

PUBLISHED RESEARCH ARTICLES

Amazon's MTurk: A Currently Underutilised Resource for Survey Researchers?

Michael Farrell¹ , Breda Sweeney² 

¹ University College Cork, ² National University of Ireland, Galway

Keywords: accounting, methods, surveys, mturk

<https://doi.org/10.52399/001c.22019>

Accounting, Finance & Governance Review

Vol. 27, 2021

A recent innovation in empirical academic research is the use of online labour markets as a source of data. One such market, Amazon's Mechanical Turk ("MTurk") has been used by studies published in high-quality accounting journals to source participants. Given the traction of this data source in high-calibre publications, it is timely to assess its current impact and future potential for accounting research. This paper examines the extent of adoption of MTurk as a data collection tool in leading accounting journals and specifically considers its adoption and suitability for survey research. Findings reveal that the use of MTurk in high-quality accounting publications is gathering momentum, with approximately the same number of articles published/accepted in 2019 as the total number of articles published in the preceding seven years. However, it is also found that nearly all the journal articles reviewed adopted MTurk for experimental research with only a limited presence in survey research. The study contributes to the literature by providing a comprehensive review of the adoption of MTurk in high-quality accounting journals by frequency, research method, and research participant type. Further, it analyses the unique methodological concerns that MTurk poses for survey-based accounting research, thereby providing researchers with guidance on its potential future usefulness and pitfalls to avoid. The paper concludes that difficulties in the availability of, and screening for, specific groups of participants may limit its potential for survey research until online labour market platforms are developed further.

Introduction

Amazon's Mechanical Turk ("MTurk"), an online labour market, has become popular among social scientists as a source of survey and experimental data (Paolacci & Chandler, 2014). A Google Scholar search by Chandler & Shapiro (2016) finds that approximately 15,000 papers containing the phrase "Mechanical Turk" were published between 2006 and 2014, including hundreds of papers published in top-ranked social science journals using data collected from MTurk. More recently, a Google Scholar search of the phrase "Mechanical Turk" with a search range from 2015 to 2019 returned over 32,000 results.¹ MTurk is not unique in its offering; numerous private companies offer researchers pre-screened research participants. However, these tend to come at a relatively higher cost and provide less control for researchers over participant screening procedures (Wessling et al., 2017). Furthermore, MTurk possesses a large, accessible market that is at least as representative as traditional participant pools (Palan & Schitter, 2018; Paolacci & Chandler,

¹ <http://bit.ly/mturk1519>, accessed in June 2020.

2014) . However, some studies have found differences between MTurk participants and traditional participants (e.g. Brink, Lee, et al., 2019; Goodman et al., 2013).

MTurk has been used across different accounting research fields including, for example, financial accounting studies that examine investors' reliance on non-financial information disclosures (Dong, 2017) and the impact on investors' judgements of corporate social responsibility reports (Elliott et al., 2017); management accounting studies investigating the effect of performance reporting frequency on employee performance (Hecht et al., 2019) and motivations for people to report honestly (Murphy et al., 2019); auditing studies focusing on manager responses to internal audit (Brown & Fanning, 2019) and standards of care required by jurors when assessing auditor negligence (Maksymov & Nelson, 2017); and taxation studies addressing decision makers' willingness to evade taxes (Brink & White, 2015) and the effect of consumer-directed tax credits on motivating purchasing behaviour (Stinson et al., 2018).

Given the growing popularity of MTurk, the first objective of this paper is to provide a timely review of the use of MTurk in high-calibre empirical accounting research. MTurk is constantly evolving as a data collection method (Hunt & Scheetz, 2019), thus creating a need for regular reviews and considerations of this research data-source. The second objective of the paper is to examine its adoption and suitability for survey-based research in accounting. In general, survey research provides researchers with the ability to tap into relatively complex, multi-faceted phenomena as they occur in their natural setting, while at the same time maintaining the degree of standardisation that is necessary for quantitative analysis and theory testing (Speklé & Widener, 2018). In addition, survey methods are suitable to map current practices in the field, which can provide insights regarding interesting research topics that have yet to be studied (Speklé & Widener, 2018). While survey research is considered the most heavily criticised method in the management accounting field (Young, 1996), the key issue has been how surveys are deployed rather than criticism of the actual research method itself (Van der Stede et al., 2005). However, survey research will remain a commonly applied research method (Van der Stede, 2014) because even critics of the method recognise the power that collective opinions have on the behaviour and functioning of individuals, organisations, and society (Van der Stede et al., 2005). On this basis, it is useful to examine the opportunities and challenges that MTurk presents for survey researchers and consider the directions, if any, that survey research using MTurk data is likely to take in the future.

The next section of this paper briefly overviews how MTurk works. It is not the primary purpose of this paper to detail technical guidance on how to use MTurk but useful information for survey researchers is provided throughout. The overview of Mturk is followed by an analysis of the use of MTurk as a

data source in leading accounting journals by journal and year of publication, research methods, purpose of the MTurk data (including whether Mturk is used as a main or supplemental data source), and type of research participant.

Findings show that in addition to a noticeable increase in publications using MTurk, experimental research is the dominant method used in these publications, with survey research having only a limited presence in four mixed-methodology papers. Furthermore, it is found that MTurk is often employed as an additional data source for supplemental empirical tests and for out-of-sample testing of research instruments rather than for main sample testing. The paper also assesses the suitability of MTurk for survey research and discusses operational details relating to validity concerns for survey researchers. In general, we find that the validity concerns for MTurk data are like those from more traditional data sources, although there is an increased risk of “survey impostors”, i.e. survey participants pretending to be someone else.

The paper concludes with a discussion on what the future holds for accounting survey research using MTurk. In the long term, with expected improvements in, and expansion of, online labour markets, this method of data collection is likely to become a mainstream tool for survey researchers. However, there are currently limitations around participant screening and the availability of specialist participant pools. Therefore, MTurk is more likely to be used in the short to medium term as a quick, cost-effective tool for out-of-sample testing of surveys (including pre-testing and pilot-testing) where the final data will be collected using more traditional methods. Furthermore, given the current debate in the management literature on the importance of replicability and reproducibility for credibility of research (e.g. Aguinis et al., 2017; Cuervo-Cazurra et al., 2016), it is foreseeable that MTurk will also become popular with survey researchers as an additional data source for supplemental and/or replication testing. While the ability to replicate a study using MTurk data in a relatively short period of time is attractive for researchers to increase the credibility of their research, we caution that it may have unexpected consequences relating to the willingness of researchers to share ideas in early-stage papers at conferences. Ideas could be empirically tested in a short period of time by other researchers using MTurk data, and this could potentially reduce the contribution of the original paper before its publication.

Overview of MTurk

Amazon Mechanical Turk (“MTurk”) is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually (Amazon, n.d.). These processes and jobs are known on MTurk as human intelligence tasks (HITs), which are broadly defined as tasks that are difficult or impossible for computers to perform (Hunt & Scheetz, 2019). Employers (called requesters) recruit employees (called workers) to complete HITs for remuneration (called a reward) (Hunt & Scheetz, 2019). MTurk is open to both companies and individuals to post a diverse variety of tasks for workers to

perform, such as verifying search results for companies like Google, analysing the content of print advertisements, transcribing audio, and taking surveys (Hunt & Scheetz, 2019).

MTurk has a vast range of uses but was never designed *specifically* for academic research. Fortunately, third-party software programs are available that use MTurk to complete HITs but offer greater functionality, particularly to academic researchers. One example of a third party intermediary useful for academic researchers is TurkPrime. TurkPrime is designed as a research tool whose aim is to improve the quality of the crowdsourcing data collection process and optimise MTurk for researchers (Litman et al., 2017). TurkPrime's core features are currently available at no additional cost to academic researchers (although there are other additional features that attract fees). For the remainder of this document, unless specifically mentioned otherwise, the use of the term MTurk refers to the use of MTurk on its own or through the TurkPrime platform. The next section examines the use of MTurk in top-ranked accounting publications.

Growth of MTurk in Accounting Research

To address the first objective of this paper and assess the current popularity of MTurk in high-calibre accounting research, we review the accounting journals ranked as 4*, 4, and 3 in the Chartered Association of Business Schools 2018 Academic Journal Guide ("The ABS Rankings"),² up to the end of 2019 (including "online early") for the presence of MTurk as a data source (please refer to Appendix A for listing of journals). We use the advanced search function in Google Scholar to filter search results by (i) journal title being reviewed, and (ii) any one of the keywords "MTurk", "Turk", or "Turkprime". For each journal, all articles appearing in the initial search results are subjected to an initial screening to assess their suitability. Several articles are excluded on the basis that they reference but do not use MTurk, or they are general methodological papers, or MTurk metastudies (although any relevant findings from these papers are discussed elsewhere in the paper). Following this initial screening, all remaining articles are reviewed to determine (i) the research methodology/methodologies, (ii) the purpose of MTurk in the research (including whether MTurk is used as a main or supplemental data source), and (iii) the MTurk participant characteristics.

[Table 1](#) summarises the articles using MTurk data by leading accounting journal and year of publication. The findings show that the frequency of use of MTurk-sourced data is increasing rapidly since 2012 (the earliest article in the sample). By 2019, MTurk studies that year had risen to 13, with a further 14 articles online early. To place some additional context on these figures, the journal with the most MTurk studies is *The Accounting Review*, with 40% of the journal articles. *The Accounting Review* typically has six issues per year

² <https://charteredabs.org/>

containing 14 original research articles per issue. Therefore, the five published MTurk articles in 2019 account for approximately 6% of all the 2019 published articles in *The Accounting Review*.

The volume of publications reveals only part of the story; it is also important to examine the research methods used and [Table 2](#) summarises these findings. Of the 55 papers, 49 papers use experiments,³ either as the sole research method (43 papers) or as part of mixed-methods research (six papers). Four papers use archival methods (Hsieh et al., 2020; Jiang et al., 2016; Jung et al., 2019; Madsen & McMullin, 2019), and two papers use mixed methods including a combination of archival, interview, and survey methods (Cao et al., 2018) and a combination of archival and survey methods (Blankespoor et al., 2017). In total, survey methods appear in just four papers, all involving mixed research methods (Blankespoor et al., 2017; Cao et al., 2018; Carcello et al., 2018; Kadous et al., 2019).

The relatively high volume of experimental methods' papers is not surprising. Paolacci et al. (2010) identify MTurk early on as an increasingly popular source of experimental data for social scientists because the MTurk population is large, readily accessible, and in relation to the U.S., at least as representative of the U.S. population as more traditional subject pools (e.g., university students). However, the availability of a broad, general population does not necessarily mean that this is the population of interest to accounting researchers. Therefore, we conduct further analysis of the articles using MTurk. [Table 3](#) summarises the type of participants recruited in these studies.

Nearly half of the participant groups fall under the category of “non-specific participant”, where researchers did not require participants to meet any specific technical qualification criteria. This is consistent with the conclusions in other studies: Hunt & Scheetz (2019) believe crowdsourcing platforms are best suited for obtaining average individuals within society; Farrell et al. (2017) conclude that online workers can be suitable proxies in accounting research that investigates the decisions of non-experts; and Buchheit et al. (2019) find that online workers are good research participants when fluid intelligence (defined in their article as general reasoning and problem-solving ability) is needed for reasonably complex experimental tasks in which incoming knowledge is not critical. However, researchers also raise potentially significant issues with the general MTurk population. Buchheit et al. (2018) observe that when compared with a more general population, several studies show that MTurk participants are younger, more computer literate, and more likely to be single, but they are less likely to be homeowners and religiously affiliated. Paolacci & Chandler (2014) also state that workers tend to be younger (about 30 years old), overeducated, underemployed, less religious, and more liberal

³ An experiment involves the manipulation of independent variables and observing their effects on dependent variables (Birnberg et al., 1990). On the other hand, a survey involves a standardised approach in collecting information from sampling units to make inferences about the population (Birnberg et al., 1990). Birnberg et al. (1990) further outline that experiments are largely intended to test causal theories while surveys are more designed to test cross-sectional theories.

than the general population. Brink, Lee, et al. (2019) further add that the MTurk population is more willing to justify unethical behaviour, more trusting in others, places lower importance on hard work, and has lower capitalist values. Goodman et al. (2013) find that MTurk samples differ from traditional samples on several dimensions such as personality measures and attention span. They also find that MTurk participants are more likely to use the internet than traditional participants and are less extraverted and have lower self-esteem. Furthermore, in relation to attention span, they find that MTurk participants perform significantly worse when survey length is long (greater than 16 minutes). However, this contrasts with previous research that finds MTurk participants are equally attentive as other participants when surveys are approximately five minutes in duration (Paolacci et al., 2010). In summary, while all the above issues may not affect the conclusions drawn in a research project, they are important considerations in the research design phase.⁴

Returning to [Table 3](#), non-professional investors are the second most popular type of research participant used in the MTurk studies examined. However, relatively few publications provide detailed insights into the research project definition or screening criteria used to recruit “non-professional investors” on MTurk. Tang & Venkataraman (2018, p. 339) is one of the few exceptions:

“To ensure that participants are reasonable proxies for non-professional investors and possess the knowledge required to complete our experimental task, we use two criteria to screen participants. First, participants must have taken at least two courses in accounting or finance to ensure that they understand the financial context of our study. Second, participants should, at a minimum, understand the difference between quarterly earnings guidance and quarterly earnings reports. To ensure that our participants meet this requirement, we screen them by testing their knowledge on whether quarterly earnings reporting, and earnings guidance, are mandatory or voluntary disclosures. Only participants who correctly answer these questions and meet the accounting/finance course requirements proceed to our experiment.”

Finally, the “other” category in [Table 3](#) includes a diverse set of workers including, for example, experienced chess players (Bentley, 2019), those having business experience with internal auditors (Carcello et al., 2018), and those possessing both crowdfunding and video game experience (Madsen & McMullin, 2019).

⁴ Brink, Lee, et al. (2019) provide further useful guidance to researchers in this regard.

Table 2: Research Methods Used for MTurk Accounting Research Published in ABS 4*, 4, and 3 Ranked Accounting Journals

Research Method(s)	Frequency	Approx. % of total	Refers to the number(s) of the journal articles (JA) in Appendix B
Experimental	43	78%	JA: 1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 38, 42, 43, 44, 45, 47, 48, 49, 50, 51, 53, 54, 55.
Archival	4	7%	JA: 36, 37, 39, 46.
Experimental/Archival	3	5%	JA: 10, 20, 52
Experimental/Survey	2	4%	JA: 19, 40
Experimental/Interviews	1	2%	JA: 41
Archival/Survey	1	2%	JA: 8
Archival/Interview/Survey	1	2%	JA: 18
Total number of Papers	55	100%	

Table 3: Participants Used for MTurk Accounting Research Published in ABS 4*, 4, and 3 Ranked Accounting Journals

General description of participant	Frequency	Approx. % of total	Refers to the number(s) of the journal articles (JA) in Appendix B
Non-specific participant	26	47%	JA: 1, 3, 4, 5, 8, 10, 11, 15, 16, 17, 21, 22, 27, 29, 31, 32, 34, 36, 37, 39, 40, 41, 43, 44, 48, 49.
Non-professional investor	16	29%	JA: 2, 9, 20, 23, 24, 25, 26, 28, 30, 35, 38, 42, 45, 50, 52, 53.
Others	13	24%	JA: 6, 7, 12, 13, 14, 18, 19, 33, 46, 47, 51, 54, 55.
Total number of Papers	55	100%	

Table 1: Journal Articles using MTurk Data Published in ABS 4*, 4, and 3 Ranked Accounting Journals

Year	ABS 4* Ranked Journals			ABS 4 Ranked Journals		ABS 3 Ranked Journals						Totals
	The Accounting Review (TAR)	Journal of Accounting and Economics (JAE)	Journal of Accounting Research (JAR)	Contemporary Accounting Research (CAR)	Review of Accounting Studies (RAS)	Behavioral Research in Accounting (BRIA)	Auditing: A Journal of Practice and Theory	Accounting Horizons	Accounting and Business Research (ABR) ⁺	Journal of the American Taxation Association	Management Accounting Research (MAR) ⁺	
2012	-	-	1 JA 50. [~]	-	-	-	-	-	-	-	-	1
2013	-	-	-	-	-	1 JA 54.	-	-	-	-	-	1
2014	1 JA 9.	-	-	-	-	-	-	-	-	-	-	1
2015	1 JA 51.	-	-	2 JA 32, 42.	-	1 JA 12.	-	-	-	1 JA 13.	-	5
2016	1 JA 11.	-	-	-	1 JA 37.	-	-	-	-	-	-	2
2017	3 JA 3, 31, 47.	1 JA 10.	1 JA 8.	1 JA 27.	-	1 JA 6.	-	-	1 JA 25.	-	-	8
2018	4 JA 2, 18, 52, 53.	1 JA 4.	1 JA 28.	1 JA 1.	-	2 JA 48, 55.	-	1 JA 19.	-	-	-	10
2019 (published)	5 JA 7, 14, 22, 24, 30.	1 JA 39.	-	4 JA 16, 40, 43, 44.	1 JA 20.	1 JA 35.	-	-	-	-	1 JA 17.	13
2019 (early online)	7 JA 5, 21, 23, 34, 41, 45, 46.	1 JA 36.	-	5 JA 15, 29, 33, 38, 49.	-	-	1 JA 26.	-	-	-	-	14
Totals	22	4	3	13	2	6	1	1	1	1	1	55

[~] Refers to the number(s) of the journal articles (JA) in appendix B.

⁺ Journals published outside of North America.

Table 4: Purpose of Mturk Sourced Research Data in Accounting Research Published in ABS 4*, 4, and 3 Ranked Accounting Journals

Use of MTurk data	Frequency	Approx. % of total	Refers to the number(s) of the journal articles (JA) in Appendix B
Primary data source for empirical tests.	37	67%	JA: 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 26, 30, 31, 33, 35, 38, 42, 43, 45, 47, 48, 49, 50, 53, 54, 55.
Secondary data source (out-of-sample testing).	10	18%	JA: 4, 22, 29, 27, 36, 37, 39, 40, 44, 46.
Secondary data source (supplemental empirical tests).	8	15%	JA: 12, 17, 28, 32, 34, 41, 51, 52.
Total number of Papers	55	100%	

A final analysis of the 55 papers examines whether researchers use MTurk as a main or supplemental source of data and [Table 4](#) summarises these findings. In 37 papers (over 65%), MTurk is used as a main data source. In ten studies (over 18%), MTurk is used as a second data source for out-of-sample testing of research instruments. In the eight remaining studies (over 14%), MTurk data is used as a second data source for supplemental empirical tests.

Finally, only two journals outside North America in the sample (Management Accounting Research and Accounting and Business Research) published MTurk papers in the 2012-2019 period and have only published one paper each. This suggests that MTurk may be a more acceptable data collection tool for North American journals. Alternatively, it may reflect that experimental research, the methodology used in the majority of MTurk studies reviewed here, is generally more established in North American journals. It may also reflect that the majority of MTurk workers are based in the United States. Analysing the demographics over the 2019 calendar year,⁵ US workers accounted for between 68% and 76% of the Mturk worker population. India is second with between 16% and 19% of the worker population.

In summary, the use of MTurk in high-quality accounting publications is increasing, particularly for experimental research, with a relatively low presence in survey research. Obviously, the low presence of MTurk data in survey studies published in high-quality accounting journals does not infer that MTurk is unsuitable for survey research. However, the finding raises questions whether issues exist with MTurk that may hinder its adoption by survey researchers. The next section addresses this question by discussing the usefulness of MTurk for empirical survey research. Specifically, the main operational details of the platform are considered with regards to the key validity concerns of survey researchers, and potential roadblocks are identified for survey researchers using MTurk.

⁵ Up-to-date pool characteristics are available at <http://www.mturk-tracker.com>. It should be noted that the worker location demographics are most likely affected by workers using Virtual Private Servers (VPS) to hide their true location (this is discussed further in the paper).

Assessing MTurk as a Data Source for Survey Research

“Mail surveys are seductive in their apparent simplicity—type up some questions, reproduce them, address them to respondents, wait for returns to come in, and then analyze the answers” (Mangione, 1995, p. 2-3, as cited by Van der Stede et al., 2005). Data collection through an online labour market survey is similarly attractive in terms of speed and ease of access to survey participants: Set up a HIT containing questions, make the HIT available to workers, collect responses and pay workers for the HIT, analyse results. In practice, research data collection will be more complicated. Most researchers will add additional screening procedures, each increasing the complexity of the overall process. Farrell et al. (2017) emphasise the need to reduce the risk of “impostors”, i.e. workers pretending to be who they are not, and “scoundrels”, i.e. workers averting effort and providing false information. Smith et al. (2016) further summarise that issues relating to sample integrity and data quality are the two main concerns of using online panels (groups of research participants). Furthermore, the authors identify that threats to data quality are created by two distinct but potentially overlapping response styles: “Speeders”, where a respondent does not thoroughly read the questions and uses minimal cognitive effort to provide answers that satisfy the question, and “Cheaters”, where a respondent intentionally answers survey questions dishonestly and in a fashion that maximises their opportunity for participation and subsequent rewards.

To respond to these validity threats, and re-narrowing the focus exclusively to survey research, Farrell et al. (2017) recommend that detailed screening of survey participants be carried out *ex-ante* (before issuing the survey), *in-survey* (while the survey is in progress), and *ex-post* (when the data collection is complete). We address each procedure in the following sections. We discuss issues more unique to MTurk in greater detail than issues that are common across all types of surveys. Also, Appendix C summarises the main steps required to use MTurk for survey research.⁶

Ex-ante Screening

Explicit steps must be taken to ensure that participants have the relevant knowledge or experience to participate in a study (Hunt & Scheetz, 2019). In general, the requirement for extensive screening procedures arises because researchers must rely on workers self-selecting into HITs based on workers' own assessments of whether they meet the HIT criteria or not. This could be especially problematic for surveys where ineligible workers might be enticed by the payments on offer to complete survey HITs. While third party providers offer *ex-ante* screening procedures at an additional cost (e.g. Qualtrics Panel, SurveyMonkey Audience, TurkPrime Panel), they do not provide researchers with detailed insight into their screening procedures, which could increase validity concerns (Wessling et al., 2017). Wessling et al. (2017) maintain that

⁶ In addition to numerous internet sources, there are peer-reviewed publications, which although fragmented, provide useful technical guidance on how to use MTurk (e.g. Buchheit et al., 2018; Hunt & Scheetz, 2019).

while these commercial companies claim confidence in their pre-screening, they offer little external verification. Wessling et al. encourage researchers who use such services to monitor and validate the quality of the screening.⁷ Typically, ex-ante screening involves the inclusion of screening questions either at the beginning of the survey or in a separate survey. For example, Hunt & Scheetz (2019) include eight unpaid screening questions at the beginning of their survey instrument and terminate participation for workers not answering in the specified manner. In their experience, they find that Institutional Review Boards (IRBs)⁸ will allow research designs using unpaid screening questions if workers are informed in the instructions to the HIT that payment depends on successfully answering the screening questions, and that they can return the HIT with no negative impact on their MTurk rating if they do not qualify. Hunt & Scheetz also state that potential worker aversion to unpaid screens has never materially impacted upon either author's ability to obtain responses.⁹

This two-survey approach asks workers to identify their characteristics when there is no motive to deceive, and then limits the second survey to those workers who have passed the initial screening (Wessling et al., 2017). Buchheit et al. (2018) suggest that if researchers want a particular kind of expertise, then they can ask pointed questions that only experts would be able to answer. In this manner, the risk of falsely claimed expertise is mitigated. Buchheit et al. also suggest that the screening questions should have a wide number of specific response options where only some (or one) meet the participation requirements. This would reduce demand effects by making the "right" choice less transparent and less subject to guessing (Buchheit et al., 2018). In the second stage, researchers could then use an invitation-only HIT (e.g., through TurkPrime) to target those participants whose answers in the first stage meet the screening criteria (Buchheit et al., 2018). As well as creating a longer "break" between screening questions and the primary task, this approach lowers the number of questions required in the second stage, thus reducing the time needed to complete the primary instrument and lowering associated risks of subject distraction or fatigue (Buchheit et al., 2018). However, Wessling et al. (2017) suggest that screening questions from the survey should be re-asked in the second survey. Their rationale is that it is important to control for possible alternative explanations for inconsistent responses between the two stages, such as take/retake reliability error and change in status or character between the two surveys.

Palan & Schitter (2018) highlight an additional screening risk whereby a population of professional survey-takers may be evolving on crowdsourcing platforms like MTurk. This could lead to loss of participant naivety (Palan

⁷ For further information, see Brandon et al. (2014) for a relatively recent comparison of some of the main commercial participant recruitment services, including MTurk.

⁸ Institutional Review Board (IRB) is a general term which refers to the institutional committee responsible for carrying out independent reviews of institution research studies using human participants. In certain countries such as Ireland, this board may be more commonly referred to as a Research Ethics Committee.

⁹ Wessling et al. (2017) also recommend not paying workers for the screening section of the survey.

& Schitter, 2018). The effect of online subjects participating in potentially hundreds of studies has yet to be quantified, but it has the potential to bias results which suffer from practice effects (Chandler et al., 2014). Chandler et al. (2014) recommend that if researchers are concerned about participant naivety, they should, at a minimum, make an effort to uncover if participants have participated in similar studies previously. Specifically related to survey studies, this would involve additional pre-screening questions. Wessling et al. (2017) provide a longer-term recommendation that involves researchers developing their own ongoing MTurk participant panels where researchers, over time, collect information that could be used to classify and build knowledge about respondents.

In-survey Validity Checks

Common MTurk in-survey checks include reverse-coded questions, instructional manipulation checks, and average completion time. None of these checks are unique to MTurk. Reverse-coded questions are common in all types of surveys as a method of detecting acquiescence bias.¹⁰ Instructional Manipulation Checks (IMCs) are also quite common in surveys.¹¹ However, Hunt & Scheetz (2019) raise another participant naivety issue whereby workers seem to have become aware of these types of checks and now have higher pass rates than traditional study participants. This means researchers should take additional care in interpreting the results of IMCs and make efforts to avoid using more typical forms of IMCs in their survey. However, Peer et al. (2014) conclude that attention-check questions are unnecessary if high-reputation workers are used.

Finally, in relation to average completion time, Elliott et al. (2018) and Brasel et al. (2016) excluded respondents who completed the required task in under a certain amount of time. Some of the dedicated online survey platforms can capture time spent on each screen of the survey or even prohibit participants from progressing until a certain amount of time has passed (Hunt & Scheetz, 2019). Finally, Litman et al. (2017) recommend monitoring the HIT dropout rate, and bounce rate,¹² as they can be important indicators that something may be wrong with the survey instrument.

¹⁰ The use of reverse-coded questions is covered in the general best practice survey design guidance of Dillman et al. (2014), the more general discussion of common method bias issues in Podsakoff et al. (2012), and the more accounting specific discussions of Speklé & Widener (2018).

¹¹ IMCs consist of embedded questions that are similar to the other questions in length and response format but unlike the other questions, an IMC asks participants to ignore the standard response format and instead provide a confirmation that they have read the instruction (Oppenheimer et al., 2009). One example is the blue-dot task which Oppenheimer et al. (2009, p. 871) describe: "*The blue dot IMC takes the form of a Likert Scale. The IMC reads, "please click on the little blue circle at the bottom of the screen. Do not click on the scale items that are labeled from 1 to 9."*

¹² The bounce rate is the percentage of total workers who preview a HIT but do not accept it.

Ex-post Validation Considerations

In general, ex-post data examination for MTurk surveys and traditional surveys is similar and according to Hair et al. (2017), can be considered as four separate assessments: Missing data assessments, suspicious response patterns, outliers, and data distributions.¹³

However, one ongoing issue relating to online panels is the use of Internet Protocol (IP) addresses as a means of identifying participants (e.g. to check that their location corresponds to the research participant requirements). Dennis et al. (2020) discuss four issues with using IP addresses as a proxy for a person's identity:

1. The dynamic assignment of IP addresses by Internet Server Providers (ISPs) often allows individuals to obtain new IP addresses on demand.
2. IP addresses identify machines, not individuals; therefore, an individual can use multiple unique machines to obtain multiple unique IP addresses at the same time.
3. Individuals can also use virtual machines on stand-alone servers (e.g., Virtual Private Servers (VPSs) or Virtual Private Networks (VPNs)) to conceal the IP address of the machine they are working on.
4. There is no official database that links IP addresses to specific locations.

In summary, the above issues can result in a single worker completing the same HIT multiple times and/or completing a HIT when they are unqualified e.g., inappropriately using a VPS in the US to make it look like they are a US worker. To address these issues, Dennis et al. (2020) recommend that researchers supplement cutting-edge IP screening procedures with an ex-post analysis of open-ended question style attention checks.¹⁴ This recommendation is based on Dennis et al.'s own empirical analyses where they found that an analysis of open-ended questions was highly effective in uncovering invalid responses.

Finally, if a worker passes initial pre-screening tests, completes the HIT, but fails in-survey or post-survey screening, there will still be an issue over whether they should be paid.¹⁵ While online participants have several motives for participating in studies, incentives is the most cited (followed by curiosity, enjoyment, and participants wanting to have their views heard) (Smith et al.,

¹³ See Hair et al. (2017) and Sarstedt & Mooi (2019) for a more detailed discussion of these issues.

¹⁴ More advanced screening procedures are being made available all the time. Dennis et al. (2020) encourage researchers to stay up to date with the latest data integrity tools as they become available and they make particular reference to recent upgrades in Turkprime.

¹⁵ Determining standard pay rates is a subjective process. Hunt & Scheetz (2019) provide a useful overview of MTurk operational issues, including pay rates and approval rating systems. Based on their review of online blogs and forums, they suggest that workers prefer tasks that pay \$6 per hour. This aligns with Buchheit et al. (2018) who suggest \$2.00 or less for twenty minute accounting tasks while Farrell et al. (2017) state that the consensus estimate of the mean effective hourly wage on MTurk is under \$5.

2016). Buchheit et al. (2018) find no consensus in the literature regarding the compensation of participants who fail screening tests; they observe researchers who provide full payment, partial payment, or no payment to such participants. However, Brasel et al. (2016) rejected payment for participants who completed the study but did not correctly answer at least 90 percent of the comprehension checks included throughout the research instrument. Furthermore, Brink, Eaton, et al. (2019) found that informing participants upfront in the HIT description about the monitoring of responses and application of penalties increased the level of honest reporting in their study.

Concluding Thoughts

The use of MTurk in leading accounting journals is gathering pace, with nearly the same number of articles published/accepted in 2019 in ABS 4*/4/3 journals as the total articles published in the preceding seven years. Experimental research is used in all but six of the 55 articles reviewed and all but two articles are published in North American journals. Given the lack of research using MTurk in journals based outside North America, there is a need for further research to examine its global acceptability as a data collection tool among academic researchers.

Van der Stede et al. (2005) observe that the quality of survey data is as weak as the weakest link in the survey data collection process. Our paper has documented guidance on participant selection and screening issues to mitigate the potential for MTurk to be the weakest link in a study. MTurk's utility depends on using best practices and carefully considering the issues raised by MTurk's many evaluators (Buhrmester et al., 2018). Regarding the potential of MTurk data for survey research, one factor that may limit its usefulness in the short term is that MTurk has been most frequently used to date to recruit non-experts. This raises a concern that the ease of accessing certain research participants may drive the type of research questions addressed. For example, in their overview of experimental audit research, Simnett & Trotman (2018) foresee that as audit practitioners become more difficult to access for experiments, audit researchers will move to topics that can use more easily accessible surrogates for auditing (including online participants). The authors see this research as generally being less informative and perceive that it will negatively affect the type of audit research conducted in the future. Also, whether sufficient numbers of more niche "expert" participants are available on the platform is not clear and even if they are, the costs of screening for them may be prohibitive. However, if the use of online labour markets continues to grow, so too will the number and variety of competitor platforms to MTurk. Like any software-adoption decision involving competing products, it may become the norm that researchers carry out their own assessment of the merits of various platforms (against each other and against more traditional sources) to determine the data source that best meets their needs and the resources they have available. Therefore, in the long term, it is still expected that MTurk and other similar platforms are likely to become more mainstream data sources for researchers.

In the short to medium term, it is anticipated that MTurk, as currently available, is more likely to be used as a quick, cost-effective tool for out-of-sample testing of surveys (including pre-testing and pilot-testing) where the final data will be collected using more traditional methods. Given that data collection for an entire study is possible in a matter of hours (Goodman et al., 2013), MTurk might also be suitable as a tool for undergraduate or Masters' dissertations where project durations are shorter, research objectives are narrower, and contributions more limited. It is also likely that MTurk will become popular as an additional data source for supplemental and/or replication tests. There is a growing debate in the management literature on the importance of reproducibility and replicability for credibility of research (Aguinis et al., 2017; Cuervo-Cazurra et al., 2016). A recent special issue in *Strategic Management Journal* devoted to replications points to the growing acceptability of replication studies in high ranked journals (Ethiraj et al., 2016). While the ability to replicate a study in a relatively short period of time using MTurk data is a welcome development for many researchers to increase the credibility of research findings, the potential for another researcher in the area to quickly build upon an early-stage paper increases encroachment risk. This may have implications for researchers' willingness to share ideas in an early-stage paper at conferences given the possibility that ideas could be replicated or built upon, and data collection completed in a period of weeks, thereby potentially reducing the contribution of the original study before its publication.

Overall, MTurk has much potential for empirical survey accounting research. However, researchers need to proceed with caution and demonstrate rigour in considering additional threats to validity that can arise from selection and screening of participants. This paper has provided an overview of key validity concerns which will be useful to survey researchers in this regard. Undoubtedly, online labour platforms will continue to grow in use by empirical accounting researchers.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Aguinis, H., Cascio, W. F., & Ramani, R. S. (2017). Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies*, 48(6), 653–663. <https://doi.org/10.1057/s41267-017-0081-0>
- Amazon. (n.d.). *Amazon MTurk*. <https://www.mturk.com/>
- Asay, H. S. (2018). Horizon-Induced Optimism as a Gateway to Earnings Management. *Contemporary Accounting Research*, 35(1), 7–30. <https://doi.org/10.1111/1911-3846.12388>
- Asay, H. S., Elliott, W. B., & Rennekamp, K. (2017). *Disclosure Readability and the Sensitivity of Investors' Valuation Judgments to Outside Information*. 92(4), 1–25. <https://doi.org/10.2308/accr-51570>
- Asay, H. S., & Hales, J. (2018). Disclaiming the Future: Investigating the Impact of Cautionary Disclaimers on Investor Judgments Before and After Experiencing Economic Loss. *The Accounting Review*, 93(4), 81–99. <https://doi.org/10.2308/accr-51924>
- Asay, H. S., Libby, R., & Rennekamp, K. (2018). Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, 65(2–3), 380–398. <https://doi.org/10.1016/j.jacceco.2018.02.002>
- Austin, C. R., Bobek, D., & LaMothe, E. G. (2019). The Effect of Temporary Changes and Expectations on Individuals' Decisions: Evidence from a Tax Compliance Setting. *The Accounting Review*.
- Bartlett, G. D., Kremin, J., Saunders, K. K., & Wood, D. A. (2017). Factors influencing recruitment of non-accounting business professionals into internal auditing. *Behavioral Research in Accounting*, 29(1), 119–130. <https://doi.org/10.2308/bria-51643>
- Bentley, J. W. (2019). Decreasing Operational Distortion and Surrogation Through Narrative Reporting. *The Accounting Review*, 94(3), 27–55. <https://doi.org/10.2308/accr-52277>
- Birnberg, J. G., Shields, M. D., & Young, S. M. (1990). The Case for Multiple Methods in Empirical Management Accounting Research (With an Illustration from Budget Setting). *Journal of Management Accounting Research*, 2(Fall), 33–66.
- Blankespoor, E., Hendricks, B. E., & Miller, G. S. (2017). Perceptions and Price: Evidence from CEO Presentations at IPO Roadshows. *Journal of Accounting Research*, 55(2), 275–327. <https://doi.org/10.1111/1475-679X.12164>
- Bonner, S. E., Clor-proell, S. M., & Koonce, L. (2014). *Mental Accounting and Disaggregation Based on the Sign and Relative Magnitude of Income Statement Items*. 89(6), 2087–2114. <https://doi.org/10.2308/accr-50838>
- Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2–3), 329–357. <https://doi.org/10.1016/j.jacceco.2017.03.002>
- Brandon, D. M., Long, J. H., Loraas, T. M., Mueller-Phillips, J., & Vansant, B. (2014). Online Instrument Delivery and Participant Recruitment Services: Emerging Opportunities for Behavioral Accounting Research. *Behavioral Research in Accounting*, 26(1), 1–23. <https://doi.org/10.2308/bria-50651>
- Brasel, K., Doxey, M. M., Grenier, J. H., & Reffett, A. (2016). Risk disclosure preceding negative outcomes: The effects of reporting critical Audit matters on judgments of auditor liability. *The Accounting Review*, 91(5), 1345–1362. <https://doi.org/10.2308/accr-51380>
- Brink, W. D., Eaton, T., Grenier, J. H., & Reffett, A. (2019). Deterring Unethical Behavior in Online Labor Markets. *Journal of Business Ethics*, 156(1), 71–88. <https://doi.org/10.1007/s10551-017-3570-y>

- Brink, W. D., & Lee, L. S. (2015). The effect of tax preparation software on tax compliance: A research note. *Behavioral Research in Accounting*, 27(1), 121–135. <https://doi.org/10.2308/bria-50977>
- Brink, W. D., Lee, L. S., & Pyzoha, J. S. (2019). Values of participants in behavioral accounting research: A comparison of the M-turk population to a nationally representative sample. *Behavioral Research in Accounting*, 31(1), 97–117. <https://doi.org/10.2308/bria-52103>
- Brink, W. D., & White, R. A. (2015). The effects of a shared interest and regret salience on tax evasion. *Journal of the American Taxation Association*, 37(2), 109–135. <https://doi.org/10.2308/atax-51196>
- Brown, T., & Fanning, K. (2019). The joint effects of internal auditors' approach and persuasion tactics on managers' responses to internal audit advice. *The Accounting Review*, 94(4), 173–188. <https://doi.org/10.2308/acct-52295>
- Bucaro, A. C., Jackson, K. E., & Lill, J. B. (2019). The Influence of Corporate Social Responsibility Measures on Investors' Judgments when Integrated in a Financial Report versus Presented in a Separate Report. *Contemporary Accounting Research*. <https://doi.org/10.1111/1911-3846.12542>
- Buchheit, S., Dalton, D. W., Pollard, T. J., & Stinson, S. R. (2019). Crowdsourcing intelligent research participants: A student versus mturk comparison. *Behavioral Research in Accounting*, 31(2), 93–106. <https://doi.org/10.2308/bria-52340>
- Buchheit, S., Doxey, M. M., Pollard, T., & Stinson, S. R. (2018). A technical guide to using amazon's mechanical turk in behavioral accounting research. *Behavioral Research in Accounting*, 30(1), 111–122. <https://doi.org/10.2308/bria-51977>
- Buhrmester, M. D., Talafar, S., & Gosling, S. D. (2018). An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use. *Perspectives on Psychological Science*, 13(2), 149–154. <https://doi.org/10.1177/1745691617706516>
- Cade, N. L., Koonce, L., Mendoza, K. I., Rees, L., & Tokar, M. B. (2019). Assets and Liabilities: When Do They Exist? *Contemporary Accounting Research*, 36(2), 553–587. <https://doi.org/10.1111/1911-3846.12479>
- Cannon, J. N., & Thornock, T. A. (2019). How do managers react to a Peer's situation? The influence of environmental similarity on budgetary reporting. *Management Accounting Research*, 44, 12–25. <https://doi.org/10.1016/j.mar.2018.11.002>
- Cao, S. S., Ma, G., Tucker, J. W., & Wan, C. (2018). Technological Peer Pressure and Product Disclosure. *The Accounting Review*, 93(6), 95–126. <https://doi.org/10.2308/acct-52056>
- Carcello, J. V., Eulerich, M., Masli, A., & Wood, D. A. (2018). The value to management of using the internal audit function as a management training ground. *Accounting Horizons*, 32(2), 121–140. <https://doi.org/10.2308/acch-52046>
- Cardinaels, E., Hollander, S., & White, B. J. (2019). Automatic summarization of earnings releases: attributes and effects on investors' judgments. *Review of Accounting Studies*, 24(3). <https://doi.org/10.1007/s11142-019-9488-0>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46(1), 112–130. <https://doi.org/10.3758/s13428-013-0365-7>
- Chandler, J., & Shapiro, D. (2016). Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annual Review of Clinical Psychology*, 12(1), 53–81. <https://doi.org/10.1146/annurev-clinpsy-021815-093623>
- Chen, C. X., Pesch, H. L., & Wang, L. W. (2019). Selection Benefits of Below-Market Pay in Social-Mission Organizations: Effects on Individual Performance and Team Cooperation. *The Accounting Review*. <https://doi.org/10.2308/acct-50982>

- Church, B. K., Jiang, W., Kuang, X. (Jason), & Vitalis, A. (2019). A Dollar for a Tree or a Tree for a Dollar? The Behavioral Effects of Measurement Basis on Managers' CSR Investment Decision. *The Accounting Review*, 94(5), 117–137. <https://doi.org/10.2308/accr-52332>
- Clor-Proell, S., Guggenmos, R. D., & Rennekamp, K. M. (2019). Mobile Devices and Investment News Apps: The Effects of Information Release, Push Notification, and the Fear of Missing Out. *The Accounting Review*. <https://doi.org/10.2308/accr-52625>
- Cuervo-Cazurra, A., Andersson, U., Brannen, M. Y., Nielsen, B. B., & Reuber, A. R. (2016). From the Editors: Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies*, 47(8), 881–897. <https://doi.org/10.1057/s41267-016-0005-4>
- Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of mturk data: A discussion of virtual private servers and the limitations of ip-based screening procedures. *Behavioral Research in Accounting*, 32(1), 119–134. <https://doi.org/10.2308/bria-18-044>
- Dennis, S. A., Griffin, J. B., & Zehms, K. M. (2019). The Value Relevance of Managers' and Auditors' Disclosures About Material Measurement Uncertainty. *The Accounting*, 94(4), 215–243. <https://doi.org/10.2308/accr-52272>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.).
- Dong, L. (2017). Understanding investors' reliance on disclosures of nonfinancial information and mitigating mechanisms for underreliance. *Accounting and Business Research*, 47(4), 431–454. <https://doi.org/10.1080/00014788.2016.1277969>
- Doxey, M. M., Hatfield, R. C., Rippy, J. A., & Peel, R. K. (2019). *Auditing: A Journal of Practice & Theory*. 39(2), 27–50. <https://doi.org/10.2308/ajpt-18-032>
- Elliott, W. B., Grant, S. M., & Hobson, J. L. (2019). Trader Participation in Disclosure: Implications of Interactions with Management. *Contemporary Accounting Research*. <https://doi.org/10.1111/1911-3846.12524>
- Elliott, W. B., Grant, S. M., & Hodge, F. D. (2018). Negative News and Investor Trust: The Role of \$Firm and #CEO Twitter Use. *Journal of Accounting Research*, 56(5), 1483–1519. <https://doi.org/10.1111/1475-679X.12217>
- Elliott, W. B., Grant, S. M., & Rennekamp, K. M. (2017). How Disclosure Features of Corporate Social Responsibility Reports Interact with Investor Numeracy to Influence Investor Judgments. *Contemporary Accounting Research*, 34(3), 1596–1621. <https://doi.org/10.1111/1911-3846.12302>
- Emett, S. A. (2019). Investor Reaction to Disclosure of Past Performance and Future Plans. *The Accounting Review*, 94(5), 165–188. <https://doi.org/10.2308/accr-52343>
- Ethiraj, S. K., Gambardella, A., & Helfat, C. E. (2016). Replication in Strategic Management. *Strategic Management Journal*, 37, 2191–2192. <https://doi.org/10.1002/smj>
- Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, 92(1), 93–114. <https://doi.org/10.2308/accr-51447>
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26(3), 213–224. <https://doi.org/10.1002/bdm.1753>
- Grenier, J. H., Pomeroy, B., & Stern, M. T. (2015). The effects of accounting standard precision, auditor task expertise, and judgment frameworks on audit firm litigation exposure. *Contemporary Accounting Research*, 32(1), 336–357. <https://doi.org/10.1111/1911-3846.12092>

- Hair, J. F., Hult, T. G., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Sage Publications.
- Hayes, M. J., & Reckers, P. (2019). The Role of Narcissistic Hypocrisy in the Development of Accounting Estimates. *Contemporary Accounting Research*. <https://doi.org/10.1111/1911-3846.12552>
- Hecht, G. W., Hobson, J. L., & Wang, L. W. (2019). The effect of Performance Reporting Frequency on Employee Performance. *The Accounting Review*. <https://doi.org/10.1017/CBO9781107415324.004>
- Holt, T. P. (2019). An examination of nonprofessional investor perceptions of internal and external auditor assurance. *Behavioral Research in Accounting*, 31(1), 65–80. <https://doi.org/10.2308/bria-52276>
- Hsieh, T.-S., Kim, J.-B., Wang, R. R., & Wang, Z. (2020). Seeing is believing? Executives' facial trustworthiness, auditor tenure, and audit fees. *Journal of Accounting and Economics*, 69(1), 101260. <https://doi.org/10.1016/j.jacceco.2019.101260>
- Hunt, N. C., & Scheetz, A. M. (2019). Using MTurk to distribute a survey or experiment: Methodological considerations. *Journal of Information Systems*, 33(1), 43–65. <https://doi.org/10.2308/isys-52021>
- Jiang, D., Kumar, A., & Law, K. K. F. (2016). Political contributions and analyst behavior. *Review of Accounting Studies*, 21(1), 37–88. <https://doi.org/10.1007/s11142-015-9344-9>
- Johnson, J., Theis, J., Vitalis, A., & Young, D. (2019). The Influence of Firms' Emissions Management Strategy Disclosures on Investors' Valuation Judgments. *Contemporary Accounting Research*. <https://doi.org/10.1111/1911-3846.12545>
- Jung, J. H., Kumar, A., Lim, S. S., & Yoo, C. Y. (2019). An analyst by any other surname: Surname favorability and market reaction to analyst forecasts. *Journal of Accounting and Economics*, 67(2–3), 306–335. <https://doi.org/10.1016/j.jacceco.2019.02.002>
- Kadous, K., Proell, C. A., Rich, J., & Zhou, Y. (2019). It Goes without Saying: The Effects of Intrinsic Motivational Orientation, Leadership Emphasis of Intrinsic Goals, and Audit Issue Ambiguity on Speaking Up. *Contemporary Accounting Research*, 36(4), 2113–2141. <https://doi.org/10.1111/1911-3846.12500>
- Kelly, K., Dinovitzer, R., Gunz, H., & Gunz, S. (2019). The Interaction of Perceived Subjectivity and Pay Transparency on Professional Judgement in a Profit Pool Setting: The Case of Large Law Firms. *The Accounting Review*. <https://doi.org/10.1017/CBO9781107415324.004>
- Koonce, L., Miller, J., & Winchel, J. (2015). The Effects of Norms on Investor Reactions to Derivative Use. *Contemporary Accounting Research*, 32(4), 1529–1554. <https://doi.org/10.1111/1911-3846.12118>
- Krische, S. D. (2019). Investment Experience, Financial Literacy, and Investment-Related Judgments. *Contemporary Accounting Research*, 36(3), 1634–1668. <https://doi.org/10.1111/1911-3846.12469>
- Lambert, T. A., & Peytcheva, M. (2019). When Is the Averaging Effect Present in Auditor Judgments? *Contemporary Accounting Research*, 37(00), 1–20. <https://doi.org/10.1111/1911-3846.12512>
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49(2), 433–442. <https://doi.org/10.3758/s13428-016-0727-z>
- Liu, Y., Huang, Z., & Williamson, M. (2019). Are Investors Warned by Disclosure of Conflicts of Interest? The Moderating Effect of Investment Horizon. *The Accounting Review*. <https://doi.org/10.2308/tar-2017-0284>

- Madsen, J. M., & McMullin, J. L. (2019). Economic Consequences of Risk Disclosures: Evidence From Crowdfunding. *The Accounting Review*. <https://doi.org/10.2308/accr-52641>
- Maksymov, E. M., & Nelson, M. W. (2017). Malleable standards of care required by jurors when assessing auditor negligence. *The Accounting Review*, *92*(1), 165–181. <https://doi.org/10.2308/accr-51427>
- Mangione, T. W. (1995). *Mail surveys: improving the quality*. Sage Publications.
- Morrow, M., Stinson, S. R., & Doxey, M. M. (2018). Tax incentives and target demographics: Are tax incentives effective in the health insurance market? *Behavioral Research in Accounting*, *30*(1), 75–98. <https://doi.org/10.2308/bria-51929>
- Murphy, P. R., Wynes, M., Hahn, T., & Devine, P. G. (2019). Why are People Honest? Internal and External Motivations to Report Honestly. *Contemporary Accounting Research*. <https://doi.org/10.1111/1911-3846.12543>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Paolacci, G., & Chandler, J. (2014). Inside the Turk. *Current Directions in Psychological Science*, *23*(3), 184–188. <https://doi.org/10.1177/0963721414531598>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. <https://doi.org/10.3758/s13428-013-0434-y>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Rennekamp, K. (2012). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research*, *50*(5), 1319–1354. <https://doi.org/10.1111/j.1475-679x.2012.00460.x>
- Rennekamp, K., Rupar, K. K., & Seybert, N. (2015). Impaired judgment: The effects of asset impairment reversibility and cognitive dissonance on future investment. *The Accounting Review*, *90*(2), 739–759. <https://doi.org/10.2308/accr-50879>
- Sarstedt, M., & Mooi, E. (2019). *A Concise Guide to Market Research*. Springer.
- Simnett, R., & Trotman, K. T. (2018). Twenty-five-year overview of experimental auditing research: Trends and links to audit quality. *Behavioral Research in Accounting*, *30*(2), 55–76. <https://doi.org/10.2308/BRIA-52138>
- 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.org/10.1016/j.jbusres.2015.12.002>
- Speklé, R. F., & Widener, S. K. (2018). Challenging Issues in Survey Research: Discussion and Suggestions. *Journal of Management Accounting Research*, *30*(2), 3–21. <https://doi.org/10.2308/jmar-51860>
- Stinson, S. R., Barnes, B. G., Buchheit, S., & Morrow, M. (2018). Do consumer-directed tax credits effectively increase demand? Experimental evidence of conditional success. *Journal of the American Taxation Association*, *40*(2), 1–19. <https://doi.org/10.2308/ATAX-51960>

- Tan, H.-T., & Yu, Y. (2018). Management's Responsibility Acceptance, Locus of Breach, and Investors' Reactions to Internal Control Reports. *The Accounting Review*, 93(6), 331–355. <https://doi.org/10.2308/accr-52077>
- Tang, M., & Venkataraman, S. (2018). How patterns of past guidance provision affect investor judgments: The joint effect of guidance frequency and guidance pattern consistency. *The Accounting Review*, 93(3), 327–348. <https://doi.org/10.2308/accr-51905>
- van der Heijden, H. (2013). Charities in competition: Effects of accounting information on donating adjustments. *Behavioral Research in Accounting*, 25(1), 1–13. <https://doi.org/10.2308/bria-50295>
- Van der Stede, W. A. (2014). A manipulationist view of causality in cross-sectional survey research. *Accounting, Organizations and Society*, 39(7), 567–574. <https://doi.org/10.1016/j.aos.2013.12.001>
- Van der Stede, W. A., Young, S. M., & Chen, C. X. (2005). Assessing the quality of evidence in empirical management accounting research: The case of survey studies. *Accounting, Organizations and Society*, 30(7–8), 655–684. <https://doi.org/10.1016/j.aos.2005.01.003>
- Wessling, K. S., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, 44(1), 211–230. <https://doi.org/10.1093/jcr/ucx053>
- Wright, A. M., & Wu, Y. J. (2018). The impact of auditor task difficulty and audit quality indicators on jurors' assessments of audit quality. *Behavioral Research in Accounting*, 30(2), 109–125. <https://doi.org/10.2308/BRIA-52081>
- Young, S. (1996). *Survey research in management accounting: a critical assessment*. In: Richardson, A. (Ed.), *Research Methods in Accounting: Issues and Debates*. CGA-Canada Research Foundation.

Appendices

Appendix A: Listing of Accounting Journals by 4*/4/3 Ranking

4 Ranked Accounting Journals*

The Accounting Review, Journal of Accounting and Economics, Journal of Accounting Research, and Accounting, Organizations, and Society.

4 Ranked Accounting Journals

Contemporary Accounting Research and Review of Accounting Studies.

3 Ranked Accounting Journals

Abacus, Accounting and Business Research, Accounting Forum, Accounting Horizons, Accounting, Auditing & Accountability Journal, Auditing: A Journal of Practice and Theory, Behavioral Research in Accounting, British Accounting Review, British Tax Review, Critical Perspectives on Accounting, European Accounting Review, Financial Accountability & Management, Foundations and Trends in Accounting, International Journal of Accounting, Journal of Accounting and Public Policy, Journal of Accounting Literature, Journal of Accounting, Auditing & Finance, Journal of Business Finance & Accounting, Journal of International Accounting, Auditing and Taxation, Journal of the American Taxation Association, and Management Accounting Research.

Appendix C: Summary of Main Steps Required to Carry Out MTurk Survey Research

1. Consider the availability of the target population on MTurk.

The probability of obtaining a suitable sample falls with the increasing specificity of a target population. Also, the requirement for a niche population will increase the amount of screening procedures required.

2. Engage TurkPrime as an intermediary for handling MTurk data.

The core features of TurkPrime are free and will reduce the overall cost of carrying out an MTurk survey.

3. If possible, use a separate specialist online-survey provider to design/ implement the survey.

Surveys can be carried out directly on MTurk without re-directing participants to a separate online survey but in general, a survey generated directly on MTurk will take longer to create and will have reduced functionality.

4. Determine the required ex-ante screening checks for the target population.

A rigorous screening process will reduce the risk of "impostors" completing the survey.

Appendix B: Alphabetical Listing of Journal Articles (JA) Referenced in Tables 1, 2, and 3

-
- | | |
|---|--------------------------------|
| 1. Asay (2018) | 29. Elliott et al. (2019)* |
| 2. Asay & Hales (2018) | 30. Emmett (2019) |
| 3. Asay et al. (2017) | 31. Farrell et al. (2017) |
| 4. Asay et al. (2018) | 32. Grenier et al. (2015) |
| 5. Austin et al. (2019)* | 33. Hayes & Reckers (2019)* |
| 6. Bartlett et al. (2017) | 34. Hecht et al. (2019)* |
| 7. Bentley (2019) | 35. Holt (2019) |
| 8. Blankespoor et al. (2017) | 36. Hsieh et al. (2020)* |
| 9. Bonner et al. (2014) | 37. Jiang et al. (2016) |
| 10. Bonsall et al. (2017) | 38. Johnson et al. (2019)* |
| 11. Brasel et al. (2016) | 39. Jung et al. (2019) |
| 12. Brink & Lee (2015) | 40. Kadous et al. (2019) |
| 13. Brink & White (2015) | 41. Kelly et al. (2019)* |
| 14. Brown & Fanning (2019) | 42. Koonce et al. (2015) |
| 15. Bucaro et al. (2019)* | 43. Krische (2019) |
| 16. Cade et al. (2019) | 44. Lambert & Peytcheva (2019) |
| 17. Cannon & Thornock (2019) ⁺ | 45. Liu et al. (2019)* |
| 18. Cao et al. (2018) | 46. Madsen & McMullin (2019)* |
| 19. Carcello et al. (2018) | 47. Maksymov & Nelson (2017) |
| 20. Cardinaels et al. (2019) | 48. Morrow et al. (2018) |
| 21. Chen et al. (2019)* | 49. Murphy et al. (2019)* |
| 22. Church et al. (2019) | 50. Rennekamp (2012) |
| 23. Clor-Proell et al. (2019)* | 51. Rennekamp et al. (2015) |
| 24. Dennis et al. (2019) | 52. Tan & Yu (2018) |
| 25. Dong (2017) ⁺ | 53. Tang & Venkataraman (2018) |
| 26. (Doxey et al., 2019) | 54. van der Heijden (2013) |
| 27. Elliott et al. (2017) | 55. Wright & Wu (2018) |
| 28. Elliott et al. (2018) | |
-

* Journal articles available early online.

⁺ Journal articles published outside North America (2 in total).

5. Determine how much to pay each worker.

Determining how much to pay participants is still a subjective process and MTurk fees (charged as a percentage of the rate paid to workers) must also be factored in.

6. Implement in-survey screening checks.

Researchers must consider that having too many in-survey screenings checks (e.g. attention checks) might increase the survey to an unsatisfactory length and cost. Therefore, there will be a certain trade-off here.

7. Implement ex-post screening checks.

Most traditional ex-post screening checks are still valid for MTurk data.