

Utilizing short version big five traits on crowdsourcing

Utilizing short
version big
five traits

Kousaku Igawa

Innovation Management, Tokyo Institute of Technology, Tokyo, Japan, and

Kunihiko Higa and Tsutomu Takamiya

Tokyo Institute of Technology, Tokyo, Japan

117

Received 7 November 2019
Revised 19 January 2020
Accepted 21 January 2020

Abstract

Purpose – The purpose of this paper is to examine the efficacy of the Japanese ten-item personality inventory (TIPI-J), a short version of the big five (BF) questionnaire, on crowdsourcing. The BF traits are indicators of personality and are said to be an effective predictor of study performance in various occupations. BF can be used in crowdsourcing to predict crowd workers' performance; however, it will be difficult to use in practice for two reasons like the time-and-effort issue and the bias issue. In this study, an empirical analysis is conducted on crowdsourcing to examine if TIPI-J can solve those issues.

Design/methodology/approach – To investigate the issues, two tasks are posted on a crowdsourcing provider. Both TIPI-J and full version BF are conducted before and after selecting crowd workers. Structural validity and convergence validity are tested with correlation analysis between before (TIPI-J) and after (full version BF) data to examine the bias issue. Additionally, those correlations are compared with previous study and significances are examined.

Findings – The correlations in “conscientiousness” is 0.45-0.50, respectively, compared with a previous study, those two correlations did not show significance. This indicates that no clear bias exists.

Originality/value – This is the first research to investigate the efficacy of TIPI-J on crowdsourcing and showed that TIPI-J can be a useful tool for predicting crowd workers' performance and thus it can help to select appropriate crowd workers.

Keywords Human resource, Quality evaluation, Work performance, Crowdsourcing, Big five, Task assignment

Paper type Research paper

Introduction

Crowdsourcing has been used in various areas and it has been recognized as an effective way of human resource utilization. For example, Upwork, one of the largest crowdsourcing providers (CSP), reported 14 million users in 180 countries with \$1b in annual freelancer billings as of March 2017 (Snagajob, 2017; Brier and Pearson, 2020). Crowd works, that is one of the largest CSP in Japan, has over two million registered crowd workers as of December 2018.

These CSPs provide project works, which have fixed objectives and deadlines. A client who wants to conduct a project work post a task description including outlining the details



© Kousaku Igawa, Kunihiko Higa and Tsutomu Takamiya. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

International Journal of Crowd
Science
Vol. 4 No. 2, 2020
pp. 117-132
Emerald Publishing Limited
2398-7294
DOI 10.1108/IJCS-11-2019-0031

of the task, a proposed reward and a deadline at a CSP site. Registered crowd workers review the posted task and submit bids if they are interested in it. The client selects his/her preferred crowd worker. When the task is completed, the worker delivers the output to the client. This type of crowdsourcing is called project type crowdsourcing. In this type of task, a client will not know if she/he has selected an appropriate worker until she/he actually sees the final output. To help a client to make a well-informed selection, CSPs provide some information of registered crowd workers. Such as profile, task experience, the number of tasks received a few messages and so on. However, because the information provided by CSPs is limited, it will not be easy for clients to make the appropriate selection.

In addition, some CSPs prohibit clients to use undesignated communication tools such as Skype, Google Hangout or phone calls to contact crowd workers to prevent clients from offering tasks to crowd workers without CSPs. In that case, it will become more difficult to select appropriate crowd workers. Moreover, when the posted task is attractive, a lot of crowd workers will apply and it will become even more difficult to select appropriate crowd workers. The quality of the output depends on the clients' ability to make a well-informed selection at the hiring stage. So, it is necessary to find an effective way to help clients to make appropriate selections.

There is a variety of indicators, which can help to predict workers' performance. For example, some researchers showed that the big five (BF) traits are related to work performance (Barrick and Mount, 1991; Barrick *et al.*, 2001; Schmidt and Hunter, 1998; Anderson and Viswesvaran, 1992). Especially one of the BF traits, "conscientiousness" is said to correlate most closely with workers' performance (Barrick *et al.*, 2001).

Regarding crowdsourcing, some research studies addressed to predict workers' performance with BF traits. Kazai *et al.* (2011) investigated the relationship between personality traits and output quality with simple, easy and low compensation tasks such as labeling, dictation and so on called microtask. The results showed some BF traits relate to output quality. Although almost research studies are focusing on microtask type crowdsourcing, Igawa *et al.* (2016) conducted experiments on project type crowdsourcing. They reported "conscientiousness" has a correlation with crowd workers' work performance. Based on these studies, the BF traits, especially "conscientiousness," is an effective indicator to predict crowd workers' work performance.

However, there are some issues applying BF to select appropriate crowd workers (Igawa *et al.*, 2016). One is the time-and-effort issue. It takes around 20-30 min to complete the BF questionnaire because it includes over 70 questions (Murakami and Murakami, 2001). It is difficult for a client to request crowd workers to answer the questionnaire before officially selecting crowd workers. Another is the bias issue. The result of BF scores may be biased if data is collected before selecting crowd workers. Because crowd workers want to show themselves better than they are to be selected. Some previous studies (Kazai *et al.*, 2011; Mourelatos and Tzagarakis, 2016) gathered BF scores before selecting crowd workers and there may be a bias issue. In this paper, Japanese ten-item personality inventory (TIPI-J) (Oshio *et al.*, 2014), that is a questionnaire to evaluate personality with only 10 questions, (therefore, it doesn't have the time-and-effort issue) is conducted and tested whether the result shows bias or not.

In this research, TIPI-J is obtained before selecting crowd workers and the BF traits (Murakami and Murakami, 2001) are conducted after selecting crowd workers as a full version of Big five (full BF). Correlations between TIPI-J and full BF are analyzed and compared with the previous study (Oshio *et al.*, 2014) to understand whether biases exist or not.

Related works

Crowdsourcing

The term “crowdsourcing” was first defined by [Jeff Howe \(2006\)](#). As the number of definitions were increased, [Enrique et al. \(2012\)](#) reported that there were more than 40 definitions about “crowdsourcing.” In general, crowdsourcing is used as outsourcing to solve problems or to create ideas by asking unspecified crowd workers via the internet.

Although there are many research about crowdsourcing, only a few research reported crowd worker selection. Those research address the issue focus on microtask crowdsourcing. [Kittur et al. \(2008\)](#) investigated crowd worker’s output quality in micro-task crowdsourcing and showed that when a job was posted with some simple check questions, gaming users were eliminated and the quality of output was increased. In view of project type crowdsourcing, [Assemi and Daniel \(2012\)](#) investigated the relationship between the crowd worker’s outcomes and profile information (i.e. the number of portfolio items, verified credentials, average recommendation and average weighted rating). They classified the profile information into two distinct categories as internal and external information. As the result, they showed that the external profile information, which was published by CSP (i.e. the number of portfolio items, skills assessed as top 10 per cent), significantly related to crowd workers’ outcomes, whilst the internal information in their profiles (i.e. ratings) does not significantly related to outcomes. [Gong \(2015\)](#) suggested the statistical model that predicts whether a crowd worker will be chosen by clients. He applied the AHP-based model to CSP information (i.e. the total volume of participants and the number of bidding) and the model was tested on a data set of 348 completed IT service crowdsourcing tasks from Chinese CSP. An analysis of the matching between the test results and the actual selection results was conducted and the result showed the overall matching rate was 88.22 per cent.

Although, these existing research studies have focused on the relationship between CSP information and crowd worker’s outcomes or the prediction of crowd workers selection by clients, research on success or failure of tasks assigned to crowd workers and the quality of selected crowd workers have not been found.

[Igawa et al. \(2016\)](#) investigated the correlation between the quality of output and other indicators such as CSP information, individual work performance (IWP) indicator and BF indicators. As a result, “counter work behavior” of IWP indicators and “consciousness” of BF showed significance. However, they reported that the time-and-effort issue of obtaining BF data makes it difficult to be used in practice.

Big five

During the past two decades, the theory that personality has consisted of five factors has been widely accepted ([Wiggins, 1996](#)). These five factors to describe personality have been reported by many researchers ([Fiske, 1949](#); [Goldberg, 1999](#)). Recently, it is called as “BF” or “five-factor model.” However, there are some differences among researchers, “extraversion,” “conscientiousness,” “emotional stability,” “agreeableness” and “openness to experience” are typically used as BF.

BF found effective in different cultures and with different languages ([Digman and Shmelyov, 1996](#)). [Murakami \(2003\)](#) reported the effectiveness of BF in Japan. He referred to the lexical analysis method of Goldberg ([Goldberg, 1990](#); [Goldberg, 1999](#)) and tested BF to 370 Japanese university students and re-reported the appearance of five factors. In addition, [Murakami and Murakami \(2001\)](#) developed a Japanese BF questionnaire that is consisted of 70 items. The questionnaire was tested for the validity and reliability of 1,166 samples in Japan.

In view of the relationship between personnel selection and BF, several research studies are conducted. In 1991, Barrick and Pearson (1991) conducted meta-analysis about the relationship between work performance and BF. They reported “conscientiousness” showed a significant correlation (0.20-0.23) with all job performance criteria and all occupational groups. In addition, Barrick *et al.* (2001) conducted quantitative summarization of the results of 15 prior meta-analytic studies. The number of samples was over 800 million. This research showed “conscientiousness” had a correlation in all job performance criteria and all occupations and “emotional stability” had correlation in all job performance criteria and some occupations.

These research studies (Barrick and Mount, 1991; Barrick *et al.*, 2001; Schmidt and Hunter, 1998; Anderson and Viswesvaran, 1992) have shown the relation between “conscientiousness” and work performance in wide job performance criteria and occupations and “emotional stability” in some area. In addition, most meta-analyzes have suggested that “conscientiousness” is somewhat more strongly correlated with overall job performance than is “emotional stability.” Moreover, Feist (1998) showed “openness” has a strong link to creativity.

“Conscientiousness” is associated with dependability, achievement striving and planfulness. Especially, it is said that low “conscientiousness” such as careless, irresponsible, lazy, impulsive and low in achievement striving is not beneficial for job performance, and therefore, employees with high scores on “conscientiousness” should also obtain higher performance at work.

Ten item personality inventory

The BF framework has become the most widely used and extensively researched model of personality (John and Srivastava, 1999). However, it has not been used universally because of the time-and-effort issue. The most comprehensive instrument is Costa and Mc Crae’s (Costa and McCrea, 1992) 240-item NEO personality inventory reviser (NEO-PI-R) and it takes about 45 min to complete. NEO-PI-R is too lengthy for many research purpose and some other shorter instruments have been used. Samuel D. Gosling *et al.* (2003) developed a very brief measure of BF, called ten-item personality inventory (TIPI), developed and evaluated in terms of convergence with BF measures, test-retest reliability, patterns of predicted external correlates and convergence between self and observer ratings. TIPI showed adequate levels of each of the criteria. Convergent correlations between TIPI and BF were from 0.65 to 0.87 and all of them were significant. Gosling *et al.* (2003) suggested that TIPI represents a sensible option when the time-and-effort is limited. TIPI takes only a few minutes to complete and it shows adequate levels of the criteria. In Japan, Oshio *et al.* (2014) developed a Japanese version of the TIPI-J based on TIPI (2003).

They examined TIPI-J reliability and validity. The participants were 902 Japanese undergraduates (376 men and 526 women) completed the TIPI-J and one of the other BF scales as follows: BF scale (Wada, 1996); five-factor personality questionnaire (Fujishima *et al.*, 2005); BF scale short version (Uchida, 2002); BF (Goldberg, 1999); or the NEO-FFI (Shimonaka *et al.*, 1999). Convergent correlation between TIPI-J and the other BF scales was investigated. Except for the correlation between TIPI-J and NEO-FFI of “openness,” the rest of all items showed high correlation and significance in 1 per cent. In addition, test-retest reliability was examined. The results generally supported the reliability and validity of the TIPI-J.

Big five on crowdsourcing. The number of research studies addressing to predict crowd workers’ performance is increasing and almost these research studies are using CSP’s information such as demographics (Ross *et al.*, 2010), workers’ gender, age and profession (Downs *et al.*, 2010) and behavioral data (i.e. a number of tasks completed, average time per task, etc.) (Kazai *et al.*, 2011).

Recently because the predictive power is limited by only using CSP's information, there are some research studies having interests in using individual personality to predict task output quality. Kazai *et al.*, 2011) investigated relations between the BF traits and work performance (accuracy) with simple labeling tasks on Amazon Mechanical Turk (MTurk). A total of 155 workers completed them and the results present that "openness" significantly relates to accuracy ($r = 0.19, p < 0.05$), "conscientiousness" and "agreeableness" also have a positive relation to accuracy ($r = 0.10$, not significant and $r = 0.16$, not significant, resp). They mention that "the behavioral data are more effective at distinguishing lower quality workers, but the personality characteristics can be useful to distinguish between good and better workers." Mourelatos and Tzagarakis (2016) investigate relations between the quality of task output and cognitive (i.e. education levels, computer literacy levels and English literacy levels) and non-cognitive (i.e. the BF traits) skill. As the result, "extraversion" shows 5 per cent significance to the quality of task output in all experiment settings and "emotional stability" shows 1 per cent significance to the quality in one specific experiment setting.

In view of competition type crowdsourcing (i.e. idea competitions, business model competitions), Faullant *et al.* (2016) investigate how personality dispositions affect potential workers' decision whether or not to enter crowdsourcing competition. An experiment is conducted on competition crowdsourcing and BF is gathered from both 69 participants and 157 non-participants. As a result, the worker will participate in a crowdsourcing competition that has a significantly high score on "openness" and "extraversion."

In the broader area, there are some studies about the relation between online user behavior and users' personality traits. For example, the relation has been shown in the area of e-commerce (Huang and Yang, 2010) and social media (Gosling *et al.*, 2007).

Recently, personality traits, especially the BF traits, are considered as effective indicators to predict crowd workers' performance besides CSP information. However, there are some studies investigating relations between work performance and personality traits, almost of them are focusing on microtask type crowdsourcing (Kazai *et al.*, 2011; Mourelatos and Tzagarakis, 2016). In addition, these studies gather personality data before officially selecting crowd workers, and therefore, there may be a bias issue that crowd workers want to show themselves better than they are to be selected.

Regarding project type crowdsourcing, Igawa *et al.* conduct experience and gather the BF traits from crowd works. The results show some correlations between output quality and the BF traits. However, the data is gathered after selecting crowd workers and it can not be used when clients select crowd workers. There still remain practical problems.

Research design

It is clear that the quality of the output depends on whether clients make a proper selection of crowd workers. Because of a wide variety of crowd workers, it is getting difficult to make an appropriate selection.

On the hand, there are a number of studies that show the BF traits, especially "conscientiousness," having a correlation with a variety of work performance (Barrick and Mount, 1991; Barrick *et al.*, 2001; Schmidt and Hunter, 1998; Anderson and Viswesvaran, 1992).

Regarding crowdsourcing, because CSP provided information is limited, personality traits, especially the BF traits, are expected as a potential predictor of crowd workers' performance.

However, recent research studies focus on microtask crowdsourcing (Kazai *et al.*, 2011; Mourelatos and Tzagarakis, 2016) and they gather the BF traits before selecting workers. Therefore, there may be a bias issue.

Igawa *et al.* (2016) investigate relations between the BF traits and workers' work performance on project type crowdsourcing and showed "consciousness" related to workers' work performance in crowdsourcing. However, there seem to be some issues such as the time-and-effort issue and the bias issue.

This research is being made to attempt to solve these issues. This section discusses the objective and design of the research.

Research objective

In this study, TIPI-J (Oshio *et al.*, 2014) is used to investigate the time-and-effort issue and the bias issue. TIPI-J consists of only 10 items and it can solve the time-and-effort issue. However, there remains a bias issue. TIPI-J is used to collect the BF traits indicators before selecting crowd workers.

To examine the bias issue, full BF (Murakami and Murakami, 2001) are also collected after officially selecting crowd workers.

Previous studies about relations between TIPI and other BF traits conducted structural validity test, convergent validity test, external test and retest to assess validity (Rammstedt and John, 2007; Oshio *et al.*, 2014).

The structural validity test is to investigate the intercorrelations among the scales of TIPI and other BF traits. Convergence validity test is to calculate correlations between TIPI and other BF traits and check these coefficients. External validity test is using also peer ratings in addition to self-report ratings and retest is to conduct the same BF survey two times some weeks later first survey was conducted and check those correlations.

In this study, structural validity tests and convergent validity tests are conducted to assess validity. Retest and external validity tests are not conducted because it is difficult to hire the same workers once the task is completed and also difficult to ask workers' peers to answer BF surveys in a crowdsourcing setting. In addition, Cronbach's alpha is calculated to assess reliability.

The correlations among the scales of TIPI-J and full BF are calculated for each BF indicators ("extraversion," "conscientiousness," "emotional stability," "agreeableness" and "openness to experience") as an assessment of structural validity.

Then, the correlations between TIPI-J and full BF are calculated for each BF indicators as assessment convergence validity. To increase validity, the correlations are compared to the previous study of TIPI-J (Oshio *et al.*, 2014) and significances between correlations of this study and the previous study are examined.

Procedure

To examine the difference between TIPI-J and full BF, tasks are posted at a CSP in Japan and crowd workers need to complete the TIPI-J questionnaire when they apply. After some crowd workers are officially selected, they need to fill out the full BF questionnaire when they start the task. The correlation between TIPI-J and full BF is examined for each BF indicator.

The brief surveys procedures are as follows:

Posting the task. The task will be posted at a CSP site. The task description includes outlining the details of the task, proposed reward, deadline, etc. It is also indicated that crowd workers who want to apply for this task need to answer the TIPI-J questionnaire.

Obtaining Japanese ten-item personality inventory data. Registered crowd workers who are interested in the task need to submit an application message with a filled TIPI-J questionnaire (Oshio *et al.*, 2014). At this point, as crowd workers have not been selected yet,

there may be a bias issue in the questionnaire data. Crowd workers may answer the questionnaire to show them better than they are.

Selecting crowd workers. Usually, clients select crowd workers by reviewing the crowd workers' information such as profile, skills, experiences, application message and so on. In this research, all crowd workers who agreed with terms like reward or deadline are selected and assigned the task.

Obtaining full big five data. After crowd workers are selected, they are requested to provide full BF (Murakami and Murakami, 2001) data through a web-based questionnaire. At this point, no bias is expected because crowd workers have been already selected, and thus, no need to show themselves better than they are.

Examining correlation among the scales of Japanese ten-item personality inventory data and full big five data for each indicator. To assess structural validity, correlations among the scales of those two data are examined and to investigate intercorrelations with coefficients.

Examining the correlation between Japanese ten-item personality inventory data and full big five data for each indicator. To assess convergence validity, correlations between those two data are examined to verify that TIPI-J scores obtained before selecting crowd workers are significantly different from full BF scores obtained after selecting crowd workers.

Investigating the significance of the difference between this research's tasks and previous study. By comparing the correlation of this research with that of the previous study, the significance of the difference between those correlations is examined. If there is no significance, TIPI-J, which has the potential to forecast full BF at the same level as the previous study, can also be used on crowdsourcing.

Sample and data collection

To increase generality, two tasks are posted at one of the largest CSP in Japan with over two million registered workers. Details of the two tasks are described as follows:

Tasks. Task 1 is to translate Japanese articles into English. The topic of the article is the introduction of Japanese sake (alcohol).

The article is as follow:

“where to start? First, sake is classified into three types as follows: junmaishu, honjozoshu and ginjoshu on the basis of their ingredients and rice-polishing ratio. And they are usually described on product labels and restaurant menus [. . .]”

A crowd worker is asked to deliver a translation, which could be easily understood by English speaking foreigners.

The delivery deadline is one week after the task is assigned and the reward is ¥6,000 (about \$54)

Task 2 is to survey US crowd workers and make short reports about them. A crowd worker assigned this task is asked to visit the US crowdsourcing site (upwork) and check some successful crowd workers' profiles. Then, the crowd worker needs to report the reason why checked workers succeeded in crowdsourcing. The report is 300-400 characters in Japanese and must include 2 or 3 concrete information from the crowd workers' profile.

The delivery deadline is one week after the task is assigned and the reward is ¥4,000 (about \$36).

Japanese version of the ten-item personality inventory. When crowd workers apply for the posted task, they are regulated to fill out TIPI-J questionnaire, which is consisted of a total of 10 questions each by seven-point Likert-scale. TIPI-J questionnaire is described as follows:

I see myself as:

- Extraverted, enthusiastic;
- Critical, quarrelsome;
- Dependable, self-disciplined;
- Anxious, easily upset;
- Open to new experiences, complex;
- Reserved, quiet;
- Sympathetic, warm;
- Disorganized, careless;
- Calm, emotionally stable; and
- Conventional and uncreative.

According to [Oshio et al. \(2014\)](#). TIPI-J shows correlations with full BF ([Murakami and Murakami, 2001](#)) from 0.47 to 0.84 and all correlations showed significance. TIPI-J requires much less time to fill out compared with the full BF questionnaire. In this research, crowd workers answer the TIPI-J questionnaire in the first message when they apply for the task.

Full version of the big five. Questionnaires conceived by [Murakami and Murakami \(2001\)](#) are used as full BF traits (full BF). These are originally written in Japanese, with the accompanying questionnaire covering 70 items such as “If anything, I am lazy,” “I don’t like talking in front of people” and so on. Respondent answers either, “I think so” or “I don’t think so.” If the respondent cannot answer the question, he/she can select “?” which means not applicable. In this research, a full BF questionnaire is conducted at a web site and workers are directed to answer the questionnaire through a web site after crowd workers are selected for the task.

Other information. Some other information is acquired to support analyzing the result.

- The number of tasks completed;
- The number of client ratings received;
- The number of thanks as follows: received from clients, which is similar to a “like” on Facebook;
- The average score of the ratings awarded by clients on a scale of 1-5;
- The number of skills claimed by the crowd workers;
- The number of skills related to the posted task;
- Crowd workers self-assessed the average score of related skills on a scale of 1-5; and
- Period of registration with the CSP.

Data analysis. After TIPI-J and full BF data are gathered from the CSP web site, at first, the Smirnov-Grubbs test is conducted to detect outliers. Respondents can select “?” which means not applicable in full BF. According to [Murakami and Murakami \(2001\)](#), respondents who answer too many n/a are not reliable because there is a possibility that they do not understand questionnaires exactly.

Second, correlations between TIPI-J and full BF are analyzed and tests of no correlation are examined to check significances.

Lastly, the significance of the difference between the correlations of this research and those of the previous study is examined.

For these analyzes, Python 3.6 with libraries such as Numpy, scipy and pandas are used.

Result

Participants

The surveys of Task 1 were conducted in December 2015 and 36 crowd workers completed the task. Then, Task 2 was conducted in June 2017 and 68 crowd workers completed the task.

Table I shows demographic data of the crowd workers who participated in Tasks 1 and 2. Regarding Task 1, nearly 60 per cent of participants were under 40 years old and 75 per cent were female. Nearly, 50 per cent of participants were under 40 years old and about 70 per cent were female in Task 2. There was no big gender and generation variance between Tasks 1 and 2.

Summary of pre/post big five traits

To detect outliers, the Smirnov–Grubbs test was applied to the number of n/a responses for both Tasks 1 and 2 because many n/a are not reliable (Murakami and Murakami, 2001). As a result, one outlier was detected in Task 2 and omitted.

Table II shows the result of pre (TIPI-J) (Oshio *et al.*, 2014)/post (full BF) (Murakami and Murakami, 2001) questionnaire. TIPI-J ranges from 2 to 14 and post BF ranges from 32 to 75.

Comparing five factors between Tasks 1 and 2, all factors score in Task 1 were higher than those of Task 2. Cronbach’s alpha reliabilities for TIPI-J are from 0.47 to 0.73 and the mean is 0.57 and for full BF is from 0.68 to 0.92 and the mean is 0.82.

Compared with previous studies, full BF reliabilities are reported from 0.72 to 0.84 (Murakami and Murakami, 2001). TIPI-J reliabilities are not reported in the study, but TIPI (English version of TIPI) reliabilities are reported from 0.40 to 0.73 (Gosling *et al.*, 2003). TIPI-J reliabilities seem to be unusually low internal consistency because TIPI scales have only two items, but the results are as same as those of the previous study.

Characteristics	Class	Task 1 (<i>n</i> = 36)		Task 2 (<i>n</i> = 68)	
		Frequency	(%)	Frequency	(%)
Age	0-19	0	0.0	1	1.5
	20-29	8	22.2	12	17.6
	30-39	13	36.1	18	26.5
	40-49	9	25.0	26	38.2
	50-59	5	13.9	10	14.7
	60-69	1	2.8	1	1.5
Sex	Male	9	25.0	20	29.4
	Female	27	75.0	48	70.6
Occupation	Part-time	2	5.6	10	14.7
	Student	2	5.6	2	2.9
	Company employee	9	25.0	14	20.6
	Self-employed	9	25.0	19	27.9
	Homemaker	6	16.7	9	13.2
	Other	8	22.2	14	20.6

Table I.
Demographic data of
crowd workers

Correlation among the scales of pre-Japanese ten-item personality inventory and post full big five questionnaire data: structural validity

The results among the scales of pre questionnaire data in Task 1 are from 0.01 to 0.41 and the mean is 0.22 and those in Task 2 are from 0.22 to 0.55 and the mean is 0.41. The results among the scales of pre questionnaire data in Task 1 are from 0.17 to 0.46 and the mean is 0.31 and those in Task 2 are from 0.17 to 0.38 and the mean is 0.29. Regarding previous studies (Rammstedt and John, 2007; Oshio et al., 2014), the correlation of 0.40 is reported as the highest intercorrelations. In this study, the scales of pre questionnaire data in Task 2 are clearly high.

Correlation between pre-Japanese ten-item personality inventory and post (full big five) questionnaire data: convergence validity

Table III shows the correlation between pre and post questionnaire data in Tasks 1 and 2. Except “agreeableness,” four factors of BF showed a significant correlation between pre and post data in Task 1. “Extraversion” showed the highest correlation of five factors in both Tasks 1 and 2. Further, all factors in Task 2 showed significance and “extraversion” showed the highest correlation both Tasks 1 and 2, “agreeableness” showed the lowest correlation.

Comparison with the previous study

Table IV shows the result of the significance of the difference between the correlations of this research and those of the previous study by Oshio et al. (2014). As a result, only “extraversion” of Task 2 shows significance in lower correlation than that of TIPI-J and no other factors show significance.

No, the clear difference is found between this result and the previous study, and therefore, it is concluded that TIPI-J can be used in place of full BF.

Table II.
Summary of pre and post questionnaire data

Big five trait	Task 1 (n = 36)						Task 2 (n = 67)					
	Pre (TIPI-J)			Post (full version)			Pre (TIPI-J)			Post (full version)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Extraversion	9.39	2.91	5-14	53.00	9.88	34-69	8.1	3.1	2-14	47.7	9.40	34-71
Agreeableness	11.72	1.28	9-14	47.47	8.99	32-67	10.4	2.0	4-14	44.9	9.04	21-67
Conscientiousness	10.75	2.21	4-14	58.14	7.14	45-70	8.9	2.5	3-14	53.5	10.72	27-70
Neuroticism	9.58	2.62	4-14	49.67	10.61	31-66	8.9	2.6	2-14	48.3	9.52	31-66
Openness	10.42	2.58	2-14	56.03	7.68	38-75	9.2	2.8	2-14	54.8	8.90	32-75

Table III.
Summary of pre and post big five trait

Big five trait	Task 1 (n = 36)	Task 2 (n = 67)
Extraversion	0.77**	0.71**
Agreeableness	0.31	0.44**
Conscientiousness	0.45**	0.50**
Neuroticism	0.56**	0.59**
Openness	0.58**	0.50**

Notes: * $p < 0.5$; ** $p < 0.01$

Discussion

In this research, the bias issue of the pre-task questionnaire was investigated for TIPI-J in crowdsourcing.

As a result of the survey, the clear bias evidence is not found. In view of structural validity, some scores of pre and post questionnaire data is higher than those of previous studies. However, convergence validity with correlations between pre and post questionnaire data shows the significance of all correlations. Moreover, those correlations are compared with the previous study (Oshio *et al.*, 2014) and show no significance except “extraversion.”

Therefore, it can be concluded that TIPI-J can be used as a pre-task questionnaire and eventually, it will be helpful to select appropriate crowd workers.

However, “extraversion” had lower significance than the previous study (Oshio *et al.*, 2014), scored 0.71 on Task 2, this is the highest correlation of all other BF scores and this showed may have enough correlation to forecast full BF (Murakami and Murakami, 2001) because the correlation of “extraversion” between TIPI-J and full BF showed significance ($p < 0.01$). In addition, “extraversion” is said to show the correlation with work performance in limited occupations like salesperson or management person (Barrick and Mount, 1991; Barrick *et al.*, 2001) and this may not affect so much to predict crowd workers’ work performance.

On the other hand, “conscientiousness” Tasks 1 and 2 show second-lowest correlations. This is because the middle of pre “conscientiousness” scores showed a low correlation with post “conscientiousness” scores. However, the higher pre “conscientiousness” score group includes higher post “conscientiousness” score crowd workers and lower pre “conscientiousness” score group includes lower post “conscientiousness” score crowd workers. Table V showed that the average of post “conscientiousness” scores for the 10 and

Big five trait	Oshio <i>et al.</i> Study (<i>n</i> = 216)	Task 1 (<i>n</i> = 36)	Task 2 (<i>n</i> = 67)
	cor	cor	cor
Extraversion	0.84	0.77	0.71*
Agreeableness	0.47	0.31	0.44
Conscientiousness	0.64	0.45	0.50
Neuroticism	0.67	0.56	0.59
Openness	0.50	0.58	0.50

Notes: * $p < 0.5$; ** $p < 0.01$

Table IV.
Correlation and
significance Task 1,
Task 2 and
previous study

Task 1	Task 1			Task 2				
	The average of post “conscientiousness” score			The average of post “conscientiousness” score				
(%)	<i>n</i>	Highest group	Lowest group	<i>p</i>	<i>n</i>	Highest group	Lowest group	<i>p</i>
10	4	62.8	48.3	0.009**	7	61.6	44.4	0.009**
20	7	59.6	52.3	0.041*	14	59.4	44.4	0.001**

Notes: * $p < 0.5$; ** $p < 0.01$

Table V.
The average
scores of post
“conscientiousness”
for the 10 and 20 per
cent with the highest
and lowest pre score.

20 per cent with the highest and lowest pre “conscientiousness” score groups. In addition, a one-sided *t*-test was examined for those averages of post “conscientiousness” scores with the highest and lowest pre “conscientiousness” score groups and *p*-values were described.

As the result, the average post “conscientiousness” scores of top and bottom 10 per cent groups showed 1 per cent significance in Tasks 1 and 2 and the average “conscientiousness” scores with top and bottom 20 per cent groups showed 1 per cent significance in Task 2 and 5 per cent significance in Task 1. Especially, in the case of 10 per cent, the difference in averages of high “conscientiousness” score group and that of low “conscientiousness” score group was more than 10 points. Therefore, the pre “conscientiousness” score may be useful to distinguish high performers from low performers.

Of course, this result of the correlation between pre and post of “conscientiousness” is useful to predict crowd workers’ performance, it would be better if a higher score can be obtained because “conscientiousness” is reported as the main indicator to estimate work performance (Barrick and Mount, 1991; Barrick *et al.*, 2001; Schmidt and Hunter, 1998; Anderson and Viswesvaran, 1992).

To find a higher correlation, multiple regression analyses had been conducted to estimate the “conscientiousness” score. Pre (TIPI-J) (Oshio *et al.*, 2014) “conscientiousness” score and several CSP information like the number of tasks, the average score of the ratings awarded by clients and the number of skills claimed by the crowd workers are used as independent variables and post (full BF) (Murakami and Murakami, 2001) “conscientiousness” score as a dependent variable.

Table VI shows the result. Adjusted R^2 was 0.525 in Task 1 and 0.389 in Task 2. Those showed a middle correlation with independent variables. However, It has not been shown that CSP variables contribute to the forecast of the post “conscientiousness” score. All CSP information shows no significance and conversely all pre “conscientiousness” score shows significance ($p < 0.01$).

CSP information may not be helpful to forecast post “conscientiousness” scores and crowd workers’ performance and also, the information may not be helpful to forecast crowd workers’ work performance.

On the other hand, there are a lot of open to public information of crowd workers such as career, appeal point or profile. Such information may be applicable to forecast full BF with the use of natural language processing analysis. This will be the next research topic.

Dependent variable	Task 1 ($n = 36$)			Task 2 ($n = 67$)		
	Post (big five) “conscientiousness” score			Post (big five) “conscientiousness” score		
Adjusted R^2	0.525			0.389		
F	4.424			6.757		
Independent variable	Co-eff	Std. err	p	Co eff	Std. err	p
Pre (TIPI-J) score	1.809	0.485	0.001**	2.013	0.483	0.000**
The number of project done	0.007	0.086	0.431	0.011	0.046	0.820
The number of tasks completed	0.006	0.060	0.926	0.000	0.000	0.435
The average score of the ratings awarded by	0.568	3.979	0.888	-8.363	4.391	0.062
The number of skills claimed	-0.452	0.305	0.154	0.573	0.326	0.085
const	36.554	18.192	0.058	74.091	22.488	0.002

Notes: * $p < 0.5$; ** $p < 0.01$

Table VI.
Multiple regression
analysis among post
big score and CSP
information

Limitation

A number of samples are 36 in Task 1 and 67 in Task 2. The larger sample size is needed to increase the validity of findings. Because both Task 1 and Task 2 are not simple tasks like translation and English web information analysis, not so many Japanese crowd workers can apply those tasks.

Regarding Task 2, to increase applicants, additional option service on CSP, that kept showing the posted task at the top of the web page, was tried. Even with this service, only 68 crowd workers applied for Task 2.

If a task is simple as a microtask type, many crowd workers will apply with a much smaller reward. Because the target of the study is workers with some skills, the higher reward may be needed to attract more crowd workers. However, it will be difficult to achieve due to budget constraints.

Conclusion

Recently, crowdsourcing has been used as a worldwide effective way of human resource utilization. However, because of the variety and overwhelming size and scale of the workforce available, it is often difficult for clients to identify appropriate crowd workers.

A number of research studies (Ross *et al.*, 2010; Downs *et al.*, 2010; Kittur *et al.*, 2008; Assemi and Schlagwein, 2012) have investigated using CSP's information to predict crowd workers' output quality. However, the predictive power is limited because of a little CSP's information. Some research studies (Kazai *et al.*, 2011; Igawa *et al.*, 2016; Mourelatos and Tzagarakis, 2016) have an interest in BF traits to predict crowd workers' work performance.

On the other hand, many research studies (Barrick and Mount, 1991; Barrick *et al.*, 2001; Schmidt and Hunter, 1998; Anderson and Viswesvaran, 1992) have shown that BF traits, especially "conscientiousness," have a correlation with work performance. Igawa *et al.* (2016) showed a correlation between "conscientiousness" and work performance on crowdsourcing experiments. BF can be helpful to select appropriate workers in various occupations and situations.

However, there are two issues like the time-and-effort issue and the bias issue with using BF on crowdsourcing. The time-and-effort issue can be solved by TIPI-J (Oshio *et al.*, 2014), which is a short version of BF and has only 10 items, however, the bias issue still exists.

In this study, the survey was conducted on crowdsourcing to examine the efficacy of TIPI-J (Oshio *et al.*, 2014). To investigate the bias issue, TIPI-J (pre) (Oshio *et al.*, 2014) questionnaire is conducted before selecting crowd workers and full BF (post) (Murakami and Murakami, 2001) is conducted after selecting them. Then, the correlation between pre and post is analyzed and those correlations are compared with the previous study.

As a result, most correlations between pre and post showed significance. TIPI-J can be used to forecast the full BF score on crowdsourcing.

In addition, comparing those correlations with the previous study, there is no significance in the correlation between this study and the previous study except "extraversion."

In a previous study (Oshio *et al.*, 2014), the TIPI-J questionnaire and full BF questionnaire (Murakami and Murakami, 2001) were conducted to undergraduates. In this study, the result showed no clear difference with the correlation between this study and the previous study. It can be said that there may be no clear the bias issue that appeared on crowdsourcing and practically TIPI-J can be helpful for a client to forecast crowd workers' BF scores. Eventually, it may conclude that TIPI-J can help to select appropriate crowd workers.

On the other hand, "conscientiousness" in Tasks 1 and 2 showed second-lowest correlations. However, when focusing on top and bottom 10 per cent pre "conscientiousness"

score groups, the averages of post “conscientiousness” score showed significance. In practice, this indicates that clients can use TIPI-J to select high “conscientiousness” crowd workers for higher performance.

In addition, multiple regression analysis had been conducted to estimate “conscientiousness” scores by using several CSP provided quantitative information; however, it has not been shown that CSP variables contribute to the estimation of full BF. There still remains a lot of qualitative information and other analysis will be expected in future studies.

In conclusion, the results have important implications for the application of TIPI-J on crowdsourcing. Some previous studies focused on applying TIPI on crowdsourcing, but there have been few studies to investigate structural validity, convergence validity and reliability. However, in this study, some results showed intercorrelation among the scores of TIPI-J and full BF and structural validity was not enough verified, convergence validity and correlations with the previous study showed high significance.

Moreover, from the practical point of view, this study contributes to understanding the practical usage of the short version BF traits. The results of the survey indicated that there was no clear bias and showed the same level of correlation with the previous study. Practically, for clients who want to post tasks on crowdsourcing, it may be useful to predict crowd workers’ performance using TIPI-J with only 10 questions.

Future studies can explore some of the issues identified in this study such as examining with larger sample data and investigating other ways to improve correlation with “conscientiousness.”

References

- Anderson, G., (1992), and Viswesvaran, C. “An update of the validity of personality scales in personnel selection: a meta-analysis of studies published after 1992”, *13th Annual Conference of the Society of Industrial and Organizational Psychology, Dallas*.
- Assemi, B. and Schlagwein, D. (2012), “Profile information and business outcomes of providers in electronic service marketplaces: an empirical investigation”, *Australasian Conference on Information Systems (ACIS)*, ACIS, pp. 1-10.
- Barrick, M.R. and Mount, M.K. (1991), “The big five personality dimensions and job performance: a meta-analysis”, *Personnel Psychology*, Vol. 44 No. 1, pp. 1-26.
- Barrick, M.R., Mount, M.K. and Judge, T.A. (2001), “Personality and performance at the beginning of the new millennium: What do we know and where do we go next?”, *International Journal of Selection and Assessment*, Vol. 9 Nos 1/2, pp. 9-30.
- Brier, E. and Pearson, R. (2020), “Upwork’s SVP of marketing explain what it takes to perfect an offering that relies on people”, available at: <https://techdayhq.com/community/articles/upwork-s-svp-of-marketing-explains-what-it-takes-to-perfect-an-offering-that-relies-on-people>, (accessed 23 December 2018).
- Costa, P.T. and McCrea, R.R. (1992), “Revised neo personality inventory (neo-pi-r) and neo-five-factor inventory (NEO-FFI)”, Psychological Assessment Resources.
- Digman, J.M. and Shmelyov, A.G. (1996), “The structure of temperament and personality in Russian children”, *Journal of Personality and Social Psychology*, Vol. 71 No. 2, pp. 341-351.
- Downs, J.S., Holbrook, M.B., Sheng, S. and Cranor, L.F. (2010), “Are your participants gaming the system?: screening mechanical Turk workers”, *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2399-2402.
- Faullant, R., Holzmann, P. and Schwarz, E.J. (2016), “everybody is invited but not everybody will come – the influence of personality dispositions on users’entry decisions for crowdsourcing competitions”, *International Journal of Innovation Management*, Vol. 20 No. 6, p. 1650044.

-
- Feist, G.J. (1998), "A Meta-analysis of personality in scientific and artistic creativity", *Personality and Social Psychology Review*, Vol. 2 No. 4, pp. 290-309.
- Fiske, D.W. (1949), "Consistency of the factorial structures of personality ratings from different sources", *The Journal of Abnormal and Social Psychology*, Vol. 44 No. 3, pp. 329-344.
- Fujishima, Y. Yamada, N. and Tsuji, H. (2005), "Construction of short form of five-factor personality questionnaire".
- Goldberg, L.R. (1990), "An alternative" description of personality": the big-five factor structure", *Journal of Personality and Social Psychology*, Vol. 59 No. 6, pp. 1216-1229.
- Goldberg, L.R. (1999), "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models", *Personality Psychology in Europe*, Vol. 7 No. 1, pp. 7-28.
- Gong, Y. (2015), "Enabling flexible IT services by crowdsourcing: a method for estimating crowdsourcing participants", *Open and Big Data Management and Innovation*, Springer, pp. 275-286.
- Gosling, S.D., Gaddis, S. and Vazire, S. (2007), "Personality impressions based on Facebook profiles", *Icwm*, Vol. 7, pp. 1-4.
- Gosling, S.D., Rentfrow, P.J. and Swann, W.B. Jr, (2003), "A very brief measure of the Big-Five personality domains", *Journal of Research in Personality*, Vol. 37 No. 6, pp. 504-528.
- Howe, J. (2006), "The rise of crowdsourcing", *Wired Magazine*, Vol. 14 No. 6, pp. 1-4.
- Huang, J.-H. and Yang, Y.-C. (2010), "The relationship between personality traits and online shopping motivations", *Social Behavior and Personality: An International Journal*, Vol. 38 No. 5, pp. 673-679.
- Igawa, K., Higa, K. and Takamiya, T. (2016), "An exploratory study on estimating the ability of high skilled crowd workers", *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 10-14 July 2016*, pp. 735-740.
- John, O.P. and Srivastava, S. (1999), "The big five trait taxonomy: history, measurement, and theoretical perspectives", *Handbook of Personality: Theory and Research*, Vol. 2 No. 1999, pp. 102-138.
- Kazai, G., Kamps, J. and Milic-Frayling, N. (2011), "Worker types and personality traits in crowdsourcing relevance labels", *Proceedings of the 20th ACM international conference on Information and knowledge management*, ACM, pp. 1941-1944.
- Kittur, A., Chi, E.H. and Suh, B. (2008), "Crowdsourcing user studies with Mechanical Turk", 2008: ACM, pp. 453-456.
- Mourelatos, E. and Tzagarakis, M. (2016), "Worker's cognitive abilities and personality traits as predictors of effective task performance in crowdsourcing tasks", *Proceedings of 5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*, pp. 112-116.
- Murakami, Y. (2003), "Big five and psychometric conditions for their extraction in Japanese", *The Japanese Journal of Personality*, Vol. 11 No. 2, pp. 70-85.
- Murakami, Y. and Murakami, C. (2001), *Big Five Handbook*, Gakugei Tosho Co., Ltd.
- Oshio, A., Abe, S., Cutrone, P. and Gosling, S.D. (2014), "Further validity of the Japanese version of the ten-item personality inventory (TIPI-J)", *Journal of Individual Differences*, Vol. 35 No. 4.
- Rammstedt, B. and John, O.P. (2007), "Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and German", *Journal of Research in Personality*, Vol. 41 No. 1, pp. 203-212.
- Ross, J., Irani, L., Silberman, M., Zaldivar, A. and Tomlinson, B. (2010), "Who are the crowdworkers?: shifting demographics in mechanical Turk", *CHI'10 extended abstracts on Human factors in computing systems*, ACM, pp. 2863-2872.

- Schmidt, F.L. and Hunter, J.E. (1998), "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings", *Psychological Bulletin*, Vol. 124 No. 2, pp. 262-274.
- Shimonaka, Y., Nakazato, K., Gondo, Y. and Takayama, M. (1999), *Revised NEO-Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual for the Japanese Version*, (In Japanese), Tokyo Shinri, Tokyo.
- Snagajob (2017), "Snagajob appoints former Upwork CEO to board of directors", available at: www.prnewswire.com/news-releases/snagajob-appoints-former-upwork-ceo-to-board-of-directors-300417689.html (accessed 31 December 2018).
- Uchida, T. (2002), "Effects of the speech rate on speakers' personality-trait impressions", *Japanese Journal of Psychology*.
- Wada, S. (1996), "Construction of the big five scales of personality trait terms and concurrent validity with NPI", *The Japanese Journal of Psychology*, Vol. 67 No. 1, pp. 61-67.
- Wiggins, J.S. (1996), *The Five-Factor Model of Personality: Theoretical Perspectives*, Guilford Press.

Further reading

- Estellés-Arolas, E. and González-Ladrón-de-Guevara, F. (2012), "Towards an integrated crowdsourcing definition", *Journal of Information Science*, Vol. 38 No. 2, pp. 189-200.

Corresponding author

Kousaku Igawa can be contacted at: kousaku.igawa@gmail.com