

University of La Verne Institutional Review Board Guidance for Data Collection Using Crowdsourcing Tools

Guidance for Data Collection Using Crowdsourcing Tools

What is Crowdsourcing?

Data crowdsourcing is the process of obtaining data from a large number of sources via online platforms such as Amazon's MTurk, Prolific Academic, Ideanote, Wazoku Crowd, Quora, and 99designs. Quatrics and SurveyMonkey have their own online panels and samples as well. These platforms appeal to researchers for their large and often diverse populations of motivated participants.

Advantages of these platforms include cost-effectiveness, speed of data collection, and built in processes for study design, participant recruitment, and integrated compensation systems. Crowdsourcing also allows a researcher access to populations that are historically hard-to-reach such as specific populations that are tailored to your research including specifics such as gender identification, sex, age, or ethnicity.

Disadvantages include data quality issues, the potential for respondents to not be who they claim to be, and the necessity of screening questions to disqualify crowdsourcing cheaters (providing improper or incorrect answers), speeders (rushing through the survey), and online bots (using bot programs to answer surveys instead of using human beings).

Ethical considerations when using crowdsourcing tools

1. Country of origin of participants: Be aware of international participation on these platforms and the privacy and confidentiality regulations that may apply for your cohort (please refer to the international research policy such as the General Data Protection Regulation (GDPR) that affects European Economic Area (EEA) residents <https://laverne.edu/irb/wp-content/uploads/sites/28/2018/10/La-Verne-IRB-International-Research-GDPR.pdf>).
2. Financial compensation for participants: Participants depending on the task requested may spend time resources and provide labor that merit greater than minimal industry standards. Financial compensation amounts may influence participation rates and bias participant composition.
3. Inclusion of vulnerable populations: Be mindful that many crowdsourcing platforms have "workforces" in developing countries and this increases the risk of inadvertently targeting economically disadvantaged persons.
4. The principles of the Belmont Report: Evaluation of the three principles, Respect for persons, Beneficence, and Justice, should initiate and guide any study design involving

human subjects. Design and thoughtful implementation of controls are essential for the use of crowdsourcing tools.

Best Practices when considering crowdsourcing your data

1. Piloting your study: Design is critical. Formative pilot studies can identify issues with design and provide insight on platform issues that may need to be addressed prior to a full study.
2. Making data quality a priority: Create data quality standards to include timeliness of task completion and response patterns.
3. Utilizing attention checks: Create attention checks to avoid speeders and bots from responding to for survey-based studies.
4. Researching the do's and don'ts of crowdsourcing tools: Research best practices for crowdsourcing within your field and professional associations and organizations. The breadth and depth of these tools has increased exponentially over the past decade, therefore, it is recommended to reach out to subject matter experts within your professional communities.
5. Install information technology standards: To ensure that participants are human, be sure to include information technology industry standards such as reCAPTCHA (<https://www.google.com/recaptcha/about/>).

Additional readings

- Afuah, A. (2018). Crowdsourcing: A primer and research framework. In creating and capturing value through crowdsourcing. In C. L. Tucci, A. Afuah, & G. Viscusi (Eds.), *Creating and capturing value through crowdsourcing* (online edition, pp.11–38). Oxford Academic. <https://doi.org/10.1093/oso/9780198816225.003.0002>
- Li, Y., Chang, L., Li, L., Bao, X., & Gu, T. (2021). Key research issues and related technologies in crowdsourcing data collection. *Wireless Communications and Mobile Computing*, 2021, 1–13. <https://doi.org/10.1155/2021/8745897>
- Rea, S. C., Kleeman, H., Zhu, Q., Gilbert, B., & Yue, C. (2020). Crowdsourcing as a tool for research: Methodological, fair, and political considerations. *Bulletin of Science, Technology & Society*, 40(3-4), 40-53. <https://par.nsf.gov/servlets/purl/10286990>
- Schlagwein, D., Cecez-Kecmanovic, D., & Hanckel, B. (2019). Ethical norms and issues in crowdsourcing practices: A Habermasian analysis. *Information Systems Journal*, 29(4), 811-837.

- Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85(1), 140–156.
<https://doi.org/10.1080/03637751.2017.1342043>
- Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, 69(8), 3139–3148.
<https://doi:10.1016/j.jbusres.2015.12.002>