

ARTICLES

Comparing Amazon's MTurk and a Sona Student Sample: A Test of Data Quality Using Attention and Manipulation Checks

Yani Zhao^{ORCID}, Sherice Gearhart^{ORCID}

Keywords: Survey Research, Amazon MTurk, Student Sample, Data Quality

<https://doi.org/10.29115/SP-2023-0005>

Survey Practice

Vol. 16, Issue 1, 2023

The need for cost-effective data collection leads researchers to explore options, especially for respondent-administered online surveys. Student samples are convenient and cheap for social scientists when students fit the target population. However, student samples are criticized for their homogeneity and lack of generalizability (Kees et al. 2017). Another low-cost option is Amazon Mechanical Turk (MTurk), a crowdsourcing platform used for collecting online data from a seemingly broader population. Despite the appeal, it is important to compare data quality. The purpose here is to compare data quality between MTurk and student samples. To control data quality, researchers rely on several tactics such as screener questions to exclude unqualified respondents (Arndt et al. 2022). Subjective manipulation check and attention-check are used to examine respondent engagement and performance. Completion speed might also indicate effort/attention. Since samples should collect data from participants resembling the target population, sample diversity also serves as an indicator of data quality in this study (Kees et al. 2017; Roulin 2015). However, it should be noted that having a diverse sample does not always guarantee higher sample quality, especially when conducting studies on a homogeneous population.

Literature Review

Amazon MTurk

Amazon founded MTurk in 2005 for crowdsourcing online tasks (Aruguete et al. 2019). It has since been used for collecting data from human subjects (Arndt et al. 2022) due to its speed (Buhrmester, Kwang, and Gosling 2011), low expense (Bates and Lanza 2013), and demographically and geographically diverse samples (Chandler, Mueller, and Paolacci 2013; Sheehan and Pittman 2016). The service is beneficial to the private sector and academic studies (Amazon Mechanical Turk 2019). Used by social scientists since 2011 (Buhrmester, Kwang, and Gosling 2011), MTurk has become a crucial resource for social scientists (Edelman 2020). Since 2012, it has been noted that more than half of the top universities in the United States have published research using MTurk to collect data about behaviors (Goodman, Cryder, and Cheema 2012).

Student Sample

Student samples are common data collection sources due to convenience and low cost (Höglinger and Wehrli 2017). However, students are often younger, geographically constrained, educated, and relatively inexperienced with the world (Kees et al. 2017), limiting generalizability. Regardless, many universities provide student participant pools. Sona Systems (i.e., Sona) is a popular platform used to create, manage, and control sign-up and credit assignment

(Sona 2022). Studies comparing student samples and MTurk have been inconclusive (Feitosa, Joseph, and Newman 2015) in terms of data quality and respondent characteristics.

Screeners or Filter Question

Screeners help identify eligible participants (Buchheit et al. 2018). Screeners are generally based on population attributes and are employed to reduce sampling coverage bias, to ensure respondents have the predetermined characteristics of the population (Fricker 2008). Screeners sift respondents to exclude invalid responses to ensure respondents possess certain characteristics (Arndt et al. 2022).

MTurkers are monetarily motivated, increasing the likelihood that they will participate regardless of eligibility. Compared to students, MTurkers demonstrate lower pass rates on location-related and demographic screeners (Arndt et al. 2022). However, MTurkers have a higher pass rate compared to alternatives (Boas, Christenson, and Glick 2018; Chandler et al. 2019). With little research comparing MTurk and student sample performance, this study explores this aspect of data quality.

Subjective Manipulation Checks

Manipulation checks are integrated to ensure participants are affected as the research design intended (Hauser, Ellsworth, and Gonzalez 2018). Different from instructional manipulation checks used to assess attentiveness, subjective manipulation checks investigate respondents' perceptions of manipulated variables (Kane and Barabas 2019). Performance on manipulation checks assess data quality and limited research compares the pass rate between MTurk and student samples. MTurkers perform better than commercial panels/services (Berry, Kees, and Burton 2022; Zhang and Gearhart 2020). However, effect sizes on framing manipulations were highest for students in a lab, followed by online student samples, and MTurkers, respectively, while professional samples showed a lower effect size (Kees et al. 2017). This study proposes that students should outperform MTurkers.

Attention-check

Another aspect of data quality is passing rate on attention-checks, sometimes referred to as 'validity checks' (Aruguete et al. 2019). Existent findings on instructional manipulation checks are conflicted. Generally, MTurkers are more attentive since they are monetarily incentivized (Hauser and Schwarz 2016). However, others indicate they are less likely to pass attention-checks compared to students (Aruguete et al. 2019; Goodman, Cryder, and Cheema 2012) or limited differences exist (Kees et al. 2017). To examine data quality, attention-checks are examined to assess differences.

Completion Time/Speed

Completion time is an indicator of data quality and respondent effort. “Participants who answer questions very quickly are unlikely to be giving sufficient effort or attention, hurting data quality” (Chandler et al. 2019). This is “rushing,” and it is assumed that respondents inadvertently or intentionally may not carefully read and/or answer questions (Leiner 2019), resulting in poor data quality (Zhang and Gearhart 2020). Conversely, long completion times may indicate a pause or disengagement (Aruguete et al. 2019). MTurkers are known to have faster completion speeds than students (Aruguete et al. 2019; Kees et al. 2017).

Sample Diversity

Sample diversity is essential when comparing to the population of interest and is commonly investigated using demographics (Buchheit et al. 2018). The demographic composition of MTurkers has been found to reach diverse ages, genders, ethnicities, income, and education levels compared to students (Casler, Bickel, and Hackett 2013; Weigold and Weigold 2022). To test diversity, this study compares the racial/ethnic makeup of both samples.

Research Questions and Hypotheses

RQ1: Will MTurk or students have a higher pass rate on screener questions?

H1a: The student sample will perform better on the context-specific frame manipulation check.

H1b: The student sample will perform better on the context-transcendent frame manipulation check.

RQ2: Will the MTurk or student respondents perform better on the attention-check?

H2: MTurk respondents have faster completion times than students.

RQ3: Will the MTurk or student sample have more ethnically/racially diverse respondents?

Methodology

With institutional review board approval, a purposive sample of female Instagram users were recruited with either a monetary or course credit incentive. Participants were instructed to complete a brief questionnaire and view a short video (i.e., 71 seconds) before finalizing participation, which was advertised as taking less than 15 minutes.

Participants and Procedure

STUDY 1: AMAZON MTURK

Data collected from MTurk came from 225 participants who began the online questionnaire on March 22, 2022. Respondents were paid \$1 for completion and an extra \$0.50 for a “premium qualification” intended to guarantee female respondents. Since requesters are urged to be flexible and not hurry participants, participants had a 5-hour time frame and were paid after 48 hours. All participants who attempted the initial screener questions were included to assess pass rate for screeners. Followingly, only participants who passed the initial screener questions (see Appendix 1) and were reported as completing the survey in the Qualtrics report were included in the final sample ($N = 211$).

STUDY 2: STUDENT SAMPLE

Study 2 was posted May 2, 2022, to collect a purposive sample of female undergraduate students using a university-affiliated Sona participant pool, in exchange for one quarter-hour credit in courses. Participation was begun by 335 individuals who were all used to assess performance on the screeners. The final sample was determined ($N = 317$) after removing respondents who did not pass the screeners or reported as unfinished.

Measures

Screener check. Two questions measured whether participants were female Instagram users. The first asked whether they currently use Instagram (0 = *no* or 1 = *yes*). The second asked self-identified gender (0 = *male/prefer not to say* or 1 = *female*).

Manipulation check. Manipulation checks aim to examine whether respondents can identify manipulated variables (i.e., message frames). The first item required respondents to discern context-specific frames (0 = *incorrect* or 1 = *correct*). The second asked respondents to identify context-transcendent frames (0 = *incorrect* or 1 = *correct*).

Attention-check. Attention-checks ensure respondents pay attention instead of selecting answers randomly. Within a set of questions, a single item asked respondents to select “strongly agree” on a 7-point Likert-type scale (1 = *strongly disagree* to 7 = *strongly agree*). Responses were dummy-coded (0 = *incorrect* or 1 = *correct*).

Race/Ethnicity. Respondents self-reported their racial/ethnic identity from a list of six options (1 = *African American/Black*, 2 = *Asian/Pacific Islander*, 3 = *Caucasian (non-Hispanic)*, 4 = *Hispanic/Latino*, 5 = *Native American/Alaskan native*, and 6 = *other*).

Table 1. Summary of Results

	MTurk Sample		Student Sample		PValue	Phi
	Pass Rate	N Passed	Pass Rate	N Passed		
Screener – Instagram User	97.2%	205	94.6%	300	.20	-.06
Screener – Gender	85.3%	180	81.4%	258	1.00	.14
Context-Specific Frame Manipulation Check	71.3%	127	82.8%	202	.01	.14
Context-Transcendent Frame Manipulation Check	48.6%	86	67.9%	165	<.001	.19
Attention-check	89.9%	161	70.9%	175	<.001	-.23
	Percentage	N	Percentage	N	PValue	phi
Race/ethnicity					<.001	.26
Caucasian	73.2%	131	63.7%	158		
Asian	3.4%	6	2.0%	5		
Native Americans	6.7%	12	0.8%	2		
Hispanics	8.4%	15	23.4%	58		
African American/black	5.6%	10	8.1%	20		
Other	2.8%	5	2.0%	5		
	Mean	Median	Mean	Median	PValue	r
Time Completion	594.89s	511s	916.73s	494.5s	.12	-.07

Results

RQ1 asked pass rates on screeners (use of Instagram and self-identified gender) between MTurkers and students (see [Table 1](#)). Results of chi-square analysis suggest no significant difference MTurkers and students on their pass rate on either the first ($\chi^2 (1, N = 528) = 1.93, p = 0.20$) or second screener ($\chi^2 (1, N = 528) = 1.38, p = 0.29$). A *z*-test of proportion indicates that MTurkers (97.2%, 85.3%, respectively) and students (94.6%, 81.4%, respectively) did not feature a different pass rate on either screener. Therefore, both provided good quality data.

H1a predicted that the student sample would perform better on the context-specific frame manipulation check. Chi-square analysis suggests that the student sample performs significantly better than do MTurkers ($\chi^2 (1, N = 422) = 7.84, p = .01$). A *z*-test of proportion reveals a significant difference between how often the students (82.8%) and the MTurkers (71.3%) correctly answered the first manipulation check. H1a was supported and students featured greater data quality.

H1b predicted the student sample would perform better on the context-transcendent frame manipulation check. Chi-square analysis suggests that students perform significantly better on the second manipulation check than do MTurkers ($\chi^2 (1, N = 420) = 15.89, p < .001$). A *z*-test for proportion indicates that students correctly answered the manipulation check more often (67.9%) than did MTurkers (48.6%). H1b was supported. Yet, pass rates indicate low data quality.

RQ2 asked whether MTurkers or students would perform better on the attention-check. Chi-square suggests that MTurkers perform significantly better on the attention-check than do students ($\chi^2 (1, N = 427) = 23.28, p < .001$). A z -test for proportion indicated MTurkers correctly responded to the attention-check more frequently (89.9%) than students (70.6%). Therefore, MTurk produced higher quality data.

H2 predicted that MTurkers have faster completion times than students. Due to outliers, medians were used to compare completion times. A Mann-Whitney U test was run to assess differences in completion time. Results indicate that time spent in seconds was not significantly different between MTurkers ($Mdn = 511; M = 692.78$) and students ($Mdn = 494.5; M = 1124.38$), $U = 30771, z = -1.56, p = .12$. Therefore, H2 was not supported.

RQ3 asked whether the MTurk or student sample has more diverse respondents. Among completions, chi-square analysis suggests that the student sample has significantly more diversity ($\chi^2 (5, N = 427) = 28.00, p < .001$). MTurkers featured a greater proportion of Caucasian (73.2%), Asian (3.4%) and Native Americans (6.7%) than did the students (63.7%, 2.0%, 0.8%, respectively). Conversely, MTurkers featured less Hispanics (8.4%) and African American/black (5.6%) than the students (23.4%, 8.1%, respectively). However, both MTurkers and students featured similar proportions of "other" (2.8%, 2.0%, respectively).

Discussion

While scholars often rely on student research participants, they are led to other options to reach heterogeneous samples. This study compared the data quality between a student sample and MTurk participants as cost-effective options for collecting data. Results showed similar pass rates on screen questions between both groups, but students outperformed MTurkers on subjective manipulation checks and had more ethnic diversity, suggesting Sona may be a good option for cost-conscious researchers.

Both groups had participants who failed screener questions, likely due to carelessness and multitasking. These results contrast previous findings that MTurkers are less likely to pass screeners (Arndt et al. 2022). Since both populations were incentivized, they may not consider the purpose or may quickly complete participation while multitasking, further disregarding instructions, which MTurkers are known to do (Necka et al. 2016). Similarly, students participated near the end of the semester, possibly making them eager to earn credit and respond carelessly.

Students performed better on subjective manipulation checks, aligning with previous work (Kees et al. 2017). This may be due to students being more likely to comply with authority (Sears 1986), while MTurkers are more focused on earning monetary incentives and may find it difficult to fully engage, especially while multitasking (Necka et al. 2016).

MTurkers performed better on the attention-check compared to students. While consistent with some (Hauser and Schwarz 2016), results conflicted with others revealing that students are more attentive (Arndt et al. 2022; Aruguete et al. 2019) or no different (Kees et al. 2017). Performance differences might be associated with motivations (Aruguete et al. 2019). Given different motivational incentives, MTurkers may pay more attention to payment (Hauser and Schwarz 2016), leading to their reputation as professional survey takers (Hillygus, Jackson, and Young 2014).

Results of the manipulation check offer some insight into the participants' focus and cognitive engagement. However, differences in performance between the MTurk and student samples highlight a challenge of using MTurk. Since MTurkers are considered professional survey responders, they may be better equipped to seek out and identify attention-checks, which are often recognizable. Alternative methods for assessing attention that can be integrated seamlessly to avoid distraction should be considered. When choosing a data source, researchers should consider the experience of the participants on the platform.

Concerning subjective manipulation check questions, many studies do not separate these from attention-checks/instructional manipulation checks. For example, Kees et al. (2017) applied one subjective manipulation check question in their attention-check/instructional manipulation check. Berry, Kees, and Burton (2022) borrowed their attention-check questions from Kees et al. (2017) and thereby had the same problem. Such an issue should be paid more attention to in future studies on data quality comparison.

The average time it took MTurkers to complete the questionnaire was within the expected maximum range of 900 seconds. However, the student sample took an average of 1124.38 seconds to finish, revealing the presence of outliers. To compare response times, median scores were used. Although there is no significant difference, the mean score indicated that students featured longer completion time compared to MTurkers on average. Longer completion time could indicate disengagement, which might explain lower attention-check pass rate in the student sample.

Students were more diverse in ethnicity/race compared to MTurk. This may be due to student body composition in the southwest of the United States, which came from a federally designated Hispanic Serving Institution where over 25% self-identify as Hispanic (U.S. Department of Education 2022). The broad MTurk network may lead researchers to believe they are reaching a diverse sample, but it may not compete with a diverse student body. The choice of sample collection method, whether using MTurk or student samples, should be guided by the target population. The ultimate goal of sample selection is to obtain a representative sample that accurately reflects the population of interest. To achieve this, researchers must take into account the target population's characteristics and consider potential sources of bias that could

affect the representativeness of the sample. It is important to note that if the population of interest is homogeneous, diversity in the sample may not necessarily result in higher sample quality. Nonetheless, even within a homogeneous population, there may be variability in relevant characteristics that are pertinent to the research question. In such cases, it may still be necessary to ensure that the sample is representative of the population in terms of these relevant characteristics.

Limitations

Limitations of this work include the unequal sample sizes, which necessitated the application of nonparametric tests. MTurkers received payment for participation, resulting in a smaller sample size than the student group, which possibly decreased the power of the analysis compared to equal group sizes. Respondents also had the potential to accidentally pass attention and manipulation checks. Random order was used, respondents had equal distribution and saw identical items, which should minimize false positives. However, respondents may choose correct answers by chance. Future studies should approach subjective manipulation checks with caution to make sure they are easily comprehended. Other variables should also be considered when assessing data quality. Finally, the comparison of sample quality was only conducted on females, which limited the generalizability of the findings.

Submitted: December 23, 2022 EST, Accepted: May 20, 2023 EST

REFERENCES

- Amazon Mechanical Turk. 2019. "Getting Started with Surveys on MTurk." *Happenings at MTurk*. Updated December 10, 2021. <https://blog.mturk.com/getting-started-with-surveys-on-mturk-e2eca524c73>.
- Arndt, Aaron D., John B. Ford, Barry J. Babin, and Vinh Luong. 2022. "Collecting Samples from Online Services: How to Use Screeners to Improve Data Quality." *International Journal of Research in Marketing* 39 (1): 117–33. <https://doi.org/10.1016/j.ijresmar.2021.05.001>.
- Aruguete, Mara S., Ho Huynh, Blaine L. Browne, Bethany Jurs, Emilia Flint, and Lynn E. McCutcheon. 2019. "How Serious Is the 'Carelessness' Problem on Mechanical Turk?" *International Journal of Social Research Methodology* 22 (5): 441–49. <https://doi.org/10.1080/13645579.2018.1563966>.
- Bates, John A., and Brian A. Lanza. 2013. "Conducting Psychology Student Research Via the Mechanical Turk Crowdsourcing Service." *North American Journal of Psychology* 15 (2): 385.
- Berry, Christopher, Jeremy Kees, and Scot Burton. 2022. "Drivers of Data Quality in Advertising Research: Differences across MTurk and Professional Panel Samples." *Journal of Advertising* 51 (4): 515–29. <https://doi.org/10.1080/00913367.2022.2079026>.
- Boas, Taylor C., Dino P. Christenson, and David M. Glick. 2018. "Recruiting Large Online Samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics." *Political Science Research and Methods* 8 (2): 232–50. <https://doi.org/10.1017/psrm.2018.28>.
- Buchheit, Steve, Marcus M. Doxey, Troy Pollard, and Shane R. Stinson. 2018. "A Technical Guide to Using Amazon's Mechanical Turk in Behavioral Accounting Research." *Behavioral Research in Accounting* 30 (1): 111–22. <https://doi.org/10.2308/bria-51977>.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk." *Perspectives on Psychological Science* 6 (1): 3–5. <https://doi.org/10.1177/1745691610393980>.
- Casler, Krista, Lydia Bickel, and Elizabeth Hackett. 2013. "Separate but Equal? A Comparison of Participants and Data Gathered via Amazon's MTurk, Social Media, and Face-to-Face Behavioral Testing." *Computers in Human Behavior* 29 (6): 2156–60. <https://doi.org/10.1016/j.chb.2013.05.009>.
- Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2013. "Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers." *Behavior Research Methods* 46 (1): 112–30. <https://doi.org/10.3758/s13428-013-0365-7>.
- Chandler, Jesse, Cheskie Rosenzweig, Aaron J. Moss, Jonathan Robinson, and Leib Litman. 2019. "Online Panels in Social Science Research: Expanding Sampling Methods beyond Mechanical Turk." *Behavior Research Methods* 51 (5): 2022–38. <https://doi.org/10.3758/s13428-019-01273-7>.
- Edelman, Josef. 2020. "How to Conduct Better Research Studies with the MTurk Toolkit." *CloudResearch*. Updated January 5, 2021. <https://www.cloudresearch.com/resources/research/how-mturk-toolkit-enhances-research-studies/>.
- Feitosa, Jennifer, Dana L. Joseph, and Daniel A. Newman. 2015. "Crowdsourcing and Personality Measurement Equivalence: A Warning about Countries Whose Primary Language Is Not English." *Personality and Individual Differences* 75 (March): 47–52. <https://doi.org/10.1016/j.paid.2014.11.017>.
- Fricker, Ronald D. 2008. "Sampling Methods for Web and E-Mail Surveys." In *The SAGE Handbook of Online Research Methods*, edited by Nigel G. Fielding, Raymond M. Lee, and Grant Blank, 195–216. Thousand Oaks, CA: SAGE Publications Ltd. <https://doi.org/10.4135/9780857020055.n11>.

- Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2012. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26 (3): 213–24. <https://doi.org/10.1002/bdm.1753>.
- Hauser, David J., Phoebe C. Ellsworth, and Richard Gonzalez. 2018. "Are Manipulation Checks Necessary?" *Frontiers in Psychology* 9 (June). <https://doi.org/10.3389/fpsyg.2018.00998>.
- Hauser, David J., and Norbert Schwarz. 2016. "Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants." *Behavior Research Methods* 48 (1): 400–407. <https://doi.org/10.3758/s13428-015-0578-z>.
- Hillygus, D. Sunshine, Natalie Jackson, and McKenzie Young. 2014. "Professional Respondents in Nonprobability Online Panels." In *Online Panel Research: A Data Quality Perspective*, edited by Mario Callegaro, Reginald P. Baker, Jelke Bethlehem, Anja S. Göritz, Jon A. Krosnick, and Paul J. Lavrakas, 219–37. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781118763520.ch10>.
- Höglinger, Marc, and Stefan Wehrli. 2017. "Measuring Social Preferences on Amazon Mechanical Turk." In *Social Dilemmas, Institutions, and the Evolution of Cooperation*, edited by Ben Jann and Wojtek Przepiorka, 527–46. Berlin, Germany: Walter de Gruyter GmbH. <https://doi.org/10.1515/9783110472974-025>.
- Kane, John V., and Jason Barabas. 2019. "No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments." *American Journal of Political Science* 63 (1): 234–49. <https://doi.org/10.1111/ajps.12396>.
- Kees, Jeremy, Christopher Berry, Scot Burton, and Kim Sheehan. 2017. "An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk." *Journal of Advertising* 46 (1): 141–55. <https://doi.org/10.1080/00913367.2016.1269304>.
- Leiner, Dominik Johannes. 2019. "Too Fast, Too Straight, Too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys." *Survey Research Methods* 13 (3): 229–48. <https://doi.org/10.18148/SRM/2019.V13I3.7403>.
- Necka, Elizabeth A., Stephanie Cacioppo, Greg J. Norman, and John T. Cacioppo. 2016. "Measuring the Prevalence of Problematic Respondent Behaviors among MTurk, Campus, and Community Participants." Edited by Jelte M. Wicherts. *PLOS ONE* 11 (6): e0157732. <https://doi.org/10.1371/journal.pone.0157732>.
- Roulin, Nicolas. 2015. "Don't Throw the Baby Out With the Bathwater: Comparing Data Quality of Crowdsourcing, Online Panels, and Student Samples." *Industrial and Organizational Psychology* 8 (2): 190–96. <https://doi.org/10.1017/iop.2015.24>.
- Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51 (3): 515–30. <https://doi.org/10.1037/0022-3514.51.3.515>.
- Sheehan, Kim Bartel, and Matthew Pittman. 2016. *Amazon's Mechanical Turk for Academics: The HIT Handbook for Social Science Research*. Irvine, CA: Melvin & Leigh, Publishers LLC.
- Sona. 2022. "Platform Recruitment." Sona Systems. August 24, 2022. <https://www.sona-systems.com/platform/participant-recruitment/>.
- U.S. Department of Education. 2022. "White House Initiative on Advancing Educational Equity, Excellence, and Economic Opportunity for Hispanics." U.S. Department of Education. 2022. <https://sites.ed.gov/hispanic-initiative/hispanic-serving-institutions-hsis/>.
- Weigold, Arne, and Ingrid K. Weigold. 2022. "Traditional and Modern Convenience Samples: An Investigation of College Student, Mechanical Turk, and Mechanical Turk College Student Samples." *Social Science Computer Review* 40 (5): 1302–22. <https://doi.org/10.1177/08944393211006847>.

Zhang, Bingbing, and Sherice Gearhart. 2020. "Collecting Online Survey Data: A Comparison of Data Quality among a Commercial Panel & MTurk." *Survey Practice* 13 (1): 1–10. <https://doi.org/10.29115/sp-2020-0015>.

Appendix 1

Screenener questions

1. Do you use Instagram?
 - a. Yes
 - b. No
2. Which gender do you most identify with?
 - a. Male
 - b. Female
 - c. Prefer not to say

Subjective Manipulation Checks

3. Thinking about the video you just saw, which of the following was emphasized?
 - a. Facial Yoga & Your Health
 - b. Facial Yoga & Your Appearance
 - c. Eating habits
 - d. I don't know
4. Thinking about the video you just saw, which of the following was emphasized?
 - a. Health will Benefit from Facial Yoga
 - b. Appearance will Benefit from Facial Yoga
 - c. Health will Suffer without Facial Yoga
 - d. Appearance will Suffer without Facial Yoga
 - e. I don't know

Attention Check

5. For this statement, please select the strongly agree.
 - a. Strongly disagree
 - b. Disagree
 - c. Somewhat disagree
 - d. Neither agree nor disagree

- e. Somewhat agree
- f. Agree
- g. Strongly agree

Race/Ethnicity

6. Please indicate the race or ethnicity with which you most identify.
- a. African American/Black
 - b. Asian/Pacific Islander
 - c. Caucasian (non-Hispanic)
 - d. Hispanic/Latino
 - e. Native American or Alaskan native
 - f. Other (please specify below) _____