



The Opportunities and Limitations of Using Mechanical Turk (MTurk) in Public Administration and Management Scholarship

Justin M. Stritch, Mogens Jin Pederson & Gabel Taggart

To cite this article: Justin M. Stritch, Mogens Jin Pederson & Gabel Taggart (2017): The Opportunities and Limitations of Using Mechanical Turk (MTurk) in Public Administration and Management Scholarship, International Public Management Journal, DOI: [10.1080/10967494.2016.1276493](https://doi.org/10.1080/10967494.2016.1276493)

To link to this article: <http://dx.doi.org/10.1080/10967494.2016.1276493>



Accepted author version posted online: 19 Jan 2017.



Submit your article to this journal [↗](#)



Article views: 54



View related articles [↗](#)



View Crossmark data [↗](#)

The Opportunities and Limitations of using Mechanical Turk (MTurk) in Public Administration and Management Scholarship

Justin M. Stritch¹, Mogens Jin Pederson², Gabel Taggart¹

¹ARIZONA STATE UNIVERSITY

²AARHUS UNIVERSITY

Abstract

Other social science fields are increasingly conducting research using Amazon's Mechanical Turk (MTurk)—an online crowdsourcing platform—but how might MTurk be useful to public administration and management research? This article provides an introduction of the platform and considers both the opportunities and limitations for using MTurk in public administration and management scholarship. We find that MTurk might be relevant for examining particular types of research questions. We identify five areas where MTurk data may complement and enhance public administration and management research: (1) exploratory analyses and survey construction; (2) measurement refinement of latent constructs; (3) experiments; (4) longitudinal research and data collection; and (5) collection of data from citizens. The article emphasizes how a key requisite conditions both the applicability of MTurk data and the validity of MTurk-based findings: The researcher must understand the boundaries and potential of the platform since the issues related to representativeness, participation, and data quality are non-trivial.

KEYWORDS: Mechanical Turk, public management, experiments, research design

INTRODUCTION

Technological developments have created new platforms for social science research. With the ubiquity of the internet, researchers can obtain data from crowdsourcing platform—i.e., crowds of people participating in an online community or web-based platforms. Participants on these web-based platforms may respond to requesters' questions, problems, and specific tasks. Indeed, the posting of discrete tasks, such as surveys and experiments, to an online community is a common use of crowdsourcing platforms by researchers (Mason and Suri 2012). Other social science fields are increasingly using Amazon's Mechanical Turk (MTurk) crowdsourcing platform for research (Paolacci and Chandler 2014). Examples include political science (Arceneaux 2012; Berinsky, Huber, and Lenz 2012; Berinsky, Margolis, and Sances 2014; Dowling and Wichowsky 2015; Healy and Lenz 2014; Thibodeau, Peebles, Grodner, and Durgin 2013), economics and behavioral science (Chandler and Kapelner 2013; Chandler, Mueller, and Paolacci 2014; Horton, Rand, and Zeckhauser 2011), and psychology (Buhrmester, Kwang, and Gosling 2011; Eriksson and Simpson 2011; Fishbach, Henderson, and Koo 2011; Gómez et al. 2011).

We are witnessing the beginning of a similar trend in public administration and management research. Over the past few years, studies conducted on the MTurk platform have thus emerged in the major public administration and management journals such as *Journal of Public Administration Research & Theory* (Marvel 2016; Jilke, Van Ryzin, and Van De Walle 2016), *Public Administration Review* (Marvel 2014; Marvel and Girth 2016; Smith 2016), *International Public Management Journal* (Marvel 2015), *Public Management Review* (Kaufman and Tummers 2016), and the *American Review of Public*

Administration (Pedersen and Stritch 2016). This development calls for a critical discussion about the utility of the MTurk platform to public administration and management research. What are the opportunities and limitations of using MTurk to advance public administration and management scholarship?

Motivated by this question, this article provides a discussion of the MTurk platform for public administration and management scholars who may not yet be familiar with the platform. Following a brief overview of the platform, we analyze the opportunities and limitations for using MTurk to examine questions important to public administration and management. Moreover, we discuss five particular areas where MTurk, as an alternative to more conventional means for data collection, may enhance aspects of public administration and management scholarship and complement existing research.

While many crowdsourcing platforms exist, we focus on MTurk due to the fact that MTurk is by far the most frequently used platform for social science research (Paolacci and Chandler 2014). Providing a how-to guide is not our article's objective. Rather, we discuss the potential advantages of crowdsourced data obtained using MTurk, areas of concern, and uses to enhance public administration research. Specific implementation issues may differ across crowdsourcing platforms, but our observations will likely apply to present and future crowdsourcing platforms similar to MTurk (e.g., InnoCentive, Threadless, Lánzanos, iStockPhoto, ModCloth, and Fiat Mio (Estellés-Arolas and González-Ladrón-de-Guevara 2012)).

Motivation

Recent developments warrant a critical review of the value of MTurk to public administration and management. In particular, calls for more experimental public administration and management research (Kelman 2007; Perry 2012) have been answered by an increasing use of experimental research designs within the discipline (Bouwman and Grimmelikhuijsen 2014). At the same time there have been increasing pushes toward more integration of the psychological and behavioral sciences in public administration and management scholarship (Olsen 2015b). More specifically, scholars use the term “Behavioral Public Administration” to describe public administration and management research drawing directly from theories of psychology and associated research methods, including experimental designs (Grimmelikhuijsen, Jilke, Olsen, and Tummers 2015). The burgeoning stream of research is evidenced by recent formation of EGPA’s Permanent Study Group on Behavioral Public Administration, preconference workshops on experimentation at the 2015 and 2016 Public Management Research Association Conference, a special forum on Behavioral Public Administration at the 2015 Academy of Management Meeting, as well as the International Public Management Journal’s special issue on experiments (Blom-Hansen, Morton, and Serritzlew 2015). Moreover, several recent publications incorporate psychological theories and experimental methods (Baekgaard 2015; Baekgaard and Serritzlew 2015; Jilke, Van Ryzin, and Van de Walle 2016; Marvel 2015, 2016; Olsen 2015a; Pedersen 2015; Riccucci, Van Ryzin, and Lavena 2014; Riccucci, Van Ryzin, and Li 2015).

At the same time, MTurk has emerged as a platform for experimental research in other fields of social science (Paolacci and Chandler 2014) because MTurk lends itself well to survey experimental studies. MTurk's large user base and functional technology make it a convenient means of data collection (Chandler et al. 2014). Analyses demonstrate that MTurk data are considerably more diverse than samples of college undergraduates (Buhrmester et al. 2011; Johnson and Borden 2012). Moreover, studies show that MTurk data exhibit good reliability and share psychometric properties with data collected by more traditional survey procedures and laboratory experiments (Buhrmester et al. 2011; Johnson and Borden 2012). Some scholars even suggest that crowdsourcing platforms, such as MTurk, are "a valid and superior partner to in-person data collection" (Casler, Bickel, and Hackett 2013, 2156).

Another motivation for a review of the value of MTurk to public administration and management relates to a need for delineation: When and where may MTurk represent a particularly relevant and useful tool for public administration and management scholars? This article emphasizes that the value of MTurk to the field is situational; it depends on the particular puzzle the researcher aims to solve.

The remainder of this article proceeds as follows. First, we describe the MTurk platform and elaborate on the strengths and opportunities of using MTurk for research. Second, we discuss the limitations of MTurk. Third, we conclude with a discussion of particular areas of research where MTurk may complement and enhance public administration research.

POTENTIAL BENEFITS FOR SCHOLARSHIP

MTurk provides a platform for users (also known as Requesters) to recruit people to perform tasks known as Human Intelligence Tasks (HITs). HITs are often short, repetitive, and small in scope (Kittur, Chi, and Suh 2008; Little, Chilton, Goldman, and Miller 2009; Novotney and Callison-Burch 2010). Examples of frequently posted HITs include transcription and linguistic tasks (Evanini, Higgins, and Zechner 2010; Fort, Adda, and Cohen 2011; Marge, Banerjee, and Rudnicky 2010). In addition, social scientists have started to post surveys and experiments with the purpose of using MTurk as a means of collecting research data.

While Requesters can post tasks, MTurk allows workers (also known as Turkers) to search for tasks for which they are qualified (see discussion later on qualifications). When a Turker is qualified for a HIT they can select and complete that HIT. Turkers submit their completed work to be reviewed by the Requestor. If the Requestor approves of the work, the Turker receives monetary compensation—at rates that vary with the type of work and with changes in the market for HITs. In social science research, Requesters typically embed a link in the HIT that directs the Turkers to an online survey platform such as Qualtrics or SurveyMonkey.

Scholars suggest a number of reasons why MTurk has emerged as a popular platform for social science research (Buhrmester et al. 2011; Mason and Suri 2012; Paolacci and Chandler 2014; Paolacci, Chandler, and Ipeirotis 2010). We review the

general advantages of collecting data from Mechanical Turk: subject access, cost, survey development and testing, and collection of longitudinal data. As we will discuss, however, each of these advantages poses tradeoffs that potentially limit the utility of the data.

Subject Access

MTurk offers public administration and management researchers access to a pool of research participants (Callison-Burch 2009; Crump, McDonnell, and Gureckis 2013; Kittur et al. 2008; Paolacci and Chandler 2014) spanning a wide range of ages, ethnicities, socio-economic backgrounds, and countries of origins (Berinsky et al. 2012; Casler et al. 2013). MTurk is thus useful for researchers with an interest in examining social phenomena among a large and diverse subject pool.

Moreover, the size and diversity of the MTurk respondent pool provide researchers with the opportunity to obtain information from subpopulations that can be notoriously hard to reach in the context of more traditional survey research, e.g., individuals with disabilities; lesbian, gay, bisexual, or transgender (LGBT) individuals; or victims of workplace stress or harassment (Smith et al. 2015). In addition to ease of subject access, Smith et al. (2015) note that the anonymity afforded participants on the MTurk platform may make it easier to obtain private information from people who might otherwise try to conceal their identity or provide incorrect responses (e.g., an LGBT individual may be hesitant about divulging his or her sexual orientation in an workplace survey, even when promised full anonymity).

MTurk may be especially beneficial to experimental survey research. Many argue that experimental research in the social sciences has long been characterized by an overreliance of student samples (Davis and Holt 1993; Kagel and Roth 1995). Although some research suggests that the experimental responses of students are seldom different from those of other subject pools (Alm, Bloomquist, and McKee 2011; Ball and Cech 1996; Plott 1987), using students as surrogates for non-students remains a controversial issue (Levitt and List 2007). MTurk entails a possible way forward; a viable tool by which experimental research can move past the use of student samples (Henrich, Heine, and Norenzayan 2010; Sears 1986).

This access, however, comes at the expense of a non-trivial tradeoff for public management and administration researchers. Since Turkers are online and participate in research remotely, the contextual realism of experiments can be limited. Second, and important with respect to generalizability, is the concern that the Turkers are not representative of any particular set of public managers or public employees. However, there are still reasons one might consider using MTurk despite these tradeoffs. First, many research questions necessitate the use of citizen data—not responses from public employees or managers. Second, it is possible restrict analyses of responses to only Turkers who are employed in the public sector, but also contrast responses from Turkers employed in other sectors. Finally, samples of non-public employees may be able to offer insights about general human cognition and behavior that can be generalized to public employees and are of interest to public management researchers.

Cost

MTurk is considered a cost-effective means of collecting data. For example, Paolacci and colleagues (2010) replicate classic experiments in judgment of decision making at a cost of \$1.71 per hour per participant. Following an Amazon price increase in July 2015, the cost remains relatively low compared to traditional means of collecting data, such as surveys or laboratory studies. Amazon requires a 20% fee on the amount paid to Turkers. HITs with 10 or more assignments are charged an additional 20% fee. A perusal through the MTurk market in late 2014 indicates that the average rate is around \$2/hour. A 15 minute survey can thus be priced at around 50¢, entailing a cost of \$700 for 1,000 responses. Importantly, research finds that pay rates do not affect the quality (reliability) of the data collected (Buhrmester et al. 2011; Mason and Suri 2012). However, MTurk is a dynamic market and researchers need to do their own due diligence on the appropriate pay rates at any given time for a particular task.

Requesters need, however, to consider several issues when setting pay rates. First, Requesters need to specify the correct amount of time required to complete the HIT to allow Turkers to assess the value of participation. Misleading Turkers about expected completion time might lead to high rates of attrition. Moreover, provision of such incorrect information is associated with ethical issues. In accordance with IRB standards, informed consent of participants requires Requesters to disclose the amount of time that HIT participation is expected to take. As with any experiment or survey, as much information about the research purpose as possible should be provided prior to participation and consent should be obtained prior to using the data. Typically, this

information is presented after a Turker selects a HIT. Upon selecting the HIT, the Turker clicks a link to an externally hosted survey platform. This introduction page describes the task and provides a general description of the content (e.g. workplace attitudes; experiences with government). Moreover, the introduction page should inform the participant of any known risks (or lack thereof) and that participants have the right to exit the study or proceed if they give their consent. Finally, disclosure should be provided if deception is used and respondents should have the opportunity to withdraw their responses after disclosure.

Survey Development And Testing

Researchers can use MTurk as a tool for development of surveys and survey experiments. By providing access to a standing pool of participants, MTurk is highly useful for pre-test and development of reliable survey items. In addition, researchers can more efficiently engage in preliminary theory testing of new ideas and concepts (Mason and Suri 2012, 2-3). Consequently, MTurk may assist scholars in “high risk, high reward” research (i.e., research with a high probability of failure, but potentially having a major research impact) and provides a platform for basic proof of concept studies before taking to instrumentation and data collection in the field or among public managers or employees. MTurk provides opportunities for development and refinement of instrumentation before a project is piloted on the sample or population of interest. MTurk, however, only provides a platform for collecting data and does not—in any circumstance—diminish the importance of careful theory development and survey construction.

Longitudinal Research

Finally, researchers can conduct longitudinal research using MTurk by (1) following up with Turkers who completed a previous HIT; (2) only allowing new HITs to be available to these Turkers, and (3) linking the obtained responses to the individual Turkers (using a unique Turker ID number) (Berinsky et al. 2012). Use of TurkPrime is another viable option—a research platform that integrates with MTurk and supports tasks that are common to the social and behavioral sciences, including longitudinal panel studies (Litman, Robinson, and Abberbock 2016). MTurk thus facilitates longitudinal data collections and analyses. Among other possibilities, MTurk can assess test-retest reliability of latent constructs. Additionally, the ability to use MTurk to follow-up with participants allows scholars to measure variables at different points in time using different measurement instruments. While such procedure may not necessarily solve the issue of common source bias—a major limitation of cross-sectional public administration and management research (Jakobsen and Jensen 2015; Meier and O'Toole 2013)—it may help reduce the risk of common method variance (Podsakoff, MacKenzie, Lee, and Podsakoff 2003; Podsakoff, MacKenzie, and Podsakoff 2012).

MAIN ISSUES OF ATTENTION

MTurk may facilitate research, but public administration and management researchers must consider a number of issues when using MTurk for data collection. We will now discuss issues related to representativeness, participation, generalization, and data quality.

Representation: Who Is A Turker?

According to Paolacci and Chandler (2014), in 2014 there were more than 500,000 Turkers residing in 190 countries. Studies indicate that the Turker pool is predominantly located in the United States and India. Ross and colleagues (2010) found that approximately 57 percent of Turkers were from the United States and 32 percent from India.

When Ross and colleagues (2010) looked only at HIT responses from Turkers located in the United States, the average Turker's age was 31, 55 percent were female, more than half had a college degree, and the median income was between \$20,000 and \$30,000. Some studies find that U.S. Turkers tend to be more educated, female, younger, and have lower salaries than the general population (Paolacci and Chandler 2014; Paolacci et al. 2010; Ross et al. 2010), while recent studies in public administration and management find that U.S. Turkers tend to be more male (Jilke et al. 2016; Marvel 2015a, 2016). That being said, the MTurk population provides a diverse response group that is more representative than the typical student sample (Buhrmester et al. 2011; Casler et al. 2013; Johnson and Borden 2012).

The demographics of Turkers is an important issue—both in relation to the generalizability of research findings and the possibilities for examining selective subpopulations. For issues of generalizability, executing an MTurk survey with numerous demographic items allows the researcher subsequently to weight the data so that results are largely generalizable to a particular population. Additionally, as in the case of much

public administration and management research, asking questions that provide high granularity with respect to work and employment context is absolutely vital.

One glaring concern is the degree to which Turkers can help researchers understand organizational phenomena (public or private). One important characteristic to consider is the employment status and work experience of Turkers. We recently collected data from 3,451 Turkers as part of the National Administrative Studies Project-Citizen data collection project (Taggart, Stritch, and Bozeman 2016). The data allow us to draw subsamples from the Turker population. In asking about their employment status, 43 percent reported working full-time for an organization. When we include part-time workers, nearly 59 percent reported working for an organization. While 14 percent reported being self-employed, the remainder of the respondents reported being homemakers (8 percent), students (21 percent), out of work—not looking for work (2 percent), out of work—looking for work (13); retired (3 percent), or unable to work due to disability (3 percent)

We asked all those who indicated working full-time or part-time at an organization to describe their current position, 19 percent indicated a management position, 11 percent indicated working in IT services, 29 percent identified as professionals, 17 percent responded as administrative support, 15 percent identified as working in sales, and 22 percent reported “Other.” We further probed the sector of an individual’s employment and found that 1,352 respondents worked in a private sector organization, 279 worked in the non-profit sector, and 314 worked in a public sector organization.

The idea that the Turker group is chronically unemployed is not an accurate depiction. The Turkers are a diverse group and there is variation in their level of employment and work experiences. We will now discuss a few technical approaches for limiting Turker participation to particular subpopulations of Turkers (e.g., public employees).

Finding Subpopulations On Mturk

Public administration and management scholars conducting MTurk research might want to focus on subpopulations within MTurk. For example, they may want to limit participation to Turkers located in a specified geographic area or to Turkers exhibiting certain key characteristics (e.g., public sector employment). In general, three broad options exist for collecting data from a specific subpopulation on MTurk.

The first option for limiting participation is a pre-selection option. The pre-selection option entails restricting access to the HIT using Amazon's Qualification Types. Qualification Types enable the Requester to specify the particular kind of Turkers who can perform the HIT—based on Amazon's general account information, the Turker's HIT history, or how the Turker performed on past HIT's of the Requester.

The second method for limiting participation is a post-selection option. The researcher slices up the collected MTurk data ex-post to focus on the subpopulation of interest—for example using the Turkers' responses to survey items capturing sample selection criteria such as country of origin, employment status, or sector of employment. This option is useful if the researcher wishes to compare the results of the subpopulation of interest to

the larger MTurk population or other subpopulations. Web-based survey tools often used in conjunction with MTurk HITs can tell the Turker's location based on his or her computer's IP address, thus providing the Requester with information about the location of the Turker. One limitation of this approach is that researchers might need a large number of participants in order to get enough respondents from the subgroup of interest.

Finally, all listed HITs in the MTurk Marketplace have a title, description, keywords, and a statement of the amount of time that Turkers have for completing the HIT. The third option of limiting participation entails that Requesters include information in the HIT description about the type of Turker they are looking for to accept the HIT. For example, a Requester may state that they want U.S. Turkers that are employed in the public sector. Functionally, this option is the easiest to execute and is cost-free, but whether the Turkers performing the HIT in fact fit the specified request in the HIT description cannot be verified.

Depending on the research design, a hybrid approach using more than one of these options may be the best to achieve the desired response from a specific subpopulation. For example, Requesters seeking a specific population from a certain geographic area can state this in the HIT description, use a standardized Qualification type based on the Turker's country on file, and further refine their sample by restricting the HIT to only those who performed a certain way on a previous HIT. However, generalizing the responses from a subpopulation on MTurk to a larger population is not the same as in traditional sampling methods, an issue we discuss in the next section.

Generalizability Using Mturk Survey Data

Researchers should be aware of the potential external validity limitations of crowdsourced data. They should also be cautious about the way they describe data obtained on MTurk so as to not misrepresent the representativeness of the data. Survey data obtained from MTurk do not represent a true random sample. Responses are provided by interested Turkers who have already self-selected themselves into the Turker population. Researchers can aim their HIT at a target population using the options we describe in the previous section but have no way of knowing who within that target population passed on their HIT. Therefore, calculation of a survey response rate for that target population is not feasible.

The issues of generalizability and representation must be handled differently in MTurk. Knowledge about the general MTurk population, the law of large numbers, and the inclusion of background variables in MTurk surveys allow MTurk researchers to understand who is responding and to what extent the data generalize to specific populations. For example, because Turkers respond online MTurk survey respondents will probably involve an underrepresentation of people without internet access. While Berinsky and colleagues (2012) observe that within the U.S., Asians are overrepresented, while both Hispanics and African-Americans are underrepresented among Turkers. Similarly, Gilke and colleagues (2016) find that Turkers are fairly diverse and nationwide in scope, but more male, white, and younger when compared to the U.S. adult population in the American Community Survey. Marvel's (2015a, 2016) take up on MTurk was

generally more male, white, educated, Democrat, and liberal than the U.S. population (see also Huff and Tingley 2015; Chandler and Shapiro 2016; Clifford, Jewell, and Waggoner, 2015).

However, even survey research using random sampling is seldom perfectly representative. Surveys that use traditional sampling methods, i.e., paper form surveys and face-to-face interview surveys, also struggle with issues of representativeness. Voluntary survey participation requires some extent of personal time allocation and effort, entailing a likely underrepresentation of low-income groups. Like in random sampling, data obtained from MTurk might benefit from different statistical techniques (e.g., data weights) that may help alleviate the issue, but full population representativeness is not feasible. This is true for survey research in general—but especially data obtained through crowdsourcing.

Furthermore, because the cost per response on MTurk is often lower relative to traditional sampling methods, a larger number of responses is often achievable on MTurk under similar research budget conditions. In some sampling cases, a small number of responses may entail that the obtained responses are unrepresentative of the target population, increasing sampling bias concerns and reducing the potential for extrapolation of results (Groves et al. 2013; White, Armstrong, and Saracci 2008). Moreover, small numbers reduce the likelihood of detecting a true effect (i.e., increasing the risk of “false negatives”) (Button et al. 2013). The problem of small sample sizes also relates to an ethical dimension, as unreliable research is inefficient and wasteful.

A recent societal development has enlarged the problem of small sample sizes in sampling-based research. For the past few decades, the response rates to surveys have been in a general decline (Curtin, Presser, and Singer 2005; De Leeuw and Heer 2002; Hansen 2007). Simultaneously, the internet and email has made designing and implementing surveys for research purposes easier in a technical sense, and electronic questionnaires are fast becoming a commonly used method for collecting information from large, geographically dispersed populations (Dillman, Smyth, and Christian 2014; Edwards et al. 2009). However, web-based and email surveys notoriously suffer from low-response rates and participation (Dillman et al. 2014). MTurk is thus a potential remedy for small response sizes in survey-based public administration and management research. However, as we discussed previously, getting a response from a specific group of people may prove difficult using MTurk. Thus, an understanding of the method of data collection in MTurk is essential for avoiding research designs that goes beyond the limitations of Turkers.

Quality Of Data

Another concern relates to the quality of data with respect to accuracy, consistency, validity, and completeness. Researchers note that some Turkers might not pay sufficient attention to the tasks they are performing or are simply clicking through surveys and experiments as quickly possible (Goodman, Cryder, and Cheema 2013). For some Turkers, survey participation may be strictly motivated by the payment they receive for completing a HIT. In consequence, some Turkers may click through an MTurk survey without paying much attention to the specified survey items. Such actions may dilute the

reliability of MTurk data and increase the chance of false negatives (Button et al. 2013). However, money does not appear to be the primary motive for HIT completion among Turkers, especially those located in the U.S. Research shows that Turkers are intrinsically motivated and participate for other reasons, such as curiosity, a way to spend free time, or simply finding the tasks enjoyable (Buhrmester et al. 2011; Kaufmann, Schulze, and Veit 2011).

One might argue that since individuals are responding outside of their own work context there is less incentive to lie. The biggest threat to quality, in our view, is people simply clicking responses to finish quickly. However, such problems exist with traditional surveys as well. In addition, the platform has mechanisms that incentivize high quality work (attention) as Turkers who fail quality checks run the risk of being denied payment and getting low Quality ratings.

Smith (2013) outlines multiple mechanisms commonly used in survey research conducted on the internet to detect issues related to data quality. First, researchers can set an upper and lower threshold for the amount of time it takes to complete a survey and reject the responses that fall above or below the thresholds. Second, researchers can also check for response patterns within a survey response to identify Turkers who were not engaged in the survey. Third, “quality check” questions can be placed throughout the survey, and used to check for both response quality and survey fatigue. For example, in a set of survey items applying Likert-type responses, researchers can include a statement saying "Select 'Strongly Disagree' on this line." This procedure helps identify Turkers who are

simply clicking through the survey without reading each statement. As another option, researchers can expose respondents to two item statements basically capturing the same underlying meaning. For example, one statement saying "Being innovative makes me excited," another statement saying "I do not get excited when I am being innovative." Respondents with identical responses (e.g., who answer "Strongly Agree" to both statements) can be flagged for quality or fatigue effects. Finally, researchers may employ a survey question that asks respondents to select particular answer choices for identifying Turkers who are not reading the survey items.

There are other mechanisms that are specific to MTurk that might be used by researchers to check quality. For instance, researchers may include a notice in the HIT description warning Turkers about such quality check questions and that HIT payment is contingent on passing these quality checks. Finally, many survey software packages allow researchers to view the IP address of respondents. When multiple responses come from the same IP address, researchers can remove these extra responses to reduce this threat to independence and internal validity (Horton, Rand, and Zeckhauser 2011; Jilke et al. 2015).

In sum, the use of MTurk data in public administration and management research entails potential issues related to sample representativeness, participation, and data quality.

While these issues are inherent to most survey research, they may be particularly troublesome to crowdsourced survey research. Different technical procedures can be used to alleviate or largely eliminate these issues. Furthermore, most studies that provide

financial compensation to respondents are subject to the same threat to data reliability if the incentive is simply to complete the task to receive compensation.

POTENTIAL APPLICATIONS IN PUBLIC MANAGEMENT AND ADMINISTRATION SCHOLARSHIP

MTurk can serve as a platform for the collection of data from broad populations for different types of studies—experimental, cross-sectional, and longitudinal—but scholars need to be cautious about when and where to use the platform: MTurk is far from relevant for all types of public administration and management questions. There are some very specific ways that the utilization of MTurk can enhance the current state of public administration research. We explain those ways below, but with caution. Specifically, we believe that crowdsourced data might enhance the current public administration scholarship in relation to the following five domains: (1) Exploratory analyses and survey construction; (2) Measurement refinement of latent constructs; (3) Experimental research; (4) Longitudinal research that can limit common method bias and be used to develop temporally stable measures issues relating to common methods bias; and (5) Research on the citizens' attitudes towards government and/or government agencies.

Exploratory Analyses And Survey Construction

One of the challenges of public management research is the fact that failure can be expensive. Generating large and high quality samples of public administrators or public sector leaders is generally time consuming and expensive. As mentioned, scholars' may thus refrain from initiating “high risk, high rewards” research. In other words, researchers

have an incentive to pursue projects with a high probability of working out, at the expense of interesting ideas that are unproven. Similarly, scholars may decide to use MTurk for its cost effectiveness and ease of access to respondents by altering existing research questions to fit with MTurk. Such an approach leans toward research produced for its own sake, and has a lower probability of furthering knowledge on relevant questions.

Prior to conducting both surveys and experiments on the target populations, use of MTurk samples may be a relatively low-cost way for constructing and testing measurement instruments. Similarly, researchers may run a pilot test of survey items using MTurk. In essence, MTurk is an inexpensive and, potentially, effective tool for testing instruments and identifying textual errors prior to a survey's distribution. In public management, scholars may benefit from using MTurk for the purposes of (1) pilot tests of surveys and survey items, (2) exploratory tests of variable associations, and (3) pretests of reliability and validity in relation to measurement scales/instruments.

Measurement Refinement

In addition to assisting in the basic building of surveys, researchers might use MTurk samples to build and refine the measurement of latent constructs. Research on public service motivation (PSM) provides a useful example for illustrating how MTurk can be useful. Since its initial formulation (Perry and Wise 1990), much PSM research has been dedicated to the refinement of PSM measurement (Coursey and Pandey 2007; Kim 2009, 2011; Perry 1996; Vandenabeele 2008; Wright, Christensen, and Pandey 2013). Today,

however, MTurk might have assisted and expedited this development and allowed for the validation of the instrument in a heterogeneous group and in multiple sectors. For instance, rather than utilizing a convenience sample of MPA student and public sector workers for an exploratory factor analysis, Perry (1996) might have considered an MTurk population to administer a survey to collect data for measurement exploration. As an example, recent work by Resh, Marvel, and Wen (2016) uses Turkers to develop an implicit measure of prosocial motivation that is not subject to the same threats to measurement posed by social desirability biases as survey items.

In short, the cost effectiveness and ease of access to MTurk samples allows researchers to focus their efforts on the validation of single measures and multiple constructs. The cost of survey research often requires researchers to hedge against null-findings by including many additional variables in surveys—decreasing the incentive to take the time necessary for appropriate measure development and validation. Furthermore, diverse, heterogeneous groups might facilitate the development of measures that are also invariant across sub-populations.

Experiments

Public management and administration research is characterized by an increasing use of experimental and quasi-experimental research designs for examining causal questions (Avellaneda 2013; Bellé 2013, 2014; Brewer and Brewer 2011; Christensen et al. 2013; Feeney 2012; Gallo and Lewis 2012; Gordon 2009; James 2009, 2011; Kelman and Friedman 2009; Kiefer, Hartley, Conway, and Briner 2015; Pedersen 2015; Moynihan

2006). While experiments on MTurk do not replace the high quality field experiments (Avellaneda 2013; Bellé 2013, 2014; James 2011) in terms of external and ecological validity, MTurk provides a useful platform for survey experiments with high levels of internal validity similar to that of laboratory studies (Casler et al. 2013). MTurk has proven to be a viable medium for implementing experimental studies on issues relevant to public administration (such as red tape, public service motivation, administrative burden, representative bureaucracy, and trust in government) as evidenced by the recent literature. Marvel (2016) used the platform to conduct an experiment examining the effect of favorable performance information on individual's implicit attitudes about the U.S. Postal Service. Jilke, Van Ryzin, and Van De Walle (2016) performed an experiment on the MTurk platform examining how an increase in the number of public service providers decreases the likelihood that an individual would switch from a poorly performing provider. The platform, however, has been largely been ignored by researchers interested in questions of internal public management (for an exception see Pedersen and Stritch 2016).

One potential reason we have not seen MTurk used for experiments related to internal management is a general skepticism about the use of a convenience sample of Turkers (Landers and Behrend 2015; Smith et al. 2015). Landers and Behrend (2015) aptly point out that all organization-based samples, in essence, are convenience samples, or a function of feasibility. The term "convenience" oversimplifies complex issues and has been taken to be synonymous with an *easy* sample. Such judgements are counterproductive. Ease of access is not a valid criteria for assessing the appropriateness

of a sample population for a specific research question. The real issue is whether or not the sample is valid for the proposed research questions. As an example, to ask Turkers to imagine themselves as public managers in an experiment might be inappropriate. On the other hand, for Turkers who have worked before or are currently working, asking them to imagine themselves as employees might be reasonable. Public management researchers can use survey experiments to learn how employees might respond to or are affected by different organizational situations and contexts. In many cases, the findings might be of interest to scholars interested in the internal management of public organizations. Extant cross-sectional studies of variable relationships in public organization are widely marked by high levels of external validity, but most suffer from internal validity concerns. Experiments using MTurk may provide valuable knowledge of causal relationships between phenomena in public organizations. In this context, MTurk experiments may significantly contribute to the identification of causal relationships among variables.

Longitudinal Research: Limiting Common Method Bias

One of the areas for scholars to advance public administration scholarship is in the area of longitudinal studies (Zhu 2013). As previously mentioned, MTurk can be used for longitudinal data analyses. By exploiting this MTurk feature, researchers may reduce concerns about common method bias. Studies relying on single surveys for the measurement of the independent and dependent variables are subject to internal validity threats in the form of common method bias (Podsakoff et al. 2003; Podsakoff and Organ 1986). Variations of common method bias, i.e., common rater or common source bias, continues to threaten the validity of much of the extant public administration research

(Favero and Bullock 2014; Jakobsen and Jensen 2015). Favero and Bullock (2014) observe that the only way to eliminate common source bias is by obtaining measures of the independent and dependent variables from separate sources (e.g., by capturing one of the variables by an “objective” measure). We agree but suggest that there are times where such procedure is close to impossible. For instance, say that we are interested in examining in the relationship between two latent constructs, e.g., self-efficacy and job attitudes. In this case, a respondent’s self-reported answers would likely be better and more accurate than if such description was provided by an outside evaluator. In instances such as these, MTurk researchers can use data from multiple points in time and different instruments, helping to limit common method bias (Podsakoff et al., 2012).

Longitudinal Research: Measurement Stability And Test-Retest Reliability

Additionally, the ability to utilize MTurk for longitudinal analyses can also help support the development of temporally stable measures. For instance, three constructs have emerged prominently in public administration research over the past twenty years: red tape (Bozeman 1993), PSM (Perry 1996), and managerial networking (Meier and O'Toole 2001). Interestingly, there has not been any inquiry into the test-retest reliability and the temporal stability of the measures. A weakness of public administration and management scholarship has been a lack of attention to construct validation and measurement (Wright, Manigault, and Black 2004). Since many of the organizational attributes (e.g., red tape) and many of the individual attributes (e.g., PSM and managerial networking) that are of interest to many public administration researchers are not necessarily sector specific, the general population of MTurk workers might be used to

form a sample that scholars could use to comprehensively examine constructs and build measures.

To date, there has been limited grand-scale investigation into the test-retest reliability of many measures developed in public administration and management research. This absence stands in stark contrast with the development of other measures of latent constructs used outside of public administration research, such as organizational commitment (e.g. Allen and Meyer 1996; Lam 1998) and the Big Five personality traits (e.g. Robins, Fraley, Roberts, and Trzesniewski 2001; Schuerger, Zarrella, and Hotz 1989). MTurk provides a reasonable means of helping to build valid and reliable measures of organizational attributes, as well as individual attitudes, motives, behaviors, and perceptions.

Citizens' Attitudes And Perceptions Of Public Management

Areas that are ripe for research include the opinions and reactions of citizens to bureaucratic organizations and bureaucrats in them. This is the focus of extant public administration and management studies conducted on the MTurk platform (Marvel 2015, 2016). Other public management literatures where citizens are the group of interest might also be appropriate for exploring using MTurk. For instance, themes such as citizen coproduction (Bovaird 2007; Brudney and England 1983; Jakobsen and Andersen 2013; Osborne 2008; Riccucci et al. 2015), citizen engagement (Reddel and Woolcock 2004; Smith, Santucci, Xu, Cox, and Henderson 2012), and administrative burden (Herd,

DeLeire, Harvey, and Moynihan 2013; Moynihan, Herd, and Harvey 2014) might all be studied using different types of experimental designs on the MTurk platform.

CONCLUSION

This article introduces both the potential and limitations of MTurk. We observe how other social science fields are increasingly using data collected from MTurk. Motivated by this observation, we ask whether, when, and how crowdsourcing may be a useful tool to public administration and management scholarship. Table 1 summarizes our discussion of the (1) possibilities of using MTurk for data collection, (2) limitations and areas of concern, and (3) particular areas where the potentials of using MTurk to enhance current public administration scholarship are especially pronounced.

Overall, we find that MTurk may be a useful tool in the toolbox of public administration and management scholars—when used appropriately and in relation to a specific set of public administration puzzles. In particular, we identified five areas where MTurk may complement and enhance public management research; these include: (1) Exploratory analyses and survey construction; (2) Measurement refinement of latent constructs; (3) Experimental research; (4) Longitudinal research that can limit common method bias and be used to develop temporally stable measures issues relating to common methods bias; and (5) Research on the citizens' attitudes towards public administration and government. MTurk may be especially relevant in relation to certain types of research and research questions, especially experimental designs where citizen perception, reactions, and responses are of interest to the researchers. Deciding whether or not it is appropriate to

use of MTurk ultimately depends on the nature of the question at hand. For researchers interested in the actions of public or nonprofit managers, MTurk data is no substitute for high quality samples of public and nonprofit managers.

Good research takes into account the applicability of MTurk data and the validity of MTurk-based findings: The researcher must understand the boundaries and potential of the platform. MTurk is in no way a panacea for empirical public administration research. The issues related to representativeness, participation, and data quality are non-trivial. However, for researchers interested in the relationship among variables for public and nonprofit managers, MTurk might be a cost effective and efficient way of building and pretesting instruments prior to distributing the survey to the target sample population. Similarly, for experimental research on executive decision making, there is no substitute for other more traditional data collection procedures—but MTurk may be useful for survey experimental research on individuals' perception of the decision making processes (e.g. how individuals perceive decisions based on different rationales and decision making process).

In the absence of readily available groups of public managers and public employees to serve as participants in exploratory, preliminary, and proof of concept studies, MTurk might be a helpful tool for testing and refining instruments. We believe that MTurk might be useful with respect to preliminary steps in research—but MTurk cannot replace samples of public managers for many of the questions fundamental to public administration. However, one can envision conducting experiments on general

populations while manipulating variables related to the public management context. This has already been done with respect to red tape (Kaufman and Feeney 2014), PSM (Brewer 2011), and program performance data (Moynihan 2006) using student samples. When implemented correctly, MTurk might provide a better population than student populations (Buhrmester et al. 2011; Casler et al. 2013; Johnson and Borden 2012).

In terms of methodology, the field of public administration and management has come a long way. Still, the examination of several important public administration questions remains marked by an overreliance of cross-sectional observational data. MTurk may be used as a platform for experimental research that can strengthen the internal validity of our current knowledge of the relationships between managerial, personnel, and/or organizational variables and phenomena.

Furthermore, a number of constructs that public administration scholars are interested in are not limited to public and nonprofit organizations. MTurk can be a way for public administration scholars to conduct more general research and help us connect to a broader audience outside of public administration. For instance, by testing constructs like PSM and red tape in experimental settings among general populations, public administration may be in a better position to export these constructs to more general literatures. In sum, MTurk provide a useful platform for advancing particular aspects of public administration research—inasmuch as MTurk researchers recognize the requisites for good crowdsourced-based survey research and know the research questions where MTurk is of relevance and use.

ABOUT THE AUTHORS

Justin M. Stritch (jstritch@asu.edu) is an Assistant Professor at Arizona State University and is a Senior Research Affiliate at the Center for Organization Research and Design (CORD). His research focuses on work motivation, public management, employee decision-making, and employee performance.

Mogens Jin Pedersen (mjp@sfi.dk) is a Postdoctoral Researcher at SFI—The Danish National Centre for Social Research and Department of Political Science, Aarhus University. His research focuses on work motivation, public management, organizational behavior and theory, and gender and racial biases.

Gabel Taggart (Gabel.Taggart@asu.edu) is a doctoral candidate in Public Administration and Policy at Arizona State University. His research interests include public management and higher education studies.

REFERENCES

- Allen, N. J., and Meyer, J. P. 1996. "Affective, Continuance, and Normative Commitment to the Organization: An Examination of Construct Validity." *Journal of Vocational Behavior* 49(3): 252-276.
- Alm, J., Bloomquist, K. M., and McKee, M. 2011. "Comparing Student and Non-Student Reporting Behavior in Tax Compliance Experiments." Pp. 93-98 in A. Plumley, ed., *Recent Research on Tax Administration and Compliance*. Washington, DC: IRS Research

Bulletin.

Arceneaux, K. 2012. "Cognitive Biases and the Strength of Political Arguments."

American Journal of Political Science 56(2): 271-285. doi:10.2307/23187099

Avellaneda, C. N. 2013. "Mayoral Decision-Making: Issue Salience, Decision Context, and Choice Constraint? An Experimental Study with 120 Latin American Mayors."

Journal of Public Administration Research and Theory 23(3): 631-661.

doi:10.1093/Jopart/Mus041

Baekgaard, M. 2015. "Performance Information and Citizen Service attitudes: Do Cost Information and Service Use Affect the Relationship?" *International Public Management*

Journal 18(2): 228-245. Doi:10.1080/10967494.2015.1022676

Baekgaard, M., and Serritzlew, S. 2015. "Interpreting Performance Information:

Motivated Reasoning or Unbiased Comprehension." *Public Administration Review* N/A-

N/A. doi:10.1111/Puar.12406

Ball, S. B., and Cech, P.-A. 1996. "Subject Pool Choice and Treatment Effects in

Economic Laboratory Research." *Research in Experimental Economics* 6(3): 239-292.

Bellé, N. 2013. "Experimental Evidence on the Relationship between Public Service

Motivation and Job Performance." *Public Administration Review* 73(1): 143-153.

Bellé, N. 2014. "Leading to Make a Difference: A Field Experiment on the Performance

Effects of Transformational Leadership, Perceived Social Impact, and Public Service

Motivation." *Journal of Public Administration Research and Theory* 24(1): 109-136.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. 2012. "Evaluating Online Labor Markets

For Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3):

351-368.

- Berinsky, A. J., Margolis, M. F., and Sances, M. W. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay attention on Self-Administered Surveys." *American Journal of Political Science* 58(3): 739-753. Doi:10.1111/Ajps.12081
- Blom-Hansen, J., Morton, R., and Serritzlew, S. 2015. "Experiments in Public Management Research." *International Public Management Journal* 18(2): 151-170. Doi:10.1080/10967494.2015.1024904
- Bouwman, Robin and Stephan Grimmelikhuijsen (2014). "Reviewing Experimental Public Administration Research: The Emergence of a Hybrid Tradition." Paper presented at the IRSPM Conference, Ottawa, Canada, April 9-11, 2014.
- Bovaird, T. 2007. "Beyond Engagement and Participation: User and Community Coproduction of Public Services." *Public Administration Review* 67(5): 846-860.
- Bozeman, B. 1993. "A Theory of Government 'Red Tape.'" *Journal of Public Administration Research and Theory* 3(3): 273-304.
- Brewer, G. A. 2011. "Parsing Public/Private Differences in Work Motivation and Performance: An Experimental Study." *Journal of Public Administration Research and Theory* 21(Suppl 3): I347.
- Brewer, G. A., and Brewer, G. A. 2011. "Parsing Public/Private Differences in Work Motivation and Performance: An Experimental Study." *Journal of Public Administration Research and Theory* 21(Suppl 3): I347-I362. doi:10.1093/Jopart/Mur030
- Brudney, J. L., and England, R. E. 1983. "Toward a Definition of the Coproduction Concept." *Public Administration Review* 43(1): 59-65.

- Buhrmester, M., Kwang, T., and Gosling, S. D. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6(1): 3-5.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14(5): 365-376.
- Callison-Burch, C. 2009. "Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk." Paper presented at the Conference on Empirical Methods in Natural Language Processing, Singapore, August 6-7.
- Casler, K., Bickel, L., and Hackett, E. 2013. "Separate But Equal? A Comparison of Participants and Data Gathered Via Amazon's Mturk, Social Media, and Face-to-Face Behavioral Testing." *Computers in Human Behavior* 29(6): 2156-2160.
- Chandler, D., and Kapelner, A. 2013. "Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets." *Journal of Economic Behavior and Organization* 90, 123-133.
- Chandler, J., Mueller, P., and Paolacci, G. 2014. "Nonnaïveté Among Amazon Mechanical Turk Workers: Consequences and Solutions For Behavioral Researchers." *Behavior Research Methods* 46(1): 112-130.
- Chandler, J. and D. Shapiro. 2016. "Conducting Clinical Research Using Crowdsourced Convenience Samples." *The Annual Review of Clinical Psychology* 12:53-81
- Christensen, R. K., Whiting, S. W., Im, T., Rho, E., Stritch, J. M., and Park, J. 2013. "Public Service Motivation, Task, and Non-Task Behavior: A Performance Appraisal Experiment with Korean MPA and MBA Students." *International Public Management Journal* 16(1): 28-52.

Clifford, S., R.M. Jewell, and P.D. Waggoner. 2015. "Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?" *Research and Politics*. DOI: 10.1177/2053168015622072

Coursey, D. H., and Pandey, S. K. 2007. "Public Service Motivation Measurement." *Administration and Society* 39(5): 547-568. doi:10.1177/0095399707303634

Crump, M. J., McDonnell, J. V., and Gureckis, T. M. 2013. "Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research." *Plos One* 8(3): E57410.

Curtin, R., Presser, S., and Singer, E. 2005. "Changes in Telephone Survey Nonresponse over the Past Quarter Century." *Public Opinion Quarterly* 69(1): 87-98.

Davis, D. D., and Holt, C. A. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.

De Leeuw, E. D., and Heer, W. D. 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." Pp. 41-54 in Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A., eds., *Survey Nonresponse*. Utrecht University.

Dillman, D. A., Smyth, J. D., and Christian, L. M. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.

Dowling, C. M., and Wichowsky, A. 2015. "Attacks without Consequence? Candidates, Parties, Groups, and the Changing Face of Negative Advertising." *American Journal of Political Science* 59(1): 19-36. doi:10.1111/Ajps.12094

Edwards, P., Roberts, I., Clarke, M., Diguseppi, C., Wentz, R., Kwan, I., Cooper, R., Feliex, L. M. and Pratap, S. (2009). "Methods to Increase Response to Postal and Electronic Questionnaires." *Cochrane Database System Review* 3, 1-12.

Eriksson, K., and Simpson, B. 2011. "Perceptions of Unfairness in Allocations between Multiple Recipients." *Cognitive Psychology* 62(3): 225-244.

Estellés-Arolas, E., and González-Ladrón-De-Guevara, F. 2012. "Towards an Integrated Crowdsourcing Definition." *Journal of Information Science* 38(2): 189-200.

Evanini, K., Higgins, D., and Zechner, K. 2010. "Using Amazon Mechanical Turk for Transcription of Non-Native Speech." Paper presented at the annual conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, June 1-6.

Favero, N., and Bullock, J. B. 2014. "How (Not) to Solve the Problem: An Evaluation of Scholarly Responses to Common Source Bias." *Journal of Public Administration Research and Theory* 25(1): 285-308.

Feeney, M. K. 2012. "Organizational Red Tape: A Measurement Experiment." *Journal of Public Administration Research and Theory* 22(3): 427-444.

Fishbach, A., Henderson, M. D., and Koo, M. 2011. "Pursuing Goals with Others: Group Identification and Motivation Resulting From Things Done Versus Things Left Undone." *Journal of Experimental Psychology* 140(3): 520-534.

Fort, K., Adda, G., and Cohen, K. B. 2011. "Amazon Mechanical Turk: Gold Mine or Coal Mine?" *Computational Linguistics* 37(2): 413-420.

- Gallo, N., and Lewis, D. E. 2012. "The Consequences of Presidential Patronage for Federal Agency Performance." *Journal of Public Administration Research and Theory* 22(2): 219-243.
- Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., and Swann Jr, W. B. 2011. "On the Nature of Identity Fusion: Insights into the Construct and a New Measure." *Journal of Personality and Social Psychology* 100(5): 918-933.
- Goodman, J. K., Cryder, C. E., and 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.
- Gordon, S. C. 2009. "Assessing Partisan Bias in Federal Public Corruption Prosecutions." *American Political Science Review* 103(4): 534-554.
- Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., and Tummers, L. 2015. "Behavioral Public Administration: Combining Insights From Public Administration and Psychology." Paper presented at the Academy of Management, Vancouver, BC, August 7-11.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. 2013. *Survey Methodology*. Hoboken, NJ: John Wiley & Sons.
- Hansen, K. M. 2007. "The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI-Study." *International Journal of Public Opinion Research* 19(1): 112-121.
- Healy, A., and Lenz, G. S. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy." *American Journal of Political Science* 58(1): 31-47. Doi:10.1111/Ajps.12053

- Henrich, J., Heine, S. J., and Norenzayan, A. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33(2-3): 61-83.
- Herd, P., Deleire, T., Harvey, H., and Moynihan, D. P. 2013. "Shifting Administrative Burden to the State: The Case of Medicaid Take- Up." *Public Administration Review* 73(S1): S69-S81.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics* 14(3): 399-425.
- Huff, C. and D. Tingley. 2015. "“Who Are These People?” Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents." *Research and Politics*. DOI: 10.1177/2053168015604648
- Jakobsen, M., and Andersen, S. C. 2013. "Coproduction and Equity in Public Service Delivery." *Public Administration Review* 73(5): 704-713. doi:10.1111/Puar.12094
- Jakobsen, M., and Jensen, R. 2015. "Common Method Bias in Public Management Studies." *International Public Management Journal* 18(1): 3-30.
- James, O. 2009. "Evaluating the Expectations Disconfirmation and Expectations Anchoring Approaches to Citizen Satisfaction with Local Public Services." *Journal of Public Administration Research and Theory* 19(1): 107-123.
- James, O. 2011. "Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments." *Journal of Public Administration Research and Theory* 21(3): 399-418.

- Jilke, S., Van Ryzin, G. G., and Van De Walle, S. 2016. "Responses to Decline in Marketized Public Services: An Experimental Evaluation of Choice Overload." *Journal of Public Administration Research and Theory* 26: 421-432.
- Johnson, D. R., and Borden, L. A. 2012. "Participants at Your Fingertips Using Amazon's Mechanical Turk to Increase Student-Faculty Collaborative Research." *Teaching of Psychology* 39(4): 245-251.
- Kagel, J. H., and Roth, A. E. 1995. *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.
- Kaufman, W., and Feeney, M. K. 2014. "Beyond the Rules: The Effect of Outcome Favourability on Red Tape Perceptions." *Public Administration* 92(1): 178-191.
- Kaufmann, W., & Tummers, L. (2016). "The Negative Effect of Red Tape on Procedural Satisfaction." *Public Management Review*. DOI: 10.1080/14719037.2016.1210907
- Kaufmann, N., Schulze, T., and Veit, D. 2011. "More than Fun and Money. Worker Motivation in Crowdsourcing—A Study on Mechanical Turk." Paper presented at the Americas Conference on Information Systems, Detroit, Michigan, August 4-7.
- Kelman, S. 2007. "Public Administration and Organization Studies." Pp. 225-67 in A. Brief and J. P. Walsh, eds., *Academy of Management Annals*. New York: Erlbaum.
- Kelman, S., and Friedman, J. N. 2009. "Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service." *Journal of Public Administration Research and Theory* 19(4): 917 -946.
- Kiefer, T., Hartley, J., Conway, N., and Briner, R. B. 2015. "Feeling the Squeeze: Public Employees' Experiences of Cutback- and Innovation-Related Organizational Changes

Following A National Announcement of Budget Reductions." *Journal of Public Administration Research and Theory* 25 (4): 1279-1305. doi:10.1093/Jopart/Muu042

Kim, S. 2009. "Revising Perry's Measurement Scale of Public Service Motivation." *The American Review of Public Administration* 39(2): 149.

Kim, S. 2011. "Testing A Revised Measure of Public Service Motivation: Reflective Versus Formative Specification." *Journal of Public Administration Research and Theory* 21(3): 521-546.

Kittur, A., Chi, E. H., and Suh, B. 2008. "Crowdsourcing User Studies with Mechanical Turk." Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, April 5-10.

Lam, S. S. K. 1998. "Test-Retest Reliability of the Organizational Commitment Questionnaire." *The Journal of Social Psychology* 138(6): 787-788.
doi:10.1080/00224549809603264

Landers, R. N. and T.S. Behrend. 2015. "An Inconvenient Truth: Arbitrary Distinctions Between Organizational, Mechanical Turk, and Other Convenience Samples." *Industrial and Organizational Psychology* 8:142-164

Levitt, S. D., and List, J. A. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *The Journal of Economic Perspectives* 21(2): 153-174.

Litman, L., J. Robinson, and T. Abberback. 2016. "TurkPrime.com: A Versatile Crowdsourcing Data Acquisition Platform for the Behavioral Sciences." *Behavior Research Methods*: 1-10.

- Little, G., Chilton, L. B., Goldman, M., and Miller, R. C. 2009. "Turkit: Tools For Iterative Tasks on Mechanical Turk." Paper Presented at the Proceedings of the ACM SIGKDD Workshop on Human Computation, Paris, France, June 28.
- Marge, M., Banerjee, S., and Rudnicky, A. I. 2010. "Using the Amazon Mechanical Turk for Transcription of Spoken Language." Paper Presented at the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas, TX, March 14-19.
- Marvel, J. D. 2016. "Unconscious Bias in Citizens' Evaluations of Public Sector Performance." *Journal of Public Administration Research and Theory* 26(1): 143-158.
- Marvel, J. D. 2015. "Public Opinion and Public Sector Performance: Are Individuals' Beliefs About Performance Evidence-Based Or the Product of Anti-Public Sector Bias?" *International Public Management Journal* 18(2): 209-227.
- Marvel, J. D. 2014. "The Boston Marathon Bombings: Who's to Blame and Why It Matters for Public Administration." *Public Administration Review* 74(6):713-25.
- Marvel, J. D., and Girth, A. M. 2016. "Citizen Attributions of Blame in Third-Party Governance." *Public Administration Review* 76(1): 96-108.
- Mason, W., and Suri, S. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44(1): 1-23.
- Meier, K. J., and O'Toole Jr, L. J. 2013. "I Think (I Am Doing Well), Therefore I Am: Assessing the Validity of Administrators' Self-Assessments of Performance." *International Public Management Journal* 16(1): 1-27.

- Meier, K. J., and O'Toole, L. J., Jr. 2001. "Managerial Strategies and Behavior in Networks: A Model with Evidence from US Public Education." *Journal of Public Administration Research and Theory* 11(3): 271-294.
- Moynihan, D., Herd, P., and Harvey, H. 2014. "Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen-State Interactions." *Journal of Public Administration Research and Theory* 25(1): 43-69.
- Moynihan, D. P. 2006. "What do We Talk about When We Talk about Performance? Dialogue Theory and Performance Budgeting." *Journal of Public Administration Research and Theory* 16(2): 151 -168.
- Novotney, S., and Callison-Burch, C. 2010. "Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription." Paper presented at the annual conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, June 1-6.
- Olsen, A. L. 2015a. "Citizen (Dis) Satisfaction: An Experimental Equivalence Framing Study." *Public Administration Review* 75(3): 469-478.
- Olsen, A. L. 2015b. "'Simon Said,' We Didn't Jump." *Public Administration Review* 75(2): 325-326.
- Osborne, S. P. 2008. *The Third Sector in Europe: Prospects and Challenges*. New York: Routledge.
- Paolacci, G., and Chandler, J. 2014. "Inside the Turk: Understanding Mechanical Turk as a Participant Pool." *Current Directions in Psychological Science* 23(3): 184-188.
- Paolacci, G., Chandler, J., and Ipeirotis, P. G. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5): 411-419.

- Pedersen, M. J. 2015. "Activating the Forces of Public Service Motivation: Evidence from a Low-Intensity Randomized Survey Experiment." *Public Administration Review* 75(5):734-46.
- Pedersen, M. J., and Stritch, J. M. 2016. "Internal Management and Perceived Managerial Trustworthiness: Evidence from a Survey Experiment." *The American Review of Public Administration*. DOI:10.1177/0275074016657179.
- Perry, J. L. 2012. "How Can We Improve Our Science to Generate More Usable Knowledge for Public Professionals?" *Public Administration Review* 72(4): 479-482.
- Perry, J. L. 1996. "Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity." *Journal of Public Administration Research and Theory* 6(1): 5-22.
- Perry, J. L., and Wise, L. R. 1990. "The Motivational Bases of Public Service." *Public Administration Review* 50(3): 367-373.
- Plott, C. R. 1987. "Dimensions of Parallelism: Some Policy Applications of Experimental Methods." Pp. 193-219 in A. E. Roth, ed., *Laboratory Experimentation in Economics: Six Points of View*. New York, NY: Cambridge University Press.
- Podsakoff, P. M., Mackenzie, S. B., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88(5): 879-903.
- Podsakoff, P. M., Mackenzie, S. B., and Podsakoff, N. P. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63, 539-569.

- Podsakoff, P. M., and Organ, D. W. 1986. "Self-Reports in Organizational Research: Problems and Prospects." *Journal of Management* 12(4): 531-544.
- Reddel, T., and Woolcock, G. 2004. "From Consultation to Participatory Governance? A Critical Review of Citizen Engagement Strategies in Queensland." *Australian Journal of Public Administration* 63(3): 75-87.
- Resh, W. G., Marvel, J. D., and Wen, B. 2016. "Do What You Mean, or Mean What You Say? When Self-Interest Crowds Out Prosocial Motivation and Mission Match." Paper presented at the 2016 Public Management Research Conference, Aarhus, Denmark, June 22-24th.
- Riccucci, N. M., Van Ryzin, G. G., and Lavena, C. F. 2014. "Representative Bureaucracy in Policing: Does it Increase Perceived Legitimacy?" *Journal of Public Administration Research and Theory* 24(3): 537-551.
- Riccucci, N. M., Van Ryzin, G. G., and Li, H. 2015. "Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study." *Public Administration Review* N/A-N/A. doi: 10.1111/puar.12401.
- Robins, R. W., Fraley, R. C., Roberts, B. W., and Trzesniewski, K. H. 2001. "A Longitudinal Study of Personality Change in Young Adulthood." *Journal of Personality* 69(4): 617-640.
- Ross, J., Irani, L., Silberman, M., Zaldivar, A., and Tomlinson, B. 2010. "Who are the Crowdworkers? Shifting Demographics in Mechanical Turk." Paper Presented at the Conference on Human Factors in Computing Systems, Atlanta, GA, April 10-15.

- Schuerger, J. M., Zarrella, K. L., and Hotz, A. S. 1989. "Factors That Influence the Temporal Stability of Personality by Questionnaire." *Journal of Personality and Social Psychology* 56(5): 777-783.
- Sears, D. O. 1986. "College Sophomores in the Laboratory: Influences of A Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51(3): 515-530.
- Smith, C., Santucci, D., Xu, S., Cox, A., and Henderson, K. A. 2012. "'I Love My Job, But ...': A Narrative Analysis of Women's Perceptions of their Careers in Parks and Recreation." *Journal of Leisure Research* 44(1): 52-69.
- Smith, J. 2016. "The Motivational Effects of Mission Matching: A Lab-Experimental Test of a Moderated Mediation Model." *Public Administration Review* 76(4): 626–637.
- Smith, N. A., Sabat, I. E., Martinez, L. R., Weaver, K. and Xu, Su. 2015. "A Convenient Solution: Using MTurk to Sample from Hard-To-Reach Populations." *Industrial and Organizational Psychology* 8: 220-228.
- Smith, S. 2013. "4 Ways to Ensure Valid Responses for your Online Survey," Qualtrics, last modified April 22, 2013, <https://www.qualtrics.com/blog/online-survey-valid-responses/>.
- Taggart, G., Stritch, J. M., and Bozeman, B. (2015). National Administrative Studies Project (NASP)—Citizen: Waive 1 [data file and codebook]. Phoenix, Arizona: Center for Organization Research and Design (CORD) at Arizona State University.
- Thibodeau, P., Peebles, M. M., Grodner, D. J., and Durgin, F. H. 2013. "The Wished-For Always Wins Until the Winner Was Inevitable All Along: Motivated Reasoning and

Belief Bias Regulate Emotion During Elections." *Political Psychology*

Doi:10.1111/Pops.12100

Vandenabeele, W. 2008. "Development of a Public Service Motivation Measurement Scale: Corroborating and Extending Perry's Measurement Instrument." *International Public Management Journal* 11(1): 143-167.

White, E., Armstrong, B. K., and Saracci, R. 2008. *Principles of Exposure Measurement in Epidemiology: Collecting, Evaluating and Improving Measures of Disease Risk Factors*. New York: Oxford University Press.

Wright, B. E., Christensen, R. K., and Pandey, S. K. 2013. "Measuring Public Service Motivation: Exploring the Equivalence of Existing Global Measures." *International Public Management Journal* 16(2): 197-223.

Wright, B. E., Manigault, L. J., and Black, T. R. 2004. "Quantitative Research Measurement in Public Administration: An Assessment of Journal Publications." *Administration and Society* 35(6): 747-764.

Zhu, L. 2013. "Panel Data Analysis in Public Administration: Substantive and Statistical Considerations." *Journal of Public Administration Research and Theory* 23(2): 395-428.
doi:10.1093/Jopart/Mus064

Table 1. Summary of Advantages, Potential Pitfalls, and Uses of Crowdsourced Data to Enhance Public Administration Research

POTENTIAL BENEFITS FOR SCHOLARSHIP
A. Subject Access
B. Cost
C. Survey Development and Testing
D. Longitudinal Research
MAIN ISSUES OF ATTENTION
A. Representation: Who is a Turker?
B. Finding Subpopulation in MTurk
C. Generalizability Using MTurk Survey Data
D. Quality of Data
POTENTIAL APPLICATIONS IN PUBLIC MANAGEMENT AND ADMINISTRATION SCHOLARSHIP
A. Exploratory Analyses and Survey Construction
B. Measurement Refinement
C. Experiments
D. Longitudinal Research: Limiting Common Method Bias, Measurement Stability and Test-Retest Reliability
E. Citizens' attitudes and Perceptions of Public Management and Administration