THE EFFECT OF RESEARCH PROJECT DESCRIPTIONS ON SAMPLES DRAWN FROM MECHANICAL TURK

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Arts in Experimental Psychology.

By

Christopher James Holden

Director: Harold A. Herzog
Professor
Psychology Department

Committee Members: Dr. David M. McCord, Psychology
Dr. Christopher Cooper, Political Science and Public Affairs

March 2012

ACKNOWLEDGEMENTS

I would like to thank my committee for their support and direction throughout this process. Specifically, I would like to thank Dr. Harold Herzog whose early interest in Mechanical Turk led me to develop this project, Dr. David McCord for his input and knowledge of personality measures, and Dr. Christopher Cooper for his help in developing the keywords used and for his knowledge of political science.

I would also like to thank Dr. James Goodwin, who provided ideas and input on items used in the questionnaire. I would also like to thank Adam D. Hicks who provided guidance academically, and as a friend. To my friends and colleagues, thank you for your encouragement, access to social activities, and time spent cycling throughout this process. Finally, I would like to thank my family for their continued enthusiasm and support.

TABLE OF CONTENTS

ABSTRACT

THE EFFECT OF RESEARCH PROJECT DESCRIPTIONS ON SAMPLES DRAWN FROM MECHANICAL TURK

Christopher James Holden, M.A.

Western Carolina University (March 2012)

Director: Dr. Harold A. Herzog

The use of online research is growing in psychology. This growth has been augmented by the development of the free online service, Mechanical Turk (MTurk) created by Amazon. While Buhrmester, Kwang, and Gosling (2011) found that MTurk is a reliable means of data collection; the MTurk data collection process may be prone to recruitment bias. This study investigated whether the keywords in project descriptions influence responses. Subjects were recruited through MTurk and selected to be in one of four keyword conditions. These subjects completed a five-factor model personality measure, a political ideology scale, and the Homophobia Scale (Wright, Adams, & Bernat, 1999). Scores on these three scales were assessed for differences based on the keyword condition. After analysis, it was determined that there were no significant differences between conditions on any of the dependent measures. Keywords do not seem to influence recruitment bias on MTurk surveys.

INTRODUCTION

Psychologists have been shifting from paper and pencil surveys to online surveys. There are different options for conducting research online. One that has grown dramatically in popularity is Amazon's Mechanical Turk (MTurk). This website was originally developed for commercial purposes, but has quickly become an outlet for psychological research. A historical perspective of methodological change is provided, followed by a discussion of online research and its limitations and benefits. Next, an overview of MTurk is provided, followed by personal account of what it is like to use MTurk. Finally, the limitations and benefits specific to MTurk are discussed.

This thesis examined the effect of project description on MTurk response patterns. The purpose of this study was to determine whether choice of keywords in the project description influences research results via self-selection of subjects. Researchers have the option of including a title and keywords for each study broadcasted via MTurk, and thus, understanding the importance of project description is crucial to developing an understanding of this data collection tool.

LITERATURE REVIEW

Historical Perspective

Online research is a recent development in psychology. Psychologists began to turn to this approach in the mid 1990's (Birnbaum & Reips, 2005). Although online research is a new modality for investigation, methods in psychology have experienced many changes in the past. In the late nineteenth century, psychology was establishing itself as a scientific discipline separate from philosophy. To do this, psychologists borrowed methodology from other fields. Much of this early work focused on fundamental cognitive processes and used some of the methods in physiology to investigate topics such as perception, sensation, and memory (Goodwin, 2003). Measures such as reaction time and difference thresholds became the primary tools for psychologists of the time period. To gather these measures, intricate machines were often needed. These were frequently made of brass, which is why the term "brass instrument" psychology is often used to refer to this period (Goodwin, 2003). Data were typically collected from small numbers of subjects.

Psychologists in the late nineteenth century also aligned themselves with structuralism, the exploration of internal structures of the mind. This led to the adoption of the introspective approach (Goodwin, 2003). Instead of taking physiological measurements, subjects and experimenters would spend time reflecting on their own thoughts and experiences. In some cases, this was done with self-report measures such as questionnaires. G. Stanley Hall was the strongest American advocate for the use of questionnaires in psychological research, and self-report measures are still commonly

used. Questionnaires could be administered to large numbers of people and did not require the use of various apparatus.

Another shift in methodology came with mental testing. This movement stemmed from the work of Binet, Galton, and Cattell. Binet's achievement was the development of the intelligence test now known as the Stanford-Binet (Siegler, 1992). Galton's approach was unique in that he saw intelligence as one of many traits that varied in humans and considered it to be a determinant of survival. Cattell applied his knowledge of brass instrument psychology to the field of mental testing (Goodwin, 2003).

Psychologists have applied the changing technologies to survey research. The development of the Likert scale and the polling of public opinion influenced psychology. As the telephone became widespread, in the 1940s, researchers began to use telephone surveys, and this approach became common in the 1960s (Dillman, Smyth, & Christian, 2009). Telephoning subjects allowed for more diverse and representative samples, and, in some cases, could expedite the process of data collection. Telephone surveys can often be used in combination with other modes, such as in-person surveys and mailed surveys. Mailing is a sampling method that has been used from the beginning of questionnaire research (Goodwin, 2003).

Another way in which psychologists have taken advantage of technology is the adaptation of computerized tests and measures. Psychologists began to computerize paper-and-pencil tests in the 1960's (Gilberstadt, Lushene, & Buegel, 1976; Lushene et al., 1974). Initially, computers were implemented in the scoring procedures for measures like the Minnesota Multiphasic Personality Inventory (MMPI) (Graham, 2006) and the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 2008). The next development was

to completely computerize the administration of these measures (Gilberstadt et al., 1976; Lushene et al., 1974).

It was found that computerized versions of the MMPI correlated highly with paper-and pencil administrations (Lushene et al., 1974). Computerized intelligence testing, more specifically, subscales of the WAIS, were also comparable to the other forms of administration (Gilberstadt et al., 1976). The computerization of tests continued into the 2000's. In one study (Williams & McCord, 2006), scores on the Raven Standard Progressive Matrices test (RSPM) were compared between subjects who were randomly placed in one of four conditions that mixed computerized versions and standard versions of the RSPM. Scores on both versions of the RSPM were found to be comparable, and there were no significant differences in anxiety between the standard or computerized versions.

Dillon (1992) and Noyes and Garland (2008) suggest that some aspects of performance may not transfer when tests or assessments are computerized. However, the results discussed above suggest that these threats are minimal. Further, these effects may diminish in those who use computers often.

Conducting Psychological Research Online

Psychologists today are conducting much of their research online. There are multiple benefits of using the Internet for surveys (Dillman et al., 2009). One is the high rate at which data can be collected. The adoption of online research is the next step in research methodology for psychologists.

Many studies have used online samples. For example, Johnson (2005) assessed the validity of personality measures using a web sample. His participants were not

actively recruited, but found the website on their own. Online participants were as consistent on personality measures as paper and pencil respondents. The author concluded that online personality measures are valid.

In an ambitious project, Ekman, Dickman, Klint, Weiderpass and Litton (2006) gathered data from over 96,000 Swedish women to determine whether epidemiological studies could be carried out online. These women were selected from a population registry and randomly assigned to groups with different types of reminders (email or mail) and variations in the options for completing the survey (paper and/or online). Of the 70% of women who responded, 41% of them did so online. However, women who responded online were more educated and less likely to smoke. The group with the lowest response was women who received a reminder by mail but were only given the option of responding online.

Kendler, Myers, Potter and Opaleskyl (2009) used an Internet sample to investigate personality, psychopathology and substance use in twins, relatives and non-relatives. Their sample was taken from the same website as the Gosling, Vazire, Srivastava, and John (2004) sample (www.outofservice.com); while the Gosling et al. sample was quite diverse, the Kendler et al. sample was largely female and white. It is interesting that large differences in sample demographics can be obtained using the same website.

Online data collection sometimes is turned to as a last resort. For example, Harrison and Christie (2004) studied young drivers. They contacted local government and Catholic schools to recruit participants. Executives at the schools were hesitant to commit to this study because of the extra paperwork it would cause them. Not wanting to

abandon their research, the authors turned to the Internet and successfully recruited 30 respondents.

The Validity of Online Research

Psychologists are often criticized for conducting studies on narrow and non-representative groups of people. More specifically, psychologists are criticized for their reliance on convenience samples of college subject pools. Sears (1986) found that 70% of the social psychology studies conducted in 1980 used undergraduates as subjects. This is problematic as people at this stage in life have a less cohesive sense of self and do not have fully developed attitudes. Generalization to the larger population is also an issue. The Internet offers a potential solution to the generalization problem because of its worldwide access to participants.

Some researchers have investigated the validity of online research. To do this, researchers compare data gathered from Internet samples to data gathered from more traditional samples, such as college subject pools. Gosling et al. (2004) used a website to obtain a large sample (N = 361,703) of self-selected participants. Demographic and personality data were then compared to data from 510 traditional samples in published research. The researchers found that:

- Internet sample are more diverse than traditional samples
- Internet subjects are not maladjusted or prone to depression
- Personality measures replicate, regardless of testing medium
- Internet subjects are motivated
- Findings of Internet research are consistent with traditional findings

One preconception was confirmed: the increased anonymity creates problems with Internet samples. With increased anonymity comes a higher likelihood for repeat responders. However, the researchers point out that steps can be taken to decrease repeat responders. Ultimately, Gosling et al. (2004) concluded that Internet samples are more diverse, though not necessarily more representative of the general population. In addition, Internet data are at least as valid and reliable as traditionally gathered data.

A study comparing the psychometric equivalency of web-based research to previously collected data by Meyerson and Tryon (2003) supports this conclusion. Meyerson and Tryon administered the Sexual Boredom Scale online, along with corresponding validation scales. Respondents to this survey were matched based on demographic information and a subsample matched to the existing data was created. Structural equation modeling was used to compare these subjects to existing data. The two samples had the same results, and the authors concluded that data collection via the Internet was valid, reliable, efficient, cost effective, and reasonably reliable.

Berrens, Bohara, Jenkins-Smith, Silva, and Weimer (2003) compared Internet samples to phone samples to determine whether political research could be carried out over the Internet. Their analysis showed no major differences between the Internet and phone sample in environmental attitudes. Negative associations between environmental attitudes and perception of threat caused by climate change were found in both samples. While the Internet sample was slightly more liberal than the phone sample, this did not affect their responses.

Graham et al. (2006) studied whether the Internet could be used to conduct smoking cessation programs. Participants first received telephone counseling and

assessment in the form of an interview. Subjects who consented to participate in the follow-up online study were retested using the same assessments. The test-retest reliabilities for the subjects who consented to the online follow-up were high, and the researchers concluded that the Internet offers a reliable mechanism for carrying out and assessing smoking cessation programs.

Lewis, Watson and White (2009) compared responses to messages about drunk driving between a sample completing a paper-and-pencil survey and a sample taking the survey online. All participants were exposed to the same messages about drunk driving. The mean change in attitude scores was the same for the two groups after exposure to drunk driving messages. There were, however, differences in the demographics of these two samples. First, more females were in the paper-and-pencil samples. Second, the Internet sample had more respondents over the age of 55, and fewer under the age of 24. Despite these differences, the authors considered the Internet sample to be more representative of the general population than the paper-and-pencil samples.

Chuah, Drasow, and Roberts (2006) gathered personality data using Internet and paper-and-pencil samples. The researchers used Goldberg's Big Five personality inventory and included an in-person computer lab condition. All participants were undergraduates randomly assigned to one of the three conditions: paper-and-pencil, computer lab, and Internet. No major differences were found between the groups. The authors concluded that the three testing modalities were equivalent.

Studies of obsessive-compulsive tendencies have also been carried out over the Internet and compared to paper-and-pencil samples. Coles, Cook and Blake (2009) administered the Obsessive-Compulsive Inventory and gathered demographic

information for both samples. There were no significant differences between the two samples. The authors note that older computers may not be able to display the information as the researchers intended.

Another method for testing the reliability of Internet samples is to replicate previous research not conducted online. Vadillo, Bárcena, and Matute (2006) investigated whether associative learning tasks could be carried out online. The researchers obtained the same effects as the original studies, but also concluded that web studies were a compliment to traditional methods rather than a replacement. Rueckert (2005) investigated whether cerebral asymmetry tasks could be carried out on the Web. She found that performance in online experiments on a chimeric face task, which is used to tease out differences in the two hemispheres of the brain, was similar to performance in the laboratory.

Benefits of Online Research

A picture of the benefits and problems with online research has emerged. Web studies offer faster responses, fewer response errors, and cost less (Dillman et al., 2009; Lyons, Cude, Lawrence, & Gutter, 2005). Online research reduces experimenter bias, and the computer allows for more control over the stimulus materials (Spyridakis, Wei, Barrick, Cuddihy, & Maust, 2005). Web studies also provide increased anonymity and afford the possibility of recruiting hard to reach populations (Berrens et al., 2003; Harrison & Christie, 2004). Electronic measures reduce data entry and allow for direct exporting into statistical software packages (Rademacher & Lippke, 2007). The need for paper copies of measures is also eliminated. Finally, samples are often larger (Ekman et

al., 2006; Reimers, 2007) and comparable with traditional samples (Gosling et al., 2004; Lewis et al, 2009).

Limitations of Online Research

There are limitations of online research. First, there is lack of control of testing environment, as subjects are not coming into a lab (Buhrmester et al., 2011; Skitka & Sargis 2006). This lack of control can create problems with the responders. Subjects may be tempted to complete the survey more than once, or they may fail to take the survey seriously and respond inconsistently. Second, online participants may come from outside the United States. While this is beneficial for conducting cross-cultural research, it can be problematic when researchers would like to focus on specific populations. In one study, over 10% of subjects came from a non-English speaking country (Reimers, 2007). A third limitation is that participants might lie. Although participants may lie in any study, web studies are less personal and more detached, increasing the possibility of deceptive responding. Individuals may take on new identities online, contributing to deceptive responding (Skitka & Sargis, 2006).

Demographic differences may also produce limitations. Skitka and Sargis (2006) found that Internet users are typically younger, wealthier and more educated. They also propose that differences in technical capabilities among computers may influence the way in which information is displayed. Harrison and Christie (2004) also discuss problems users could have downloading or uploading information and the rate of dropouts in online studies. Rueckert (2005) suggests that the likelihood of dropping out may be increased by the lack of social pressure exerted on the subject. The way in which participants respond may also be problematic. Subjects may be less motivated (Vadillo et

al., 2006), may answer inconsistently (Reimers, 2007) or may answer in a way that is inattentive, careless, or hurried. Further, subjects may respond more than once (Johnson, 2005). Manipulation checks can be built into online studies to counter this problem (Oppenheimer, Meyvis, & Davidenko, 2009), but this requires an additional step in the survey construction process.

An Overview of MTurk

An increasing number of psychologists are turning to Amazon's Mechanical Turk (MTurk) for online data collection (Paolacci, Chandler, & Ipeirotis, 2010). This is a free service provided by Amazon where a potential user can set up two types of accounts. The first type is the Worker account. As a Worker, an individual can log on and choose from various Human Intelligence Tasks (HITs) (surveys, experiments, photo categorization, etc.), which they complete for a monetary reward. These HITs typically consist of tasks that are not so difficult that they require specialization, but would be impossible for a computer to carry out. The type of HIT found on MTurk varies, but a substantial proportion of HITs are devoted to academic research. HITs are usually simple and take a small amount of time to complete. Rewards are also typically small, usually about 15 cents for a 15-minute survey. Despite this fact, many Workers complete thousands and even hundreds of thousands of surveys, and some rely on MTurk for their income (Ipeirotis, 2010).

In the second account, (Requester account) Requesters post HITs and hire the Workers. Requesters typically use a credit card account to pay for their study, and Amazon reimburses the Workers either through their Amazon account, or through a direct deposit (if the Worker is a resident of the United States). Well-known Requesters

have built a reputation on MTurk (e.g. Dolores Labs, Channel Intelligence, Casting Words), while others only use MTurk occasionally.

Researchers sign up for a Requester account. This grants them access to built-in survey design and monitoring tools and gives the Requester the option of posting HITs using external links and survey software such as Qualtrics or SuveyMonkey. The Requester account grants the user access to the large group of potential Workers. Once this account is set up, multiple HITs can be posted, allowing for simultaneous data collection. Requesters can select the number of Workers needed, and once this number has been reached, the HIT will automatically be taken down. Thus, the data are easily collected and analyzed, making MTurk a one-stop data shop for researchers.

An on-line community has developed around MTurk. Although there are no face-to-face interactions between the Workers or between Workers and Requesters, there is communication on message boards dedicated to MTurk users, or "Turkers." This communication is primarily between workers, but can be between Workers and Requesters. More about the MTurk community is included in the discussion section of this thesis.

Personal Experience as a Turker

Prior to starting this project, I spent time on MTurk completing HITs over two weeks. I completed 30 HITs and made nine dollars in the process. Immersing myself in the world of MTurk allowed me to gain experience and shaped this project. At first, completing HITs on MTurk (or Turking) was daunting. A great deal of information was presented to me at once. Each HIT has a title and keywords (see Appendix A). This information becomes quite helpful in deciding which HITs to take. I also had to learn

which HITs are good (pay the most for effort) and which ones to avoid (scams or HITs that require more work than they are worth). Once I had developed this "MTurk sense," Turking became easier and faster.

MTurk can be addictive. Once I realized that I could complete HITs in a timely manner, my focus shifted toward completing as many HITs as possible. When I made this step, Turking became mindless. I began to spend little time reading descriptions and instructions in order to further expedite the Turking process. It became easy to sit at the computer Turking for great lengths of time, and at about 10 cents per HIT, I almost had to do this to make any money. This approach is beneficial for the Turker, but detrimental to the Requester.

Demographics Specific to MTurk

The demographics of Turkers are important. Initially, most Turkers were from the United States. Now Americans only constitute about half of the MTurk population (Ipeirotis, 2010). While Americans are still the largest group, Indians now constitute 34% of the Turkers compared to 49% for Americans. Amazon recently allowed Turkers to be paid in rupees (Ipeirotis, 2010). Other sources show that a majority of Turkers live in India (46%), as opposed to the United States (39%) (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010).

Most Turkers in the United States are female, but the opposite is true of Turkers in India. Turkers in India tend to be younger than American Turkers (Ipeirotis, 2010). There are conflicting findings over education and income. In both countries, Turkers seem to be more educated than the general population, but they have lower incomes than the general population. Indian Turkers report significantly lower incomes than their

American counterparts (Ipeirotis, 2010). Turkers in both countries are predominantly single and do not have children. More Turkers in India report being unemployed and rely on MTurk for income. While MTurk is more diverse than the traditional college sophomore sample, we may be trading one flawed sampling modality for another.

Benefits of MTurk

MTurk offers many of the benefits as other online research outlets. Some studies indicate that Turkers are more diverse than college subject pool samples and other Internet samples (Buhrmester et al., 2011; Kendler et al., 2009). It is claimed that MTurk data is as reliable as data from traditional methods (Buhrmester et al., 2011). Thus, Buhrmester et al. conclude that MTurk is a viable means for quickly and inexpensively gathering quality data. This conclusion is supported by Chilton, Sims, Goldman, Little, and Miller (2009). The researchers posted games of rock-paper-scissors and other cooperative games on MTurk and received responses from 22 human pairs of players in less than 4 minutes. What is interesting about this study is that it required real-time interaction between Turkers. If cooperation like this can occur online, it seems likely that this a group of people who are completing tasks in an effortful manner.

Buhrmester et al., (2011) also determined that many of the Workers on MTurk are self-motivated and enjoy the tasks that they complete online. If Turkers were self-motivated, one would expect that they are providing the researcher with quality data and that they are focusing on the task at hand.

Limitations of MTurk

Previous research Christopher Cooper and I conducted on MTurk supports many of these limitations (Holden & Cooper, n.d.). Our subjects provided inconsistent

responses and used terminology such as “nil” rarely encountered in the United States. Despite the survey supposedly being limited to Turkers in the United States, this suggests that subjects were coming from other countries. Subjects also completed surveys much faster than expected. In some cases, participants completed 15-minute surveys in less than 5 minutes. Further limitations will be included in the discussion section of this thesis.

Keywords on MTurk

MTurk screens are laid out like forums and message boards. Often, so much information is displayed that it is easy to become distracted (Appendix B). Workers must wade through the pages of HITs. There are filters, such as sorting by highest paid HITs, to aid in this process. It is, however, the keywords and titles that are most important in helping a Turker decide which HITs to take. No matter how they are sorted, each HIT appears in its own box in which most of the space is dedicated to the title. If worded properly, a title can provide a Turker with the all the necessary information. The keywords help to clarify and expand upon this. HITs can also be searched for by their keyword, so Requesters often use some of the same words in the title and keywords, increasing the likelihood that Turkers will find and take their HIT. Topical words are typically used, as are words describing the type of HIT and the length of time needed to complete the hit (Buhrmester, 2010).

PURPOSE

This study investigated the potential bias caused by keywords on MTurk. Keywords and titles of HITs on MTurk are crucial in the recruitment of Turkers. Therefore, it seems likely that this information can skew results via recruitment bias. For example, Turkers with different attitudes or personalities might be more likely to participate in different studies based on titles or keywords. Previous work has shown that the wording used in hyperlinks influences the responses people have to the links (Wei et al., 2005), and hyperlinks act in much of the way as keywords. To test this, a series of manipulations of wording were carried out.

These wording manipulations were carried out at the keyword level, with the title being held constant. This approach was chosen to amplify the effects of the keyword manipulation on the types of subjects that complete the surveys. The title "20 Minute Survey on Personality and Attitudes" was used across four conditions. The title was associated with each HIT on MTurk, and the keywords appeared below the title after the subjects clicked on the HIT. The keywords were manipulated across four conditions. In the first condition, neutral keywords were used. In the second condition, keywords related to sexuality were used. The final two conditions contained keywords associated with political ideology. One condition contained keywords related to conservative political ideology, and the other, liberal political ideology. This keyword manipulation was the independent variable of this study.

Subjects across the four conditions completed demographic information, as well as three scales. The first scale was a political ideology scale, which was followed by a

five-factor model of personality scale, and finally a scale assessing the subjects' homophobia. The political ideology scale and the homophobia scale were chosen because of their relation to the keyword manipulations. The personality scale assessed the effect of personality on task selection. Specific hypotheses about the effects of the manipulation are discussed below.

Hypothesis 1

A revised version of the Wilson-Patterson (Wilson & Patterson, 1968) scale, a measure of political values, was used. It was hypothesized that the wording would draw different subsamples of Turkers. The conservative keyword condition would attract more conservative Turkers, and more liberal Turkers would be attracted to the liberal keyword condition.

Hypothesis 2

The Homophobia Scale (Wright et al., 1999) measures negative attitudes toward gays and lesbians. It was hypothesized that Turkers who are familiar with, or are involved in the Lesbian, Gay, Bisexual, and Transgender (LGBT) community, would have lower scores on the Homophobia Scale. Turkers searching for sexuality related keywords will be less homophobic than others searching for more neutral or politically related keywords.

Hypothesis 3

Based on the findings of Buhrmester et al. (2011) it was hypothesized that the personality data will remain consistent between groups. A full explanation of these groups and the wording used is provided in the methods section. The M5-120 (Johnson, 2001) was used to assess personality.

METHOD

Subjects

Subjects were recruited through MTurk and were limited to Turkers from the United States. None reported being less than 18 years of age. Subjects participated in a Human Intelligence Task (HIT) that varied only by keywords. In all, 266 subjects were recruited. Subjects were paid 25 cents for completing the questionnaire. This level of remuneration is typical of HITs of this type on MTurk.

Materials

An online questionnaire consisting of four parts was administered to subjects. This questionnaire was presented using Qualtrics survey software. A unique link to the questionnaire was generated using Qualtrics, and was included on the HIT posted to MTurk. Subjects completed all four measures on Qualtrics after reading a consent form (Appendix E), and were only required to insert a random number generated in the questionnaire to the HIT on MTurk. The first part of this questionnaire included a basic demographics form, followed by the political ideology scale, the M5-120 personality inventory, and the Homophobia Scale.

Demographics. Subjects first completed nine demographic questions: the year they were born, their gender, level of education, annual income, their state of residence, their race/ethnicity, and their country of origin. Finally, subjects were asked to identify the criteria they used to find the HIT. If subjects indicated that they used a keyword search to find the HIT, they were asked to include the words they used in the search. The

complete listing of demographic questions, as they were administered appears in Appendix F.

Political ideology scale. Political ideology was measured with revised version of a Wilson-Patterson Scale (1968) developed by Oxley et al. (2008) (see Appendix G). Two ambiguous items were removed from the Oxley et al. (2008) scale, leaving 26 items, each with three response options (Agree, Disagree, Uncertain). These 26 items were grouped into either a liberal or conservative category (13 items for each category). This produced a range in scores of 0 (Liberal) to 52 (Conservative).

M5-120. The third part of this questionnaire was the M5-120, a Big-Five personality scale developed by Johnson (2001). This 120-item scale consists of five-point Likert-type items corresponding with each of the five personality factors. The M5-120 is listed in Appendix H.

Homophobia scale. The Homophobia Scale (HS) measures negative attitudes toward gays and lesbians (Wright et al., 1999). This is a 25-item scale with 5-point Likert type items that has three subscales (Behavior/Negative Affect, Affect/Behavioral Aggression, Cognitive Negativism). This scale was included because it has been validated, and comparisons can be made to the original dataset. The HS is provided in Appendix I.

Procedure

Eight keywords associated with the HITs on MTurk were manipulated. The title of each HIT was held constant for each condition to heighten the effect of the keyword manipulation. The title was "20 Minute Survey on Personality and Attitudes." Keywords were manipulated to fit into the neutral, sexuality related, conservative or

liberal condition. Four keywords were held constant throughout the conditions to further heighten the effect of the manipulation. This keyword manipulation acted as the independent variable and had four conditions.

Neutral control condition. Keywords in this condition were modeled after suggestions in Buhrmester's "Amazon Mechanical Turk Guide for Social Scientists" (2012), and were designed to attract an unbiased sample of Turkers. The following keywords were used: *survey, experiment, questionnaire, and 20-minute*. Four keywords were included in this condition that were repeated in the following three conditions. These keywords were: *research, fast, easy, and psychology*.

Sexuality related condition. For this condition, the following keywords were used: *sexuality, homosexuality, gay, and lesbian*. It is likely that Turkers searching for these keywords are familiar with people of varying sexualities, and thus would be less homophobic than those searching for other keywords. The four repeated keywords were used in this condition to provide subjects with information about the questionnaire.

Conservative condition. Keywords in this condition aligned with conservative political ideology. These keywords were: *republican, conservative, pro-life, and small government*. It was hypothesized that Turkers in this condition would have conservative ideology. The four repeated keywords were included in this condition.

Liberal condition. In the final condition, keywords were used that highlighted liberal political ideology. These keywords were: *liberal, democrat, pro-choice, and big government*. It was hypothesized that Turkers in this condition would have liberal ideology. This condition also included the four repeated keywords.

Posting of HITs. Each of the four conditions (sexuality related keywords, conservative keywords, liberal keywords, and neutral keywords) were posted on MTurk as separate HITs. These HITs were published sequentially and were kept active until the desired number (50) of subjects had been reached. One HIT was taken down before the next was posted. The sequence followed the order of conditions; neutral keywords were used first, followed by sexuality keywords, conservative keywords, and finally liberal keywords. The data collection process lasted approximately two months.

Subjects were told that the HIT was to be posted multiple times, at regular intervals and were asked not to participate if they have already participated in a previous condition. Subjects were further warned that if they did participate in more than one condition, they would not be rewarded for their participation. Responses from repeat responders were removed from analysis, and were not paid. IP addresses of each respondent were used to identify repeat responders. IP addresses are typically associated with a single computer, but can be used by multiple computers on the same connection. Therefore, a comparison of reported demographic information was made in the case of repeated IP addresses.

Manipulation checks. Manipulation checks were used to assess the attentiveness of responders (Oppenheimer et al., 2009). These manipulation checks mimicked items on the questionnaire, and were placed periodically throughout the survey (Appendix J). For example, one manipulation check read: “If you are reading this, please answer inaccurate.” Three manipulation checks were used. Responses on the three manipulation checks were combined to provide an overall pass or fail rating. Having two out of three correct responses was considered passing.

In all, 260 responses were collected. Seven matching IP addresses were identified. (Six of the subsequent attempts from these IP were deleted, as there was a variance in the demographic information provided by the seventh matching IP address.) Seventeen repeated Worker IDs were identified. One Worker completed the survey three times, generating 18 deleted responses. Responses from 24 participants were excluded from data analysis after failing manipulation checks. An additional 19 responses were deleted, because these participants provided incomplete data. This generated a final sample size of 193 (a total loss of 26%). In the analysis, there were 51 subjects in the neutral control condition, 49 in the sexuality related, 52 in the conservative, and 41 in the liberal keyword condition.

RESULTS

Political Ideology

A one-way between subjects analysis of variance was used to assess the effect of keywords on the political ideology scale scores. There was no significant difference between the groups in political ideology: F (3,155) = 1.75, p = .16. Thus, hypothesis 1, that choice of keywords will influence political ideology scores, was not supported. Means and standard deviations for total political ideology by keyword condition are included in Appendix A.

Homophobia

A one-way between subjects analysis of variance was used to assess effects of keywords on the homophobia scale scores. There was no significant difference between the groups in homophobia: F (3,159) = .21, p = .89. Thus, hypothesis 2, that choice of keywords will influence homophobia scores, was not supported. Means and standard deviations for homophobia by keyword condition are in Appendix A.

Personality

A one-way between subjects MANOVA was conducted to assess differences between keyword groups in personality. Each of the five personality factors generated by the M5-120 (Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience) were the dependent variables. There was no significant overall effect for the multivariate model [F (15,508)= .91, p =.557, Wilk's lambda =.93, partial eta squared = .02]. Hypothesis 3, that choice of keywords will not influence responses on personality scales, was supported. The keywords did not attract people of significantly different

personalities. Means and standard deviations by condition of the five personality factors are shown in Appendix B.

Keywords Searched

In all, 40 subjects reported searching by keyword to find the HIT placed on MTurk. Twenty-six subjects (65%) reported searching for the keyword *survey*. Four subjects reported searching for the keyword *sex*, a close match to the *sexuality* keyword. A similar case was seen with the keyword *minute*, which was searched for once, and contains a portion of the keyword *20-minute*. The keywords *easy* and *fast* were both searched for three times by different subjects. *Research, psychology,* and *gay* were each searched for once by different subjects. Different subjects searched for the keywords, *quick,* and *personality* once, but these do not match the keywords provided.

DISCUSSION

As revealed by the lack of significant differences between conditions in political ideology and total homophobia, keywords associated with a HIT on MTurk have no effect on the samples of subjects drawn, and thus the responses on the dependent measure or measures. A similar finding was found for personality. Different keywords do not draw Turkers with different personalities.

Implications of Current Findings

The current research was designed to test whether HIT keywords affects MTurk samples. Keywords were manipulated across four conditions: neutral keywords, sexuality keywords, liberal keywords, and conservative keywords. There were no significant differences in scores on the three dependent measures (M5, HS, and PIS). This finding suggests that choice of keywords for a HIT did not affect responses. The lack of differences in scores between conditions suggests keywords do not adversely affect survey outcomes.

Only 40 out of 193 subjects (20.7%) searched for keywords to find the HIT. Of those who searched for keywords, a majority reported searching for a keyword in the neutral control condition (*survey*), with five reporting searching for keywords in the sexuality related condition (*gay* and *sex*). These findings suggest that while Turkers may search by keyword to find HIT, they are more concerned about the type of HIT than the subject matter of the HIT.

Limitations of the Study

The keywords used for this study were related to the items on the questionnaire, regardless of condition. Although neutral keywords were used in one condition, it could be argued that words such as "psychology" are related to the materials. Future research could investigate the effects of keywords that are not matched with the survey materials. Another option would be to investigate the effects of not including keywords on HIT responses. This would be a particularly stringent test of the importance of keywords.

Other limitations to this study could be related to the way in which the HITs were posted. In order to maintain the "freshness" of the HITs for this study, a "batching" approach was used. It is hard to determine whether this procedure expedited data collection, but it made organization and management of the data cumbersome. Because MTurk produces a data file for each HIT, multiple files had to be placed together and labeled to analyze the data in aggregate. Along with increased data management demands, this batching procedure may have flooded the MTurk HIT market, and deterred some Turkers from completing the HITs.

Finally, posting batches of HITs increases the occurrence of repeat responders (Rand, 2011). Repeat responders can be identified by their Worker IDs, demographic information, and IP addresses if a program such as Qualtrics is used that can log IP addresses. These procedures were used to identify repeat responders in the sample, and the responses from 23 participants were removed from analysis. Posting one HIT per condition would have eliminated this issue. Further, IP addresses should be crosschecked against Worker IDs and demographics, as the IP addresses are an imprecise tool for identifying individual participants. The IP address could be shared among family members and housemates, or in public places such as business complexes and libraries.

What's Good About MTurk?

MTurk provides a more diverse sample than otherwise obtained in college sophomore samples, and valid data can be obtained rapidly and inexpensively (Buhrmester et al., 2011), with proper parameters. However, Buhrmester et al., as well as other sources (Mason & Suri, 2011), suggest that MTurk is not representative of any one population and may not be representative of the Internet population. Payments on MTurk are less than what would typically be given to participants in live experiments. Furthermore, the experimenter does not have to directly interact with the participants, and multiple participants can complete tasks simultaneously. This inability to directly interact presents some limitations, which will be discussed in the section below.

One benefit of MTurk is the ease with which data can be managed. Surveys built within MTurk allow for data to be exported directly to Excel. Many external survey design programs (Qualtrics, SurveyMonkey) also allow data to be exported to Excel or SPSS. This eliminates the need for manual data entry and eradicates data entry errors. Although the data entry process is eliminated, researchers will very likely have to spend time cleaning the data.

Limitations of MTurk

While there are benefits to MTurk, there are limitations. Research on MTurk lacks control otherwise offered in lab settings (Buhrmester et al., 2011, & Rand, 2011). While this helps to ameliorate experimenter bias (Spyridakis et al. 2005), some research designs require experimenter presence. These experiments would not be possible on MTurk. The experience of each participant also cannot be controlled (Coles et al., 2009, & Skitka and Sargis, 2006). Participants may also decide to complete HITs in noisy

environments, or could have uncontrolled distractions. They may also decide to stop at any point in the survey (Rand, 2011) or pause and shift between the HIT and other tasks, diminishing their attention. Therefore, when environmental control is a crucial design feature, researchers would be wise to pursue other options.

Behavioral observations or physiological measures would also be impossible on MTurk (Rand, 2011). Presentation of visual information such as pictures and videos can be accomplished on MTurk, but the experience of each participant with this information is likely to vary. Finally, studies requiring follow up contact with the participants may be difficult. The researcher would have to rely on email, or the systems within MTurk, to contact participants for follow-up. Multiple projects have shown this to be possible (Buhrmester et al., 2011, & Holden, Dennie, & Hicks, n.d.), but only a portion of the total respondents are likely to agree to follow-up. In one study (Holden et al., n.d.), a sample of 281 participants was recruited, but only 67 agreed to participate in the follow-up. Both studies (Buhrmester et al., 2011, Holden et al., n.d.) obtained strong test-retest reliability. This strong test-retest reliability suggests that the same individuals are completing the survey. However, if an MTurk account were to be shared between two or more persons, it is possible for the follow-up data to be inaccurate.

Culturally sensitive measures may not translate if the HIT is broadcasted internationally. This could cause items to be misinterpreted. Data can be skewed in other ways as well. For example, inattentive responders can provide bad data (Osborne & Blanchard, 2011), and without the presence of an experimenter, participants are more likely to be tempted to coast through the survey, or multitask as they complete HITs. In

fact, participants in one study reported completing HITs in their spare time when they may be engaged in other activities, such as watching TV (Ipeirotis, 2010).

Finally, there may be a difference in results obtained between universities using MTurk. For example, schools with more funding can afford to pay more Turkers compared to schools with less funding (Holden et al., 2011). A larger sample size can increase statistical power (Hinkle, Wiersma, & Jurs, 2003) and lessen the impact of responses lost to data cleaning. Furthermore, it is recommended that researchers collect data from a large sample of Turkers or to have multiple Turkers complete the same task in order to increase accuracy ("Cooking Tip #5", 2012). While this is sound advice, such practices may not be feasible at smaller universities because of funding limitations. For example, a small university may only be able to pay 200 workers to complete a HIT. If the researchers had to delete roughly 25% of their data (as was the case in this thesis), they would experience a loss of 50 subjects, and would be unable to gather more data from MTurk to counter this loss. This effect would be amplified in surveys that take longer to complete, or if Turkers were to demand higher pay.

Conclusions

The lack of significant differences between conditions on the political, sexual attitudes, and personality measures suggests that choice of keywords does not influence responses on questionnaires by drawing in different samples of people. MTurk is still in its beta version, and will continue to change. Researchers may also develop new techniques for MTurk research, making their findings stronger and more reliable. It is also possible that the use of MTurk will decline and disappear in the next few years. Competing online survey websites have already emerged such as SocialSci and YouGov

(https://www.socialsci.com/, & http://corp.yougov.com/scientific-research/), both of which are more tailored to scientific research, and do not contain the commercial aspects seen on MTurk. These sites will continue to develop and psychologists will need to be aware of these developments.

REFERENCES

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C., & Weimer, D. L. (2003). The advent of Internet surveys for political research: A comparison of telephone and Internet samples. *Political Analysis*, *11*(1), 1-22.

Birnbaum, M. H., & Reips, U. D. (2005). Behavioral research and data collection via the Internet. In R. W. Proctor, & K. P. L. Vu (Eds.), *Handbook of human factors in Web design* (pp. 471-491). Mahwah, NJ: Erlbaum.

MD Buhrmester. (2012, January 18). Amazon Mechanical Turk guide for social scientists [Web log post]. Retrieved from http://homepage.psy.utexas.edu/homepage/students/buhrmester/MTurk%20Guide.htm

Buhrmester, M. D., Kwang, T., & 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.

Chilton, L. B., Sims, C. T., Goldman, M., Little, G., & Miller, R.C. (2009). Seaweed: a Web application for designing economic games. *Proceedings of the Human Computation Workshop.* Paris: ACM, June 28, 2009.

Chuah, S. C., Drasow, F., & Roberts, B. W. (2006) Personality assessment: does the medium matter? No. *Journal of Research in Personality*, *40*, 359-376.

Coles, M. E., Cook, L. M., & Blake, T. R. (2007) Assessing obsessive compulsive symptoms and cognitions on the Internet: Evidence for the comparability of paper and Internet administration. *Behavior Research and Therapy*, *45*, 2232-2240.

Cooking tip #5 ask multiple Workers to complete a HIT. (2012, January, 26). Retrieved from http://mechanicalturk.typepad.com/blog/tips-tricks/.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: the Tailored Design Method* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Dillon, A. (1992). Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics,* 35, 1297-1326.

Ekman, A., Dickman, P. W., Klint, A., Weiderpass, E., & Litton, J. (2006) Feasibility of using Web-based questionnaires in large population-based epidemiological studies. *European Journal of Epidemiology*, *21*(2), 103-111.

Gilberstadt, H., Lushene, R., & Buegel, B. (1976). Automated assessment of intelligence: The TAPAC test battery and computerized report writing. *Perceptual and Motor Skills, 43,* 627-635.

Goodwin, C. J. (2003). Psychology's experimental foundations. In S.F. Davis, *Handbook of research methods in experimental psychology* (pp. 3-23). Malden, MA: Blackwell Publishing Ltd.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, *59*(2), 93-104.

Graham, A. L., Papandonatos, G. D., Bock, B. C., Cobb, N. K., Baskin-Sommers, A., Niaura, R., & Abrams, D. B. (2006). Internet vs. telephone-administered questionnaires in a randomized trial of smoking cessation. *Nicotine & Tobacco Research*, *8*(1), 49-57.

Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology.* New York: Oxford University Press.

Harrison, W., & Christie, R. (2004). *Using the Internet as a survey medium: Lessons learnt from a mobility survey of young drivers.* Road Safety Research Policing and Education Conference, Perth, Australia.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences*. (5th Ed.). Boston, MA: Houghton Mifflin Company.

Holden, C. J. & Cooper, C. A. A cautionary tale: Reassessing the efficacy of MTurk for social science research. Manuscript submitted for publication.

Holden, C.J., Dennie, T., & Hicks, A.D. The validity of the M5-120 on Amazon's Mechanical Turk. (Available from the author at: cjholden2@catamount.wcu.edu).

P Ipeirotis. (2010, March 9). The New Demographics of Mechanical Turk [Web log post]. Retrieved from http://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html.

Johnson, J. A. (2001, May). *Screening massively large data sets for non-responsiveness in web-based personality inventories*. Invited talk to the joint Bielefeld-Groningen Personality Research Group, University of Groningen, The Netherlands.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103-129.

Kendler, K. S., Myers, J., Potter, J., & Opaleskyl, J. (2009). A Web-based study of personality, psychopathology and substance use in twin, other relative and relationship pairs. *Twin Research and Human Genetics*, *12*(2), 137-141.

Lewis, I., Watson, B., & White, K. (2009). Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participant responses to health-related messages. *Australian Journal of Psychology*, *61*, 107-116.

Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. *Journal of Personality Assessment, 38,* 353-361.

Lyons, A. C., Cude, B., Lawrence, F. C., & Gutter, M. (2005). Conducting research online: Challenges facing researchers in family and consumer sciences. *Family and Consumer Sciences Research Journal*, *33*(4), 341-356.

Mason, W., & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods.* Doi: 10.3758/s13428-011-0124-6.

Meyerson, P., & Tryon, W. W. (2003). Validating Internet research: A test of psychometric equivalence of Internet and in-person samples. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 614-620.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent?. *Ergonomics,* 51(9), 1352-1375.

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.

Osborne, J. W., & Blanchard, M. R. (2011). Random responding from participants is a threat to the validity of social science research results. *Frontiers in Psychology*, *1*, 1-7.

Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller J. L., Scalora, M. J., Hatemi, P. K., Hibbing, J. R. (2008). Political attitudes vary with physiological traits. *Science, 321,* 1667-1670.

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.

Rademacher, J. M., & Lippke, S. (2007). Dynamic online surveys and experiments with the free open-source software dynQuest. *Behavior Research Methods*, *39*(3), 415-426.

Rand, D. G. (In Press). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*. doi: 10.1016/j.jtbi.2011.03.004.

Reimers, S. (2007). The BBC Internet study: general methodology. *Archives of Sexual Behavior*, *36*(2), 147-161.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B., (2010). Who are the crowdworkers?: Shifting demographics in Mechanical Turk. CHI EA'10: *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, (pp. 2863-2872). doi: 10.1145/1753846.1753873.

Rueckert, L. (2005). A Web-based study of cerebral asymmetry for perception of emotion. *Behavior Research Methods*, *37*(2), 271-276.

Sears, D.O. (1986). College sophomores in the lab: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51,* 515-530.

Siegler, R. S. (1992). The other Alfred Binet. *Developmental Psychology*, *28(2),* 179.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology*, *57*, 529-555.

Spyridakis, J. H., Wei, C., Barrick, J., Cuddihy, E., & Maust, B. (2005). Internet-Based Research: Providing a Foundation for Web-Design Guidelines. *IEEE Transactions on Professional Communication*, *48*(3), 242-260.

Vadillo, M. A., Bárcena, R., & Matute, H. (2006). The Internet as a research tool in the study of associative learning: An example from overshadowing. *Behavioural Processes*, *73(1),* 36-40.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition*. San Antonio, TX: Pearson.

Wei, C. Y., Evan, M. B., Eliot, M., Barrick, J., Maust, B., & Spyridakis, J. H. (2005). Influencing Web-browsing behavior with intriguing and informative hyperlink wording. *Journal of Information Science*, *31*(5), 433-445.

Williams, J. E., & McCord, D. M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices test. *Computers in Human Behavior*, 22, 791-800.

Wilson, G. D., & Patterson, J. R. (1968). A new measure of conservatism. *British Journal of Social and Clinical Psychology, 1,* 264-269.

Wright, Jr., Adams, H., & Bernat, J. (1999). Development and validation of the homophobia scale. *Journal of Psychopathology and Behavioral Assessment*, 21, 337-347.

APPENDIX A

Means and Standard Deviations of Total Political Ideology and Homophobia Scores by Keyword Condition

	Neutral Keywords	Liberal Keywords	Conservative Keywords	Sexuality Keywords
Political Ideology	22.6 (7.7)	19.5 (6.8)	22.7 (6.4)	22.9 (7.9)
Homophobia	81.1 (18.2)	81.8 (15.5)	83.3 (17.6)	80.2 (20.7)

APPENDIX B

Means and Standard Deviations of Personality Factors by Condition

	Neutral Keywords	Liberal Keywords	Conservative Keywords	Sexuality Keywords
Extraversion	75.2 (15.9)	78 (13.3)	73.2 (16)	72.6 (13.1)
Agreeableness	90.2 (14.8)	90.4 (14.6)	92.6 (13.1)	91.4 (12.3)
Conscientiousness	88 (16.2)	89.4 (13.3)	93.4 (15.1)	88 (15.5)
Neuroticism	66.2 (17.7)	66.3 (14.1)	63.6 (19.3)	68 (17.3)
Openness	79.4 (12.6)	84.5 (12.3)	82.4 (15.1)	83.2 (13.5)

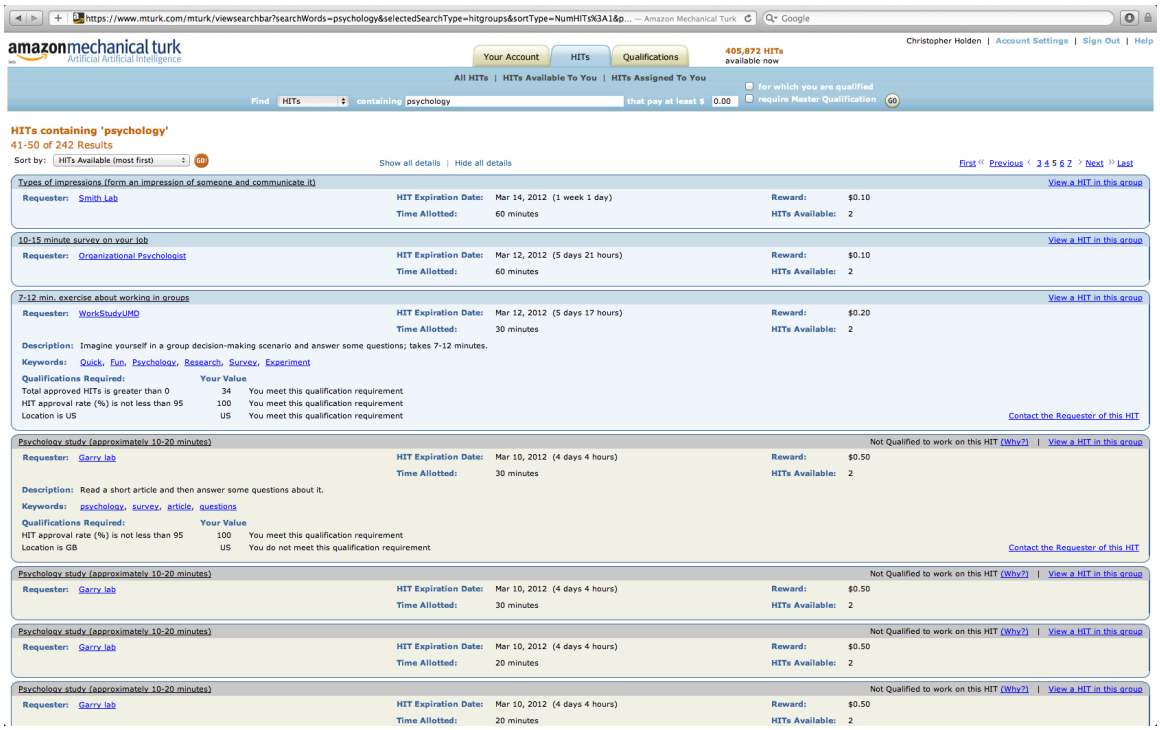
APPENDIX C

Sample HIT, Showing Title and Keywords

Note. Photo obtained from https://www.mturk.com/mturk/viewsearchbar?searchWords=psychology&selectedSearchType=hitgroups&sortType=NumHITs%3A1&pageNumber=5&searchSpec=HITGroupSearch%23T%234%2310%23-1%23T%23%21Reward%216%21rO0ABXQABDAuMDA-%21keyword_list%212%21rO0ABXQACnBzeWNob2xvZ3k-%21%23%21NumHITs%211%21%23%21.

APPENDIX D

Sample HIT Listing



Note. Photo obtained from https://www.mturk.com/mturk/viewsearchbar?searchWords=psychology&selectedSearchType=hitgroups&sortType=NumHITs%3A1&pageNumber=5&searchSpec=HITGroupSearch%23T%234%2310%23-1%23T%23%21Reward%216%21rO0ABXQABDAuMDA-%21keyword_list%212%21rO0ABXQACnBzeWNob2xvZ3k-%21%23%21NumHITs%211%21%23%21.

APPENDIX E

Informed Consent Form
Project Title: Personality and Attitudes Survey

What is the purpose of this research?

The purpose of this research is to better understand personality and how it affects everyday attitudes.

What will be expected of me?

You will be asked to complete a 180-item questionnaire consisting of demographic information, personality information, and information on attitudes that you may or may not have. You also must be at least 18 years old to participate in this study.

How long will the research take?

The questionnaires should take approximately 20 minutes to complete.

Will my answers be anonymous?

Yes. In no way will your name be linked to the data. The only form of identification will be the random number used for payment.

Can I withdraw from the study if I decide to?

Absolutely. You can withdraw from the research at any time and ask that your answers not be used.

Can I take the survey more than once?

No. This study will be posted on MTurk multiple times in order gather information from as many workers as possible. Please do not take this survey more than once. You will NOT be paid for any attempts other than the first.

Is there any harm that I might experience from taking part in the study?

There is no inherent risk of harm when participating in this study.

How will I benefit from taking part in the research?

You will receive 25 cents as payment for the completion of this survey. Additionally, you will have the satisfaction of knowing that you contributed to the body of knowledge in psychology.

Who should I contact if I have questions or concerns about the research?

Contact me (Christopher Holden) via email at cjholden2@catamount.wcu.edu . You can also contact the Western Carolina University IRB Chair at (828) 227-7212.

Do you consent to the study above?

_____ Yes, I agree to participate and I understand that my participation is voluntary. I understand that I will be able to stop taking this survey at any point in time. I also understand that there will be no consequences for not completing the survey.

_____ No, I do not agree to participate.

APPENDIX F

Demographic Information

In what year were you born? _____

Do you consider yourself to be:

- ❍ Male
- ❍ Female
- ❍ Transgender
- ❍ Other
- ❍ Prefer not to answer

What is the highest level of education you have completed?

- ❍ Less than High School
- ❍ High School / GED
- ❍ Some College
- ❍ 2-year College Degree
- ❍ 4-year College Degree
- ❍ Masters Degree
- ❍ Doctoral Degree
- ❍ Professional Degree (JD, MD)

What is your annual income range?

- ❍ Below $20,000
- ❍ $20,000 - $29,999
- ❍ $30,000 - $39,999

- ❍ $40,000 - $49,999
- ❍ $50,000 - $59,999
- ❍ $60,000 - $69,999
- ❍ $80,000 - $89,999
- ❍ $90,000 or more
- ❍ $70,000 - $79,999

In what state/territory do you currently reside? (Drop-down responses were used)

Do consider yourself to be:

- ❍ White/Caucasian
- ❍ Black/African American
- ❍ Hispanic
- ❍ Asian
- ❍ Native American
- ❍ Pacific Islander
- ❍ Mixed/multiracial
- ❍ Other

What is your country of origin? (Drop-down responses were used)

What criteria did you use to find this HIT? (Please check all that apply)

- ☐ HIT Creation date
- ☐ HITs Available
- ☐ Reward amount
- ☐ Expiration date
- ☐ Title

- ❑ Time allotted
- ❑ Keyword search
- ❑ Recommendation by other MTurk user
- ❑ Did not do anything specific to find this HIT

In the blank below, please provide the keywords you entered into the search.

_____

APPENDIX G

Political Ideology Scale (The Revised WP-I)

1. Please indicate whether you agree or disagree with the topic listed below: School prayer

   1 – Agree 2 – Disagree 3 – Uncertain

2. Please indicate whether you agree or disagree with the topic listed below: Pacifism

   1 – Agree 2 – Disagree 3 – Uncertain

3. Please indicate whether you agree or disagree with the topic listed below: Death Penalty

   1 – Agree 2 – Disagree 3 – Uncertain

4. Please indicate whether you agree or disagree with the topic listed below: Socialism

   1 – Agree 2 – Disagree 3 – Uncertain

5. Please indicate whether you agree or disagree with the topic listed below: Patriot Act

   1 – Agree 2 – Disagree 3 – Uncertain

6. Please indicate whether you agree or disagree with the topic listed below: Pornography

   1 – Agree 2 – Disagree 3 – Uncertain

7. Please indicate whether you agree or disagree with the topic listed below: Patriotism

1 – Agree 2 – Disagree 3 – Uncertain

8. Please indicate whether you agree or disagree with the topic listed below: Women's equality

   1 – Agree 2 – Disagree 3 – Uncertain

9. Please indicate whether you agree or disagree with the topic listed below: Biblical truth

   1 – Agree 2 – Disagree 3 – Uncertain

10. Please indicate whether you agree or disagree with the topic listed below: Premarital sex

    1 – Agree 2 – Disagree 3 - Uncertain

11. Please indicate whether you agree or disagree with the topic listed below: Iraq War

    1 – Agree 2 – Disagree 3 - Uncertain

12. Please indicate whether you agree or disagree with the topic listed below: Gay marriage

    1 – Agree 2 – Disagree 3 - Uncertain

13. Please indicate whether you agree or disagree with the topic listed below: Welfare spending

    1 – Agree 2 – Disagree 3 - Uncertain

14. Please indicate whether you agree or disagree with the topic listed below: Abortion rights

    1 – Agree 2 – Disagree 3 - Uncertain

15. Please indicate whether you agree or disagree with the topic listed below: Tax cuts

    1 – Agree 2 – Disagree 3 - Uncertain

16. Please indicate whether you agree or disagree with the topic listed below: Evolution

    1 – Agree 2 – Disagree 3 - Uncertain

17. Please indicate whether you agree or disagree with the topic listed below: Gun control

    1 – Agree 2 – Disagree 3 – Uncertain

18. Please indicate whether you agree or disagree with the topic listed below: Globalization

    1 – Agree 2 – Disagree 3 – Uncertain

19. Please indicate whether you agree or disagree with the topic listed below: Military spending

    1 – Agree 2 – Disagree 3 – Uncertain

20. Please indicate whether you agree or disagree with the topic listed below: Pollution control

    1 – Agree 2 – Disagree 3 - Uncertain

21. Please indicate whether you agree or disagree with the topic listed below: Warrantless searches

    1 – Agree 2 – Disagree 3 – Uncertain

22. Please indicate whether you agree or disagree with the topic listed below: Small government

1 – Agree 2 – Disagree 3 - Uncertain

23. Please indicate whether you agree or disagree with the topic listed below: School standards

    1 – Agree 2 – Disagree 3 - Uncertain

24. Please indicate whether you agree or disagree with the topic listed below: Foreign aid

    1 – Agree 2 – Disagree 3 - Uncertain

25. Please indicate whether you agree or disagree with the topic listed below: Obedience

    1 – Agree 2 – Disagree 3 - Uncertain

26. Please indicate whether you agree or disagree with the topic listed below: Compromise

    1 – Agree 2 – Disagree 3 – Uncertain

Scoring the revised WP-I: Odd- and even-numbered items are scored differently for the Agree and Disagree response. On odd-numbered items, an Agree response is scored as a 2 and a Disagree response is scored as a 0. The opposite is true with even-numbered items, where an Agree response is scored as a 1 and a Disagree response is scored as a 2. All Uncertain responses are scored as a 1. This produces a range in scores of 0 (Liberal) to 52 (Conservative).

APPENDIX H

M5-120

M5-120 Questionnaire						Page 2
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate
1	Worry about things.	O	O	O	O	O
2	Make friends easily.	O	O	O	O	O
3	Have a vivid imagination.	O	O	O	O	O
4	Trust others.	O	O	O	O	O
5	Complete tasks successfully.	O	O	O	O	O
6	Get angry easily.	O	O	O	O	O
7	Love large parties.	O	O	O	O	O
8	Believe in the importance of art.	O	O	O	O	O
9	Use others for my own ends.	O	O	O	O	O
10	Like to tidy up.	O	O	O	O	O
11	Often feel blue.	O	O	O	O	O
12	Take charge.	O	O	O	O	O
13	Experience my emotions intensely.	O	O	O	O	O
14	Love to help others.	O	O	O	O	O
15	Keep my promises.	O	O	O	O	O
16	Find it difficult to approach others.	O	O	O	O	O
17	Am always busy.	O	O	O	O	O
18	Prefer variety to routine.	O	O	O	O	O
19	Love a good fight.	O	O	O	O	O
20	Work hard.	O	O	O	O	O
21	Go on binges.	O	O	O	O	O
22	Love excitement.	O	O	O	O	O
23	Love to read challenging material.	O	O	O	O	O
24	Believe that I am better than others.	O	O	O	O	O
25	Am always prepared.	O	O	O	O	O
26	Panic easily.	O	O	O	O	O
27	Radiate joy.	O	O	O	O	O
28	Tend to vote for liberal political candidates.	O	O	O	O	O
29	Sympathize with the homeless.	O	O	O	O	O
30	Jump into things without thinking.	O	O	O	O	O
31	Fear for the worst.	O	O	O	O	O
32	Feel comfortable around people.	O	O	O	O	O
33	Enjoy wild flights of fantasy.	O	O	O	O	O
34	Believe that others have good intentions.	O	O	O	O	O
35	Excel in what I do.	O	O	O	O	O
36	Get irritated easily.	O	O	O	O	O
37	Talk to a lot of different people at parties.	O	O	O	O	O
38	See beauty in things that others might not notice.	O	O	O	O	O
39	Cheat to get ahead.	O	O	O	O	O
40	Often forget to put things back in their proper place.	O	O	O	O	O
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate

M5-120 Questionnaire						Page 3
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate
41	Dislike myself.	O	O	O	O	O
42	Try to lead others.	O	O	O	O	O
43	Feel others' emotions.	O	O	O	O	O
44	Am concerned about others.	O	O	O	O	O
45	Tell the truth.	O	O	O	O	O
46	Am afraid to draw attention to myself.	O	O	O	O	O
47	Am always on the go.	O	O	O	O	O
48	Prefer to stick with things that I know.	O	O	O	O	O
49	Yell at people.	O	O	O	O	O
50	Do more than what's expected of me.	O	O	O	O	O
51	Rarely overindulge.	O	O	O	O	O
52	Seek adventure.	O	O	O	O	O
53	Avoid philosophical discussions.	O	O	O	O	O
54	Think highly of myself.	O	O	O	O	O
55	Carry out my plans.	O	O	O	O	O
56	Become overwhelmed by events.	O	O	O	O	O
57	Have a lot of fun.	O	O	O	O	O
58	Believe that there is no absolute right or wrong.	O	O	O	O	O
59	Feel sympathy for those who are worse off than myself.	O	O	O	O	O
60	Make rash decisions.	O	O	O	O	O
61	Am afraid of many things.	O	O	O	O	O
62	Avoid contacts with others.	O	O	O	O	O
63	Love to daydream.	O	O	O	O	O
64	Trust what people say.	O	O	O	O	O
65	Handle tasks smoothly.	O	O	O	O	O
66	Lose my temper.	O	O	O	O	O
67	Prefer to be alone.	O	O	O	O	O
68	Do not like poetry.	O	O	O	O	O
69	Take advantage of others.	O	O	O	O	O
70	Leave a mess in my room.	O	O	O	O	O
71	Am often down in the dumps.	O	O	O	O	O
72	Take control of things.	O	O	O	O	O
73	Rarely notice my emotional reactions.	O	O	O	O	O
74	Am indifferent to the feelings of others.	O	O	O	O	O
75	Break rules.	O	O	O	O	O
76	Only feel comfortable with friends.	O	O	O	O	O
77	Do a lot in my spare time.	O	O	O	O	O
78	Dislike changes.	O	O	O	O	O
79	Insult people.	O	O	O	O	O
80	Do just enough work to get by.	O	O	O	O	O
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate

M5-120 Questionnaire						Page 4
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate
81	Easily resist temptations.	O	O	O	O	O
82	Enjoy being reckless.	O	O	O	O	O
83	Have difficulty understanding abstract ideas.	O	O	O	O	O
84	Have a high opinion of myself.	O	O	O	O	O
85	Waste my time.	O	O	O	O	O
86	Feel that I'm unable to deal with things.	O	O	O	O	O
87	Love life.	O	O	O	O	O
88	Tend to vote for conservative political candidates.	O	O	O	O	O
89	Am not interested in other people's problems.	O	O	O	O	O
90	Rush into things.	O	O	O	O	O
91	Get stressed out easily.	O	O	O	O	O
92	Keep others at a distance.	O	O	O	O	O
93	Like to get lost in thought.	O	O	O	O	O
94	Distrust people.	O	O	O	O	O
95	Know how to get things done.	O	O	O	O	O
96	Am not easily annoyed.	O	O	O	O	O
97	Avoid crowds.	O	O	O	O	O
98	Do not enjoy going to art museums.	O	O	O	O	O
99	Obstruct others' plans.	O	O	O	O	O
100	Leave my belongings around.	O	O	O	O	O
101	Feel comfortable with myself.	O	O	O	O	O
102	Wait for others to lead the way.	O	O	O	O	O
103	Don't understand people who get emotional.	O	O	O	O	O
104	Take no time for others.	O	O	O	O	O
105	Break my promises.	O	O	O	O	O
106	Am not bothered by difficult social situations.	O	O	O	O	O
107	Like to take it easy.	O	O	O	O	O
108	Am attached to conventional ways.	O	O	O	O	O
109	Get back at others.	O	O	O	O	O
110	Put little time and effort into my work.	O	O	O	O	O
111	Am able to control my cravings.	O	O	O	O	O
112	Act wild and crazy.	O	O	O	O	O
113	Am not interested in theoretical discussions.	O	O	O	O	O
114	Boast about my virtues.	O	O	O	O	O
115	Have difficulty starting tasks.	O	O	O	O	O
116	Remain calm under pressure.	O	O	O	O	O
117	Look at the bright side of life.	O	O	O	O	O
118	Believe that we should be tough on crime.	O	O	O	O	O
119	Try not to think about the needy.	O	O	O	O	O
120	Act without thinking.	O	O	O	O	O
		Innacurate	Moderately Innacurate	Neither	Moderately Accurate	Accurate

APPENDIX I

Homophobia Scale

This part of the questionnaire is designed to measure your thoughts, feelings, and behaviors with regard to homosexuality. It is not a test, so there are no right or wrong answers. Answer each item by circling the number after each question as follows:

1. Strongly disagree
2. Disagree
3. Neither disagree nor agree
4. Agree
5. Strongly Agree

1. Gay people make me nervous.

   1 2 3 4 5

2. Gay people deserve what they get.

   1 2 3 4 5

3. Homosexuality is acceptable to me.

   1 2 3 4 5

4. If I discovered a friend was gay I would end the friendship.

   1 2 3 4 5

5. I think homosexual people should not work with children.

   1 2 3 4 5

6. I make derogatory remarks about gay people.

1 2 3 4 5

7. I enjoy the company of gay people.

1 2 3 4 5

8. Marriage between homosexual individuals is acceptable.

1 2 3 4 5

9. I make derogatory remarks like “faggot” or “queer” to people I suspect are gay.

1 2 3 4 5

10. It does not matter to me whether my friends are gay or straight.

1 2 3 4 5

11. It would not upset me if I learned that a close friend was homosexual.

1 2 3 4 5

12. Homosexuality is immoral.

1 2 3 4 5

13. I tease and make jokes about gay people.

1 2 3 4 5

14. I feel that you cannot trust a person that is homosexual.

1 2 3 4 5

15. I fear homosexual persons will make sexual advances towards me.

1 2 3 4 5

16. Organizations which promote gay rights are necessary.

1 2 3 4 5

17. I have damaged property of gay persons, such as “keying” their cars.

1 2 3 4 5

18. I would feel comfortable having a gay roommate.

1 2 3 4 5

19. I would hit a homosexual for coming on to me.

1 2 3 4 5

20. Homosexual behavior should not be against the law.

1 2 3 4 5

21. I avoid gay individuals.

1 2 3 4 5

22. It does not bother me to see two homosexual people together in public.

1 2 3 4 5

23. When I see a gay person I think, “What a waste.”

1 2 3 4 5

24. When I meet someone I try to find out if he/she is gay.

1 2 3 4 5

25. I have rocky relationships with people that I suspect are gay.

1 2 3 4 5

APPENDIX J

Manipulation Checks Used in the Second, Fourth, and Sixth Section of the M5-120

1. If you are reading this, please answer inaccurate.
2. If you are reading this, please answer neither.
3. If you are reading this, please answer accurate.