Evaluating Reddit as a Crowdsourcing Platform for Psychology Research Projects

Raymond Luong[1] and Anna M. Lomanowska[2] (iD)

Abstract

Background: Online crowdsourcing platforms, such as Amazon Mechanical Turk (MTurk), have become popular alternatives to the ubiquitous student samples used in psychology research. r/SampleSize, an alternative pool on the website Reddit, allows for online participant recruitment without compulsory or immediate payment, making it potentially useful for students, research trainees, and course instructors. **Objective:** The current study sought to assess the viability of using r/SampleSize as a participant pool by comparing its data characteristics to MTurk and existing lab samples. **Method:** Two hundred and fifty-six MTurk workers and 277 r/SampleSize participants completed identical questionnaires on demographics, participation motivations, and standard psychology scales. **Results:** Participants recruited through r/SampleSize reported diverse ages, education levels, income, and employment, although White ethnic background and US residence were predominant. r/SampleSize participants were more internally motivated than MTurk to participate in research and had greater need for cognition but did not differ significantly in altruism or motivation to gain self-knowledge. r/SampleSize data reliability and quality were comparable to MTurk and lab samples across most analyses. **Teaching Implications:** r/SampleSize can be used to recruit relatively large and diverse samples for undergraduate research projects with minimal setup, labor, and cost. **Conclusion:** The findings suggest that r/SampleSize is a diverse and viable participant pool.

Keywords

MTurk, Reddit, r/SampleSize, participant pool, online recruitment, crowdsourcing, online sample

Online crowdsourcing has become a popular method of recruiting research participants for studies in psychology. The ease and timeliness of data collection online as well as access to relatively large samples make this recruitment strategy particularly attractive for students, research trainees, and course instructors in psychology (Sciutto, 2015). Unlike traditionally used university student samples that suffer from problems with generalizability and representativeness (Hanel & Vione, 2016), online recruitment offers greater demographic diversity and data quality (Buhrmester et al., 2011; Peer et al., 2017; Rouse, 2015; Sciutto, 2015). These features can offer trainees the opportunity to pose more diverse research questions and enable improved training in research methodology by accounting for factors such as sufficient samples for statistical power (Vankov et al., 2014), as well as generalizability and representativeness in data collection (Henrich et al., 2010). The importance of training students to promote better research practices is particularly relevant in the context of the replicability crisis in psychology (Morling & Calin-Jageman, 2020; Perlman & McCann, 2005).

Most dedicated online crowdsourcing platforms, such as Amazon Mechanical Turk (MTurk), are designed for participant recruitment in return for financial compensation (i.e., pay-per-survey). The cost of online recruitment (e.g., $6 an hour recommended for MTurk studies; Moss, 2019) reflects increasing demands for higher compensation rates (Keith et al., 2017; Peer et al., 2017; Rouse, 2015). This recruitment strategy is thereby financially prohibitive for in-class projects, independent studies, and other research training where supervisors and trainees do not typically have access to considerable funds (e.g., Kierniesky, 2005). The use of social media platforms, such as Facebook or Twitter, is one accessible alternative; however, this recruitment approach poses some ethical and methodological limitations. For example, Sciutto (2015) indicates that Facebook participants are known to the researchers, which may increase risks of socially desirable responding and coercion.

[1] Department of Psychology, McGill University, Montreal, Quebec, Canada
[2] Department of Psychology, University of Toronto Mississauga, Ontario, Canada

Corresponding Author:
Anna M. Lomanowska, Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada L5L 1C6.
Email: anna.lomanowska@utoronto.ca

Another avenue for a more financially accessible online recruitment platform is Reddit, a social news and media aggregation website where users can browse and post content anonymously. Based on a review of preliminary studies which recruited participants through specific communities on Reddit (*subreddits*), Shatz (2017) suggested that Reddit can be a potential source of fast, reliable, and diverse data. In addition, researchers have the option to provide compensation to Reddit participants through flexible compensation methods, such as gift card raffles, which have the benefit of minimizing financial losses to researchers. One subreddit known as r/SampleSize is a dedicated community of over 165,000 registered members who voluntarily complete online surveys ("r/SampleSize," 2012). r/SampleSize regularly moderates the posted surveys based on standardized guidelines, creating an infrastructure that supports equal opportunities for researchers. r/SampleSize therefore appears to show unique promise as a large, broadly accessible, and moderated participant pool that may be particularly attractive for students, research trainees, and course instructors in psychology.

Several recent studies have provided insights into the characteristics of the r/SampleSize participant pool, including greater demographic diversity in terms of age, educational level, and equal gender representation compared to traditionally-used samples (Brickman & Silva, 2017; Jamnik & Lane, 2017; Luong et al., 2019; Record et al., 2018). Brickman and Silva (2017) also reported that r/SampleSize participants completed surveys largely due to internal motivations over external motivations. Furthermore, Jamnik and Lane (2017) demonstrated that scale reliabilities were similar between r/SampleSize and an undergraduate sample. The responses of r/SampleSize participants also successfully replicated previous findings on psychological well-being (Jamnik & Lane, 2017) and the fundamental attribution error (Luong et al., 2019).

Although promising, the above studies are limited in their scope regarding the psychological and psychometric qualities of the r/SampleSize sample, which may discourage its use by psychology instructors, students, and research trainees. The psychological implications of the greater internal motivations of the sample observed by Brickman and Silva (2017) need to be investigated to facilitate the accurate interpretation of studies using r/SampleSize samples when they are compared to other samples that may not share these psychological characteristics. Specifically, altruism may underlie the internally motivated and voluntary participation of r/SampleSize participants (e.g., Burns et al., 2006), which would suggest that they are more altruistic than the externally motivated participants from a pay-per-survey platform like MTurk. Moreover, voluntarily seeking out research studies without compensation, which typically involve complex thought or tasks, appears characteristic of greater need for cognition (Cacioppo et al., 1984). It is important that researchers investigating any of the numerous commonly-studied constructs associated with altruism and need for cognition, such as personality traits (e.g., Furnham et al., 2016; Sadowski & Cogburn, 1997), know about

such differences—or lack thereof—when interpreting findings from r/SampleSize.

Additionally, no published research to date has directly investigated the data quality obtained from r/SampleSize. More importantly, no research to date has compared r/SampleSize to paid platforms like MTurk that are already used as reliable alternatives to undergraduate pools. The purpose of the current study was to examine the demographics of the r/SampleSize participant pool and assess the viability of r/SampleSize as an alternative participant pool to MTurk by addressing two research questions:

> R1. Participant motivations: What are the motivations for participating on r/SampleSize compared to MTurk? In accordance with the rationale presented above, we hypothesized that r/SampleSize participants would be more internally motivated to participate than MTurk workers, more altruistic compared to MTurk workers, and have a higher need for cognition compared to MTurk workers.
> R2. Data quality: How does the quality and reliability of data collected from r/SampleSize compare to MTurk?

Method

Participants

Two hundred and seventy participants responded to the MTurk survey of which 14 were excluded for withdrawal from the study as they did not reach the debriefing page of the survey. Four hundred seventy-eight participants responded to the r/SampleSize survey, of which 194 were excluded for withdrawal and seven for being under 18 years old. Of withdrawn participants, 193 did not reach the debriefing page, while only one withdrew their data at debriefing. Missing data were excluded from the final dataset on a case-wise basis (indicated per analysis if applicable).

Power analyses conducted prior to data collection using G*Power 3.1.9.2 (Faul et al., 2009) indicated that our planned sample sizes of 250 in each group were sufficient to detect, at a minimum, conventionally small effects (ds > 0.34; Cohen, 1988) across all planned participant motivations and data quality analyses at $\beta = .20$ and $\alpha = 0.00278$, excluding the reliability analyses. This minimum is based on the participant motivation analyses, which were the least sensitive among the planned analyses. For the reliability analyses, these sample sizes were sufficient to detect, at a minimum, a difference ratio in Cronbach's α of approximately 1.30 at $\beta = .20$ and $\alpha = .003$ (corrected for 15 potential pairwise comparisons) as per Bonett (2002).

Measures

All measures, analysis scripts, data, and preregistration details for the study are available on the Open Science Framework (OSF; see Transparency and Openness Statement).

Demographics. We asked participants about their gender, age, ethnicity, country of residence, household income (USD), employment status, education, and marital status. To measure political orientation, we used the 12-item Social and Economic Conservatism Scale (SECS; Everett, 2013) to measure two dimensions of social and economic political conservatism. Participants were asked to rate their positivity or negativity toward an issue (e.g., abortion). Scores of 0 indicated greater negativity toward the issue (i.e., less conservatism), whereas scores of 100 indicated greater positivity toward the issue (i.e., greater conservatism).

Participation motivations. We measured internal and external participation motivations by adapting five questions from Buhrmester et al. (2011) for use with r/SampleSize ("Why do you use r/SampleSize?") regarding interest, passing time, having fun, making money, and gaining self-knowledge. We also added an additional item on helping with research ("To help with research"). Participants rated each motivation on 7-point Likert scales (1 = *Strongly Disagree* to 7 = *Strongly Agree*).

Altruism. We measured altruism using the 10 altruism questions from the 300-item International Personality Item Pool (IPIP-NEO; Goldberg, 1999; Goldberg et al., 2006). Participants were asked to rate how accurately each item described them (e.g., "I love to help others") on 5-point Likert scales (1 = *Very Inaccurate* to 5 = *Very Accurate*). We also added 10 randomly selected items from the IPIP-NEO to reduce hypothesis guessing (see OSF materials).

Need for cognition. We measured need for cognition using the 18-item Need for Cognition Scale (NFC; Cacioppo et al., 1984). Participants were asked to rate how well each statement described them (e.g., "I would prefer complex to simple problems") on 5-point Likert scales (1 = *Extremely Uncharacteristic* to 5 = *Extremely Characteristic*).

Data quality
Reliability. Based on Peer et al. (2017), we used the IPIP-NEO, Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965) and NFC as standard scales for reliability analysis due to their ubiquity and demonstrated reliability in standardized samples. Similarly, we conducted reliability analyses on the 16-item version of the Social Desirability Scale-17 (SDS-17; Stöber, 2001) and four-item Perceived Awareness of the Research Hypothesis Scale (PARH; Rubin, 2016) due to their known reliability in previous samples.

Attention checks. We interspersed three attention check questions, one in each of the SECS, IPIP-NEO, and NFC respectively as measures of data quality. Participants were asked to respond how the question asked (e.g., "To validate your continuing participation, please select 70.").

English fluency. We assessed self-reported English fluency using a 5-point Likert item on English comprehension (1 = *Not well at all*, 5 = *Extremely Well*).

Social desirability and demand characteristics. Self-report measures can be affected by social desirability (e.g., Holtgraves, 2004) and demand characteristics (e.g., Sharpe & Whelton, 2016), so we measured social desirability using the SDS-17 and demand characteristics using the PARH.

Participant naivety. After each standardized scale (SECS, IPIP-NEO, RSES, NFC), we asked participants if they had ever answered that questionnaire before. We also asked participants if they were familiar with MTurk or r/SampleSize and if they had ever completed academic studies on that platform before.

Procedure

All procedures were approved by the University of Toronto Mississauga Research Ethics Board. Participants accessed their version of the questionnaire through links posted on the MTurk task list and r/SampleSize subreddit. MTurk data collection was split into three time blocks between July 27–31, 2018 (5 days). r/SampleSize data collection was performed between July 27–August 22, 2018 (27 days). Following r/SampleSize guidelines, we reposted the questionnaire every 24 hours if the post had fallen off the front page. Participants from either participant pool were restricted from duplicate responses through the Qualtrics ballot stuffing feature and MTurk worker ID verification.

Participants completed the survey questions in the following order: motivations, SECS, IPIP-NEO altruism items, RSES, NFC scale, SDS-17, demographics questions, PARH scale, and open-ended questions on their thoughts on the research. Participants were then debriefed and received compensation. MTurk participants were paid $1.50 USD whereas r/SampleSize participants were eligible to enter a raffle for one of five $75 USD gift cards.

Results
Demographics
Descriptive statistics of demographics for both samples are reported in Table 1. r/SampleSize participants reported equal male-female gender representation and diverse ages, education levels, and income ranges. They also reported predominantly Caucasian/White ethnic backgrounds and residence in the United States.

Preregistered Planned Analyses
Participant motivations. One-way ANCOVAs were conducted to compare the motivations of MTurk and r/SampleSize participants, controlling for social desirability and demand characteristics and correcting for multiple comparisons using the Bonferroni correction (adjusted $\alpha = 0.00278$). We used conventional guidelines from Cohen (1988) to guide interpretation of effect sizes. As shown in Figure 1, r/SampleSize participants indicated greater motivation than MTurk participants to complete interesting tasks $F(1, 523) = 35.88$, $p < .001$, $\eta_{\mathrm{p}}^2 = .06$,

Table 1. r/SampleSize and MTurk Sample Demographics.

Factor	r/SampleSize (%, n = 277)	MTurk (%, n = 256)
Gender		
Male	47.65	59.38
Female	47.65	40.23
Other	4.70	0.39
Prefer not to say	0.00	0.00
Age		
18–24	61.01	14.06
25–34	30.32	46.09
35–44	6.50	22.27
45–54	1.08	8.98
55–64	0.72	7.03
65–74	0.00	1.56
75 or above	0.00	0.00
Prefer not to say	0.36	0.00
Ethnic Background		
African/Black	1.81	9.38
Arab	1.08	0.39
Asian	6.86	17.19
Caucasian/White	80.14	66.41
Hispanic/Latino	2.89	4.30
Indigenous/Native	0.36	0.39
Other/Self-Described	6.14	1.95
Prefer not to say	0.72	0.00
Country of Residence (Top 5)		
United States of America	51.67	80.32
United Kingdom	11.90	0.40
Canada	5.20	0.40
Germany	5.20	0.00
India	0.00	13.65
Other	26.02	5.22
Income (USD)		
Less than $25,000	19.86	19.92
$25,000 to $49,999	18.41	31.64
$50,000 to $74,999	17.69	21.09
$75,000 to $99,999	9.39	13.67
$100,000 to $149,999	11.91	8.98
$150,000 or greater	6.14	3.52
Prefer not to say	16.61	1.17
Employment Status		
Employed full-time	28.16	62.11
Employed part-time	7.22	10.94
Self-employed	1.44	16.41
Not employed for pay	6.86	2.34
Home-maker/Caregiver	1.81	1.95
Student	46.93	1.56
Retired	0.36	2.34
Other	5.05	1.56
Prefer not to say	2.17	0.78
Education		
Some high school	2.53	0.39
High school/equivalent	38.99	19.14
Associate's degree	5.42	13.28
Bachelor's degree	27.80	39.45
Some graduate school	7.22	5.47
Master's degree	10.11	17.58
Other professional degree	0.72	3.13
Doctoral degree	2.53	0.78

(continued)

Table 1. (continued)

Factor	r/SampleSize (%, n = 277)	MTurk (%, n = 256)
Other	2.89	0.78
Prefer not to say	1.805	0.00
Marital Status		
Single (never married)	76.90	42.58
Married	20.22	49.61
Widowed	0.00	0.39
Divorced	1.81	7.03
Separated	0.36	0.39
Prefer not to say	0.72	0.00

$d = 0.47$, 95% CI [0.29, 0.64], to pass time $F(1, 523) = 163.75$, $p < .001$, $\eta_p^2 = .24$, $d = 1.13$, 95% CI [0.95, 1.31], have fun ($F(1, 523) = 177.01$, $p < .001$, $\eta_p^2 = .25$, $d = 1.11$, 95% CI [0.94, 1.30]), and help with research ($F(1, 523) = 78.73$, $p < .001$, $\eta_p^2 = .13$, $d = 0.70$, 95% CI [0.53, 0.88]). Conversely, r/SampleSize participants indicated much less motivation for participating to make money than MTurk participants ($F(1, 523) = 1584.67$, $p < .001$, $\eta_p^2 = .75$, $d = -3.51$, 95% CI [−3.78, −3.24]). The difference in motivation for gaining self-knowledge was not significant ($F(1, 523) = 2.81$, $p = .0940$, $\eta_p^2 < .01$, $d = 0.095$, 95% CI [0.075, 0.27]). Social desirability was a significant covariate only of motivation to help with research ($p < .001$, $\eta_p^2 = .02$). Demand characteristics were a significant covariate of motivation to complete interesting tasks ($p < .001$, $\eta_p^2 = .08$), have fun ($p < .001$, $\eta_p^2 = .04$), gain self-knowledge ($p < .001$, $\eta_p^2 = .06$), and help with research ($p < .001$, $\eta_p^2 = .05$). Mann-Whitney U tests, which were conducted to address normality and homoscedasticity violations for the analyses of motivations to pass time and make money, converged with the results from the one-way ANCOVAs.

Altruism. One-way ANCOVAs revealed that the difference between r/SampleSize and MTurk participants in self-reported altruistic personality was not significant, $F(1, 522) = 1.54$, $p = .215$, $\eta_p^2 < .01$, $d = 0.035$, 95% CI [−0.14, 0.21], but with social desirability being a significant covariate ($p < .001$, $\eta_p^2 = .04$; Figure 2A).

Need for cognition. r/SampleSize participants reported slightly higher need for cognition than MTurk participants, $F(1, 522) = 8.56$, $p = .004$, $\eta_p^2 = .02$, $d = 0.23$, 95% CI [0.058, 0.40], with only demand characteristics as a significant covariate ($p = .007$, $\eta_p^2 = .014$; Figure 2B).

Data quality
Reliability analyses. We used the cocron package in R (Diedenhofen & Musch, 2016) to conduct significance tests for Cronbach's alphas between r/SampleSize, MTurk, and existing samples as done previously (e.g., Peer et al., 2017). As shown in Table 2, the differences in Cronbach's alphas across r/SampleSize, MTurk, and existing samples for the IPIP-NEO, NFC, or SDS-17 were not significant, but the PARH
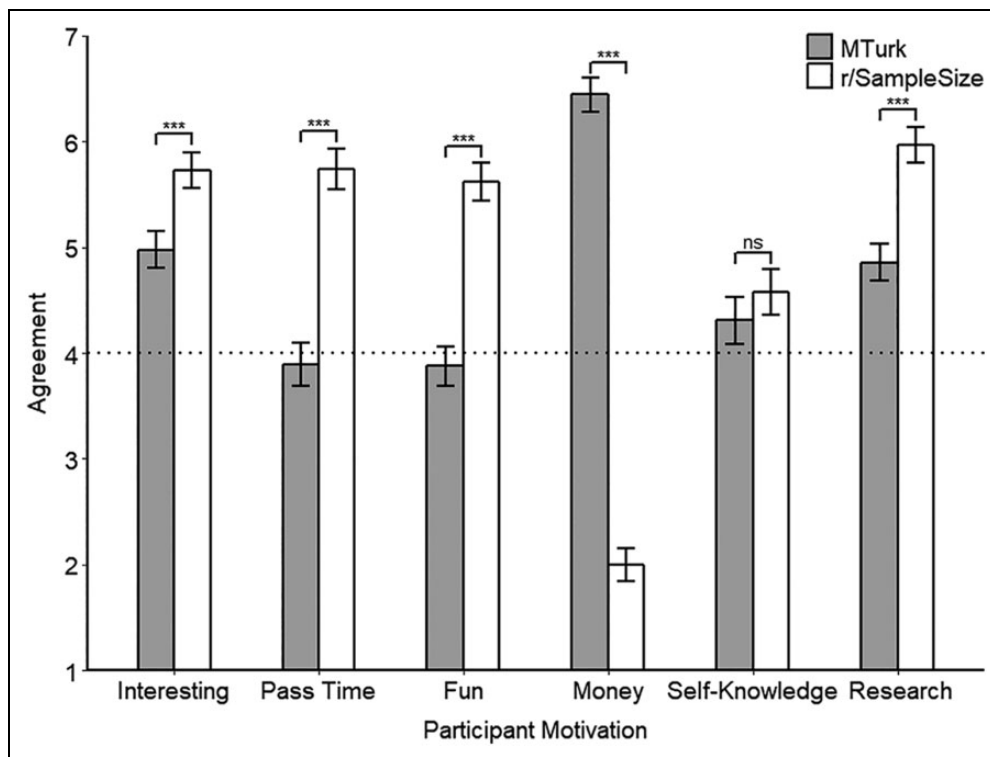
Figure 1. Estimated marginal mean motivations between r/SampleSize and MTurk participants, controlling for social desirability and demand characteristics. Error bars represent 95% CIs. The dotted line represents the midpoint of the agreement scales. ***$p < 0.001$.

and RSES did differ significantly. As shown in Table 3, post hoc analyses for the PARH showed that the differences between r/SampleSize, MTurk, and the adult sample were not significant, but MTurk was significantly lower than the adult sample. For the RSES, differences between r/SampleSize and MTurk or MTurk and the student sample were not significant, but r/SampleSize was significantly greater than the student sample.

Attention checks. Welch's *t*-test indicated that r/SampleSize participants ($M = 2.92$, $SD = 0.28$) correctly answered more attention checks on average than MTurk participants ($M = 2.77$, $SD = 0.60$), $t(352.73) = 3.60$, $p < .001$, $d = 0.32$, 95% CI [0.15, 0.49].

Social desirability and demand characteristics. Welch's *t*-tests indicated that socially desirable responding was slightly greater for r/SampleSize than MTurk participants $t(509.92) = -3.80$, $p < .001$, $d = 0.33$, 95% CI [0.16, 0.50] (Figure 2C), and differences in demand characteristics scores between r/SampleSize and MTurk participants were not significant, $t(511.67) = 0.20$, $p = .845$, $d = 0.02$, 95% CI [−0.15, 0.19] (Figure 2D).

English fluency. A one-way ANCOVA indicated that the differences in self-reported English comprehension scores between r/SampleSize and MTurk participants were not significant, $F(1, 523) = 0.31$, $p = .578$, $\eta_p^2 < .01$, $d = -0.046$, 95% CI [−0.22, 0.12], with no significant contribution of social desirability or demand characteristics.

Non-Preregistered Exploratory Analyses

Duration of questionnaire completion. Examination of histograms and boxplots of questionnaire completion times revealed multiple upper outliers. Such outliers were plausible as the questionnaire was not intended to force completion times. A Mann-Whitney *U* test (without outlier removal) indicated that the median completion time for r/SampleSize participants ($Mdn = 13.02$ min) was nearly five minutes greater than MTurk participants ($Mdn = 8.02$ min), $U = 55488$, $p < .001$, 95% CI [3.98, 5.70].

Social and economic conservatism. Welch's *t*-tests indicated that social conservatism was largely greater for MTurk than r/SampleSize participants, $t(495.45) = 11.89$, $p < .001$, $d = 1.04$, 95% CI [0.86, 1.23], and economic conservatism was moderately greater for MTurk than r/SampleSize participants, $t(509.91) = 7.61$, $p < .001$, $d = 0.67$, 95% CI [0.49, 0.84]. As shown in Table 2, differences in Cronbach's alphas across r/SampleSize, MTurk, and Everett's (2013) original SECS validation sample were not significant.

Participant naivety. Sixteen percent of MTurk participants reported familiarity with r/SampleSize, with 75% reporting no familiarity and 9% unsure. Of the MTurk participants who reported familiarity with r/SampleSize, 67% reported completing academic studies on r/SampleSize, 31% reported not completing studies, and 2% were unsure. Thirty percent of r/SampleSize participants reported familiarity with the MTurk

platform, whereas 66% reported no familiarity and 5% were unsure. Of the r/SampleSize participants who reported familiarity with MTurk, 13% reported completing academic studies on MTurk and 87% reported not completing studies.

There was no evidence that familiarity with the SECS, RSES, or SDS-17 scales differed significantly between MTurk

Figure 2. Estimated marginal means for altruistic personality (A) and estimated sum scores for need for cognition (B) between MTurk and r/SampleSize participants, controlling for social desirability and demand characteristics. Social desirability sum scores (C) and mean demand characteristics scores (D). Error bars represent 95% CIs. The dotted lines represent the midpoints of the scales. $^{**}p < 0.01$, $^{***}p < 0.001$.

and r/SampleSize participants (Table 4). More MTurk participants reported familiarity with the IPIP-NEO altruism subscale than r/SampleSize, $\chi^2(1, N = 466) = 23.26$, $p < .001$, 95% CI [0.13, 0.31]. Conversely, more r/SampleSize participants reported familiarity with the NFC than MTurk, $\chi^2(1, N = 482) = 37.42$, $p < .001$, 95% CI [−0.36, −0.19]. Given multiplicity across the five analyses, results were interpreted at $\alpha = .01$ after Bonferroni correction.

Table 4 shows the percentages of participants who failed one, two, or three attention checks. If participants were to be excluded due to their performance on the attention check, more MTurk participants would be excluded under any criteria than r/SampleSize participants.

Discussion

The current study assessed the viability of r/SampleSize as an online participant pool by comparing characteristics of the data obtained from this sample to MTurk and existing lab samples in terms of participant motivation, data quality, and demographics. The results demonstrate some differences in motivation and socially desirable responding between r/SampleSize and MTurk participants, but overall data quality and reliability is comparable between r/SampleSize, MTurk, and lab samples. The r/SampleSize participant pool also shows relative demographic diversity, with some differences compared to MTurk. These findings indicate that r/SampleSize is a diverse and viable option for participant recruitment and therefore provides an accessible alternative participant pool for psychology research courses and independent research projects.

In the present study, r/SampleSize participants were more motivated to participate by internal factors (e.g., interest, fun) with slightly higher need for cognition and were largely less motivated by making money than MTurk participants. However, both r/SampleSize and MTurk participants were generally internally motivated to participate in research, were altruistic, and had high need for cognition as responses in both groups were greater than the midpoint of the scales. These results are consistent with previous research on r/SampleSize and MTurk participant motivations (Brickman & Silva, 2017; Buhrmester
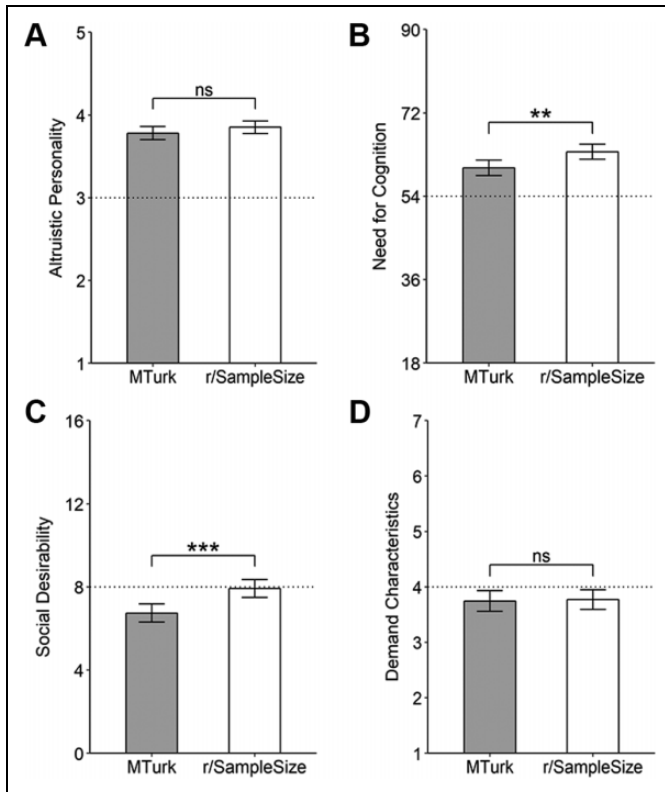
Table 2. Omnibus Tests for Cronbach's α Across r/SampleSize, MTurk, and Community Samples.

Scale	r/SampleSize α (95% CI)	MTurk α (95% CI)	Lab Sample α (95% CI)	Lab Sample Source (n)	$\chi2(2)$	p
IPIP-NEO	.80 [.76, .83]	.83 [.80, .86]	.77 [.74, .80]	Goldberg, 1999 (501)[†]	5.76	.056
NFC	.90 [.88, .92]	.92 [.91, .93]	.90 [.89, .91]	Cacioppo et al., 1984 (527)[†]	4.76	.093
SDS-17	.73 [.68, .77]	.79 [.75, .83]	.80 [.75, .84]	Stöber, 2001 (179)[†]	4.96	.084
PARH	.90 [.88, .92]	.86 [.83, .89]	.93 [.91, .95]	Wang et al., 2016 (185)[†]	15.99	< .001
RSES	.92 [.91, .93]	.91 [.89, .93]	.88 [.86, .90]	Robins et al., 2001 (508)[†]	10.56	.005
SECS—Social	.85 [.82, .88]	.88 [.86, .90]	.87 [.85, .89]	Everett, 2013 (293)[‡]	2.55	.280
SECS—Economic	.67 [.60, .73]	.65 [.58, .71]	.70 [.64, .75]	Everett, 2013 (293)[‡]	1.09	.579

Note. r/SampleSize n = 277 and MTurk n = 256 unless otherwise specified. IPIP-NEO = International Personality Item Pool (MTurk n = 255); NFC = Need for Cognition (MTurk n = 255); SDS-17 = Social Desirability Scale-17 (MTurk n = 255, r/SampleSize n = 272); PARH = Perceived Awareness of the Research Hypothesis Scale; RSES = Rosenberg Self-Esteem Scale (MTurk n = 255, r/SampleSize n = 276); SECS = Social and Economic Conservatism Scale (Social: r/SampleSize n = 267, Economic: r/SampleSize n = 266)
[†] p values were interpreted at $\alpha = .01$ as per the Bonferroni correction for five comparisons (planned).
[‡] Original Validation Sample of the SECS; p values were interpreted at $\alpha = .025$ as per the Bonferroni correction for two comparisons (exploratory).

Table 3. Post Hoc Tests for Cronbach's α Across r/SampleSize, MTurk, and Community Samples.

	r/SampleSize − MTurk			r/SampleSize—Lab			MTurk—Lab		
Scale	χ2(1)	p	Δα	χ2(1)	p	Δα	χ2(1)	p	Δα
PARH	4.48	.034	.04	4.07	.0436	−.03	14.49	< .001	−.07
RSES	0.75	.387	.01	11.44	< .001	.04	5.53	.019	.03

Note. p values were interpreted at α = .00833 as per the Bonferroni correction for six comparisons. PARH = Perceived Awareness of the Research Hypothesis Scale; RSES = Rosenberg Self-Esteem Scale.

et al., 2011). Furthermore, social desirability did not influence most motivations, but demand characteristics did significantly influence the majority of motivations. Social desirability did have a small influence on self-reported altruism as would be expected from the altruism measurement literature (e.g., Erten, 2015). Both participant pools exhibited low socially desirable responding and demand characteristics, indicating acceptable data quality for both r/SampleSize and MTurk. However, socially desirable responding was slightly higher for r/SampleSize participants, highlighting the need to control for social desirability when using r/SampleSize.

Most analyses showed no evidence of reliability differences when comparing r/SampleSize, MTurk, and in-person samples. Importantly, all scales but the SECS-Economic demonstrated acceptable reliability as per the .80 guideline for basic research tools (Nunnally, 1978) across MTurk and r/SampleSize. For the PARH, only MTurk participants demonstrated lower reliability than the lab sample, replicating past research (Rouse, 2015). Furthermore, the RSES scale reliability was higher for r/SampleSize participants than the student sample. These results and the comparable levels of demand characteristics suggest that r/SampleSize data can match and even exceed MTurk data quality. Additionally, MTurk workers completed the questionnaire 5 minutes faster than the r/SampleSize participants and showed poorer performance on attention checks, suggesting that MTurk participants were less attentive.

Exploratory findings on participant naivety suggest that r/SampleSize data quality is generally high. Most of the r/SampleSize participants were unaware of MTurk and vice versa, suggesting that MTurk and r/SampleSize are largely independent participant pools. Furthermore, across the five tested scales, there was no evidence that r/SampleSize participants differed in familiarity with the scales from the MTurk participants, with the exceptions of the IPIP-NEO, with which MTurk participants were more familiar, and the NFC, with which r/SampleSize participants were more familiar. However, across both samples, a considerable proportion of participants demonstrated familiarity with most of the scales.

Consistent with previous studies (Luong et al., 2019; Record et al., 2018), the present sample has equal gender representation as well as large age, educational, and income ranges. However, as in previous studies, participants largely reported being Caucasian/White and residing in the United States. There

Table 4. Scale Familiarity and Attention Check Performance of r/SampleSize and MTurk Participants.

Scale	r/SampleSize (%, n = 277)	MTurk (%, n = 256)
SECS		
Yes	55.60	58.59
No	25.27	31.64
Not sure	19.13	9.77
IPIP-NEO		
Yes	33.94	55.08
No	51.99	33.98
Not sure	14.08	10.94
RSES		
Yes	24.19	30.08
No	66.43	65.23
Not Sure	9.39	4.69
NFC		
Yes	53.43	31.25
No	33.94	62.50
Not Sure	12.64	6.25
SDS-17		
Yes	47.65	41.80
No	36.46	47.66
Not Sure	15.88	10.55
Attention checks		
with 1 failed	8.30	10.94
with 2 failed	0.00	3.13
with 3 failed	0.00	1.95

Note. SECS = Social and Economic Conservatism Scale; IPIP-NEO = International Personality Item Pool; RSES = Rosenberg Self-Esteem Scale; NFC = Need for Cognition; SDS-17 = Social Desirability Scale-17.

were meaningfully higher levels of social and economic conservatism among MTurk workers compared to both r/SampleSize participants and the neutral point of the SECS, but we caution that this finding was exploratory, and the SECS-Economic showed reliability values far below the .80 guideline for basic research tools (Nunnally, 1978). Nevertheless, the combination of demographic characteristics indicates a diversity of participants that are otherwise underrepresented in traditional psychological research with undergraduate student samples (Hanel & Vione, 2016).

Limitations and Future Research

In the present study, compensation was necessary for r/SampleSize participants to ensure fairness across the two recruitment platforms, which limits the generalizability of our findings to "truly voluntary" participants. However, the form of compensation for r/SampleSize was an optional raffle which did not guarantee compensation. Indeed, we found that only approximately half of r/SampleSize participants had opted into the raffle. r/SampleSize participants also reported greatly lower financial motivation and higher internal motivations, suggesting that any sample-selection effects from providing financial compensation would be mitigated by these characteristics. Future studies concerned with the influence of financial motivation should disclose information regarding compensation

after study completion rather than in the study description. Future studies should also collect r/SampleSize demographics to assess the replicability of the demographic makeup because it may change over time, as with all participant panels. Participant withdrawal rates might also vary from study to study. Although withdrawal is naturally expected for volunteers, it may depend on factors not investigated here such as study content, which could have implications for the representativeness of a given r/SampleSize sample. Ongoing validation of these sample characteristics is thus essential, and future research can use the characteristics and effect sizes observed here as a starting point for study planning.

Educational Implications

r/SampleSize provides course instructors and research supervisors with an accessible online participant recruitment platform that can be used to enhance the research training of students in psychology while maintaining acceptable data quality and addressing the limitations frequently observed with traditional university student samples. r/SampleSize allows for recruitment of relatively large and diverse samples in a short period of time. Although not as fast as MTurk data collection, in the present study it took under one month to collect our sample of 277 participants, a time period that would easily fit into a single-term psychology lab course or undergraduate research project. The platform also allows for optional compensation and flexible alternatives, such as gift card raffles, which is especially appealing given the financial constraints faced by many instructors and research supervisors.

Overall, the current study supports r/SampleSize as an alternative participant pool that matches data quality levels of MTurk and lab participants. The findings emphasize the usefulness of controlling for social desirability and demand characteristics when using this participant pool. We hope that psychology instructors, students, and trainee researchers can harness the potential of this online platform.

Declaration of Conflicting Interests

Funding

ORCID iD

Anna M. Lomanowska https://orcid.org/0000-0003-0847-6153

Transparency and Openess Statement

All measures, analysis scripts, and data are available on the Open Science Framework at https://osf.io/n6vhx. Preregistration details are also available at https://osf.io/nfk6m. Preregistration occurred shortly after data collection began due to technical issues. Data were not examined or analyzed prior to preregistration.

Open Practices

This article has received badges for Open Data, Preregistered and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, *27*, 335–340. https://doi.org/10.3102/10769986027004335

Brickman, J., & Silva, D. E. (2017). Free opinions: What a popular survey-taking forum can tell researchers about recruiting participants. https://www.researchgate.net/profile/Aaron_Cargile/post/Im_looking_for_participants_for_my_online_questionnaire-Please_help/attachment/59d6537979197b80779ab5f8/AS%3A517485024813056%401500389452312/download/Brickman+Silva+Poster+copy.pdf

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Burns, D. J., Reid, J. S., Toncar, M., Fawcett, J., & Anderson, C. (2006). Motivations to volunteer: The role of altruism. *International Review on Public and Non Profit Marketing*, *3*(2), 79–91. https://doi.org/10.1007/BF02893621

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Diedenhofen, B., & Musch, J. (2016). Cocron: A web interface and R package for the statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science*, *11*, 51–60. https://www.ijis.net/ijis11_1/ijis11_1_diedenhofen_and_musch.pdf

Erten, İ. H. (2015). Social desirability bias in altruistic motivation for choosing teaching as a career. *H. U. Journal of Education*, *30*, 77–89. http://www.efdergi.hacettepe.edu.tr/shw_artcl-22.html

Everett, J. A. C. (2013). The 12 item Social and Economic Conservatism Scale (SECS). *PLoS One*, *8*, e82131. https://doi.org/10.1371/journal.pone.0082131

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Furnham, A., Treglown, L., Hyde, G., & Trickey, G. (2016). The bright and dark side of altruism: Demographic, personality traits, and disorders associated with altruism. *Journal of Business Ethics*, *134*(3), 359–368. https://doi.org/10.1007/s10551-014-2435-x

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. *7*, pp. 7–28). Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The international personality item pool and the future of public domain personality measures. *Journal of Research in Personality*, *40*, 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Hanel, P. H. P., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS One*, *11*, e0168354. https://doi.org/10.1371/journal.pone.0168354

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, *30*(2), 161–172. https://doi.org/10.1177%2F0146167203259930

Jamnik, M. R., & Lane, D. J. (2017). The use of Reddit as an inexpensive source for high-quality data. *Practical Assessment, Research & Evaluation*, *22*(5), 1–10. https://doi.org/10.7275/swgt-rj52

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology*, *8*, 1359. https://doi.org/10.3389/fpsyg.2017.01359

Kierniesky, N. C. (2005). Undergraduate research in small psychology departments: Two decades later. *Teaching of Psychology*, *32*, 84–90. https://doi.org/10.1207/s15328023top3202_1

Luong, R., Butler, K., & Plaks, J. (2019, February). Educational passages do not reduce the fundamental attribution error [Poster presentation]. *Annual convention of the society for personality and social psychology*, Portland, OR, United States. https://doi.org/10.17605/OSF.IO/G7XUF

Morling, B., & Calin-Jageman, R. J. (2020). What psychology teachers should know about open science and the new statistics. *Teaching of Psychology*, *47*(2), 169–179. https://doi.org/10.1177/0098628320901372

Moss, A. (2019, January 3). *Five things you should be doing in online data collection*. CloudResearch. https://www.cloudresearch.com/resources/blog/five-things-you-should-be-doing-in-online-data-collection/

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Perlman, B., & McCann, L. I. (2005). Undergraduate research experiences in psychology: A national study of courses and curricula. *Teaching of Psychology*, *32*(1), 5–14. https://doi.org/10.1207/s15328023top3201_2

Record, R. A., Silberman, W. R., Santiago, J. E., & Ham, T. (2018). I sought it, I Reddit: Examining health information engagement behaviors among Reddit users. *Journal of Health Communication*, *23*, 470–476. https://doi.org/10.1080/10810730.2018.1465493

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, *27*, 151–161. https://doi.org/10.1177/0146167201272002

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior*, *43*, 304–307. https://doi.org/10.1016/j.chb.2014.11.004

Rubin, M. (2016). The perceived awareness of the research hypothesis scale: Assessing the influence of demand characteristics. *Figshare*. https://doi.org/10.6084/m9.figshare.4315778.v2 "r/SampleSize". (2012). https://www.reddit.com/r/SampleSize/

Sadowski, C. J., & Cogburn, H. E. (1997). Need for cognition in the big-five factor structure. *The Journal of Psychology*, *131*(3), 307–312. https://doi.org/10.1080/00223989709603517

Sciutto, M. J. (2015). Using Facebook to supplement participant pools for class research projects: Should we like it? *Teaching of Psychology*, *42*(2), 157–162. https://doi.org/10.1177/0098628315573140

Sharpe, D., & Whelton, W. J. (2016). Frightened by an old scarecrow: The remarkable resilience of demand characteristics. *Review of General Psychology*, *20*(4), 349–368. https://doi.org/10.1037/gpr0000087

Shatz, I. (2017). Fast, free, and targeted: Reddit as a source for recruiting participants online. *Social Science Computer Review*, *35*, 537–549. https://doi.org/10.1177/0894439316650163

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17). *European Journal of Psychological Assessment*, *17*, 222–232. https://doi.org/10.1027//1015-5759.17.3.222

Vankov, I., Bowers, J., & Munafò, M. R. (2014). On the persistence of low power in psychological science. *The Quarterly Journal of Experimental Psychology*, *67*, 1037–1040. https://doi.org/10.1080/17470218.2014.885986

Wang, X., Chen, Z., Poon, K. T., Teng, F., & Jin, S. (2016). Self-compassion decreases acceptance of own immoral behaviors. *Personality and Individual Differences*, *106*, 329–333. https://doi.org/10.1016/j.paid.2016.10.030