



Assessing the reliability of the M5-120 on Amazon's mechanical Turk



Christopher J. Holden^{*,1}, Trevor Dennie¹, Adam D. Hicks¹

Department of Psychology, Western Carolina University, 91 Killian Building Lane, Room 302B, Cullowhee, NC 28723, United States

ARTICLE INFO

Article history:

Available online 27 March 2013

Keywords:

Mechanical Turk
MTurk
Personality
Online research
M5-120
Replication

ABSTRACT

Amazon's online service, Mechanical Turk (MTurk) has become a popular option for data collection among social scientists. Early work (Buhrmester, Kwang, & Gosling, 2011) indicated that data collection through MTurk was faster and less expensive than traditional collection methods (undergraduate human subject pool), as well as being reliable when administered at different dates. Building on their work, we sought to extend this investigation of reliability to a larger measure. For the current research we chose a 120-item measure of personality. After collecting data through MTurk, it was determined that our MTurk sample had strong test–retest reliability, indicating that they did not significantly change between administration dates.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

There have been many theories behind the composition of human personality throughout history. Many older theories, such as Freudian theory, were constructed around the presence of aberrant behavior in individuals (Feist & Feist, 2006). However, much of social psychology is not concerned with abnormal behavior; instead, the focus lies on what many term “normal” personality, or specific components of personality that are common throughout the human population (Costa & McCrae, 1992a). The theories that quickly filled this niche are often referred to as trait and factor theories, because they are based on personality factors created through the statistical process of factor analysis (Cattell, 1947; Costa & McCrae, 1995; Feist & Feist, 2006). These factor theories have become very popular as they are well suited for research because their measures are quantifiable and often rely on self-report methods.

Of these factor theories, the five-factor model (FFM) has become one of the more favored. (Digman, 1990). Two of the most prolific and well-known pioneers in the field of FFM research are Costa and McCrae; and their instrument, the Revised NEO Personality Inventory (NEO-PI-R; 1992b), is one of the most widely utilized in trait personality research. The NEO-PI-R breaks the factors of personality into five specific domains with each domain being composed of six underlying facets. For a full list of the domains and facets, see Fig. 1. One drawback of the NEO-PI-R is the fact that it is copyrighted, which inhibits any type of customization and adds additional monetary constraints to researchers.

In response to these concerns, Goldberg (1999) developed the International Personality Item Pool (IPIP) in 1996. The entire IPIP is not copyrighted and is in the public domain, which allows researchers to create unique personality inventories based on the topic of their study. In fact, several proxy instruments, based on well-known instruments such as: the California Psychological Inventory (CPI; Cloninger, 1994); the NEO-PI-R (Costa & McCrae, 1992b); the Sixteen Personality Factor Scale (16PF; Conn & Rieke, 1994); and the Hogan Personality Inventory (HPI; Hogan and Hogan (1992)), have been created using the IPIP.

One such proxy instrument, created to resemble Costa and McCrae's NEO-PI-R (1992b), is referred to as the M5 Questionnaire (Johnson, 2001). Because of its versatility, the M5 Questionnaire comes in several different forms based on survey length, but the most popular form is the M5-120 because it is relatively short (120 items) and is able to provide reliable measures of all five domains and 30 underlying facets (Johnson, 2001). The M5-120 has been compared to the NEO-PI-R and has been shown to be a reliable substitute. For a full list of the corrected correlations between the IPIP and the NEO-PI-R, see Table 1.

1.1. Overview of MTurk

Many researchers are using online data collection platforms (Reimers, 2007), and Amazon's Mechanical Turk (MTurk) has become a popular option. Amazon offers this service free-of-charge, with two types of accounts. The first is the worker account; when a worker logs in, they can choose from a variety of Human Intelligence Tasks (HITs), each of which offer monetary rewards. These HITs consist of tasks that are too complex to be computerized, yet not sophisticated enough that they require specialization on the part of the workers. Typically, HITs are brief and pay only a

* Corresponding author. Address: 492 Timberlea Drive Apartment 101, Rochester Hills, MI 48309, United States. Tel.: +1 919 880 2233.

E-mail addresses: cjholden2@catamount.wcu.edu (C.J. Holden), tmdennie1@catamount.wcu.edu (T. Dennie), adhicks@wcu.edu (A.D. Hicks).

¹ Tel.: +1 828 227 7361; fax: +1 828 227 7005.

Neuroticism
Anxiety
Angry Hostility (<i>Anger</i>)
Depression
Self-Consciousness
Impulsiveness (<i>Immoderation</i>)
Vulnerability
Extraversion
Warmth (<i>Friendliness</i>)
Gregariousness
Assertiveness
Activity
Excitement-Seeking
Positive Emotions (<i>Cheerfulness</i>)
Openness to Experience
Fantasy (<i>Imagination</i>)
Aesthetics (<i>Artistic Interest</i>)
Feelings (<i>Emotionality</i>)
Actions (<i>Adventurousness</i>)
Ideas (<i>Intellect</i>)
Values (<i>Liberalism</i>)
Agreeableness
Trust
Straightforwardness (<i>Morality</i>)
Altruism
Compliance (<i>Cooperation</i>)
Modesty
Tender-Mindedness
Conscientiousness
Competence (<i>Self-Efficacy</i>)
Order
Dutifulness
Achievement Striving
Self-Discipline
Deliberation (<i>Cautiousness</i>)

Table 1

Corrected correlations between the IPIP scale (M5 Questionnaire) and the NEO-PI-R.

Domains Facets	Corrected correlations
Neuroticism	.93
Anxiety	.90
Angry hostility (<i>Anger</i>)	.91
Depression	.92
Self-consciousness	.94
Impulsiveness (<i>Immoderation</i>)	.98
Vulnerability	.96
Extraversion	.88
Warmth (<i>Friendliness</i>)	.91
Gregariousness	.98
Assertiveness	.99
Activity	.98
Excitement-seeking	.95
Positive emotions (<i>Cheerfulness</i>)	.95
Openness to Experience	.92
Fantasy (<i>Imagination</i>)	.90
Aesthetics (<i>Artistic Interest</i>)	.95
Feelings (<i>Emotionality</i>)	.90
Actions (<i>Adventurousness</i>)	.99
Ideas (<i>Intellect</i>)	.95
Values (<i>Liberalism</i>)	.86
Agreeableness	.90
Trust	.95
Straightforwardness (<i>Morality</i>)	.86
Altruism	.90
Compliance (<i>Cooperation</i>)	.97
Modesty	.95
Tender-Mindedness	.90
Conscientiousness	.88
Competence (<i>Self-Efficacy</i>)	.89
Order	.99
Dutifulness	.87
Achievement striving	.97
Self-discipline	.92
Deliberation (<i>Cautiousness</i>)	.95

Note: facet names italicized and in parentheses denote name changes in the M5 Questionnaire.

Fig. 1. The five broad domains and 30 underlying facets. The wording of the facets presented in this figure represents the original wording by Costa and McCrae's Revised NEO-PI-R (1992b). Those words appearing italicized and in parentheses denote facet name changes within the M5 Questionnaire.

few cents. MTurk was originally developed for commercial use, but a growing number of HITs are dedicated to academic research.

The second option is the requester account. Requesters provide the HITs for the workers to complete, and this is the account used by researchers. This account provides the researcher access to all of the built-in survey tools and to the entire population of workers. Requesters can post multiple HITs at once, select the number of workers desired and collect data simultaneously. Payment for each HIT is determined by the requester, and is then multiplied by the number of workers desired to produce the total cost to the requester (Amazon adds a 10% commission to this cost for their service of paying the workers individually). Thus, the data collection process is streamlined, allowing the researcher to focus on survey design and data analysis.

1.2. Previous personality work using MTurk

In their 2011 article, Buhrmester, Kwang and Gosling (hereafter BKG) take a detailed look at MTurk. After providing a brief overview of MTurk, BKG demographically compare the MTurk samples to other Internet samples, and to samples derived from traditional data collection methods. It was found that workers on MTurk come from over 50 countries, making the sample much more diverse than traditional college samples, and more diverse than other Internet samples (Buhrmester et al., 2011). Keeping in mind that workers must be paid and requesters' preference for inexpensive data collection, BKG manipulated length of task and compensation amount to test whether these two factors contribute to the speed

at which data is collected and to the quality of data obtained. While participants were recruited faster for shorter tasks, as well as tasks with higher compensation, data quality was acceptable (as measured by reliability alpha) even in the lowest of payment conditions (Buhrmester et al., 2011).

Most importantly, BKG used the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991) to compare the personality data gathered via MTurk to personality data from traditional samples. This comparison was made using reliability alphas, all of which were found to be acceptable (Buhrmester et al., 2011). Furthermore, test-retest reliabilities were used to assess data quality. Personality measures were distributed to participants who consented to both waves of the research, with a three-week period between the two distributions. Test-retest reliabilities ranged from $r = .86$ to $r = .94$ for the five-factors, each of which exceeded the test-retest reliability found in previous literature for the BFI (Gosling, Rentfrow, & Swann, 2003). With such sound psychometrics, BKG considered data collection via MTurk to be promising.

This article by BKG has inspired many researchers to use MTurk for data collection (353 citations in Google Scholar as of February 27th, 2013). As this use of MTurk continues to develop, researchers will begin to use new measures on this online platform, making it necessary to have a full understanding of the reliability of popular measures in their applications on MTurk. However, the BKG paper utilized the BFI which consists of 44 items. Compared to other personality scales such as the NEO-PI-R, which is made up of 240 items, the BFI is a relatively short measure of personality.

Download English Version:

<https://daneshyari.com/en/article/351078>

Download Persian Version:

<https://daneshyari.com/article/351078>

[Daneshyari.com](https://daneshyari.com)