INCREASING SURVEY DATA QUALITY USING SCREENING VALIDITY QUESTIONS

Lucas Schalewski, Jamie Utt, and Bryant Valencia, University of Arizona

Self-report surveys are used frequently in fraternity organizations to collect information from students. A lack of thoughtful or truthful answers on survey instruments threatens the validity of results. The current study evaluates if identifying and omitting invalid responders using screening validity questions improves data quality on two scales among fraternity men: the Illinois Rape Myth Acceptance and the Conformity to Masculinity Norms Inventory. Results indicate invalid responders bias results suggesting using screen validity questions improves data quality. This strategy can help fraternity professionals ensure their programming responds more closely to their member perceptions, attitudes, and experiences.

Increased attention dedicated to understanding and preventing sexual violence within college campuses has become a critical issue often centered on fraternity members. This shift has generated a need for additional research and assessment exploring college student experiences and perceptions of sexual violence. A notable example is Wood, Sulley, Kammer-Kerwick, Follingstad, and Busch-Armendariz's (2016) detailed overview of various administered sexual assault campus climate surveys across U.S. postsecondary institutions that often measured fraternity membership as a factor for analysis. Student affairs administrators on campuses also have increasing expectations to assess experiences and perceptions of fraternity members along with evaluating sexual violence prevention programs for accountability and improvement purposes. One common and convenient method supporting these goals is the use of self-report surveys in assessment and research.

Surveys are able to produce credible results based on the quality of instrument used (Saunders & Cooper, 2009). If a survey is poorly designed, the subsequent results may be biased, leading to misinterpretations or inappropriate uses for informing changes to policy and practices. There is limited research on how survey design and responses among fraternity members should be conducted to ensure valid results. Furthering a need for research within survey development on fraternity members is how their survey participation commonly measures sensitive issues surrounding their health, sexual violence, alcohol and drug use, and other personal behaviors. Research has shown how sensitive topics on surveys may lead to issues in response rates and misreporting outcomes (Kays, Gathercoal, & Buhrow, 2012; Tourangeau & Yan, 2007). This article investigates the use of validity screening questions as a potential method to improve data quality for surveys measuring sensitive topics among fraternity members.

Literature Review

Fraternity Membership: Rape Myth Acceptance and Conformity to Masculinity Norms

The increased attention of fraternity members in regards to sexual violence is partly due to research demonstrating fraternity members' greater acceptance towards rape myths than non-fraternity members (Bleecker & Murnen, 2005; Canan, Jozkowski, & Crawford, 2016; Murnen & Kohlman, 2007). Foubert, Brosi, and Bannon (2011) further emphasize the importance of such findings by showing rape myth acceptance increases the intent to commit sexual assault among fraternity members.

Past research has also demonstrated a

relationship between fraternity members reporting higher acceptance of rape myths and their commitments to traditional notions of masculinity. Seabrook, Ward, and Giaccardi (2016) found conformity to masculinity norms mediate the relationship of fraternity involvement and sexual violence approval. Further research illustrates fraternity members on college campuses who obey particular masculine norms (i.e., alcohol use, risk-taking) are more likely to validate rape myths and sexually aggressive behavior (Capraro, 2000; Iwamotoa, Cheng, Lee, Takamatsu, & Gordon, 2011; Turrisi, Mallett, Mastroleo, & Larimer, 2010). There is thus a demonstrated need to accurately measure both masculinity and rape supportive attitudes among fraternity members within a commitment to addressing sexual violence on college campuses. Additionally, the survey development process should consider the degree of sensitive survey questions and incorporate strategies to increase data quality.

Surveys

Surveys have become a commonly used tool in meeting numeric data needs in higher education and student affairs. The cost-effective method of collecting quantitative data enables budget-constrained offices to conduct important assessment initiatives (Schuh, Biddix, Dean, & Kinzie, 2016; Saunders & Cooper, 2009). Surveys that are nationally recognized and implemented also tend to have the added benefits of validity and reliability, reducing the burdensome process of developing an instrument (Bresciani, Gardner, & Hickmott, 2012). Further, surveys are used to support generalizability through appropriate sampling techniques, allowing for meaningful results with a selected population (Bresciani et al., 2012).

Although surveys are useful in measuring a variety of college student outcomes, it is critical they do so in valid ways. Saunders and Cooper (2009) posit, "The credibility of an assessment depends, in part, on the quality of the measurement instrument" (p. 122), and the authors further discussed validity as one way to evaluate if an instrument is of quality and thereby supportive of good assessment practices. Messick (1989) described validity in surveys as the degree to which a survey accurately measures what it claims to, based on evidence and theory. One way a survey may lack validity is measurement error, defined as the difference of an estimated value compared to its true value from inaccurate survey responses (Dillman, Smyth, & Christian, 2014). Inaccurate answers causing measurement error may be due to poor survey design, issues in the data collection phase, or a respondent's inability or unwillingness to respond in a truthful manner may result in error (Dillman et al., 2014).

Research has shown sensitive topics in surveys influence the degree of respondents' self-disclosure on question items. Respondents are less likely to respond to questions the more sensitive the questions are perceived to be (Kays, Gathercoal & Buhrow, 2012; Tourangeau & Yan, 2007). Men, as compared to women, have lower self-disclosure rates on sensitive survey questions when there is a lack of perceived privacy (Joinson, Paine, Buchanan, & Reips, 2008). These studies demonstrate a need to further explore strategies within the survey design process when asking about sensitive topics. The use of screening validity questions has been a method recently explored by other researchers.

Screening Validity Questions

Intentional survey questions measuring survey respondent degree of truthfulness and carefulness in responses have been used in middle and high schools to identify and omit invalid responders in order to increase validity of survey results. For instance, Jia, Konold, Cornell, and Huang (2016) added the screening questions, "I am telling the truth on this survey" and "How many of the questions on the survey did you answer truthfully" to a survey assessing high school student outcomes and bullying to identify invalid

respondents in Virginia public high schools with 52,012 respondents. They found 6.88% of student respondents (n=3,579) in their sample were flagged through the screening questions as invalid responders. The invalid respondents biased the survey findings when included in the analysis. The aggregated data that included invalid responses reported higher risk behaviors, lower GPA, school engagement, and depression compared to when the invalid respondents were removed.

Assessment by Cornell, Klein, Konold, and Huang (2012) on school climate among middle school students (N=7,801) had consistent findings to Jia et al.'s (2016) study. Question items "I am telling the truth on this survey," "I am not paying attention to how I answer this survey," and "The answers I have given on this survey are true" were added to the instrument and analyzed to determine the impact invalid responders may have on measuring school climate outcomes. The survey flagged 11.77% of student respondents (n=918) as invalid and similarly skewed the overall findings toward support of risky behaviors. Cornell et al. conducted a second study using two screening validity questions instead of three on a survey given to ninth grade high school students (N=7,246) assessing general school safety outcomes. Respondents who failed at least one of the screening questions were considered invalid and represented 4% (n=281) of the sample. Results indicated valid responders were more likely to have positive perceptions of the school climate and have similar views as their teachers compared to invalid responders. Cornell, Lovegrove, and Baly's (2014) found similar findings when using the same two out of the three screening validity questions on a longitudinal survey measuring risk behavior, victimization, and school climate among middle school students.

The presented research has demonstrated consistent results that adding screening items on surveys measuring sensitive topics among middle and high school students improves survey data quality. Common findings within the studies indicates invalid responders are more likely male, students of Color, and report a higher likelihood of endorsing or experiencing risky behaviors (Cornell et al., 2012; Cornell et al., 2014; Jia et al., 2016). No published research or nationally recognized student affairs surveys were identified using similar screening items in surveys of fraternity members or students in U.S. postsecondary education. The present study extends the research by examining the use of screening validity questions in a higher education context. Given that invalid responders are more likely to be men and more likely to commit risky behaviors, the present research used screening validity questions in a survey measuring rape myth acceptance and conformity to masculinity norms among fraternity members.

Method

Screening validity questions were added to a survey distributed to fraternity members at a large, public research university seeking to measure their degree of rape myth acceptance and conformity to masculinity norms. The survey was administered in-person at a series of two-hour workshops on sexual violence prevention in fall 2016 where fraternity chapters were required to have half of their members attend. All attendees were invited to participate in the survey but could choose to opt out of participation. A total of 585 surveys were returned out of 975 fraternity members who attended the workshops from among 1,872 total active fraternity members, indicating a 60% response rate among those who attended the program and 31% of the total fraternity population.

Demographic questions were also asked on the survey. Questions measured live-in status with the fraternity housing, academic class standing, race/ethnicity, and sexual orientation. Of the sample, 28% were considered "live-in" through their fraternity chapter. Most students were

in lower academic classes with 32% first-year students, 36% sophomores, 22% juniors, and 10% seniors. Additional demographics reported 26% of the respondents were students of Color and 5% identified as gay, bisexual, or queer.

Screening validity questions flag invalid responders who indicate they are not reading the survey carefully or telling the truth. Statistical differences between valid and invalid responses were evaluated to determine if adding screening validity questions supports data quality among fraternity members. Based on these differences, inflation rates are calculated which depicts the degree of change in the aggregated responses due to invalid responses.

The present research is guided by the following research question: Do screening validity questions improve data quality of Illinois Rape Myth Acceptance and Conformity to Masculinity Norms Inventory? Previous research indicates screening validity questions are more likely to identify males who engage in risky behaviors as invalid responders. Consequently, we hypothesize invalid responders will have higher degrees of Conformity to Masculinity Norms Inventory and Illinois Rape Myth Acceptance compared to valid respondents.

Measures

Screening validity questions. The questions asked to screen for the validity of outcomes included: "I am reading this survey carefully," "I am telling the truth on the survey," and "The answers I have given on the survey are true." Selected questions were the same or similar questions to previous research (Cornell et al., 2012; Cornell et al., 2014; Jia et al., 2016). The first two questions were asked using a five-point agreement Likert scale with a binary label (1 = Strongly Agree, 5 = Strongly Disagree). The third screening question was asked with a "Yes" or "No" response option.

Illinois Rape Myth Acceptance Scale. The scale was revised by Payne, Lonsway, and Fitzgerald (1999), supporting a shorter version scale which was later updated with language relevant to college students by McMahon and Farmer (2011), leading to the version utilized in the present study. The four subscales which make up the Rape Myth Acceptance Scale include (a) She Asked For It ("When girls go to parties wearing slutty clothes, they are asking for trouble"); (b) He Didn't Mean To ("Rape happens when a guy's sex drive gets out of control"); (c) It Wasn't Really Rape ("If a girl doesn't physically fight back, you can't really say it was rape"); and, (d) She Lied ("A lot of times, girls who claim they were raped just have emotional problems"). The scale is 22-questions with each subscale representing five or six question items on a five-point agreement Likert scale (Strongly Agree = 1; Strongly Disagree=5). Lower scores indicate a higher degree of acceptance of rape myths.

Conformity to Masculinity Norms Inventory. The Conformity to Masculinity Norms Inventory (CMNI) was also included to measure fraternity men's relationship with masculinity norms. The inventory was developed by Mahalik et al. (2003) to measure how men adhere to gender role norms negotiated by one's experiences. The present research used a revised version by Parent and Moradi (2011). The CMNI-46 is structured by nine distinct factors: (a) Winning; (b) Emotional control; (c) Primacy of work; (d) Risk-taking items; (e) Violence items; (f) Heterosexual self-presentation; (g) Playboy; (h) Self-reliance; and (i) Power over women. Questions were assessed utilizing a four-point agreement Likert scale (Strongly Agree= 0; Strongly Disagree 3). After recoding reverse scored survey items, lower scores represent respondents who are more conformed to masculinity norms based on a four-point agreement Likert scale.

Results

Descriptive results indicate the percent of students who failed a screening question for each item and the total percent who failed at least one of the three (see Table 1). Students who did not "strongly agree" with the statement "I am reading

Table 1
Screening Validity Questions Descriptive Statistics

Screening Validity Question	% failed	n
1) I am reading this survey carefully	17.96%	104
2) I am telling the truth on the survey	10.18%	58
3) The answers I have given on the survey are true	1.58%	9
Total respondents who failed at least one question	21.37%	125

the survey carefully" or "I am telling the truth on the survey," failed the screening questions. Respondents who did not answer "yes" to the third question, "The answers I have given on the survey are true," were marked as invalid and failed the screening question. A total of 21.37% (n=125) of survey respondents failed at least one or more of the questions which flagged their responses as invalid.

To better understand the relationship between screening validity questions, correlations between the three questions were calculated. The correlations included $r = .44$ between the first and second question, $r = .13$ between the first and third question, and $r = .11$ between the second and third (see Table 2). The positive intercorrelations between the three questions were statistically significant ($p < .01$). Results

Table 2
Correlations between Screening Validity Questions

Screening Validity Question	1	2	3
1) I am reading this survey carefully			
2) I am telling the truth on the survey	.44***		
3) The answers I have given on the survey are true	.13***	.11***	

*$p < .10$, **$p < .05$, ***$p < .01$

Table 3
Demographics of Valid and Invalid Respondents

Demographic	Valid respondents		Invalid respondents		
	n	%	n	%	Chi Square
Ethnicity					0.03
White	331	78.62%	90	21.38%	
Students of Color	113	77.93%	32	22.07%	
Living situation					1.74
Live-in	132	81.99%	29	18.01%	
Live-out	317	76.94%	95	23.05%	
Class academic standing					5.61**
Freshmen & Sophomores	294	75.77%	94	24.23%	
Juniors & Seniors	166	84.26%	31	15.74%	

*$p < .10$, **$p < .05$, ***$p < .01$

suggest respondents who failed one question were also likely to fail additional screening questions.

Demographic responses were examined for differences between valid and invalid respondents. Juniors and senior students were combined into one category to meet sample size assumptions of the chi-square hypothesis testing (Berman & Wang, 2012). For this same reason, sexual orientation was excluded from significance testing due to only a few invalid respondents identifying as gay, bisexual, or queer. Results indicate freshmen and sophomore students had a higher likelihood of failing the screening questions compared to juniors and seniors, $X^2(1, N = 585) = 5.61$, $p = .018$ (see Table 3).

To answer the research question of invalid respondents' potential impact on rape myth acceptance and conformity to masculinity norms, mean differences were calculated comparing valid and invalid responses. Inflation rates are also reported which is the percent change of total sample responses that is due to the invalid

Table 4
Invalid Responses Impact on Illinois Rape Myth Acceptance

IRMA Subscale	Total sample			Valid responses			Invalid responses					
	N	M	SD	N	M	SD	N	M	SD	Inflation Rate	Diff.	p-value
1: She asked for it	562	3.94	0.7	441	3.99	0.7	121	3.74	0.72	-1.27%	0.25	.0004***
2: He didn't mean to	553	3.47	0.71	434	3.51	0.71	119	3.32	0.72	-1.15%	0.19	.0119**
3: It wasn't really rape	545	4.49	0.65	431	4.58	0.53	114	4.15	0.91	-2.00%	0.43	.0000***
4: She lied	549	3.16	0.92	429	3.18	0.94	120	3.07	0.86	-0.63%	0.11	0.23
IRMA total	511	3.76	0.56	404	3.81	0.55	107	3.57	0.58	-1.33%	0.24	.0001***

*$p < .10$, **$p < .05$, ***$p < .01$

responses. Table 4 reports mean differences between valid and invalid respondents for the IRMA scale and subscales and inflation rates. The overall mean IRMA for fraternity members is 3.76 with a -1.33% inflation rate. A significant difference of 0.24 (p =.0001) was found between IRMA mean valid (3.81) and invalid (3.57) responses.

This negative inflation indicates that invalid responders are more likely to have attitudes and beliefs supportive of rape myths. Three IRMA subscales out of the four were statistically significant at conventional levels. Subscale 1: She asked for it (0.25, p =.0004), Subscale 2: He didn't mean to (0.19, p = .0119), and Subscale 3: It wasn't really rape had the largest difference (0.43, p = .0000) and a related inflation of -2.00%.

Table 5 reports invalid responders have higher conformity to masculinity norms with a near conventional level of significant difference between valid responses and invalid responses of .06 (p = .0610) and an inflation rate of -0.40%. Two CMNI factors reported approaching significant differences between valid and invalid responses. Playboy (0.11, p = .0598) and Self-reliance (0.09, p =.0802) with a ranging inflation rate of -0.70% to -0.82%.

Limitations

Screening validity questions measure self-

Table 5

Invalid Responses Impact on Conformity to Masculine Norms Inventory

CMNI sub-scale	Total sample			Valid responses			Invalid responses					
	N	M	SD	N	M	SD	N	M	SD	Inflation Rate	Diff.	p-value
Winning	568	2.06	0.5	444	2.07	0.51	124	2.03	0.45	-0.49%	0.04	0.5263
Emotional control	564	2.59	0.49	443	2.6	0.5	121	2.54	0.43	-0.39%	0.06	0.2256
Work	568	2.32	0.54	449	2.32	0.56	119	2.36	0.47	-	-0.04	0.4649
Risk	565	2.45	0.41	446	2.47	0.43	119	2.4	0.36	-0.82%	0.07	0.1248
Violence	555	2.19	0.54	437	2.2	0.55	118	2.15	0.44	-0.46%	0.05	0.419
Hetero Presentation	546	2.48	0.59	431	2.49	0.6	115	2.47	0.56	-0.40%	0.02	0.8129
Playboy	564	2.45	0.57	444	2.47	0.58	120	2.36	0.52	-0.82%	0.11	.0598*
Self-reliance	564	2.87	0.47	443	2.89	0.48	121	2.8	0.43	-0.70%	0.09	.0802*
Power	555	3.02	0.51	438	3.04	0.51	117	2.96	0.53	-0.66%	0.08	0.1088
CMNI total	470	2.48	0.26	368	2.49	0.27	102	2.43	0.2	-0.40%	0.06	.0610*

*$p < .10$, **$p < .05$, ***$p < .01$

reported respondent accuracy to answer survey questions truthfully. An unintended outcome of the questions also discussed by Cornell et al. (2012), Cornell et al. (2014) and Jia et al. (2016) may reflect respondents' lack of motivation and attention to read the survey directions, questions, or statements carefully or at all. These screening validity questions do not only measure respondents' truthfulness but also those who may simply write down values without any thought, resulting in inaccurate data. Therefore, it may be possible the invalid respondents are those who fall into a process of satisficing. Krosnick (1991, 1999) describes this process as the level of engagement a respondent has with completing a survey dependent on their motivation. If there is a lack of motivation to accurately review the question, understand what it is asking, reflect on an answer, select the best answer, the respondent may choose to instead select responses that do not have a relationship with the question itself. Screening validity questions still accomplishes the goal of identifying responses that are not thoughtfully answered and truthful.

The present research included one fraternity community at a large public university and therefore lacks generalizability to other populations outside the institution. The research could be easily replicated at a national level by adding similar questions on nationally recognized surveys assessing fraternity experience to overcome this limitation.

Discussion

The purpose of the study was to evaluate the impact of using screening validity questions on self-report surveys measuring rape myth acceptance and conformity to masculinity norms among fraternity members. Screening questions added to a survey measuring IRMA and CMNI helped flag invalid responses that inflated these constructs. Similar to other studies (e.g., Cornell et al., 2012; Cornell et al., 2014; Jia et al., 2016), screening questions were found to impact the measured outcome and could then

be omitted before additional statistical analysis to improve the validity of results. This study expanded on the past research by including screening questions with undergraduate college students and different measures.

The only significant difference within demographic groups measured on the survey was freshman and sophomore students had a higher number of invalid respondents than expected when compared to juniors and seniors. This was consistent with past research (i.e., Jia et al., 2016) indicating younger grade levels are more likely to be invalid respondents. Previous research indicated consistent findings with middle or high school students of Color reporting more invalid response than White students. The present research did not report a significant difference in this case.

This study was conducted on a population of fraternity men, in part, due to the consistent findings indicating men were more likely than women to be flagged as invalid responders. Results expand on this finding showing men who were invalid responders are more likely to have higher conformity to masculinity norms compared to valid responders. Qualitative research may be well suited to look into why and how masculinity is positively related to responding to self-report surveys in ways that indicate they are not reading carefully or telling the truth.

Cornell et al. (2014) describe one reason why invalid respondents may occur is because, "Immature and rebellious adolescents may be tempted to offer inflated reports of risky behaviors or they may not take a survey seriously and mark it haphazardly" (p. 1). Given immature and adolescent behaviors often times observed among fraternity organizations, this too may be a reason self-report surveys identify invalid respondents. Fraternity members who do not take questions seriously and give a thoughtful answer, future surveys being administered to fraternity members should consider using screening validity questions.

Implications for Practice

College campuses today, more than ever, have seen an increase in needs for accurate information on sexual violence experiences and perceptions (Woods et al., 2016). Fraternity professionals are simultaneously using surveys to assess student programs and services in order to effectively advocate for continued resources and improvement purposes. The present research indicates some fraternity members may not be reading carefully or telling the truth on surveys, which has implications for fraternity organizations that aim to measure attitudes and experiences on sensitive topics.

After all, fraternity professionals will continue to have difficulty designing and evaluating impactful programs on sensitive topics such as sexual violence prevention if participants' lack of truthfulness on assessment measures distorts our understandings of the problem. If including simple validity questions can help practitioners weed out those respondents who are inflating survey results, it is likely that programming or policy decisions drawn from survey results can more effectively meet the target population where they actually are. Thus, research and assessment practices using survey methodology will likely benefit from adopting screening questions to flag invalid responders. Before additional statistical analysis and reporting of the results begin with collected survey data, invalid responders should be omitted from the results to improve data quality and validity of the findings.

Further, the findings of this study highlight the particular importance of validity screening for more sensitive topics in surveys. After all, the IRMA asks significantly more sensitive questions than the CMNI, as the IRMA questions delve deeply into taboo attitudes about rape and sexual assault, and the rates of inflation were considerably more concerning on the IRMA than the CMNI. The present study, then, highlights that professionals ought to be wary of trusting survey data on the most sensitive of subjects when

that data has not been screened for validity in a manner like the method advocated here. After all, in the present study, including invalid responders in the dataset would lead practitioners to think there is a significantly more serious concern about rape supportive attitudes in the population than might be accurate, as those with higher self-reported rates of rape supportive attitudes are the same students who openly admitted they were not reading carefully or telling the truth.

Poor data quality from surveys has considerable negative consequences as the subsequent results may be used to inform policy and practices. More research and consideration should be given to survey design and implementation within fraternity community assessment and research to counter these issues. Surveys are not likely to decrease given their low-cost and ease of collecting quantitative data in assessment and research. Adding screening validity questions may be one simple, yet effective, way to reduce bias and increase data quality.

Simply put, practitioners rely on assessment tools to guide the direction of programs and services, but when those assessment tools are biased by invalid responses, fraternity professionals may very well be directing their resources toward the most problematic responses even though those responses may be a result of lying or inattention. The simple act of including three validity items into surveys and screening out those who fail the validity check can help ensure fraternity professionals are utilizing their limited time and energy in the best and most efficacious ways possible by increasing survey data quality.

References

Berman, E. M., & Wang, X. (2012). *Essential statistics for public managers and policy analysts* (3rd ed.). Thousand Oaks, CA: SAGE.

Bleecker, E.T., & Murnen, S. K. (2005). Fraternity membership, the display of degrading sexual images of women, and rape myth acceptance. *Sex Roles, 53*(7-8), 487-493. https://doi.org/10.1007/s11199-005-7136-6

Bresciani, M. J., Gardner, M. M., & Hickmotf, J. (2012). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs.* Sterling, VA: Stylus.

Canan, S. N., Jozkowski, K. N., & Crawford, B. L. (2016). Sexual assault supportive attitudes: Rape myth acceptance and token resistance in Greek and non-Greek college students f o r m two university samples in the United States. *Journal of Interpersonal Violence, 33*(22), 3502 to 3530. https://doi.org/10.1177/0886260516636064

Capraro, R. L. (2000). Why college men drink: Alcohol, adventure, and the paradox of masculinity. *Journal of American College Health, 48*(6), 307-315. https://doi.org/10.1080/07448480009596272

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*(1), 21-35. https://doi.org/10.1037/a0024824

Cornell, D., Lovegrove, P. J., & Baly, M. W. (2014). Invalid survey response patterns among middle school students. *Psychological Assessment, 26*(1), 277-287. https://doi.org/10.1037/a0034808

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). I*nternet, phone, mail, and mixed-mode surveys.* Hoboken, NJ: John Wiley & Sons.

Foubert, J. D., Brosi, M. W., & Bannon, R. S. (2011). Pornography viewing among fraternity men: Effects on bystander intervention, rape myth acceptance and behavioral intent to commit sexual assault. *Sexual Addiction and Compulsivity, 18*(4), 212-231. https://doi.org/10.1080/10720162.2011.625552

Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior, 28*(1), 251-256. https://doi.org/10.1016/j.chb.2011.09.007

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567. https://doi.org/10.1146/annurev.psych.50.1.537

Jia, Y., Konold, T. R., Cornell, D., & Huang, F. (2016). The impact of validity screening on associations between self-reports of bullying victimization and student outcomes. *Journal of Educational and Psychological Measurement,* 1-23. https://doi.org/10.1177/0013164416671767

Joinson, A. N., Paine, C., Buchanan, T., & Reips, U. R. (2007). Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior, 24*(5), 2158-2171. https://doi.org/10.1016/j.chb.2007.10.005

Iwamoto, D. K., Cheng, A., Lee, C. S., Takamatsu, S., & Gordon, D. (2011). "Man-ing" up and getting drunk: The role of masculine norms, alcohol intoxication and alcohol-related problems among college men. *Addictive Behaviors, 36*(9), 906-911. https://doi.org/10.1016/j.addbeh.2011.04.005

Mahalik, J. R., Locke, B. D., Ludlow, L. H., Diemer, M. A., Scott, R. P. J., Gottfried, M., & Freitas, G. (2003). Development of the conformity to masculine norms inventory. *Psychology of Men & Masculinity, 4*(1), 3-25. https://doi.org/10.1037//1524-9220.4.1.3

McMahon, S., & Lawrence, G. F. (2011). An updated measure for assessing subtle rape myths. *SocialWork Research, 35*(2), 71-81. https://doi.org/10.1093/swr/35.2.71

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.

Murnen, S. K., & Kohlman, M. H. (2007). Athletic participation, fraternity membership, and sexual aggression among college men: A meta-analytic review. *Sex Roles, 57*(1-2), 145-157. https://doi.org/10.1007/s11199-007-9225-1

Parent, M. C., & Moradi, B. (2009). Confirmatory factor analysis of the conformity to masculine norms inventory and development of the conformity to masculine norms inventory-46. *Psychology of Men & Masculinity, 10*(3), 175-189. https://doi.org/10.1037/a0015481

Payne, D. L., Lonsway, K.A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality, 33*, 27-68. https://doi.org/10.1006/jrpe.1998.2238

Saunders, K., & Cooper, R. M. (2009). Instrumentation. In J. H. Schuh (Eds.), *Assessment methods for student affairs* (pp. 107-130). San Francisco, CA: Jossey-Bass.

Seabrook, R. C., Ward, L. M., & Giaccardi, S. (2016). Why is fraternity membership associated with sexual assault? exploring the roles of conformity to masculine norms, pressure to uphold masculinity, and objectification of women. *Psychology of Men and Masculinity, 19*(1), 3-13. https://doi.org/10.1037/men0000076

Schuh, J. H., Biddix, J. P., Dean, L. A., & Kinzie, J. (2016). *Assessment in student affairs* (2nd ed.). San Francisco, CA: Jossey-Bass.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin,133*(5), 859-883. https://doi.org/10.1037/0033-2909.133.5.859

Turrisi, R., Mallett, K. A, Mastroleo, N. R., & Larimer, M.E. (2010). Heavy drinking in college students: Who is at risk and what is being done about it? *Journal of General Psychology, 133*(4), 401-420. https://doi.org/10.3200/genp.133.4.401-420

Wood, L., Sulley, C., Kammer-Kerwick, M., Follingstad, D., & Busch-Armendariz, N. (2016). Climate surveys: An inventory of understanding sexual assault and other crimes of interpersonal violence at institutions of higher education. *Violence Against Women, 23*(10).https://doi.org/10.1177/1077801216657897

Author Biographies

Lucas Schalewski is the Associate Director for Assessment, Research, & Grant Development at University of Arizona and a Ph.D. student in the Center for the Study of Higher Education.

Jamie Utt is a doctoral student in Teaching, Learning, and Sociocultural Studies at the University of Arizona where he studies intersections of race and schooling through critical mixed methodologies.

Bryant Valencia is a Ph.D. candidate in the Center for the Study of Higher Education and holds a graduate assistant in the University of Arizona's Early Academic Outreach department as the coordinator of masculinity outreach initiatives.