. As noted earlier, measurement validity is an iterative process that
requires continued replication, and much remains unknown about the nuances of these
consequences.

Thus, further research should continue to examine the validity of this measure and extend
on this work by utilizing this measure in our understanding of consequences of bystander
behavior. Specifically, results of our CFA diverged slightly from Banyard et al.’s (2019) study;
therefore, future research should continue to explore the optimal factor structure of this measure.
In addition, although we found support for a negative consequences factor in Study 1, we did not
find evidence of criterion for the negative consequences subscale in Study 2. Future research
should attempt to replicate our findings under different conditions (i.e., with a longer time frame
or college students from other years) to continue to assess the validity of this subscale.

Furthermore, there are likely important factors of consequences of bystander behavior
that influence their association with bystander behavior. For one, the current study recruited
students during their first semester of college, a time when students may uniquely respond to
consequences. For example, the first semester of college is an important time in cultivating new
peer relationships and building a social support network (Oswald & Clark, 2003). As this time is
characterized by the need to form a new social support system, students may be especially
susceptible to positive and negative consequences from their peers in order to maintain these
initial social connections. On the other hand, older students are likely to have solidified social networks in college, and may therefore be less susceptible than first year students to a single consequence from a peer. Future work should extend the current findings to other groups of students to explore whether consequences exert a greater influence on bystander behavior early in college.

Additionally, theory on peer influence notes the importance of considering status and relational proximity of peers in their magnitude of influence over others (Brown et al., 2008). Previous work suggests that individuals viewed as “close peers” exert a greater influence over students compared to other peers (Paek, 2009). Furthermore, peer influence seems to be moderated by peer popularity, such that peers viewed as popular hold greater influence over others than unpopular peers (Peters et al., 2010). Based on this, it may be the case that consequences perpetrated from close friends or influential peers (e.g., student athletes, peer leaders in Greek life) may have a greater effect on students than consequences from less influential peers. Furthermore, it is conceivable that individual consequences differ in the magnitude of their associations with bystander behavior. For example, being physically hurt after engaging in bystander behavior may discourage future bystander behavior to a greater degree than viewing negative posts on social media. Therefore, future research should work to further refine and evaluate this measure.

Given that the current research represents only the second attempt to validate the General Outcomes of Bystander Action Scale, this instrument may be viewed as a rough measure of consequences. Despite this, the current research demonstrates that this measure is related to future bystander behavior as theory would predict, while controlling for baseline bystander behavior and known determinants of bystander behavior. Hence, consequences of engaging in
bystander behavior are likely a much stronger predictor of bystander behavior than the current 
research suggests. To date, very little work has examined the impact of consequences of 
bystander behavior in our understanding and encouragement of bystander behavior. Future 
research should extend our findings by examining when positive and negative consequences are 
most likely to occur, what bystander behaviors may be most likely to result in consequences, and 
how bystander training programs can reduce the occurrence of negative consequences and 
increase the occurrence of positive consequences.

Conclusion

The current research provides additional construct validity for the General Outcomes of 
Bystander Action Scale and supports the notion that the positive consequences subscale is a valid 
measure for assessing positive consequences of engaging in bystander behavior. In addition, our 
results highlight the importance of considering consequences in our understanding of bystander 
behavior. However, consequences of bystander behavior likely act in nuanced ways that remain 
unknown to us. Based on the current research, the General Outcomes of Bystander Action Scale 
seems an appropriate measure to further investigate when students are likely to experience 
consequences, in what context consequences are most influential, and in what ways 
consequences affect future behavior.
Appendix A: Supplementary Results

Although a dichotomous item scoring method resulted in the best CFA model fit, we attempted to replicate our predictive analyses with the original continuous scoring method as we felt this was the most conservative test of criterion validity. Table 8 displays means, standard deviations, and Kendall’s tau correlations among study variables. The three-way interaction between positive consequences, negative consequences, and moral courage was not significant, $b = -0.003, SE = 0.001, p = .07$, suggesting that the effect of positive consequences on the relation between negative consequences and bystander behavior does not depend on levels of moral courage (and vice versa). However, we retained the three-way interaction in the model as is recommended when an interaction term is theoretically meaningful to interpret the effects of other predictors (Tabachnick & Fidel, 2019) We therefore investigated these findings using the same procedure as described earlier.

At low levels of moral courage, the two-way interaction between positive and negative consequences was not significant, $b = 0.01, SE = 0.01, p = .14$, suggesting that the effect of one type of consequence on bystander behavior does not depend on the other type of consequence. Negative consequences were not related to bystander behavior at the 2-month follow up, regardless of levels of positive consequences, $b = -0.12, SE = 0.13, p = .35$. Positive consequences were also unrelated to bystander behavior, regardless of levels of negative consequences, $b = 0.01, SE = 0.01, p = .75$. Thus, these findings suggest that students with low
levels of moral courage are unaffected in their bystander behavior by either positive or negative consequences.

When moral courage was centered at its mean, the interaction between positive and negative consequences was not significant, \( b = 0.001, \ SE = 0.04, \ p = 0.76 \); therefore, the relation between positive consequences and bystander behavior is unaffected by negative consequences, and vice versa. The main effect of negative consequences on bystander behavior was not significant, \( b = 0.04, \ SE = 0.06, \ p = .45 \). Positive consequences were also unrelated to bystander behavior at the 2-month follow up, although this effect was trending in the correct direction, \( b = 0.02, \ SE = 0.01, \ p = .07 \). For students with average levels of moral courage, neither positive consequences nor negative consequences were related to future bystander behavior.

Finally, for high levels of moral courage, there was not a significant interaction between positive and negative consequences, \( b = -0.01, \ SE = 0.005, \ p = .08 \). Contrary to our hypotheses, negative consequences were positively associated with bystander behavior at the 2-month follow up, regardless of level of positive consequences: \( b = 0.19, \ SE = 0.07, \ p = .01 \). Positive consequences were also related to greater bystander behavior, regardless of level of negative consequences, \( b = 0.03, \ SE = 0.01, \ p = .01 \). In summary, these findings suggest that both positive and negative consequences act to promote bystander behavior among individuals high in moral courage.

As expected, bystander behavior at the baseline assessment predicted bystander behavior at the 2-month follow up, \( b = 0.08, \ SE = 0.01, \ p < .001 \). Sex, \( b = 0.13, \ SE = 0.11, \ p = .22 \), and bystander efficacy, \( b = 0.01, \ SE = 0.01, \ p = .18 \), were unrelated to bystander behavior at the 2-month follow up. Condition was also not associated with bystander behavior at the 2-month follow up, \( b = -0.05, \ SE = 0.10, \ p = .60 \). University was associated with bystander behavior, such
that students at the Midwest university reported greater bystander behavior at the 2-month follow up than students at the Northeast university, \( b = -0.28, SE = 0.13, p = .03 \). Students at the South university did not differ from those at the Northeast or Midwest universities: \( b = -0.22, SE = 0.12, p = .07 \) for South vs. Northeast; \( b = -0.06, SE = 0.12, p = .63 \) for South vs. Midwest.

**Exploratory Analyses**

As we did in our primary analyses, we also explored the effect of moral courage on bystander behavior when consequences were scored with the original response scale. As before, we examined the effect of moral courage on bystander behavior at varying levels of consequences.

In the absence of positive and negative consequences, moral courage was unrelated to bystander behavior at the 2-month follow up, \( b = 0.01, SE = 0.01, p = .71 \). Similarly, in the absence of negative consequences, moral courage was unrelated to bystander behavior at mean levels of positive consequences, \( b = 0.01, SE = 0.01, p = .31 \), and at high levels of positive consequences, \( b = 0.03, SE = 0.02, p = .20 \). Similarly, in the absence of positive consequences, moral courage was unrelated to greater bystander behavior at mean levels of negative consequences, \( b = 0.02, SE = 0.02, p = .19 \), or at high levels of negative consequences, \( b = 0.08, SE = 0.04, p = .06 \). Moral courage was not predictive of future bystander behavior at mean levels of co-occurring negative and positive consequences, \( b = 0.03, SE = 0.01, p = .08 \). However, at high levels of co-occurring consequences, moral courage predicted greater bystander behavior, \( b = 0.06, SE = 0.03, p = .02 \). Thus, these results suggest that moral courage is associated with bystander behavior, but only at high levels of both positive and negative consequences.
References


GENERAL OUTCOMES OF BYSTANDER ACTION SCALE


GENERAL OUTCOMES OF BYSTANDER ACTION SCALE


GENERAL OUTCOMES OF BYSTANDER ACTION SCALE


Table 1

Descriptive statistics for each item of the General Outcomes of Bystander Action Scale

<table>
<thead>
<tr>
<th>Items</th>
<th>M (SD)</th>
<th>Range</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Consequences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. People praised me for what I did</td>
<td>0.75 (1.75)</td>
<td>0-8</td>
<td>21%</td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>1.29 (2.40)</td>
<td>0-8</td>
<td>28.1%</td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>0.65 (1.68)</td>
<td>0-8</td>
<td>17.8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2.68 (5.06)</td>
<td>0-24</td>
<td>31.7%</td>
</tr>
<tr>
<td><strong>Negative Consequences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>0.09 (0.68)</td>
<td>0-7</td>
<td>2.4%</td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>0.07 (0.47)</td>
<td>0-6</td>
<td>3.2%</td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>0.02 (0.35)</td>
<td>0-6</td>
<td>0.6%</td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action (e.g., charged with underage drinking).</td>
<td>0.02 (0.25)</td>
<td>0-5</td>
<td>0.6%</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time (e.g., time it took to intervene, being interviewed later, talking with friends about what I did)</td>
<td>0.15 (0.77)</td>
<td>0-8</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.35 (1.77)</td>
<td>0-25</td>
<td>8.3%</td>
</tr>
</tbody>
</table>

*Note.* Items on the scale are scored 0 to 8+ times. Total scores for each subscale are calculated by summing the items within the subscale, such that higher scores represent more consequences experienced. Prevalence estimates indicate the percentage of students who reported experiencing at least one instance of each consequence.
Table 2

Descriptive statistics for each item of the General Outcomes of Bystander Action Scale by sex

<table>
<thead>
<tr>
<th>Items</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD) Prevalence</td>
<td>M (SD) Prevalence</td>
</tr>
<tr>
<td><strong>Positive Consequences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. People praised me for what I did</td>
<td>0.94 (2.00) 23.1%</td>
<td>0.65 (1.60) 19.9%</td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>1.47 (2.54) 30.8%</td>
<td>1.19 (2.33) 26.7%</td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>0.83 (1.94) 19.5%</td>
<td>0.56 (1.52) 16.7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3.24 (5.63) 33.7%</td>
<td>2.39 (4.73) 30.7%</td>
</tr>
<tr>
<td><strong>Negative Consequences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>0.11 (0.71) 2.9%</td>
<td>0.09 (0.66) 2.1%</td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>0.10 (0.63) 4.1%</td>
<td>0.05 (0.36) 2.8%</td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>0.04 (0.18) 1.2%</td>
<td>0.02 (0.33) 0.3%</td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action (e.g., charged with underage drinking).</td>
<td>0.03 (0.39) 0.6%</td>
<td>0.01 (0.12) 0.6%</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time (e.g., time it took to intervene, being interviewed later, talking with friends about what I did)</td>
<td>0.23 (0.91) 7.7%</td>
<td>0.11 (0.69) 4.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.50 (2.32) 11.2%</td>
<td>0.69 (1.41) 6.7%</td>
</tr>
</tbody>
</table>

*Note.* Items on the scale are scored 0 to 8+ times. Total scores for each subscale are calculated by summing the items within the subscale, such that higher scores represent more consequences experienced. Prevalence estimates indicate the percentage of students who reported experiencing at least one instance of each consequence.
Table 3

Sample size per response option by item on the General Outcomes of Bystander Action Scale

<table>
<thead>
<tr>
<th>Items</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Consequences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. People praised me for what I did</td>
<td>391</td>
<td>27</td>
<td>11</td>
<td>16</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>356</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>407</td>
<td>19</td>
<td>16</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Negative Consequences</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>483</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>479</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>492</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action (e.g., charged with underage drinking)</td>
<td>492</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time (e.g., time it took to intervene, being interviewed later, talking with friends about what I did)</td>
<td>467</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4

*Kendall's tau correlation coefficient between items on the General Outcomes of Bystander Action Scale*

<table>
<thead>
<tr>
<th>Items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. People praised me for what I did</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>.71**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>.59**</td>
<td>.52**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>.24**</td>
<td>.21**</td>
<td>.27**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>.27**</td>
<td>.25**</td>
<td>.26**</td>
<td>.64**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>.15**</td>
<td>.13**</td>
<td>.17**</td>
<td>.33**</td>
<td>.29**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action</td>
<td>.08</td>
<td>.05</td>
<td>.10*</td>
<td>.16**</td>
<td>.14**</td>
<td>.33**</td>
<td>-</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time</td>
<td>.30**</td>
<td>.33**</td>
<td>.34**</td>
<td>.19**</td>
<td>.30**</td>
<td>.32**</td>
<td>.21**</td>
</tr>
</tbody>
</table>

* p < .05  
** p < .01
**Table 5**

*Maximum likelihood estimates with robust standard errors of factor loadings for enhanced model of the General Outcomes of Bystander Action Scale*

<table>
<thead>
<tr>
<th>Positive Consequences</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. People praised me for what I did</td>
<td>1.00</td>
<td>0.89</td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>1.27</td>
<td>0.82</td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>0.72</td>
<td>0.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative Consequences</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>1.03</td>
<td>0.99</td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>0.52</td>
<td>0.66</td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action (e.g., charged with underage drinking).</td>
<td>0.23</td>
<td>0.43</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time (e.g., time it took to intervene,</td>
<td>0.38</td>
<td>0.23</td>
</tr>
<tr>
<td>being interviewed later, talking with friends about what I did)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Robust model fit indices reported. RMSEA = 0.08 (95% CI [0.05, 0.12]), CFI = .94, TLI = .89, SRMR = 0.09
### Table 6

*Diagonally weighted least squares estimators with robust standard errors of factor loadings for the General Outcomes of Bystander Action Scale*

<table>
<thead>
<tr>
<th>Positive Consequences</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. People praised me for what I did</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>2. A friendship was strengthened because of what I did</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>3. People said positive things about me to others about what I did</td>
<td>0.90</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative Consequences</th>
<th>Unstandardized</th>
<th>Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. I saw negative posts on social media about what I did</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>5. I was harassed because of what I did</td>
<td>0.99</td>
<td>0.95</td>
</tr>
<tr>
<td>6. I was physically hurt because of what I did</td>
<td>0.86</td>
<td>0.82</td>
</tr>
<tr>
<td>7. I got in trouble as a result of my action (e.g., charged with underage drinking)</td>
<td>0.72</td>
<td>0.69</td>
</tr>
<tr>
<td>8. My involvement ended up costing me a lot of time (e.g., time it took to intervene, being interviewed later, talking with friends about what I did)</td>
<td>0.95</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Note.* Robust model fit indices reported. RMSEA = .03 [.05, .06] CFI = .99, TLI = .99, SRMR = .09
### Table 7

Means, standard deviation, and Kendall’s tau correlations of study variables to assess criterion validity.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Negative consequences</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. Positive consequences</td>
<td>.46**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3. Moral courage</td>
<td>.03</td>
<td>.08</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4. Sex</td>
<td>-.07</td>
<td>-.03</td>
<td>.06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5. Bystander efficacy</td>
<td>.01</td>
<td>.02</td>
<td>.35**</td>
<td>.12**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6. School</td>
<td>-.01</td>
<td>.03</td>
<td>.12**</td>
<td>.14**</td>
<td>.06</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7. Condition</td>
<td>-.05</td>
<td>-.09</td>
<td>-.04</td>
<td>-.10</td>
<td>-.03</td>
<td>.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8. Baseline bystander behavior</td>
<td>.19**</td>
<td>.31**</td>
<td>.08</td>
<td>.03</td>
<td>.08</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>9. 2-month bystander behavior</td>
<td>.19**</td>
<td>.22**</td>
<td>.13**</td>
<td>.11*</td>
<td>.10**</td>
<td>.03</td>
<td>-.07</td>
<td>.34**</td>
<td>-</td>
</tr>
</tbody>
</table>

**M(SD)**

<table>
<thead>
<tr>
<th></th>
<th>0.15</th>
<th>0.72</th>
<th>24.02</th>
<th>-</th>
<th>390.93</th>
<th>-</th>
<th>-</th>
<th>3.32</th>
<th>3.27</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.49)</td>
<td>(1.11)</td>
<td>(4.56)</td>
<td>(87.30)</td>
<td>(2.73)</td>
<td>(3.54)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Negative consequences ranged from 0 to 5, while positive consequences ranged from 0 to 3. Moral courage ranged from 0 to 28, bystander efficacy ranged from 0 to 500, and bystander behavior (both at baseline and the 2-month follow up) ranged from 0 to 19. Sex was coded 0 = male, 1 = female. School was coded 0 = South, 1 = Midwest, 2 = Northeast. Condition was coded 0 = control, 1 = bystander training program.

* *p < .05

** **p < .01
Table 8

Means, standard deviation, and Kendall’s tau correlations of study variables to assess criterion validity when utilizing the original item scoring method of the GOBAS.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Negative consequences</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Positive consequences</td>
<td>.46**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Moral courage</td>
<td>.03</td>
<td>.07</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Sex</td>
<td>-.07</td>
<td>-.04</td>
<td>.06</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Bystander efficacy</td>
<td>.01</td>
<td>.01</td>
<td>.35**</td>
<td>.12**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. School</td>
<td>-.01</td>
<td>.04</td>
<td>.12**</td>
<td>.14**</td>
<td>.06</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Condition</td>
<td>-.05</td>
<td>-.10*</td>
<td>-.04</td>
<td>-.10</td>
<td>-.03</td>
<td>.02</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Baseline bystander behavior</td>
<td>.19**</td>
<td>.29**</td>
<td>.08</td>
<td>.03</td>
<td>.08</td>
<td>.01</td>
<td>.01</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>9. 2-month bystander behavior</td>
<td>.19**</td>
<td>.20**</td>
<td>.13**</td>
<td>.11*</td>
<td>.10**</td>
<td>.03</td>
<td>-.07</td>
<td>.34**</td>
<td>-</td>
</tr>
<tr>
<td>M(SD)</td>
<td>0.40</td>
<td>2.82</td>
<td>24.02</td>
<td>-</td>
<td>390.93</td>
<td>-</td>
<td>-</td>
<td>3.32</td>
<td>3.27</td>
</tr>
<tr>
<td></td>
<td>(1.60)</td>
<td>(5.08)</td>
<td>(4.56)</td>
<td>(87.30)</td>
<td>(2.73)</td>
<td>(3.54)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Negative consequences ranged from 0 to 40, while positive consequences ranged from 0 to 24. Moral courage ranged from 0 to 28, bystander efficacy ranged from 0 to 500, and bystander behavior (both at baseline and the 2-month follow up) ranged from 0 to 19. Sex is coded 0 = male, 1 = female. School is coded 0 = South, 1 = Midwest, 2 = Northeast. Condition is coded 0 = control, 1 = bystander training program.

*p < .05

**p < .01
Figure 1. Two-factor model of the General Outcomes of Bystander Actions Scale
Figure 2. Visual representation of the regression coefficient between consequences and bystander behavior for individuals low in moral courage. Black slopes represent significant coefficients, while grey slopes represent non-significant coefficients.
Figure 3. Visual representation of the regression coefficient between consequences and bystander behavior for individuals with average levels of moral courage. Black slopes represent significant coefficients, while grey slopes represent non-significant coefficients. As the interaction between positive consequence and negative consequences was non-significant, only one slope is shown per graph.
Figure 4. Visual representation of the regression coefficient between consequences and bystander behavior for individuals high in moral courage. Black slopes represent significant coefficients, while grey slopes represent non-significant coefficients. As the interaction between positive consequence and negative consequences was non-significant, only one slope is shown per graph.