Is Amazon’s Mechanical Turk (MTurk) a Comparable Recruitment Source for Trauma Studies?
Krista Engle, Margaret Talbot, and Kristin W. Samuelson
Online First Publication, August 5, 2019. http://dx.doi.org/10.1037/tra0000502
Is Amazon’s Mechanical Turk (MTurk) a Comparable Recruitment Source for Trauma Studies?

Krista Engle, Margaret Talbot, and Kristin W. Samuelson
University of Colorado at Colorado Springs

Objective: Trauma researchers have recently begun using Amazon’s Mechanical Turk (MTurk) as a data collection platform that is both time- and cost-efficient. Research is needed to determine the utility, generalizability, and validity of MTurk as a recruitment source for trauma-exposed samples. Method: Data were collected from 266 trauma-exposed MTurk participants on several clinical and psychological constructs relevant to trauma research. The mean scores, prevalence rates, and correlation strengths of the MTurk sample were compared to those reported in previously published studies of undergraduate, community, and treatment-seeking samples. Results: Findings indicated that prevalence rates of post-traumatic stress disorder (PTSD) and depression were not significantly different from comparison samples, but prevalence rates of generalized anxiety were significantly higher than that of a community sample. The MTurk sample showed significantly lower mean scores of PTSD, depression, and generalized anxiety symptoms than all comparison samples. Correlations were examined to determine whether established relationships between common trauma-related constructs were correlated for MTurk participants as they were in other samples. Finally, MTurk participants who met criteria for probable PTSD scored significantly higher on measures of depression, generalized anxiety, and PTSD, and lower on CSE-T, than those without probable PTSD. Conclusions: Future trauma researchers utilizing MTurk should consider potential similarities and differences between MTurk samples and community, clinical, and undergraduate samples when interpreting the generalizability of findings.

Clinical Impact Statement
Findings from the present study may be relevant to researchers interested in using MTurk to study clinical phenomena related to trauma. Recruitment sources may lead to differences in the clinical symptoms reported by research participants after trauma exposure, which researchers should consider when making inferences about the clinical application of their findings. However, the present study also supports that MTurk participants’ self-report data are similar in important ways to the data provided by participants recruited from more traditional sources. Utilizing MTurk more frequently could increase the rate at which clinical research is conducted and the findings applied to real-world settings.

Keywords: MTurk, trauma, PTSD, depression, anxiety

Traditionally, psychology researchers have utilized convenience samples such as undergraduate subject pools to collect data. This data collection method poses a limitation for clinical research, including trauma studies, as the undergraduate student population is not necessarily representative of clinical populations. Collecting data from clinical populations, however, is typically time-consuming, financially burdensome, and oftentimes not a viable option for researchers. Additionally, the range of traumatic experiences of clinical populations within a given research study is often limited due to the convenience of recruiting from a single site (e.g., a Department of Veterans Affairs hospital or domestic violence shelter). This recruitment tactic may limit how generalizable the results are to all trauma survivors and prevents the comparison of different trauma types within a single study.

An increasingly popular alternative to these more traditional data collection methods is Amazon’s Mechanical Turk (MTurk), an online platform through which participants complete surveys for small incentives. MTurk facilitates data collection from a large, diverse sample (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). MTurk participants appear to be more representative of the general population than traditional
college samples. Furthermore, MTurk participants have been found to be older, to have more work experience, and to be more ethnically diverse than undergraduates (Behrend, Sharek, Meade, & Wiebe, 2011). Researchers have observed that MTurk participants are less susceptible to socially desirable reporting (Aust, Diedenhofen, Ullrich, & Musch, 2013). This may be particularly important for researchers collecting participants’ trauma histories, which can be stigmatizing (Miller, Canales, Amacker, Backstrom, & Gidycz, 2011; van Stolk-Cooke et al., 2018).

MTurk participants have also demonstrated equal or higher levels of effort compared to participants recruited elsewhere. MTurk participants were as likely as undergraduates to answer all questions of a survey, and they showed a similar catch-trial failing rate (e.g., how often participants incorrectly answer questions that ask participants to choose specific answers; Paolacci et al., 2010). The amount of compensation received by MTurk participants does not appear to significantly impact the quality of data collected, as internal consistency levels were similar across three compensation levels (Buhrmester et al., 2011). Test–retest reliabilities ranged from good to excellent and were comparable to those observed in samples collected using traditional methods (Buhrmester et al., 2011).

MTurk has already been used to conduct large-scale studies of trauma-related disorders. Two studies found probable posttraumatic stress disorder (PTSD) rates of 22% (based on a Posttraumatic Stress Disorder Checklist for DSM–5 (PCL-5) score ≥ 38; Frewen, Brown, Steuve, & Lanius, 2015) and 23% (based on PCL-5 ≥ 33; Price, Legrand, Brier, & Hébert-Dufresne, 2019) among trauma-exposed adults recruited through MTurk. Additionally, 32% of Price et al.’s (2019) study sample met criteria for probable depression and 27% met criteria for generalized anxiety disorder (GAD). Frewen, Brown, and Lanius (2017) also found mean PCL scores between 20.93 and 23.44 in three MTurk samples. Although trauma researchers are already utilizing MTurk, there is a continued need to assess the generalizability of MTurk samples and whether their clinical characteristics are comparable to other recruitment sources.

Several studies have examined clinical characteristics of MTurk participants and compared them to national prevalence rates. Shapiro, Chandler, and Mueller (2013) assessed stressful life events and associated clinical symptoms in 443 MTurk participants in the U.S. In some ways, rates of clinical syndromes were comparable to rates in the general population; 66% of MTurk participants reported exposure to trauma, 5% reported depression, and 2% reported physiological anxiety, all rates similar to those of the general population. In contrast, 51% of the sample reported clinical levels of social anxiety, and 37% screened positive for possible substance abuse, rates higher than in the general U.S. population (Kessler et al., 2005). In a sample of 1,098 MTurk participants, Arditte, Çek, Shaw, and Timpano (2016) found elevated rates of clinically significant social anxiety symptoms (49%), depression (32%), and physiological anxiety (28%) compared to previously reported nonclinical means. These results suggest that MTurk participants exhibit anxiety and depression symptoms at higher rates than community or epidemiological samples.

Van Stolk-Cooke and colleagues (2018) recently compared the prevalence rates of PTSD, depression, anxiety, and the mean PTSD symptom severity reported by 822 trauma-exposed MTurk participants to the prevalence rates and mean scores of other samples. The rate of PTSD in the MTurk sample was higher than the PTSD rate reported in a systematic review of epidemiological studies (Santiago et al., 2013). Twenty percent of the MTurk sample met criteria for a probable PTSD diagnosis, defined as a PCL-5 score of 33 or higher. The mean PTSD symptom severity for the MTurk sample was significantly lower than that of a veteran sample (Bovin et al., 2016), but significantly higher than that of an undergraduate sample (Blevins, Weathers, Davis, Witte, & Domino, 2015). They also found that depression and anxiety rates (28% and 22%, respectively) were significantly higher than those in the general population (Pietrzak, Goldstein, Southwick, & Grant, 2011).

Possibly more informative than comparing the prevalence rates of various disorders, however, is determining whether established relationships between clinical constructs are similarly correlated for MTurk participants and other samples. For example, research has documented relationships between PTSD symptoms and both posttraumatic cognitions (PTCs) and trauma coping self-efficacy (CSE-T). Dysfunctional PTCs have been found to mediate the relationship between trauma exposure and the subsequent development of PTSD (Foa & Rothbaum, 1998), and PTCs have been found to improve after PTSD treatment (Foa, 1997; Tolin & Foa, 1999). Similarly, CSE-T, or an individual’s perceived ability to manage posttrauma recovery demands (Benight & Bandura, 2004), is central to posttrauma adaptation because it provides a sense of control within an unpredictable environment. The predictive capacity of CSE-T for posttrauma recovery has been demonstrated in a variety of populations (e.g., Benight, Harding-Taylor, Midboe, & Durham, 2004). Understanding the relationships of clinical symptoms with outcomes like CSE-T and PTCs is relevant to identifying the practical implications of trauma exposure. MTurk would have to provide similar strengths of relationships between PTSD and constructs related to trauma exposure for it to be a valid data collection platform. Finally, establishing MTurk’s validity for studying posttrauma adaptation requires that researchers are able to differentiate between participants with and without PTSD. Data collected from participants with PTSD should be significantly different from those without PTSD.

The purpose of the present study is to extend prior research, which has only compared trauma-exposed MTurk samples and other convenience samples on mean PTSD symptom severity and prevalence rates of various diagnoses. We assessed constructs of importance to the trauma research community and systematically examined four characteristics of the current MTurk sample. First, in order to replicate prior findings of van Stolk-Cooke et al. (2018), we compared the prevalence rates of probable PTSD, depression, and generalized anxiety in the MTurk sample to rates of these disorders found in other samples. We hypothesized that the MTurk sample would display higher rates of PTSD, depression, and generalized anxiety symptoms compared to community and undergraduate samples given previous findings (Arditte et al., 2016; van Stolk-Cooke et al., 2018). MTurk participants may also be more likely to endorse symptoms in the absence of stigma related to reporting trauma histories and related outcomes in person. Second, we compared the mean scores of the MTurk participants to those of undergraduate and treatment-seeking samples on measures of PTSD, depression, and generalized anxiety symptoms. We hypothesized that the MTurk participants would display sim-
ilar mean scores to undergraduate samples, but lower mean scores than treatment-seeking samples. Given that they are seeking care for impairing symptoms, treatment-seeking samples are likely to be experiencing more severe trauma symptoms. Third, we compared the correlations between PTSD symptom severity and measures of posttrauma adaptation (i.e., PTCs and CSE-T) in the MTurk sample to treatment-seeking and undergraduate samples. We expected that the correlation strengths in the MTurk sample would not be significantly different from those in published studies regardless of recruitment source. Fourth, we divided the MTurk participants into two groups: those who met criteria for probable PTSD and those who did not meet these criteria. We hypothesized that MTurk participants who met criteria for probable PTSD would show significantly more impaired scores on clinical and post-trauma adaptation variables than MTurk participants without PTSD.

Method

Participants

All study procedures were approved by the Institutional Review Board at University of Colorado at Colorado Springs. Data from 266 trauma-exposed participants recruited via MTurk were analyzed. The data were previously published as part of an investigation of predictors of perceptions of cognitive problems in trauma-exposed adults; PCL-5, CSE-T, and PTCI scores and their relationship to cognitive problems were reported in that paper (Samuelson, Bartel, Valadez, & Jordan, 2017; see Appendix). Participants were required to be at least 18 years old, be fluent in English, live in the United States, and endorse the following question:

Have you ever experienced a stressful or traumatic life event where you believed your life was in danger, or the life of a loved one was in danger, or you were in danger of being seriously injured? For example, a serious accident, disaster, physical or sexual assault, combat, domestic violence, or childhood physical or sexual abuse.

Of the initial 470 adults who took the survey, 378 responded “yes” to this question endorsing a trauma history and subsequently completed an online consent form and self-report measures. They were compensated $2.50 for completing the study measures, which took between 30 and 60 min to complete. Prior to completion of the PCL-5, participants were asked to describe their worst traumatic event and respond to the PCL-5 with that event in mind. These event descriptions were reviewed by the authors to verify that they met Criterion A of the PTSD diagnosis in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM–5; American Psychiatric Association, 2013). Analysis of participants’ responses to the write-in question also helped eliminate bots and invalid completers. This resulted in a total of 266 participants included in the current analyses.

Published means, standard deviations, and clinical cutoffs for each of the measures of interest were identified in comparison studies. Samples were selected if they had a large N (greater than 100 participants); were from the United States; were drawn from an undergraduate, treatment-seeking, or community sample; and used the same measures as the current study.

Demographics for the current sample are presented in Table 1. Participants primarily identified as male, Caucasian, and not Hispanic or Latino. The mean age of participants was 34.04 years (SD = 9.55). The worst traumatic event reported by each participant was categorized into one of five categories: accidents and disasters (39%), adult interpersonal violence (17%), child maltreatment (15%), combat-related trauma (2%), or other types of trauma (27%).

Measures

Brief Trauma Questionnaire (BTQ). The BTQ (Schnurr, Spiro, Vielhauer, Findler, & Hamblen, 2002) is a 10-item self-report measure of exposure to traumatic events and assesses whether the individual was seriously injured or believed that his or her life was threatened during these events. The BTQ has been found to have strong reliability and validity (Schnurr et al., 2002).

Following administration of the BTQ, participants were instructed...
to describe the worst traumatic event in an open-ended response format.

**PTSD Symptom Checklist for DSM–5 (PCL-5).** The PCL-5 (Weathers et al., 2013) is a 20-item self-report measure of PTSD symptom severity. Participants provide responses on a scale of 0 (not at all) to 4 (extremely) indicating how often they experienced symptoms of PTSD in the past 30 days, allowing for a total score between 0 and 80. The PCL-5 has strong internal consistency and convergent validity (Blevins et al., 2015).

**Patient Health Questionnaire (PHQ-9).** The PHQ-9 (Kroenke, Spitzer, & Williams, 2001) is a self-report measure of depressive symptoms reflecting the criteria for major depressive disorder included in the DSM–5 (American Psychiatric Association, 2013). Participants report how often within the past two weeks they have been bothered by the symptoms described in each question by selecting a score between 0 (not at all) and 3 (nearly every day). Total scores can range from 0 to 27. The PHQ-9 has been shown to have excellent internal consistency and test–retest reliability, and strong construct and criterion validity (Kroenke et al., 2001).

**Generalized Anxiety Disorders Questionnaire (GAD-7).** The GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) is a 7-item self-report measure of symptoms of GAD (American Psychiatric Association, 2013). Participants rate on a scale from 0 (not at all) to 3 (nearly every day) how often they have been bothered by each symptom associated with GAD in the past two weeks. The GAD-7 has been shown to have strong criterion, construct, and discriminant validity; excellent internal consistency; and good test–retest reliability (Spitzer et al., 2006).

**Posttraumatic Cognitions Inventory (PTCI).** The PTCI (Foa, Ehlers, Clark, Tolin, & Orsillo, 1999) is a 33-item measure of trauma-related cognitions about the self, the world, and self-blame. Sample items include “I am a weak person” (self), “The world is a dangerous place” (world), and “Somebody else would have stopped the event from happening” (self-blame). Participants answer using a 7-point Likert scale (1 [totally disagree] to 7 [totally agree]). The PTCI has excellent internal consistency, test–retest reliability, and validity (Foa et al., 1999).

**Trauma Coping Self-Efficacy Scale.** The Trauma Coping Self-Efficacy Scale (Benight et al., 2015) is a 9-item measure that assesses self-perception of trauma-related coping self-efficacy. Items follow a prompt asking respondents to rate how they feel on a scale of 1 (I’m not at all capable) to 7 (I’m totally capable), for instance, “Accep[t]ing what happened” and “Manag[ing] distressing dreams or images about the trauma.” Test–retest reliability, internal consistency, and convergent and discriminant validity have been established (Benight et al., 2015).

**Data Analyses**

Statistical Package for the Social Sciences (SPSS) Version 25.0 was used for analyses. First, chi-square tests were used to compare the prevalence rates of probable PTSD, depression, and anxiety in the current MTurk sample and published comparison samples. Next, independent t tests were run to compare the mean scores of the MTurk sample for the primary variables of interest (i.e., PTSD, depression, and generalized anxiety symptom severity) to those of undergraduate and treatment-seeking samples. Cronbach’s alphas for the primary measures were also compared using chi-square tests to determine if the internal consistency of measures was comparable in the MTurk and published samples. Third, Fisher r-to-z transformations were conducted to compare the correlation strengths between PTSD symptom severity and key posttrauma adaptation variables in the MTurk, undergraduate, and treatment-seeking samples. Lastly, independent t tests were used to compare participants from the current sample who met criteria for probable PTSD (defined as PCL-5 scores ≥ 33) to participants who did not meet criteria for probable PTSD on the primary variables. Due to the number of comparisons planned, a significance level of p < .01 was used in order to decrease the chance of Type I error.

Prevalence rates of probable PTSD, depression, and generalized anxiety were calculated for the current MTurk sample using the clinical cutoff scores in the comparison studies. Santiago et al. (2013) and Blevins et al. (2015) used PCL-5 cutoff scores of 33 and 37, respectively. Both 33 (Bovin et al., 2016) and 37 (Blevins et al., 2015) have been shown to differentiate between those with and without probable PTSD. Pietrzak et al. (2011) and Felix, Dowdy, and Green (2018) used a PHQ cutoff score of 10, which has adequate sensitivity and specificity (Manea, Gilbody, & McMillan, 2012). Pietrzak et al. (2011) used a GAD cutoff score of 10, whereas Felix and colleagues (2018) used 9. Spitzer et al. (2006) found a cutpoint of 10 to show ideal specificity and sensitivity. However, cutoff scores between 7 and 10 were found to have similar specificity and sensitivity (Plummer, Manea, Trespel, & McMillan, 2016).

Prior to data analysis, missing data were managed using person-mean imputation; average scores on completed items replaced missing data points. Person-mean imputation was chosen due to the low proportion of missing data. No participants endorsed only the extreme ends of all the measures of interest. This suggests that no one simply “clicked through” the questionnaires without reading each item.

**Results**

**Comparison of Prevalence Rates**

**PTSD prevalence rates.** Santiago and colleagues (2013) conducted a review of multiple studies of community samples. Some samples comprised community-recruited participants, whereas others comprised individuals seeking medical treatment. They found a probable PTSD prevalence rate of 17%. The current MTurk sample had a probable PTSD prevalence rate of 11%. These prevalence rates were not significantly different, χ²(1) = 1.50, p = .22. The undergraduate comparison sample was 278 psychology students who had experienced at least one traumatic event (Blevins et al., 2015). Applying their more stringent cutoff criteria, 10% of Blevins et al.’s (2015) undergraduate sample and 8% of the current MTurk sample met criteria for probable PTSD. These prevalence rates were not significantly different, χ²(1) = 0.05, p = .82.

**Depression prevalence rates.** Pietrzak et al. (2011) collected data from a community sample of trauma-exposed adults and found a probable depression prevalence rate of 17% for the PHQ-8, a modified version of the PHQ-9 that excludes the last item of the questionnaire. This scoring method resulted in a prevalence rate of 21% in the current MTurk sample. These rates were not significantly different, χ²(1) = 0.52, p = .47. The probable depression prevalence rate of 26% in the MTurk sample was not
significantly different from the prevalence rate of 20% in 593 undergraduates (Felix et al., 2018), χ^2(1) = 1.02, p = .31.

**Generalized anxiety prevalence rates.** Pietrzak et al. (2011) found a clinical anxiety prevalence rate of 8% in their community sample using the GAD-7. The MTurk sample had a prevalence rate of 22%, which was significantly higher than that of the community sample, χ^2(1) = 7.69, p = .006. The MTurk sample’s prevalence rate was not significantly different from an undergraduate sample (Felix et al., 2018; 25% and 17%, respectively), χ^2(1) = 1.93, p = .16.

**Comparison of Mean Scores**

The assumptions for running independent t tests were met. Each of the variables included in the analyses was approximately normally distributed. The skewness values (−0.63 to 1.40) and kurtosis values (−0.62 to 1.31) were within reasonable limits of normality. Table 2 includes descriptive statistics for the PCL-5, PHQ-9, GAD-7, PTCI, and CSE-T of the current MTurk sample; the means, standard deviations, effect sizes, and alphas for the PCL-5, PHQ-9, and GAD-7 of the comparison samples; and the results of the independent t tests and chi-square tests conducted to compare these samples’ clinical characteristics.

**PCL-5.** The MTurk sample’s PCL-5 internal consistency was not significantly different from that of a trauma-exposed undergraduate sample studied by Blevins et al. (2015). The MTurk sample’s mean PCL-5 score was significantly lower than that of Blevins et al.’s (2015) undergraduate sample. However, the effect size for this difference was small. The comparison treatment-seeking sample consisted of 119 trauma-exposed adults seeking mental health care for the first time (Erwin et al., 2018). The MTurk sample’s mean score was significantly lower than the mean score of this treatment-seeking sample. The effect size of this difference was large.

**PHQ-9.** The PHQ-9 internal consistencies for the current MTurk and comparison samples were not significantly different. The MTurk sample’s mean PHQ-9 score was significantly lower than that of 202 trauma-exposed undergraduate students (Van Dusen, Tiamiyu, Kashdan, & Elhai, 2015), with a small effect size. The mean score of the MTurk sample was also significantly lower than that of a trauma-exposed treatment-seeking sample (Erwin et al., 2018), which was a medium to large effect size.

**GAD-7.** The undergraduate comparison sample consisted of 251 trauma-exposed students (Blylesby, Charak, Durham, Wang, & Elhai, 2016). The internal consistency of the GAD-7 for the undergraduate sample was significantly lower than that of the MTurk sample. The MTurk sample’s mean GAD-7 score was significantly lower than the undergraduate sample’s mean score, representing a medium effect size. The comparison treatment-seeking sample comprised 178 trauma-exposed adults seeking care at health or mental health clinics (Ghafoori, Barragan, & Palinkas, 2014). The treatment-seeking sample had significantly lower internal consistency for the GAD-7. The MTurk sample scored significantly lower on the GAD-7 than the treatment-seeking sample, which corresponds to a medium to large effect size.

**Correlation Strength Comparisons for PCL-5, CSE-T, and PTCI**

Fisher r-to-z test transformations were used to compare the correlation strengths between PTSD symptom severity (PCL-5 or PCL-C), PTCIs, and CSE-T in the MTurk sample and other trauma-exposed samples.

**PCL-5 and PTCI.** The correlation between the PCL-5 and PTCI was not significantly stronger in the current MTurk sample (r = .67) than in 500 trauma-exposed undergraduate students (Barton, Boals, & Knowles, 2013), r = .58, z = 1.95, p = .03. The correlation strength in the MTurk sample was also not significantly different from the strength in a treatment-seeking sample of outpatient abuse survivors (Barton et al., 2013), r = .67 versus .64, z = 0.34, p = .37. Similar correlation strengths were also observed when compared to a sample of hospitalized trauma patients (Be-night et al., 2015), r = .67 versus .76, z = −1.39, p = .08.

**PCL-5 and CSE-T.** The correlation strength of the PCL-5 and CSE-T in the current sample (r = .48) was not significantly different from that between the PCL-C and CSE-T of a trauma-

<table>
<thead>
<tr>
<th>Measure/sample</th>
<th>N</th>
<th>M (SD)</th>
<th>α</th>
<th>t</th>
<th>df</th>
<th>p</th>
<th>Cohen’s d</th>
<th>χ^2(1)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PCL-5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTurk</td>
<td>266</td>
<td>11.61 (13.65)</td>
<td>.95</td>
<td>.94</td>
<td>3.13</td>
<td>542</td>
<td>.002</td>
<td>.22</td>
<td>2.04</td>
</tr>
<tr>
<td>Undergraduate⁶</td>
<td>278</td>
<td>15.42 (14.72)</td>
<td>.94</td>
<td>6.74</td>
<td>383</td>
<td>&lt;.001</td>
<td>.74</td>
<td>8.98</td>
<td>.003</td>
</tr>
<tr>
<td>Treatment-seeking⁶</td>
<td>119</td>
<td>24.05 (22.13)</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PHQ-9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTurk</td>
<td>266</td>
<td>6.33 (6.14)</td>
<td>.91</td>
<td>.88</td>
<td>2.84</td>
<td>466</td>
<td>.005</td>
<td>.26</td>
<td>3.82</td>
</tr>
<tr>
<td>Undergraduate⁶</td>
<td>202</td>
<td>7.93 (5.92)</td>
<td>.88</td>
<td>5.20</td>
<td>383</td>
<td>&lt;.001</td>
<td>.58</td>
<td>.45</td>
<td>.50</td>
</tr>
<tr>
<td>Treatment-seeking⁶</td>
<td>119</td>
<td>9.99 (6.89)</td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GAD-7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTurk</td>
<td>266</td>
<td>5.56 (5.62)</td>
<td>.94</td>
<td>.90</td>
<td>5.13</td>
<td>515</td>
<td>&lt;.001</td>
<td>.45</td>
<td>12.48</td>
</tr>
<tr>
<td>Undergraduate⁶</td>
<td>251</td>
<td>8.10 (5.63)</td>
<td>.90</td>
<td>7.02</td>
<td>442</td>
<td>&lt;.001</td>
<td>.68</td>
<td>24.34</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Treatment-seeking⁶</td>
<td>178</td>
<td>9.40 (5.70)</td>
<td>.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* PCL-5 = PTSD Checklist for DSM-5; PHQ-9 = 9-item Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder 7-item Scale; chi-square tests were used to compare the Cronbach’s alphas of the PCL-5, PHQ-9, and GAD-7 for the MTurk sample and the published comparison samples.

exposed undergraduate sample (Benight et al., 2015), $r = -.60$, $z = 1.90$, $p = .03$, or of a sample of hospitalized trauma patients (Benight et al., 2015), $r = -.66$, $z = 2.02$, $p = .02$.

**Probable PTSD Versus No Probable PTSD Comparisons**

The MTurk participants were divided into those who met criteria for probable PTSD (PCL-5 $\geq 33$) and those who did not. The PTSD-positive participants had significantly higher scores on the PHQ-9, GAD-7, and PTCI, and significantly lower scores on the CSE-T than the PTSD-negative individuals (see Table 3).

**Discussion**

We sampled 266 trauma-exposed MTurk participants in order to examine the comparability of an MTurk sample to typical trauma study recruitment sources (i.e., treatment-seeking, community, and undergraduate samples). First, we compared prevalence rates of PTSD, depression, and generalized anxiety in our MTurk sample to samples from more traditional recruitment sources. PTSD and depression rates were similar across samples. The MTurk sample’s rate of generalized anxiety was significantly higher than that of a community sample, but not significantly different from that of an undergraduate sample. Although prevalence rates were similar, the MTurk sample had significantly lower mean scores of PTSD, depression, and anxiety compared to treatment-seeking and undergraduate samples.

We extended previous MTurk comparison study findings (i.e., Arditte et al., 2016; Shapiro et al., 2013; van Stolk-Cooke et al., 2018), which have primarily compared prevalence rates of various disorders between samples. Confirming our hypothesis for the similarity of samples, we did not find statistically different prevalence rates of various disorders between samples. Comparing our MTurk sample to previous findings, however. Specifically, van Stolk-Cooke et al. (2018) and Arditte et al. (2016) found higher rates of depression in their MTurk samples compared to undergraduate samples. Van Stolk-Cooke et al. (2018) also found that the PTSD mean score of their sample was higher than that of a college student sample.

In regard to the differences between the current and undergraduate samples, the present MTurk participants were likely older than typical undergraduate participants. Thus, time since the worst traumas may have been greater for the MTurk participants. Greater time since trauma is associated with lower levels of clinical symptoms (Ozer, Best, Lipsey, & Weiss, 2003). The differences between the current findings and those of Arditte et al. (2016) could be due to the differences in sampling methods. The MTurk sample collected by Arditte and colleagues (2016) was a general sample; they were not selected due to their trauma history as in the current study.

Several findings were consistent with results of previous MTurk analyses. Van Stolk-Cooke and colleagues (2018) found that their mean PTSD score was significantly lower than that of a treatment-seeking veteran sample. Van Stolk-Cooke et al. (2018) and Arditte et al. (2016) both found higher rates of anxiety in their MTurk samples compared to undergraduate and community samples. Finally, Shapiro et al. (2013) found no significant differences in the depression prevalence rates of their MTurk sample and a treatment-seeking sample. The present findings contradicted some previous findings, however. Specifically, van Stolk-Cooke et al. (2018) and Arditte et al. (2016) found higher rates of depression in their MTurk samples compared to undergraduate samples. Van Stolk-Cooke et al. (2018) also found that the PTSD mean score of their sample was higher than that of a college student sample.

### Table 3

**Comparison of MTurk Participants With and Without Probable PTSD**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Probable PTSD ($n = 30$)</th>
<th>Non-Probable PTSD ($n = 236$)</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9</td>
<td>9.13 (6.86)</td>
<td>5.98 (5.97)</td>
<td>2.68</td>
<td>.008</td>
</tr>
<tr>
<td>GAD-7</td>
<td>8.57 (6.10)</td>
<td>5.18 (5.45)</td>
<td>3.16</td>
<td>.002</td>
</tr>
<tr>
<td>PTCI</td>
<td>143.27 (31.73)</td>
<td>81.12 (36.36)</td>
<td>8.94</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CSE-T</td>
<td>44.07 (13.87)</td>
<td>54.63 (12.10)</td>
<td>-4.43</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*Note.* Participants met criteria for probable PTSD if they received a score of $\geq 33$ on the PTSD Checklist for DSM-5 (PCL-5). PHQ-9 = 9-item Patient Health Questionnaire; GAD-7 = Generalized Anxiety Disorder 7-item Scale; PTCI = Posttraumatic Cognitions Inventory; CSE-T = Trauma Coping Self-Efficacy.
Criterion A. This procedure resulted in the exclusion of 21% (Price et al., 2019) and 30% (current study) of the original MTurk samples. This process allows for careful exclusion of participants who cannot meet PTSD criteria. Additionally, the MTurk participants in the current study and Price et al.’s (2019) study could have experienced different types of trauma (although not reported in the latter study). The most common trauma type reported in the present study was accidents and disasters (39%), which are associated with lower PTSD symptom severity relative to other types of trauma (Kelley, Weathers, McDevitt-Murphy, Eakin, & Flood, 2009). In order to address generalizability of findings in future research, we recommend that researchers carefully describe the trauma types endorsed by their participants.

Differences in prevalence rates would be of greater concern for epidemiological research than for studies examining the relationships between PTSD and other constructs. Our additional correlation comparisons support the idea that trauma-exposed MTurk participants are similar in important ways to participants recruited via other methods despite the differences in prevalence rates. More specifically, the relationships between PTSD symptom severity and indicators of posttrauma adaptation are similar in the current MTurk sample, undergraduate samples, and treatment-seeking samples.

There are a number of limitations in this study that can be improved upon in future research. First, we did not include consistency checks, completion time monitoring, or instructional manipulation checks, which assist researchers in detecting invalid data. However, this study was not intended to address MTurk issues such as bots and fakers. Second, a measure of social desirability would have confirmed that the current MTurk participants accurately reported their experiences. As prior research on social desirability is inconsistent (see Aust et al., 2013; Behrend et al., 2011), future researchers may benefit from measuring social desirability in MTurk samples. Third, there are a variety of suggested PCL-5 cutoff scores throughout the trauma literature that are population-specific. As such, it is difficult to compare rates of PTSD using cutoff scores across heterogeneous samples. Thus, the comparisons may be influenced by differences in sample characteristics.

Despite these limitations, our findings provide evidence that trauma-exposed MTurk samples are similar in key ways to more traditional samples. Future research should attempt to replicate the current analyses in an effort to clarify discrepancies between the present findings and van Stolk-Cooke and colleagues’ (2018) findings. Future research may also examine additional variables to determine if the findings generalize to other trauma-relevant constructs.

References


Appendix

Narrative Description

The rates of the current MTurk sample have previously been published as part of another study. In the previously published work, the variables analyzed included PTSD symptom severity, PTCs, CSE-T, perceptions of cognitive problems, and quality of life. The study demonstrated that CSE-T and posttraumatic appraisals influenced perceptions of cognitive problems in a traumatized sample (Samuelson et al., 2017). The present study, in contrast, compared some of these variables, including PTSD symptom severity, PTCs, and CSE-T, to other published rates in the literature to determine the comparability of MTurk to other recruitment sources and is thus novel to the present investigation. The relationships that were studied in the present manuscript have not been studied in previously published work.

Received April 15, 2019
Revision received June 30, 2019
Accepted July 3, 2019