The Reliability of Crowdsourcing: Obtaining Psychometric Data through Mechanical Turk

Matthew P. Baucum

Pepperdine University
The Reliability of Crowdsourcing: Obtaining Psychometric Data through Mechanical Turk

The use of the Internet for administering self-report measures has become increasingly common (Luce et al., 2007), and it has generally been shown to provide similar test data as in-person administration (Bates & Cox, 2008; Björnsdotter, Enbrink, & Ghaderi, 2013; Bonini Campus, Zucoloto, Sampaio Bonafé, Jordani, & Maroco, 2011; Bressani & Downs, 2002; Brock et al., 2014; Luce et al., 2007). As psychologists have continued examining the viability of online psychological assessment, one particular platform, known as Mechanical Turk (or MTurk), has received considerable research support (Rouse, 2015). MTurk is an online crowdsourcing service where workers complete online tasks in exchange for payment, and research has found that personality assessment responses obtained through the service are generally reliable (Buhrmester, Kwang, & Gosling, 2011; Holden, Dennie, & Hicks, 2013; Rouse, 2015).

Such findings may hold promise for psychological test developers, who, when constructing a self-report assessment, must obtain test data for the purposes of estimating its reliability (a necessary metric for the publication of any psychological test). MTurk could serve as a valuable way for test developers to reach out to participants and obtain such test data. However, while MTurk has generally been found to be a reliable source of self-report data, we found only one study that has actually used responses from MTurk workers in the test development process (specifically, in the development of a male body image scale) (Sladek, Engeln, & Miller, 2014). This may be partly due to the fact that some believe that more research is needed on the validity of MTurk data (Buhrmester et al., 2011; Holden et al., 2013; Rouse, 2015). Buhrmester et al. even specifically recommends against the use of previously untested psychological measures on Mechanical Turk, warning that “caution is still advised…especially with instruments that have little psychometric information available” (2011). The goal of the
present study is to move towards validating MTurk as a platform for the initial psychometric assessment of self-report measures. Specifically, since existing research has focused on the reliability of Mechanical Turk for administering personality assessments, we will focus on three other types of self-report measures that have received less research support: cognitive abilities, psychopathology, and affect.

**Basics of Mechanical Turk**

Amazon’s Mechanical Turk is an online marketplace where workers complete online tasks in exchange for payment. The tasks, dubbed “Human Intelligence Tasks” or “HIT’s”, may include activities such as object recognition, natural language processing, or survey completion, and are posted by “requesters” who offer payment to those who complete them. Tasks typically pay less than one dollar, and workers will, on average, work for about $1.40 an hour (Paolacci, Chandler, & Ipeirotis, 2010). One of the advantages of MTurk is that it provides access to a large and diverse workforce; in 2014, there were more than 500,000 registered Mechanical Turk workers from more than 190 countries (Paolacci & Chandler, 2014).

**Use of Mechanical Turk for Social Science Research**

Given the diversity and convenience of MTurk, several studies have investigated its viability as a platform for research, and there is growing empirical support that it may be an inexpensive yet reliable source of research participants. For example, one study that compared an MTurk sample to a community sample found that both exhibited similar levels of task attention and cognitive ability (Goodman, Cryder, & Cheema, 2013). Another study found that MTurk workers exhibited several well-documented heuristic biases (Paolacci et al., 2010). Another study compared the results of a tool-selection task (in which participants were asked which tool from a list would be ideal for certain hypothetical situations) between an in-lab sample and an
MTurk sample, and found no significant differences between the two groups’ results (Casler, Bickel, & Hackett, 2013). It has also been demonstrated that a matched comparison group obtained through MTurk can replicate the behavior of a randomly selected control group in an academic motivation experiment (Azzam & Jacobson, 2013).

We found five studies that specifically looked at the reliability of psychological tests administered with MTurk. Rouse (2015) administered a personality questionnaire to an MTurk sample and found that the alpha coefficient closely matched the alpha value previously obtained from a community sample. Buhrmester et al. (2011) administered a personality questionnaire to Mechanical Turk workers and found retest correlations for all scales to vary only slightly from those established through laboratory sampling. Holden et al. (2013) also examined the retest reliability of a personality inventory administered over MTurk and found high retest correlations for each of the Big Five personality trait scales, but they did not comment on how these correlations compared to the reliability of in-lab samples.

The scope of these studies, however, was limited to self-report personality assessment; we were able to find only two studies that looked at the reliabilities of non-personality measures administered through MTurk. Johnson and Borden (2012) administered measures of trait and state empathy to both an MTurk sample and an in-lab sample, and found the internal consistencies of all assessment facets to vary at most by .14 between the two samples; however, such differences were not tested for statistical significance. Shapiro, Chandler, and Mueller (2013) studied psychiatric symptomology in MTurk workers and found high retest reliability for the Beck Depression Inventory (r=.87) and high internal consistency on the Beck Depression Inventory (α > .9), Beck Anxiety Index (α = .93), and Liebowitz Social Anxiety Scale (α = .97), but only moderate internal consistency for the Altman Self-Rating Scale for Mania (α = .68).
However, none of these values were compared to established reliability metrics for those tests, nor were they compared to responses from a non-MTurk sample. Thus, for non-personality assessments, empirical comparisons between the reliability of MTurk and non-MTurk test responses is lacking, and will thus be the focus of the present study.

**Psychometric Focus of Study**

The psychometric basis of this study will be the reliability of the psychological measures administered. Under Classical Test Theory, test responses are comprised of a respondent’s true score and measurement error, and a test’s reliability is the degree to which variations in scores are attributable to variations in respondents’ true scores (Furr & Bacharach, 2013). We will focus specifically on Cronbach’s alpha (a measure of internal consistency) and retest correlation coefficients (which compare respondents’ scores on two different administrations of the same test). Buhrmester et al. (2011), Holden et al. (2013), Johnson & Borden (2012), and Rouse (2015) all reported at least one of these statistics; thus, focusing on both will allow us to compare our study to past findings, and will ensure that the results are relevant to the process of psychological test development (in which internal consistency and retest reliability are important metrics).

**Current Study**

In order to add to the body of literature supporting the use of Mechanical Turk as a research platform, the present study will focus on three domains of psychological assessment that have not undergone reliability testing with Mechanical Turk: psychopathology, cognitive abilities, and affect. We plan to compare the reliability metrics obtained through MTurk to reliability metrics obtained through online administration to a university sample, to explore the
degree to which MTurk samples can provide accurate estimates of internal consistency and retest reliability.

We plan to study the reliability of a psychopathology measure because MTurk may in fact be a valuable platform for studying psychiatric symptomology. Shapiro et al. (2013) found that a high proportion of MTurk workers screened positive for substance abuse problems and for social anxiety, and that depression, general anxiety, and exposure to trauma were just as common in the MTurk population as the general population. As such, the researchers suggest that MTurk may be useful for studying clinical populations. However, they also found high rates of feigning symptoms (10.1% of the sample), and cite such malingering as a concern for MTurk data quality. Our study aims to identify whether MTurk is an accurate source of reliability metrics for psychopathology assessments (as compared to a university sample) in spite of this finding, especially considering the promise that MTurk holds for research with clinical populations (Shapiro et al., 2013).

We plan to study the reliability of a cognitive abilities measure because research focusing on cognitive assessment over MTurk is generally lacking (Paolacci & Chandler, 2014). We were able to find only one study that administered a cognitive assessment over MTurk; Paolacci et al. (2010) administered the Subjective Numeracy Scale to an MTurk worker sample, but did not report any reliability metrics for the test. Furthermore, Goodman et al. (2013) found that in an anchoring and adjustment task, some MTurk workers obtained answers from the internet, even when explicitly asked not to, and the researchers cautioned against the use of factual questions in MTurk survey research. This finding calls into question the reliability of MTurk for administering cognitive abilities scales, since many items designed to measure crystallized intelligence have factual answers. Still, cognitive abilities assessment remains a very prominent
RELIABILITY OF CROWDSOURCING

domain of psychological testing, so validating MTurk as a platform for estimating the reliability of cognitive abilities measures will be one of the foci of this study.

Lastly, we plan to investigate the reliability of MTurk workers’ responses to a measure of affect. According to Goodman et al. (2013), a sample of MTurk workers who completed a personality questionnaire scored significantly lower on emotional stability than a student sample, which calls into question whether the MTurk population is a reliable subject pool for testing an emotion-related measure. Furthermore, the personality tests previously studied on MTurk (as well as the cognitive abilities and psychopathology measures used in this study) are designed to measure stable constructs, and thus are designed to have high retest reliability. If MTurk is truly a reliable platform for estimating the reliability of psychological tests, it should be able to replicate low retest reliabilities as well as high ones. Thus, an affect measure was also selected because it is specifically meant to have low retest reliability (since affect is not a stable construct), and the degree to which an MTurk sample can replicate this will contribute to the evidence for or against it as a platform for psychometric estimation.

Thus, the goal of this study is to move towards a more comprehensive understanding of the usability of MTurk for the purposes of psychometric estimation of self-report measures. Given past research support for the reliability of Mechanical Turk self-report data, we hypothesize that the reliability metrics for these tests obtained through Mechanical Turk will match closely the reliability metrics obtained from a university sample.

Method

Participants

We plan to recruit two samples: one obtained through MTurk and one obtained through the affiliated university’s research participation pool. Both samples will consist of 150
participants. For the MTurk sample, only workers with a 95% approval rating (meaning that 95% of requesters have found their work acceptable) and who speak English (since questionnaires will be administered in English) will be included in the study. The gender breakdown, nationalities, percentage of native English speakers, and mean and standard deviation of age will be reported for each sample.

**Procedures**

Participation in this study will take place entirely online for both samples, and participants will be recruited via convenience sampling methods. For the MTurk sample, we will create a HIT explaining the nature of the experiment, making it available to any English-speaking workers with a 95% approval rating. For the university sample, we will post an invitation to participate in the affiliated university’s research recruitment site.

Participants in both samples will be asked to fill out an online informed consent form before beginning the study (Appendix A). Participants who choose to not provide consent will be redirected to a page thanking them for their time and consideration. Those who choose to continue will be redirected to the experiment survey page.

Participation in this study will involve two identical administrations of three self-report measures: an affect scale, an intelligence scale, and a psychopathology scale. For the first administration, participants will be asked to provide their age, gender, nationality, and whether English is their first language. Participants in the university sample will also be asked for their name, course for which they are seeking credit, and the name of their professor (to whom we will provide confirmation of their participation), as well as an email address with which they can be contacted for the follow-up administration. They will then complete the three surveys, which should take approximately 15 minutes. Two weeks after each participant completes this phase of
the experiment, they will be contacted through either MTurk or by email (for the university sample), requesting their participation in an identical administration of the same surveys. At the beginning of both administrations, participants will be instructed to answer all items honestly and to the best of their ability. They will be informed that the focus of the study is the reliability of the assessments, and that they should not make any conclusions or assumptions about themselves based on their test responses. Participants will also be presented a validity check question asking for the honesty of their responses (Rouse, 2015), and will be asked whether they had prior exposure to any of the self-report measures (this will be asked on the first administration only). These questions can be found in Appendix E. Participants in the Mechanical Turk sample will be paid $.50 for the first test administration and $.75 for the second administration. Participants in the university sample will be awarded one research participation credit (a requirement for the classes they will be enrolled in) for their participation in both survey administrations.

We are aiming to base retest reliability calculations on samples of at least 75 participants per sample; thus, with initial sample sizes of 150, we have planned for an attrition rate of approximately 50% between survey administrations. To account for such attrition, internal consistency coefficients will be calculated based only on data from the first administration of each measure; thus, the responses from participants who drop out of the experiment will still be included in the analysis, and attrition will only affect the availability of retest reliability data. Since participants will be required to respond to all items before being allowed to submit their data, we do not expect issues with missing test responses. However, in the unlikely event of missing data, assessment responses will be deleted list wise for any participants with missing data.
Materials

Participants will be asked to provide their age, gender, and nationality prior to beginning the study. They will then complete three self-report psychological measures: the PANAS, the Personality Inventory for DSM-5—Brief Form, and the International Cognitive Ability Resource Verbal Reasoning Scale.

**PANAS.** The PANAS is a self-report measure of positive and negative affect (Watson, Clark, & Tellegen, 1988). The test presents respondents with twenty emotions (such as “guilty” or “proud”) and asks them to indicate the extent to which they have felt that emotion that day. Ratings range from 1 (“very slightly or not at all”) to 5 (“extremely”). The test consists of a positive affect subscale and a negative affect subscale, which each consist of 10 items and whose scores can range from 10 to 50. The internal consistency coefficients of the test’s positive affect and negative affect scales are .90 and .87, respectively, while the eight-week retest reliability coefficients are .47 and .31, respectively. Other established measures of distress and psychopathology correlate positively with the PANAS’ negative affect scale and correlate negatively with its positive affect scale, supporting the test’s external validity (Watson et al., 1988). The full PANAS assessment can be found in Appendix B.

**Personality Inventory for DSM-5—Brief Form.** The Personality Inventory for DSM-5—Brief Form (PID-5-BF) is a twenty five item self-report measure that assesses personality pathology (Krueger, Derringer, Markon, Watson, & Skodol, 2013). Items consist of various statements (such as “I feel like I act totally on impulse”) that respondents rate on a scale of 0 (“Very False or Often False”) to 3 (“Very True or Often True”). Items pertain to one of five personality trait domains: Negative Affect, Detachment, Antagonism, Disinhibition, and
Psychoticism. The possible scores for each scale range from 0 to 15, for a total score range of 0 to 75. The full PID-5-BF assessment can be found in Appendix C.

International Cognitive Ability Resource: Verbal Reasoning Scale. The International Cognitive Ability Resource (ICAR) is a collection of open source intelligence items that have been developed for research purposes (Condon & Revelle, 2014). For the purposes of this study, we will be using a 16-item verbal reasoning subscale of ICAR items. A verbal reasoning scale was chosen (as opposed to other types of cognitive abilities scales) to allow us to examine participants’ responses to a cognitive abilities measure while still being able to use text-based items. The scale will be scored according to the percentage of items answered correctly, so total scores will range from 0 to 100.

Proposed Analyses

Descriptive Statistics

We will report descriptive statistics for four separate subsets of participants: the MTurk workers and university students who participated in the first survey administration (whose data will be used for internal consistency analysis), and the MTurk workers and university students who participated in both survey administrations (whose data will be used for retest reliability analysis). For each of these sample subsets, we will report total sample size, mean age, standard deviation of age, gender breakdown by percentages, and nationality breakdown by percentages (for the top five countries of origin, if applicable). We will also present the percentage of each sample that identified as non-native English speakers, that confirmed their own responses as dishonest, and that reported having prior exposure to at least one of the self-report measures. This data will be presented in the Participants section of the paper.
For each of the scales administered, we will report Cronbach’s alpha for each of its subscales, and will report this value separately for the Mechanical Turk and university samples. This will result in sixteen alpha values; eight alpha values for each sample, pertaining to the Positive Affect and Negative Affect subscales of the PANAS, the Negative Affect, Detachment, Antagonism, Disinhibition, and Psychoticism subscales of the PID-5-BF, and the entire ICAR Verbal Reasoning Scale (no subscales). We will also report a Pearson correlation coefficient for each measure, correlating participants’ scores on the first and second administration of the scale, as a measure of retest reliability. This will be reported separately for the two samples, resulting in six Pearson correlations coefficients (three per sample, one pertaining to each measure).

Furthermore, we plan to report all reliability metrics before and after removing the data of non-native English speakers (to investigate the degree to which this affects results). We will also report any reliability metrics that change after removing the data of participants who reported prior exposure to a given assessment. For all reliability metrics, data form participants who confirmed their responses as dishonest in the validity check will be eliminated. This data will appear in the Results section of the paper.

**Inferential Statistics**

To compare the surveys’ internal consistencies between the MTurk and university samples, we will use a Feldt test for the comparison of alpha coefficients. This is a common procedure for the comparison of alpha values, and it uses the ratio of $1-\alpha^2$ between two measures as an F statistic to test for significant differences between the reliabilities. We will compare the alpha values of each of the eight survey scales/subscales between the MTurk and university sample, resulting in eight Feldt tests. All tests will be conducted at the .05 level.
To compare retest reliabilities between the MTurk and university samples, we will create 95% confidence intervals around the difference between the Pearson correlation coefficients for each test. Doing this will allow us to see if zero is included in each interval (which would indicate a nonsignificant difference between the two samples’ retest reliabilities), and will also allow us to see the magnitude of the potential difference between them.

We plan to conduct all analyses using the total reliability of each sample (without excluding non-native English speakers). However, if the removal of non-native English speakers’ responses substantially reduces any differences between the reliability metrics of the MTurk and university samples, then the same statistical tests mentioned above will be performed on these adjusted reliabilities as well.
References


Appendix A

Informed Consent Form

You are invited to participate in a study which aims to understand the usefulness of psychological tests administered online. You must be at least 18 years old to participate.

Participation in this study involves two parts. If you decide to participate, you will be asked to provide basic demographic information and to complete three self-report assessments (one measuring mood, one measuring cognitive ability, and one measuring personality) at two different time points: now, then again in two weeks. Completion of these surveys should take approximately 10 minutes at each administration, for a total of 20 minutes.

Please answer honestly and to the best of your ability, without seeking outside assistance. Any information that is obtained in connection with this study will remain password protected and shared only between the principal investigators of the study. It will be disclosed only with your permission or as required by law.

You will be awarded [$0.50 for completion of these surveys, and $0.75 for completion of the follow-up tests two weeks from now] one research participation credit for your participation (i.e. completion of these surveys now and in two weeks). Potential risks of this study could include mild discomfort due to the personal nature of some of the questions.

If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without penalty.

If you have any questions, please contact the researcher at matthew.baucum@pepperdine.edu

Questions regarding the rights of research participants may be directed to the Chair of Seaver College’s IRB (susan.helm@pepperdine.edu).

☐ I have read and understood the above information and give my consent to participate

☐ I do not give my consent to participate
Appendix B

Positive Affect Negative Affect Schedule (PANAS)

This scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate to what extent you have felt this way today. Use the following scale to record your answers.

<table>
<thead>
<tr>
<th></th>
<th>Very slightly or not at all</th>
<th>A little</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Extremely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interested</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Distressed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Excited</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Upset</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Strong</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Guilty</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Scared</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Hostile</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Enthusiastic</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>Proud</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Irritable</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>Alert</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>Ashamed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>Inspired</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>Nervous</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>Determined</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>Attentive</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>Jittery</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>Active</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>Afraid</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix C

Personality Inventory for DSM-5—Brief Form (PID-5-BF)

**Instructions:** This is a list of things different people might say about themselves. We are interested in how you would describe yourself. There are no right or wrong answers. So you can describe yourself as honestly as possible, we will keep your responses confidential. We’d like you to take your time and read each statement carefully, selecting the response that best describes you.

<table>
<thead>
<tr>
<th></th>
<th>Very False or Often False</th>
<th>Sometimes or Somewhat False</th>
<th>Sometimes or Somewhat True</th>
<th>Very True or Often True</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>People would describe me as reckless.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>I feel like I act totally on impulse.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Even though I know better, I can’t stop making rash decisions.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>I often feel like nothing I do really matters.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Others see me as irresponsible.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>I’m not good at planning ahead.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>My thoughts often don’t make sense to others.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>I worry about almost everything.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>I get emotional easily, often for very little reason.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>I fear being alone in life more than anything else.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>I get stuck on one way of doing things, even when it’s clear it won’t work.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>I have seen things that weren’t really there.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>I steer clear of romantic relationships.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>I’m not interested in making friends.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>I get irritated easily by all sorts of things.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>I don’t like to get too close to people.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>It’s no big deal if I hurt other people’s feelings.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>I rarely get enthusiastic about anything.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>I crave attention.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>I often have to deal with people who are less important than me.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>I often have thoughts that make sense to me but that other people say are strange.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>I use people to get what I want.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>23</td>
<td>I often “zone out” and then suddenly come to and realize that a lot of time has passed.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>Things around me often feel unreal, or more real than usual.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>It is easy for me to take advantage of others.</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Appendix D

International Cognitive Abilities Resource Verbal Reasoning Scale

Please answer the following questions to the best of your ability, without any assistance from others. Please do not use any outside aids, such as a calculator, a dictionary, or the internet to complete this survey. Your compensation will NOT be affected by the correctness of your responses.

1. What number is one fifth of one fourth of one ninth of 900?
   (1) 2   (2) 3   (3) 4   (4) 5   (5) 6   (6) 7   (7) None of these   (8) I don't know

2. Please mark the word that does not match the other words:
   (1) Sycamore (2) Buckeye (3) Elm (4) Daffodil (5) Hickory
   (6) Sequoia (7) They all match (8) I don't know

3. The opposite of a "stubborn" person is a "______" person.
   (1) Flexible (2) Passionate (3) Mediocre (4) Reserved
   (5) Pigheaded (6) Persistent (7) None of these (8) I don't know

4. Michelle likes 96 but not 45; she also likes 540 but not 250. Which does she like?
   (1) 86   (2) 93   (3) 98   (4) 128   (5) 132   (6) 140   (7) None of these   (8) I don't know

5. Adam and Melissa went fly-fishing and caught a total of 32 salmon. Melissa caught three times as many salmon as Adam. How many salmon did Adam catch?
   (1) 7   (2) 8   (3) 9   (4) 10   (5) 11   (6) 12   (7) None of these   (8) I don't know

6. Zach is taller than Matt and Richard is shorter than Zach. Which of the following statements would be most accurate?
   (1) Richard is taller than Matt
   (2) Richard is shorter than Matt
(3) Richard is as tall as Matt
(4) It's impossible to tell
(5) Richard is taller than Zach
(6) Zach is shorter than Matt
(7) None of these
(8) I don't know

7. Joshua is 12 years old and his sister is three times as old as he. When Joshua is 23 years old, how old will his sister be?
   (1) 35  (2) 39  (3) 44  (4) 47  (5) 53  (6) 57  (7) None of these  (8) I don't know

8. The sixth month of the year is:
   (1) September  (2) July  (3) May  (4) August
   (5) June  (6) April  (7) None of these  (8) I don't know

9. If the day after tomorrow is two days before Thursday then what day is it today?
   (1) Friday  (2) Monday  (3) Wednesday  (4) Saturday
   (5) Tuesday  (6) Sunday  (7) None of these  (8) I don't know

10. Please mark the word that does not match the other words:
    (1) Buenos Aires  (2) Melbourne  (3) Seattle  (4) Cairo
    (5) Morocco  (6) Milan  (7) None of these  (8) I don't know

11. The opposite of an "affable" person is a(n) "_____" person.
    (1) Angry  (2) Sociable  (3) Gracious  (4) Frustrated
    (5) Reserved  (6) Ungrateful  (7) None of these  (8) I don't know
12. Isaac is shorter than George and Phillip is taller than George. Which of the following statements is most accurate?
   (1) Phillip is taller than Isaac
   (2) Phillip is shorter than Isaac
   (3) Phillip is as tall as Isaac
   (4) It is impossible to tell
   (5) Isaac is taller than George
   (6) George is taller than Phillip
   (7) None of these
   (8) I don't know

13. If the day before yesterday is three days after Saturday then what day is today?
   (1) Thursday (2) Saturday (3) Wednesday (4) Friday
   (5) Sunday (6) Tuesday (7) None of these (8) I don't know

14. The opposite of an "ambiguous" situation is a(n) "_____" situation.
   (1) suspicious (2) vague (3) unequivocal (4) intelligent
   (5) dubious (6) genuine (7) None of these (8) I don't know

15. How many total legs do three cows and four chickens have?
   (1) 16 (2) 18 (3) 20 (4) 21 (5) 22 (6) 24 (7) None of these (8) I don't know

16. The 4th planet from the sun is:
   (1) Jupiter (2) Saturn (3) Pluto (4) Earth
   (5) Mars (6) Venus (7) None of these (8) I don't know
Appendix E

These items will follow the administration of the three self-report surveys. The first question will be asked on the first survey administration only, while the second question will be asked at the conclusion of both administrations.

Prior to this study, did you have prior exposure to any of the surveys that you just completed? That is, had you ever seen or taken any of them? Please check all that apply.

☐ Positive Affect Negative Affect Schedule (PANAS)
☐ Personality Inventory for DSM-5—Brief Form (PID-5-BF)
☐ International Cognitive Abilities Resource Verbal Reasoning Scale

Realistically, I know some MTurk respondents do not pay close attention to the questions they are answering. This affects the quality of my data. Please select one of the following honestly. Your answer is confidential. It will not affect whether or not you receive payment and will not affect any rating given to you for your work. Did you pay attention and answer honestly?

☐ Yes, keep my data
☐ No, delete my data