Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*.

Ostroff, C. (1993). The effects of climate and personal influences on individual behavior and attitudes in organizations. *Organizational Behavior and Human Decision Processes, 56*(1), 56–90.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, 13*(2), 238–241.

Thorndike, E. L. (1939). On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them. *American Journal of Psychology, 52*, 122–124.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika, 2*(2), 121–134.

A Convenient Solution: Using MTurk To Sample From Hard-To-Reach Populations

Nicholas A. Smith
The Pennsylvania State University

Isaac E. Sabat
George Mason University

Larry R. Martinez, Kayla Weaver, and Shi Xu
The Pennsylvania State University

We agree with Landers and Behrend's (2015) proposition that Amazon's Mechanical Turk (MTurk) may provide great opportunities for organizational research samples. However, some groups are characteristically difficult to recruit because they are stigmatized or socially disenfranchised (Birman, 2005; Miller, Forte, Wilson, & Greene, 2006; Sullivan & Cain, 2004; see Campbell, Adams, & Patterson, 2008, for a review). These groups may include individuals who have not previously been the focus of much organizational research, such as those of low socioeconomic status; individuals with disabilities; lesbian, gay, bisexual, or trans-

Nicholas Smith, School of Hospitality Management, The Pennsylvania State University; Isaac Sabat, Department of Psychology, George Mason University; Larry Martinez, Kayla Weaver, and Shi Xu, School of Hospitality Management, The Pennsylvania State University.

Correspondence concerning this article should be addressed to Nicholas Smith, School of Hospitality Management, The Pennsylvania State University, 201 Mateer Building, University Park, PA 16802. E-mail: nas277@psu.edu

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

gender (LGBT) individuals; or victims of workplace harassment. As Landers and Behrend (2015) point out, there is an overrepresentation of research using "Western, educated, industrialized, rich, and democratic" participants. It is important to extend research beyond these samples to examine workplace phenomena that are specific to special populations. We contribute to this argument by noting the particular usefulness that MTurk can provide for sampling from hard-to-reach populations, which we characterize as groups that are in the numerical minority in terms of nationwide representation. To clarify, we focus our discussion on populations that are traditionally hard to reach in the context of contemporary organizational research within the United States.

We organize our response to the focal article by first discussing the differences between using MTurk and other traditional data-collection methods that have been used to recruit individuals from hard-to-reach populations. Second, we discuss the potential advantages and disadvantages associated with using MTurk to access such groups. Third, we review past research that has successfully utilized MTurk to investigate hard-to-reach populations. Fourth and finally, we provide specific recommendations to researchers who may wish to use MTurk to conduct research on hard-to-reach populations. We hope that this response will encourage future thoughts and discussions on the use of MTurk as a viable sampling tool and persuade researchers, editors, and reviewers to be more accepting of this data-collection method.

Comparing MTurk With Other Data-Collection Techniques for Sampling From Hard-To-Reach Populations

With the advent of the Internet, the ways in which researchers recruit participants and administer surveys and experiments have changed dramatically. One particularly new method for data collection is Amazon's MTurk, in which individuals can opt-in to research studies in return for nominal compensation. One benefit of utilizing MTurk is that it maintains participants' anonymity and confidentiality. As such, MTurk has a variety of strengths that lend themselves particularly well to recruiting participants who belong to hard-to-reach groups, especially in comparison with other forms of data collection traditionally used to sample from these populations.

As outlined by Landers and Behrend (2015), a great deal of organizational research relies on samples collected from student populations or from single organizations. Student samples may be useful because they are easy to reach and require relatively little compensation. Despite these benefits, student samples are not likely to comprise many employees, particularly individuals from special populations (e.g., LGBT, immigrant populations). Also, student samples tend to be more homogenous than are random national samples in that the participants tend to be mostly White and between the

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

ages of 18 and 22 (e.g., Eisenberg & Wechsler, 2003). Organizational samples ensure that data are more representative of employee experiences than student-based responses are, but these samples are also limited in the extent to which they comprise special populations.

Other researchers have reached special populations of employees through recruitment via community-based organizations (Campbell et al., 2008), via snowball sampling techniques (Law, Martinez, Ruggs, Hebl, & Akers, 2011), or by randomly approaching participants in a densely traveled public area (McCormack, 2014). Using MTurk may provide solutions to potential threats to validity and generalizability associated with each of these techniques. With respect to using community-based organizations and snowball sampling, the data are traditionally obtained through personal contacts and connections, which presents limitations as outlined in the focal article. Furthermore, the researchers often do not control who is invited to participate in the research when these strategies are utilized. Thus, the veracity of the identities of the participants cannot be confirmed, and the participants may all share a common characteristic or be located in a common geographical region. Furthermore, these strategies require participants to publicly self-identify as the identity of interest, which may result in sampling bias (e.g., only those who identify strongly will be contacted). With respect to randomly approaching participants, these participants will necessarily be geographically constrained, and they may or may not be employees.

A particular strength of MTurk in collecting data from hard-to-reach populations is that it affords complete anonymity to participants. As a consequence, it is easier to collect data on participants who represent highly stigmatized, potentially concealable identities (e.g., HIV+, psychological disabilities) as these participants would be more likely to reveal their true identities when asked in such an anonymous format. Indeed, previous scholars have noted the benefits to collecting data through online and computer-based survey methodologies due to the fact that participants are more comfortable in sharing private information through these anonymous platforms (Gosling, Vazire, Srivastava, & John, 2004; Levine, Ancill, & Roberts, 1989; Locke & Gilbert, 1995). Along these lines, researchers have also found more diversity in responses related to sexual behaviors and health-related issues when using these anonymous computer-based reporting techniques (Turner et al., 1998).

Past Research That Has Utilized MTurk To Sample From Hard-To-Reach Populations

Although very few studies have used MTurk to sample from hard-to-reach populations, the studies that have been conducted suggest that this is a

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

viable method of accessing these individuals. A study by Tenenbaum, Byrne, and Dahling (2014) recruited a sample of 153 individuals with physical disabilities in the United States by paying each participant $0.80. The authors recorded the type of disability, the age of onset, and the severity of disability of participants and then examined the independent and interactive effects of these factors on the RIASEC self-efficacy inventory (Lenox & Subich, 1994). The authors utilized several screening procedures to ensure the participants had a physical disability. The final sample included a large number of unique disabilities, such as mobility limitations, amputations, inflammatory diseases, and sensory injuries.

Vaughn, Cronan, and Beavers (2015) collected data on MTurk from individuals who identified as LGB by using quota sampling to ensure equal representation among groups and offering participants $0.50 in exchange for participation. In particular, these authors advocate for more research to establish the veracity of using MTurk to sample from LGB populations. Papa, Lancaster, and Kahler (2014) examined three different types of loss: bereavement, job loss, and divorce. Data were collected from 424 MTurk participants who met the following inclusion criteria: loss of a job for at least 6 months, death of a close loved one, or divorce within the last 12 months. Upon completion of the survey, participants were compensated $1.50. The authors report that the demographic results of this study were similar to other MTurk studies.

In Carr's (2014) study of adult cancer survivors ($n = 111$), she utilized a two-part study design with the explicit intention to examine the viability of utilizing MTurk as an effective sampling technique. For the initial survey, participants were paid $0.50, and for the followup, they were paid $0.70. This study found that participants were honest in their responses (as determined using a malingering scale), test–retest reliabilities were high, and approximately 90% of participants fully completed both surveys. This study represents one of the more rigorous examinations of the viability of MTurk to sample from hard-to-reach populations, and this should be replicated in other samples.

Finally, one of the authors recently collected qualitative data from cancer survivors using both MTurk and in-person surveys during a Susan G. Komen cancer fundraiser race. This study examined disclosure of cancer history in interview contexts. These data demonstrate that the MTurk participants reported disclosing at relatively similar rates as their non-MTurk counterparts. Due to the open-ended nature of the data, statistical comparisons cannot be reported; however, substantive differences between the results of these two sampling techniques were inconsequential.

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

Advantages of Using MTurk To Sample From Hard-To-Reach Populations

In general, MTurk has demonstrated several potential strengths in recruiting participants for organizational research, as outlined by the focal article. Again, participants recruited in this way are typically compensated much less than traditional panels are, and they represent employees from a variety of different locations and types of organizations or industries. Studies that utilize data from a single workplace limit the ability for researchers to examine the impact of contextual factors (e.g., diversity climate, perceived organizational support) on the experiences of minority employees; therefore, researchers have collected data from multiple organizations to employ cross-level analyses across organizational units (Gonzalez & DeNisi, 2009; McKay, Avery, & Morris, 2009) to examine such issues. However, collecting samples from multiple organizations, if even possible, can be time-consuming and expensive. Also, a call has been made to incorporate data from smaller organizations, as most organizational research focuses exclusively on very large and profitable organizations (Storey, 1994). Indeed, studying small businesses is important as this sector represents nearly 55 million jobs (48.5% of all jobs; U.S. Census Bureau, 2011). Additionally, with regard to diversity-related outcomes, many companies within this sector are so small that they are not required to abide by Equal Employment Opportunity Commission regulatory policies. Thus, it is important to understand the experiences of employees within these less-regulated organizations. We argue that MTurk may excel as a data-collection method in such circumstances by providing data from a large number and variety of organizations.

Additionally, users can restrict survey settings in MTurk to filter participants by certain characteristics. For example, MTurk can filter participants by IP address in order to restrict access to only U.S. participants. We recommend that researchers utilize this feature when conducting studies where variables of interest are about U.S. minorities (e.g., stereotypes that are culture-specific). Additionally, results of a recent study by Feitosa, Joseph, and Newman (2015) found measurement invariance on a Big Five personality measure across responses in an organizational field sample and an MTurk sample, but only when the sample was restricted to U.S. IP addresses. Thus, researchers need to consider both ecological validity (i.e., are participants representative of the population of interest?) and whether cultural perspectives impact differences in how respondents interpret the measures themselves.

Potential Problems (and Solutions) Associated With Using MTurk To Sample From Hard-To-Reach Populations

Although it is clear that there are several benefits to using MTurk samples for collecting data on diverse, hard-to-reach participants, there are also some

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

potential concerns, specifically with regards to the ability to lie about one's demographic background (Skitka & Sargis, 2005). Indeed, it is true that the increased anonymity afforded to MTurk users and the fact that participants are paid for taking part in the studies increases the chance that they may lie about their identities in order to participate in the study and receive payment. While we do acknowledge that this is a concerning issue regarding the use of MTurk data pools for studying hard-to-reach populations, we propose several potential remedies for this problematic issue, namely, (a) screening participants before stating the eligibility and purpose of the study, (b) filtering participants out of the study or into different studies based on their individual characteristics or identities, and (c) offering modest payments for study participation.

First, a novel way to ensure respondents are not lying about being a member of a minority group (that the authors are currently using with success) is to ask participants to self-identify before they know the purpose of the research. The participants are first asked to respond to a short series of demographic questions, with a question identifying them as part of the group of interest embedded into the beginning of the survey. These questions should be presented before the participants learn about the explicit purpose of the study, so that they are not primed with any particular manner in which to respond. For example, if interested in gathering a sample of workers who identify as LGB, one question may ask the respondent to report their sexual orientation from a list of possible options. Those who are not a member of the group of interest (e.g., self-identify as heterosexual) could be excluded from full participation. This technique would help to thwart participants from successfully fabricating an identity in order to participate and receive payment because they would not know what identity is required to qualify for participation in the study. This idea is promising, yet some researchers may feel hesitant to reject such a large number of participants in this way.

Thus, a second possible strategy is to collect data for two studies simultaneously: one that requires a specific minority group and one that does not. Similar to the technique described above, participants would be presented with a question regarding the specific demographic characteristic of interest (e.g., a question asking participants to indicate their sexual orientation). Then, those participants who do identify with the minority group will be directed to the primary study, whereas those who do not identify with that minority group will be directed to a different study. Of course, this method could only be used when the base rate for the minority sample of interest is not too low. Otherwise, this method would quickly become too costly. One solution to this problem could involve careful monitoring of the participation rates and coordinating multiple secondary studies when enough participants have been filtered through. Also, this technique would require

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

that data-collection on the second study not be adversely affected by not having participants from the specific minority group in question. This strategy, however, may be especially beneficial for collecting dual perspectives from both stigmatized targets and nonstigmatized employees (e.g., collecting data using LGB participants for the primary study and filtering non-LGB participants into a study specific to heterosexual individuals).

Third, researchers should provide monetary incentives that are appealing but not overly enticing to participants, to reduce the likelihood that participants will lie about their characteristics in order to participate. There is typically an abundance of studies for participants to complete, and thus, providing modest to low compensation for participation and telling participants that they must belong to a specific identity group will likely reduce the chances that participants would lie in order receive payment. This strategy is supported by previous research showing that the amount of compensation provided for completion was not related to response quality. That is, paying more does not necessarily mean that responses will be of higher quality, but it may be related to longer data-collection periods (Buhrmester, Kwang, & Gosling, 2011), especially when searching for highly specific groups.

In general, other best practices should be taken into account when using data collected from MTurk, such as incorporating manipulation checks, establishing thresholds for missing data, ensuring that there is enough variability in responses, deleting responses with identical IP addresses, and matching key demographic characteristics (Gosling et al., 2004). These strategies would allay fears brought forth in the focal article on repeated participation and missing data.

Conclusion

Although MTurk is a fairly new data-collection method, the existing research suggests that it is a reliable way to sample from hard-to-reach populations. However, this research is in its infancy, and researchers should continue to address the validity of this sampling method, in terms of both the psychometric qualities of responses and differences in the outcomes measured. We echo the call of the authors in the focal article to empirically consider the impact that the sampling method has on the research question at hand. However, we believe that for collecting data on hard-to-reach populations, MTurk provides a useful and valid tool that scholars may use.

References

Birman, D. (2005). Ethical issues in research with immigrants and refugees. In J. E. Trimble & C. B. Fisher (Eds.), *The handbook of ethical research with ethnocultural populations and communities* (pp. 155–178). Thousand Oaks, CA: Sage.

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

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*, 3–5. doi:10.1177/1745691610393980

Campbell, R., Adams, A. E., & Patterson, D. (2008). Methodological challenges of collecting evaluation data from traumatized clients/consumers: A comparison of three methods. *American Journal of Evaluation, 29*, 369–381. doi:10.1177/1098214008320736

Carr, A. (2014). *An exploration of Mechanical Turk as a feasible recruitment platform for cancer survivors* (Unpublished undergraduate honors theses, Paper 59). Retrieved from http://scholar.colorado.edu/cgi/viewcontent.cgi?article=1058&context=honr_theses

Eisenberg, M., & Wechsler, H. (2003). Substance use behaviors among college students with same-sex and opposite-sex experience: Results from a national study. *Addictive Behaviors, 28*, 899–913. doi:10.1016/S0306-4603(01)00286-6

Feitosa, J., Joseph, D. L., & Newman, D. A. (2015). Crowdsourcing and personality measurement equivalence: A warning about countries whose primary language is not English. *Personality and Individual Differences, 75*, 47–52. doi:10.1016/j.paid.2014.11.017

Gonzalez, J. A., & DeNisi, A. S. (2009). Cross-level effects of demography and diversity climate on organizational attachment and firm effectiveness. *Journal of Organizational Behavior, 30*, 21–40. doi:10.1002/job.498

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93–104. doi:10.1037/0003-066X.59.2.93

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*. Advance online publication. doi:10.1017/iop.2015.14

Law, C. L., Martinez, L. R., Ruggs, E. N., Hebl, M. R., & Akers, E. (2011). Trans-parency in the workplace: How the experiences of transsexual employees can be improved. *Journal of Vocational Behavior, 79*, 710–723. doi:10.1016/j.jvb.2011.03.018

Lenox, R. A., & Subich, L. M. (1994). The relationship between self-efficacy beliefs and inventoried vocational interests. *The Career Development Quarterly, 42*, 302–313. doi:10.1002/j.2161-0045.1994.tb00514.x

Levine, S., Ancill, R. J., & Roberts, A. P. (1989). Assessment of suicide risk by computer-delivered self-rating questionnaire: Preliminary findings. *Acta Psychiatrica Scandinavica, 80*, 216–220. doi:10.1111/j.1600-0447.1989.tb01330.x

Locke, S. D., & Gilbert, B. O. (1995). Method of psychological assessment, self-disclosure, and experiential differences: A study of computer, questionnaire, and interview assessment formats. *Journal of Social Behavior & Personality, 10*, 255–263. Retrieved from http://psycnet.apa.org/psycinfo/1995-31530-001

McCormack, M. (2014). Innovative sampling and participant recruitment in sexuality research. *Journal of Social and Personal Relationships, 31*, 475–481. doi:10.1177/0265407514522889

McKay, P. F., Avery, D. R., & Morris, M. A. (2009). A tale of two climates: Diversity climate from subordinates' and managers' perspectives and their role in store unit sales performance. *Personnel Psychology, 62*, 767–791. doi:10.1111/j.1744-6570.2009.01157.x

Miller, R. L., Forte, D., Wilson, B. D. M., & Greene, G. J. (2006). Protecting sexual minority youth from research risks: Conflicting perspectives. *American Journal of Community Psychology, 37*, 341–348. doi:10.1007/s10464-006-9053-4

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press

Papa, A., Lancaster, N. G., & Kahler, J. (2014). Commonalities in grief responding across bereavement and non-bereavement losses. *Journal of Affective Disorders, 161*, 136–143. doi:10.1016/j.jad.2014.03.018

Skitka, L. J., & Sargis, E. G. (2005). Social psychological research and the Internet: The promise and peril of a new methodological frontier. In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the Internet* (pp. 1–26). Oxford, United Kingdom: Oxford University Press.

Storey, D. J. (1994). *Understanding the small business sector*. New York, NY: Routledge.

Sullivan, C. M., & Cain, D. (2004). Ethical and safety considerations when obtaining information from or about battered women for research purposes. *Journal of Interpersonal Violence, 19*, 603–618. doi:10.1177/0886260504263249

Tenenbaum, R. Z., Byrne, C. J., & Dahling, J. J. (2014). Interactive effects of physical disability severity and age of disability onset on RIASEC self-efficacies. *Journal of Career Assessment*, *22*, 274–289. doi:10.1177/1069072713493981

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science, 280*, 867–873. doi:10.1126/science.280.5365.867

U.S. Census Bureau. (2011). *Statistics of U.S. businesses (SUSB) main* [Data file]. Retrieved from http://www2.census.gov/econ/susb/data/2011/us_state_naicssector_small_emplsize_2011.xls

Vaughn, A. A., Cronan, S. B., & Beavers, A. J. (2015). Resource effects on in-group boundary formation with regard to sexual identity. *Social Psychological and Personality Science, 6*, 292–299. doi:10.1177/1948550614559604

Stop Apologizing for Your Samples, Start Embracing Them

Xiaoyuan (Susan) Zhu, Janet L. Barnes-Farrell, and Dev K. Dalal
University of Connecticut

Landers and Behrend (2015) call for editors and reviewers to resist using heuristics when evaluating samples in research as well as for researchers to cautiously consider choosing the samples appropriate for their research questions. Whereas we fully agree with the former conclusion, we believe the latter can be extended even further to encourage researchers to embrace the strengths of their samples for understanding their research rather than simply defending their samples. We believe that samples are not inherently better or worse but rather better suited for different research objectives. In

Xiaoyuan (Susan) Zhu, Janet L. Barnes-Farrell, and Dev K. Dalal, Department of Psychology, University of Connecticut.

Correspondence concerning this article should be addressed to Xiaoyuan (Susan) Zhu, Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020. E-mail: xiaoyuan.zhu@uconn.edu

https://doi.org/10.1017/iop.2015.29 Published online by Cambridge University Press