# Warning Crowdsourced Study Participants About Possible Consequences for Inattentive Participation Relates to Informed Consent, Regardless of Effects on Data Quality

**Jon Agley, PhD, MPH, FAAHB**
**Casey Mumaw, MA, CIP**

**Abstract**

Brühlmann and colleagues recently examined the effects of "warning statements" on insufficient effort responding (IER) in crowdsourced research, specifically Amazon's Mechanical Turk (MTurk). Among other findings, they showed that passively reading a warning statement did not significantly reduce careless responding relative to a control condition. This brief essay discusses the context around IER on crowdsourced data collection platforms, finding Brühlmann's contribution valuable. It then points out how crowdsourced studies by U.S. researchers using mechanisms like attention checks to control for IER plausibly should include warning statements regardless of their effect on IER due to federal research regulations around payment transparency.

**Keywords:** Crowdsourced studies, Mechanical Turk, data quality, informed consent

* Corresponding author may be reached at jagley@indiana.edu

## Introduction

A meaningful amount of behavioral and health-based research is now completed online. This trend is unsurprising: approximately 95% of US adults reported using the internet in 2023, up from 52% in 2000 (Pew Research Center, 2024). A common mechanism for conducting online studies is to embed experiments or surveys into crowdsourcing ("micro-work") platforms, such as Amazon's Mechanical Turk (MTurk) (Amazon, 2024) or Prolific (Prolific, 2024). Estimates of growth in crowdsourced research vary by methodology, field, and timeframe, but include "an almost tenfold increase" in "highly cited applied psychology journals" from 2006-10 to 2011-16 (Walter et al., 2018) and an increase of over 2000% from 2012 to 2019 in management research (Aguinis et al., 2021). At present, even comparatively niche research foci have growing bodies of evidence derived through such methods. For example, a recent systematic review of MTurk studies assessing major depressive disorder, generalized anxiety disorder, or alcohol use disorder using specific standardized tools (PHQ-9, GAD-7, and AUDIT) found 27 distinct studies from 2017 to 2021 alone (Kolc et al., 2023).

While online and crowdsourced research approaches have multiple advantages (such as rapid access to large sample sizes), they also present several methodological challenges. One of the most recognized concerns is "careless or insufficient effort responding (IER)" among workers (Brühlmann et al., 2024). IER is one of several known data quality issues that can affect the validity of data obtained through online studies, along with (but not limited to) deliberate misrepresentation or dishonesty (e.g., Hydock, 2018) or the use of "bots" to automate survey completion (e.g., Buchanan & Scofield, 2018). The issue of IER is important to study because poor data quality can affect the degree to which scholarly findings derived from crowdsourced studies are valid. As health behavior research increasingly occurs using online platforms, attention to these issues is warranted. Invalid study conclusions might slow the development of theories and models, affect the perceived reliability of item scales, or have other adverse effects on researchers' progress toward understanding, explaining, or predicting phenomena.

Researchers often use a type of question colloquially known as an "attention check" to screen study participants for inattention (e.g., Hauser & Schwarz, 2016; Saravanos et al., 2021). Such questions may, for example, insert an item into a Likert-type scale stating, "Please select 'Strongly Disagree' for this item," or might ask a basic question with an extremely obvious answer for those who are paying attention, such as a true/false question, "I have previously been eaten by a shark." However, workers on crowdsourcing platforms often have extensive experience completing social and behavioral science studies, and it is possible that they learn to heuristically process such checks over time, especially if similar wording is used across many studies (Hauser & Schwarz, 2016). Thus, researchers have investigated additional ways to reduce the impact of IER on data quality.

In that light, we write this Current Issues submission in response to Brühlmann et al. (2024), who recently published a preregistered, randomized, controlled trial that investigated the effects of two different applications of a warning statement (issued prior to the start of the study) against a control condition on IER. A causal study of this preventive approach to IER is novel and we would like to commend Brühlmann and colleagues for their publication, and for working to identify mechanisms to reduce IER. At the same time, the "warning statement" mechanism, which we unpack subsequently, also touches on ethical requirements for U.S.-based research with human subjects, which adds a layer of complexity to their inclusion and composition.

## Warning Statements and IER

In their study, warning statements were conceptualized broadly as information "mak[ing] participants aware of the consequences of inattentive responding" (Brühlmann et al., 2024). As they point out in their introduction, the content of warning statements used in crowdsourced research can vary considerably (e.g., positive or negative valence, as in Toich et al., 2022) and conclusions around the effectiveness of such statements in averting IER have been mixed. They specifically note, "…one frequently used technique aiming to reduce carelessness is to make participants aware of the consequences of inattentive responding. For instance, researchers may tell participants that they will withhold incentives if participants respond inattentively…" (Brühlmann et al., 2024).

As their randomized trial shows, it is plausible that differential findings around warning statements relate at least partly to the presentation of the statement. They used the language, "IMPORTANT: This survey contains various mechanisms used to catch those who respond carelessly or randomly to the survey questions. If you are caught doing so, you will NOT be given credit for your participation and may be subject to additional penalties" (Brühlmann et al., 2024). The study arm requiring participants to type out the warning statement ("active") had more participants drop out prior to taking the survey and showed reduced IER for some (but not all) metrics relative to the control condition. In contrast, the arm only requiring the statement to be read ("passive") (Brühlmann et al., 2024) did not improve IER relative to control for any carelessness metric.

## The Broader Context of Warning Statements

It would seem, then, that including warning statements for participants to read with the intention of reducing IER may not be a fruitful approach. However, our reason for writing this essay is to prompt dialogue around the situations in which a warning statement may be warranted *regardless* of whether it reduces the prevalence of IER. This issue came to our attention when we noted that Brühlmann et al. (2024) wrote in their Discussion, "This finding is important because researchers often implement warnings about the potential loss of compensation into their surveys in the hope that it will dissuade careless behavior (e.g., in Agley et al., 2022)."

For the referenced study (Agley et al., 2022), we (the authors of this essay) were the first author (Agley) and the IRB study reviewer (Mumaw), respectively. The study did include a warning statement in the study information sheet ("This survey may include checks to screen out bots and individuals who are not eligible to take the survey. If you are screened out in this manner, the survey will end, and you should return the HIT in order to avoid your work being rejected"). However, even though Agley et al. (2022) was a study about data quality on MTurk, it did not examine the impact of the warning statement on IER, nor was it hypothesized that the statement would have an effect (i.e., the warning statement was present regardless of study arm, and was not one of the conditions to which participants were randomized).

Instead, this warning statement was included pursuant to federal guidance for human subject protections in the U.S, which notes, in part, that:

> IRBs and investigators should ensure that the consent process includes a detailed account of the terms of payment, including a description of the conditions under which a subject would receive partial or no payment (e.g., what will happen if he or she withdraws part way through the research or the investigator removes a subject from the study for medical or noncompliance reasons) (U.S. Department of Health and Human Services, n.d.).

This guidance applies only to "non-Exempt" research, and many crowdsourced surveys are "Exempt" from IRB review, typically under 45 CFR 46.104(d)(2) or (d)(3), which apply to educational tests, surveys, interviews, or observation of public behavior, and benign behavioral interventions, respectively (U.S. Department of Health and Human Services, 2020). At the same time, human subject protection programs will often apply these ethical principles to all research, since providing accurate information about a study is consistent with the Belmont principle of "Respect for Persons" (U.S. Department of Health, Education, and Welfare, 1979).

Granted, many researchers work outside the US and may be subject to different policies, laws, and regulations (and indeed, Brühlmann and colleagues were all at University of Basel in Switzerland at the time of publication). However, researchers within the U.S. should likely consider whether their crowdsourced survey procedures create circumstances in which a subject may, through their own actions (such as failing an attention check question or providing a dishonest response), be removed from a study or fail to receive payment for a study. In such cases, a warning statement that details the conditions in which a worker (on MTurk or an alternative crowdsourcing platform) might be denied payment would ideally be communicated to subjects in some way before they begin participating.

In addition, U.S. researchers may also wish to avoid warning statements that use vague language about consequences (e.g., "…subject to additional penalties"). For "non-Exempt" research, any costs to subjects are required to be clearly listed, and even for "Exempt" studies, the idea that subjects could incur ambiguous costs for participating in research might raise concerns during review. So, for instance, if the possible penalties are restricted to (a) being unable to submit the task for payment, or (b) a rejection (and reduction in a worker's rating) if the task is submitted anyway, then this should likely be explicitly stated. Doing so provides fair warning to the worker, and it also protects the researcher if a worker contacts their IRB after they are removed from the study for IER. Based on this, a warning statement should likely include at least four elements when used in U.S.-based studies:

1. A notification that the study includes various questions or other mechanisms to control for eligibility or data quality, or for related reasons.
2. A description of what will happen if a participant fails the checks (e.g., "the survey will end").
3. Instructions for what to do if a participant fails the checks (e.g., "return the HIT," though terminology and language may vary by crowdsourcing platform).
4. A clear indication of what will happen if a participant fails the checks and does not adhere to the instructions (e.g., "your work will be rejected," which on MTurk means that the participant's completion percentage will be reduced, which is an additional adverse consequence beyond not being paid and occurs if they fail the checks but submit the incomplete work for payment anyway).

There are also additional, related considerations around ethical construction of crowdsourced studies, such as minimizing the amount of uncompensated work (e.g., survey questions) completed before a decision to remove a worker and/or to deny payment is reached, but they are beyond the scope of this essay.

Finally, we want to reiterate this is not a critique of Brühlmann et al. (2024), but rather an attempt to foster additional dialogue. Their work added meaningfully to the knowledge base on data quality in online studies, and we concur with their assessment that novel and innovative ways to reduce the likelihood of IER on crowdsourced survey are both needed and have the potential to be useful to multiple fields of inquiry.

## References


Agley, J., Xiao, Y., Nolan, R., & Golzarri-Arroyo, L. (2022). Quality control questions on Amazon's Mechanical Turk (MTurk): A randomized trial of impact on the USAUDIT, PHQ-9, and GAD-7. *Behavior Research Methods, 54*(2), 885-97. https://doi.org/10.3758/s13428-021-01665-8

Aguinis, H., Villamor, I., & Ramani, R.S. (2021). MTurk research: Review and recommendations. *Journal of Management, 47*(4), 823-837. https://doi.org/10.1177/0149206320969787

Amazon. (2024). Amazon Mechanical Turk: Access a global, on-demand, 24x7 workforce. Retrieved from: https://www.mturk.com/

Brühlmann, F., Memeti, Z., Aeschbach, L.F., Perrig, S.A.C., & Opwis, K. (2024). The effectiveness of warning statements in reducing careless responding in crowdsourced online surveys. *Behavior Research Methods.*

https://doi.org/10.3758/s13428-023-02321-z

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods, 50*, 2586-2596. https://doi.org/10.3758/s13428-018-1035-6

Hauser, D.J. & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*, 400-407. https://doi.org/10.3758/s13428-015-0578-z

Hydock, C. (2018). Assessing and overcoming participant dishonesty in online data collection. *Behavior Research Methods, 50*, 1563–1567. https://doi.org/10.3758/s13428-017-0984-5

Kolc, K.L., Tan, Y.X.K., Lo, A.Z.Y., Shvetcov, A., Mitchell, P.B., & Perkes, I.A. (2023). Measuring psychiatric symptoms online: A systematic review of the use of inventories on Amazon Mechanical Turk (mTurk). *Journal of Psychiatric Research, 163*, 118-126. https://doi.org/10.1016/j.jpsychires.2023.05.027

Pew Research Center. (2024). Internet, broadband fact sheet. Retrieved from: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Prolific. (2024). Easily find vetted research participants and AI taskers at scale. Retrieved from: https://www.prolific.com/

Saravanos, A., Zervodakis, S., Zheng, D., Stott, N., Hawryluk, B., & Delfino, D. (2021). The hidden cost of using Amazon Mechanical Turk for research. In: Stephanidis, C., et al. [Editors]. *HCI International 2021 – Late Breaking Papers: Design and User Experience.* Lecture Notes in Computer Science, 13094. https://doi.org/10.1007/978-3-030-90238-4_12

Toich, M.J., Schutt, E., & Fisher, D.M. (2022). Do you get what you pay for? Preventing insufficient effort responding in MTurk and student samples. *Applied Psychology, 71*(2), 640-661. https://doi.org/10.1111/apps.12344

U.S. Department of Health, Education, and Welfare. (1979). The Belmont Report. Retrieved from: https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf

U.S. Department of Health & Human Services, Office for Human Research Protections. (2020). Human subject regulations decision charts: 2018 requirements. Retrieved from: https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts-2018/index.html

U.S. Department of Health & Human Services, Office for Human Research Protections. (n.d.). Informed consent FAQs. Retrieved from: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html

Walter, S.L., Seibert, S.E., Goering, D., & O'Boyle Jr, E.H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology, 34*, 425-452. https://doi.org/10.1007/s10869-018-9552-y