2016 IEEE 24th International Requirements Engineering Conference

Acquiring Creative Requirements from the Crowd

Understanding the Influences of Personality and Creative Potential in Crowd RE

Pradeep K. Murukannaiah, Nirav Ajmeri, and Munindar P. Singh
Department of Computer Science, North Carolina State University, Raleigh, NC 27695-8206, USA
pmuruka@ncsu.edu, najmeri@ncsu.edu, m.singh@ieee.org

***Abstract*—As a creative discipline, Requirements Engineering (RE), lends importance to understanding the associated human factors. *Crowd RE*, the approach of acquiring requirements from members of the public—the so-called *crowd*—emphasizes human factors further. We investigate how human personality and creative potential influence a requirement acquisition task. These factors are of specific importance to Crowd RE because (1) crowd workers are generally not trained in RE, and (2) a key motivation in engaging them is to benefit from their creativity.**

We propose a sequential Crowd RE process, where workers in one stage review requirements from the previous stage and produce additional requirements. To reduce potential information overload in this process, we propose strategies for selecting requirements from one stage to expose to workers in later stages.

We conducted a study on Amazon Mechanical Turk tasking 300 workers with creating requirements via the above sequential process (in the domain of smart home applications for concreteness) and tasking an additional 300 workers to rate the creativity (novelty and usefulness) of those requirements.

Our findings offer insights on how to carry out Crowd RE effectively. First, we find that a crowd worker's (1) creative potential, and personality traits of openness and conscientiousness have significant positive influence on the novelty of the worker's ideas; and (2) personality traits of agreeableness and conscientiousness have significant positive influence, but extraversion has significant negative influence on the usefulness of the worker's ideas. Second, we find that exposing a worker to ideas from previous workers cognitively stimulates the worker to produce creative ideas. Third, we identify effective strategies based on personality traits and creative potential for selecting a few requirements from a pool of previous requirements to stimulate a worker.

***Index Terms*—Idea generation; crowdsourcing; Crowd RE; creativity; personality; smart home**

I. Introduction

Requirements Engineering (RE) is a creative problem-solving process [1] and humans are its centerpiece—creative intelligence remains out of the reach of the artificial. Crowd RE [2] is an emerging avenue for soliciting human intelligence for RE tasks from the members of the public also known as the *crowd*. It offers the potential benefits of cost reduction and broader coverage compared to traditional approaches involving a few trained experts [3]. However, the scope of crowdsourced tasks or *microtasks* in current settings (both in RE and in general) is typically limited to basic human abilities such as visual recognition and language understanding.

Our **main objective** is to facilitate crowdsourcing of RE tasks that require humans to exercise creativity, which is an advanced cognitive ability. We focus on the creative task of *idea generation*, wherein stakeholders come up with useful and novel ideas, eventually to be expressed as requirements [4]. As Maiden et al. [1] note, many existing RE techniques are designed to explore a search space of known—i.e., *not* novel—requirements; thus, idea finding is an area that promises a high potential for importing established creativity techniques into RE. Further, we focus on acquiring requirements for products for which a crowd member can naturally play the role of a stakeholder during idea generation—potentially all consumer products fit the bill.

Idea generation is often understood as a social process [5]. It involves group work, where creativity depends on how well the group members exchange and process each others' ideas (*attention*) and reflect on the exchanged ideas (*incubation*) [6]. Despite recent efforts [7], [8], facilitating group work on crowdsourcing platforms remains challenging: members of the crowd are often geographically dispersed, work at distinct times, and have disparate attitudes and objectives. Groups can sometimes be detrimental to creativity, e.g., when group members experience *evaluation apprehension* (being afraid of negative evaluation) and *social loafing* (feeling that one's effort is not needed by the group) [5]. Although such threats can be reduced by recruiting groups carefully, doing so is more difficult in crowdsourcing than in traditional settings.

The **first challenge** in achieving our objective is to understand how to achieve *cognitive stimulation* in crowdsourced tasks where group work is not viable. Our motivation is that cognitive stimulation produced by exposure to others' ideas is key to an individual's creativity in a group [9]. A potential approach to facilitating creativity in nongroup settings is a *sequential* task design, wherein ideas from workers in one stage are exposed to other workers in later stages. Recent studies suggest that a sequential design can be effective for crowdsourcing creative tasks [10], and may even perform better than simultaneous task designs in some cases [8]. A challenge with sequential design, though, is selecting ideas from previous stages to expose to a worker in the current stage. Tasking a worker to process all previous ideas (potentially thousands) is not viable economically and because of the associated cognitive overload.

The **second challenge** is to understand how to select a few ideas from a pool of previously acquired ideas to cognitively stimulate a worker. We consider two factors for selecting stimulating ideas in a sequentially crowdsourced creative task: *personality traits* and *creative potential* of the workers. First, personality traits influence how one perceives one's environ-

2332-6441/16 $31.00 © 2016 IEEE
DOI 10.1109/RE.2016.68

RE 2016, Beijing, China
Research Paper

ment and interacts within it. Evidence on both traditional [11] and crowdsourced [12] groups indicates that the group's personality composition influences its performance. Thus, it is possible that personality traits influence how one generates ideas and how one processes ideas originated by others. Second, exposure to others' ideas can be stimulating as well as distracting, depending upon the extent to which one connects to the exposed ideas [13]. Thus, it is conceivable that a worker's ability to make such connections depends on his or her creative potential.

The foregoing intuitions prompt us to systematically investigate how personality traits and creative potential of workers influence crowdsourced creative tasks. Specifically, we seek to answer the following research questions.

Q_1 How do a worker's personality traits and creative potential influence the creativity of the ideas he or she produces?

Q_2 What influence do the personality traits and creative potentials of previous workers, whose ideas a worker is exposed to, have on the ideas this worker produces?

Q_3 How effective is a sequential task design for acquiring creative ideas via Crowd RE?

To answer these questions, we designed an empirical study and conducted it on Amazon Mechanical Turk (MTurk) with 600 participants (workers). In this design, we measure personality traits and creative potential for all workers. In the first phase, we collect ideas expressed as user stories from half of them for a familiar but future setting (we adopted smart homes as the setting in our study) in three sequential stages. In the second phase, we employ the remaining workers to rate acquired ideas in terms of clarity and creativity.

Contributions: (1) We describe a sequential task design for facilitating creative requirements acquisition from the crowd. The crux of our design is idea selection based on workers' personality traits and creative potentials that reduces cognitive overload on workers. (2) We conduct an empirical study on MTurk involving 600 participants to validate our hypotheses about the influences of crowd workers' personality traits and creative potential on the creativities of their ideas. Our findings have a bearing not only upon Crowd RE but also upon crowdsourcing sequential tasks, in general.

Novelty: Our work is novel for two reasons. First, although still under debate [14], an increasing amount of evidence suggests that creativity is domain-specific [15]. From this perspective, our study is the first to offer empirical insights on how personality traits and creative potential influence the specific task of creative requirements acquisition via crowdsourcing (Q_1). Second, our study is the first, across domains, to understand how exposing a user to others' ideas based on their respective personalities and creative potentials influences the creativity of the user's ideas (Q_2 and Q_3).

Significance: The public's importance in RE is increasingly recognized. Accordingly, recent efforts seek to scale Crowd RE via designing crowd workflows [3] and employing data-driven techniques for processing the requirement-related content generated by the crowd such as application reviews [16]. In contrast to these, we offer a fresh perspective on scaling Crowd RE by exploiting human factors, specifically, personality traits and creative potential. We show that these factors can be inexpensive to measure, yet prove quite effective in facilitating creative requirements acquisition from the crowd.

Organization: Section II describes our method. Section III describes our hypotheses and the analyses we perform. Sections IV and V present and discuss our results, respectively. Section VI reviews related work and we conclude with some future directions in Section VII.

II. Method

To answer the research questions above, we conducted a two-phase study on MTurk. The study was approved by our university's Institutional Review Board (IRB).

In the first phase, 300 participants (workers) generate ideas for smart home applications. In the second phase, an additional 300 workers rate the ideas generated in the first phase.

We choose smart home applications as the domain for idea generation for three reasons. First, smart home applications are still emerging, providing space for creative idea generation. Second, although practical applications are nascent, the concept of smart homes is not new. Thus, the members of the crowd are likely to possess some background knowledge on smart homes, making it a viable topic for idea generation from the crowd. Third, since the crowd is the eventual end user of smart home applications, we imagine that the members of the crowd would find generating ideas for smart home applications both interesting and worthwhile.

A. Phase 1: Idea Generation

We designed a three-stage sequential process for acquiring ideas from the crowd. Workers in the first stage generate an initial set of ideas. Workers in the second stage review ideas from the first stage and generate their ideas. Similarly, workers in the third stage review ideas from the second stage and generate their ideas. Figure 1 shows an overview of the two phases and three stages of the first phase of our study.

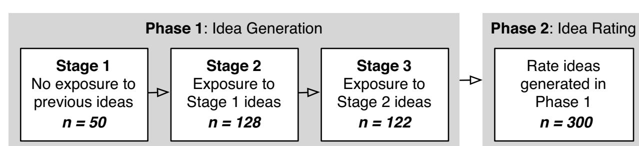

Fig. 1: An overview of our two-phase study design.

Each worker in the idea generation phase answers a personality survey, a creativity survey, and completes the idea generation task. The idea generation task involves a worker reviewing selected ideas from previous workers and generating new ideas. The idea selector, which applies only to the second and third stages, chooses a few ideas from the pool of previous ideas based on one of six strategies (Section II-A2).

1) Personality and Creativity Surveys: Before acquiring ideas, we assess each worker's personality and creative potential. First, we employ the Mini-IPIP (International Personality Item Pool) [17] scales to measure a worker's Big

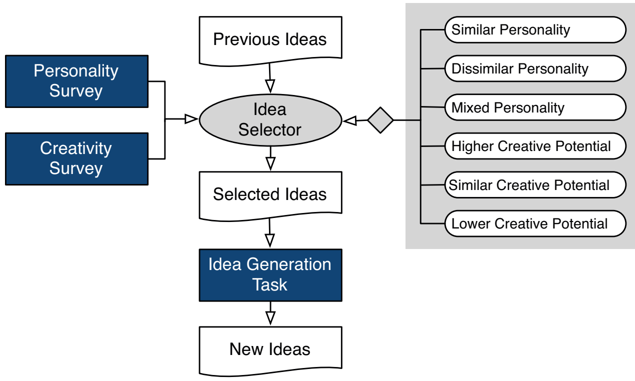

Fig. 2: An overview of our task design for the idea generation phase.

Five personality traits of Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to experience (O). The Mini-IPIP scales consist of 20 items (11 negative items)—four items for each Big Five trait. A worker responds to each of the 20 items on a Likert scale of 1 (strongly disagree) to 5 (strongly agree). Then, we compute a score for each Big Five trait as the mean of the positive and reverse-scored negative items corresponding to the trait.

Second, we employ the Creative Personality Scale (CPS) [18] for assessing a worker's creative potential. The CPS is a 30-item adjective list, consisting of 18 positively scored (e.g., capable, unconventional, and snobbish) and 12 negatively scored (e.g., conservative, honest, and narrow interests) items. A worker answers whether each of those items describes them on a Likert scale of 1 (strongly disagree) to 5 (strongly agree). We then compute a single creative potential score as the mean of the positive and reverse-scored negative items.

Both Mini-IPIP and CPS are well-known scales. An important reason why we choose these two scales is their compactness. If longer alternatives were to be used, they would increase the likelihood that research workers drop out before getting to the main task [17].

2) Idea Generation Task: The main task for workers in this phase is to come up with ideas (for potential requirements) for smart home applications. We provided basic information about smart homes, and encouraged workers to conduct additional research on smart homes. We described the task and incentivized workers to be creative as follows.

Task: Come up with smart home requirements. To do so, imagine what you would expect from a smart home.

Creativity bonus: Be creative! We provide you some sample requirements and your objective is to come up with requirements that are more creative than the samples. A creative requirement leads to products that are both useful and novel. For each requirement you produce that is more creative than the ones shown to you (as judged by us), you will receive a bonus of 20 cents (up to a maximum of USD 1).

We tasked each worker to produce at least ten ideas, each in a user story format, as shown in Figure 3. We asked workers to produce ideas distinct from the samples shown to them. However, we encouraged combinational creativity [19] by telling workers that they can make creative connections among the samples shown to them to come up with new ideas. Finally, we asked workers to choose an application category (health, safety, energy, entertainment, or other) and, optionally, provide a few descriptive tags for each idea.

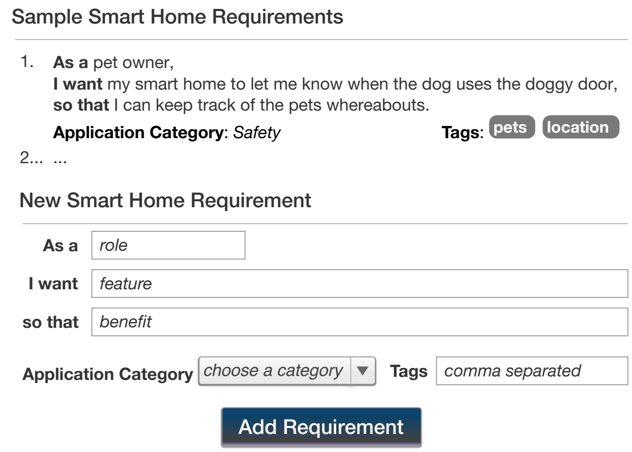

Fig. 3: A screen mockup of the ideas survey (user story format).

For each worker in the first stage, we showed one example idea that we produced. For each worker in the second and third stages, first, we selected workers from the previous stage—via one of six strategies, as detailed below. Next, we showed ten randomly selected ideas to the worker that were produced by the selected previous-stage workers.

Three of the six strategies are based on personality and three on creative potential.

a) Personality-based idea selection: Given a worker in stage m, to apply a personality-based selection strategy, we first compute a *personality distance* from the worker to each worker of the previous stage, $m-1$. The personality distance between two workers i and j is the Euclidean distance between their corresponding personality trait scores as shown below.

$$\text{Personality distance}^2(i,j) = \sum\nolimits_{\text{trait}\in\{\text{O,C,E,A,N}\}} (\text{trait}_i - \text{trait}_j)^2,$$

where O, C, E, A, and N are the Big Five personality traits.

Next, for each worker, we select previous workers based on their personality distances, as follows.

Similar personality: Three previous workers with the smallest personality distances (most similar) to the worker ($n_2 = n_3 = 20$).

Dissimilar personality: Three previous workers with largest personality distances ($n_2 = n_3 = 20$).

Mixed personality: Two previous workers with smallest and two with largest personality distances ($n_2 = n_3 = 20$).

Here, n_2 and n_3 are the numbers of workers to whom the corresponding selection strategy was applied in the second and third stages of the idea generation phase, respectively.

b) Creative potential-based idea selection: To select ideas based on the creative potential of a worker in stage m, we first group the creative potentials of workers from the previous stage $m-1$ into tertiles (three quantiles). We then choose a few workers from one of the tertiles, based on the selection strategy, as follows.

Higher creative potential: Three previous workers from the tertile above the one to which the worker's creative potential belongs ($n_2 = 23$; $n_3 = 20$).

Similar creative potential: Three previous workers from the same tertile as the one to which the worker's creative potential belongs ($n_2 = 25$; $n_3 = 22$).

Lower creative potential: Three previous workers from the tertile below the one to which the worker's creative potential belongs ($n_2 = 19$; $n_3 = 21$).

Note that all strategies may not be applicable to a worker. Specifically, the higher and lower creative potential-based strategies do not apply to workers whose creative potentials are in the first and third tertiles, respectively.

B. Phase 2: Idea Rating

Workers in the first phase generated a total of 2,966 ideas, after we excluded ideas unrelated to smart homes or identical to a sample idea shown to the worker. However, to answer our research questions, we must compare the creativity of the generated ideas. Thus, in the second phase of our study, we asked additional members of the crowd to rate the creativities of the ideas from the first phase.

We seek multiple raters for each idea so as to exclude ideas with unreliable ratings from further analysis. However, considering the number of ideas to be rated, it is not feasible to ask each rater to rate all ideas or a different rater to rate each idea. Thus, our rating study design stakes a middle ground.

First, we group the ideas by their application category indicated by the ideas generators (first phase). Next, within each category, we create bundles of ten randomly selected ideas. We then ask each rater to rate three randomly selected idea bundles. We acquired ratings from a total of 300 workers (distinct from the first phase workers), such that at least three raters rated each idea bundle.

Just as we did in Phase 1, we provided background information about smart homes to Phase 2 workers. We tasked them to generate at least three smart home ideas before proceeding to the rating task. We did so to encourage workers to think about smart home applications before rating others' ideas.

Our objective in this phase is to assess the creativities of ideas produced in the previous phase. Creativity, according to a widely accepted definition [20], entails *novelty* and *usefulness*. Further, we recognize that *clarity* is crucial: ratings of unclear ideas would confound our analysis. Thus, as shown in Figure 4, we ask workers to rate each idea for its novelty, usefulness, and clarity, described to each worker as follows.

Clarity: A clear requirement is unambiguous and provides an appropriate level of detail.

Usefulness: A useful requirement leads to products that provide value or utility to their users.

Novelty: A novel requirement is something that a user finds original and unexpected, i.e., something that is not commonplace, mundane, or conventional.

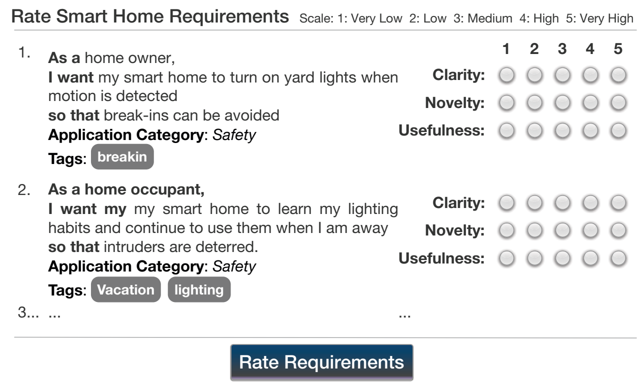

Fig. 4: A screen mockup of an idea rating screen.

We asked workers to rate each idea shown to them for each criterion above on a Likert scale of 1 (very low) to 5 (very high). We showed three rating screens to each worker—one for each bundle of ten ideas assigned to the worker—with the intuition that showing ideas from an application category together makes comparison across ideas, and thus rating, easy.

Table I shows the mean value of times spent by Phase 1 and Phase 2 workers. We provided a USD 3 base pay to each worker who completed all assigned tasks. As mentioned earlier, Phase 1 workers could earn an extra dollar as a creativity bonus. Considering the difficulty of the task, we provided the bonus to more than half of Phase 1 workers.

TABLE I: Times spent by workers and our payments.

Phase	Main task time	Other tasks time	Base pay	Bonus pay
1	29 minutes	5 minutes	USD 3	USD 1
2	16 minutes	5 minutes	USD 3	0

Table II summarizes the demographics information we collected from Phase 1 and Phase 2 workers in a presurvey. As shown in the last row of the table, a majority of our workers reported medium or higher familiarity with smart homes.

TABLE II: Demographics of our study workers.

Gender	Male: 52.6%, Female: 46.7%, Other: 0.7%
Age	18–24: 14.3%, 25–34: 52.4%, 35–45: 23%, 45–54: 6.4%, 55 or older: 3.9%
Education	Graduate degree: 14.3%, Bachelor's degree: 42.2%, Some college but no degree: 30%, High school: 13%, Less than high school: 0.5%
Familiarity with smart homes	Very low: 8%, Low: 23.5%, Medium: 44.3%, High: 20%, Very high: 4.2%

III. Evaluation

We state our hypotheses (refutable claims) and describe the analyses we perform to evaluate those hypotheses.

A. Hypotheses

To answer Q_1, we evaluate the following hypotheses.

H1A$_{\textbf{null}}$ (Null hypothesis): A worker's personality trait has no influence on the creativity of the ideas the worker produces.

H1A$_{\textbf{alternative}}$ (Alternative hypothesis): A worker's personality trait influences the creativity of the worker's ideas.

Here, personality trait can refer any of the Big Five personality traits, and creativity can refer to novelty or usefulness. We evaluate a similar pair of hypotheses for creative potential.

H1B$_{\textbf{null}}$ A worker's creative potential has no influence on the creativity of the ideas the worker produces.

H1B$_{\textbf{alternative}}$ A worker's creative potential influences the creativity of the ideas the worker produces.

To answer Q_2, we evaluate the following hypotheses.

H2A$_{\textbf{null}}$ The creativities of a worker's ideas are the same whether the worker is exposed to others' ideas via similar, dissimilar, or mixed personality-based strategy.

H2A$_{\textbf{alternative}}$ The creativities of a worker's ideas differ depending on whether the worker is exposed to others' ideas via similar, dissimilar, or mixed personality-based strategy.

We evaluate a similar pair of hypotheses for the creative potential-based idea selection strategies.

H2B$_{\textbf{null}}$ The creativities of a worker's ideas are the same whether the worker is exposed to others' ideas via higher, lower or similar creative potential-based strategy.

H2B$_{\textbf{alternative}}$ The creativities of a worker's ideas differ depending on whether the worker is exposed to others' ideas via higher, lower, or similar creative potential-based strategy.

To answer Q_3, we evaluate the following hypotheses.

H3$_{\textbf{null}}$ The creativities of the ideas produced in Stages 1, 2, and 3 of our sequential Crowd RE process are the same.

H3$_{\textbf{alternative}}$ The creativities of the ideas produced in Stages 1, 2, and 3 of our sequential Crowd RE process are different.

We deliberately state each of our alternative hypotheses as two-sided (e.g., extraversion influences novelty) as opposed to one-sided (e.g., increase in extraversion increases novelty). We do so because there is no previous evidence (from a setting similar to ours and at the granularity we desire) to suggest which one-sided alternative to employ. Thus, when a significant influence is found, we further explore the associated input and output variables to infer the direction of the influence.

B. Analyses

We analyze ideas generated in the first phase and rated in the second phase of our study.

1) Preprocessing: We preprocess our data to reduce noise. First, we exclude about 8% of idea ratings, with an associated clarity rating of less than medium. Next, we compute interrater reliability (IRR) for ratings of each idea via the intraclass correlation (ICC). The ICC is a commonly-used statistic for assessing IRR for ordinal data [21] such as ours. We exclude about 19% of ratings with very low IRR (ICC < 0.3) as unreliable. Note that we choose a low ICC cutoff so as to not exclude too many ideas from further analysis.

2) Multiple Regression: To test our hypotheses about personality traits and creative potential, we treat our data as a set of observations—each idea corresponds to an observation. Further, within each observation, we treat the idea producer's personality traits and creative potential as *factors* of interest and the mean novelty rating or usefulness rating of the idea as the *outcome*, depending upon the hypothesis being tested.

We can compute the correlation between a factor and an outcome to understand their relationship. We can then test each of our hypotheses based on the correlation between the corresponding factor and outcome. However, doing so assumes that the factors are independent of each other, failing which the conclusions can be spurious. The independence assumption may not be valid for our factors [22].

We employ multiple regression (MR) models [23] instead of pairwise correlations. An MR model helps understand each factor's partial influence on the outcome over and above the other factors. We are mainly interested in two pieces of information from an MR model: (1) the **significance** of factor influences, indicated by p-values, and (2) the **effect size** of each factor, indicated by the value of its regression coefficient.

The statistical significance of a factor influence indicates only whether a relationship exists. We reject the null hypothesis about a factor influence at the 5% significance level. In contrast, the effect size of a factor influence indicates the strength of the relationship between the factor and the outcome. Specifically, in an MR model, the regression coefficient of a factor indicates the extent to which the factor accounts for the variance in the outcome. Thus, the greater the magnitude of a factor's coefficient the greater its influence on the outcome. Further, the sign of a coefficient indicates the directionality of the corresponding factor's influence on the outcome (e.g., a positive coefficient indicates that an increase in the factor is associated with an increase in the outcome).

We also report an R^2 value for each MR model we fit, which indicates the proportion of the outcome's variance shared with the optimally weighted composite of the factors [23]. That is, considering Y as the actual outcome and $\widehat{Y}$ as the outcome predicted from optimally weighted factors, $\mathrm{R}^2 = \mathrm{sd}^2_{\widehat{Y}}/\mathrm{sd}^2_{Y}$. The R^2 value indicates how good an MR model is as a predictive model. However, it is important to note that the R^2 value in itself does not affect the conclusion we may derive about the significance and the effect sizes of factors.

3) Kruskal-Wallis Tests: To test our hypotheses about idea selection strategies and ideas from different stages, we perform the Kruskal-Wallis tests [24]. The Kruskal-Wallis test is an extension of the Wilcoxon ranksum test and a nonparametric version of the one-way ANOVA (and, thus, avoids the assumptions that populations have normal distributions).

Again, we reject the null hypothesis that ratings of all compared samples come from the same distribution at the 5% significance level. If the null hypothesis is rejected, we further perform the multiple comparison tests to determine the specific pairs of compared variables that are significantly different.

We perform all our analyses (ICC, multiple regression, Kruskal-Wallis, and multiple comparison) via Matlab [25].

IV. RESULTS

Figure 5 shows the distributions of our workers' personality traits and creative potentials. In these and other boxplots below, the diamond dots represent the mean of the distribution, and the $\times$ marks outside box indicate outliers.

The distributions of individual traits in Figure 5 show that our data represents a variety of personalities. Also, the means and standard deviations of the traits in our data follow a pattern similar to those reported in previous studies [17], [26].

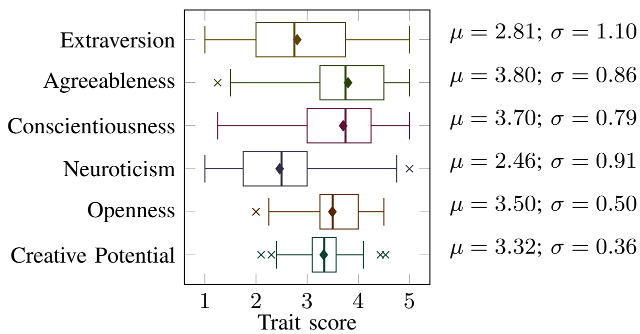

Fig. 5: Distributions of personality traits and creative potential.

A. *Factor Influences on the Outcomes (H1A and H1B)*

Table III summarizes the results from two multiple regression models (one for each outcome) that we fit to our data. In this and other tables, we highlight significant results as bold.

TABLE III: Regression coefficients indicating the influences of personality traits and creative potential on the creativity of ideas.

Variable	Novelty		Usefulness	
	Coefficient	p-value	Coefficient	p-value
Extraversion	−0.035	0.16	**−0.097**	0.03
Agreeableness	0.029	0.34	**0.163**	0.01
Conscientiousness	**0.064**	0.04	**0.151**	0.01
Neuroticism	0.025	0.37	−0.062	0.25
Openness	**0.099**	0.04	−0.116	0.18
Creative Potential	**0.174**	0.03	−0.052	0.71
R^2	0.012		0.020	

First, we find that a worker's creative potential, openness, and conscientiousness have significant influences (p-value) on the novelty of the worker's ideas. Further, the coefficient values (effect size) indicate that the creative potential has the highest influence on novelty followed by openness.

Second, we find that a worker's agreeableness has the highest influence on the usefulness of the worker's ideas, closely followed by conscientiousness. Further, we find that extraversion has a significant, though negative, influence on the usefulness of a worker's ideas.

Finally, we note that the R^2 values of both the models we fit are small. This indicates that a linear model is likely to perform poorly in predicting the outcomes from factors in our data. Nonetheless, as we alluded to in Section III-B2, R^2 values do not affect our conclusions about the trends in the data, specifically, the significance of factor influences.

B. *Idea Selection Strategies (H2A and H2B)*

Table IV and Figure 6 compare the three personality-based idea selection strategies for Stages 2 and 3, separately. These results suggest that workers exposed to others' ideas via the mixed personality-based strategy produce more novel ideas than those exposed via the other personality-based strategies. We find that this result is consistent for both stages. However, we do not find sufficient evidence to reject the null hypothesis about the effect of personality-based strategies on the usefulness of a worker's ideas.

TABLE IV: Comparing personality-based idea selection strategies.

Stage	Idea Selection Strategy	Novelty		Usefulness	
		Mean (SD)	p	Mean (SD)	p
2	Similar Personality	3.22 (1.20)		3.80 (1.12)	
	Dissimilar Personality	2.98 (1.37)	0.02	3.92 (1.16)	0.31
	Mixed Personality	**3.51** (1.31)		3.97 (1.10)	
3	Similar Personality	3.07 (1.28)		3.95 (1.21)	
	Dissimilar Personality	3.06 (1.31)	0.04	3.68 (1.27)	0.15
	Mixed Personality	**3.42** (1.24)		3.89 (1.19)	

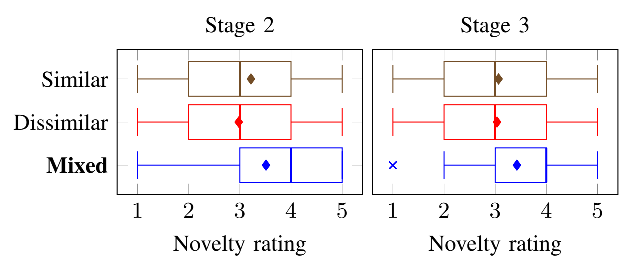

Fig. 6: Novelty ratings of ideas generated by workers exposed to others' ideas via three personality-based selection strategies.

Table V and Figure 7 compare our creative-potential based idea selection strategies. Although the higher creative potential-based strategy yields most novel ideas in both stages, we observe that the differences in novelty ratings are not significant in Stage 2. In Stage 3, however, we find that the novelty ratings for both higher and similar creative potential-based strategies are significantly higher than those for the lower creative potential-based strategy. However, in both stages, we fail to reject the null hypothesis about the effect of creative potential-based strategies on usefulness ratings.

TABLE V: Comparing creative potential-based selection strategies.

Stage	Idea Selection Strategy	Novelty		Usefulness	
		Mean (SD)	p	Mean (SD)	p
2	Higher Creative Potential	3.92 (1.16)		3.73 (1.22)	
	Lower Creative Potential	3.06 (1.12)	0.21	3.73 (1.22)	0.43
	Similar Creative Potential	3.20 (1.43)		3.91 (1.16)	
3	Higher Creative Potential	**3.57** (1.27)		3.82 (1.12)	
	Lower Creative Potential	3.09 (1.38)	0.03	3.60 (1.25)	0.36
	Similar Creative Potential	**3.41** (1.20)		3.76 (1.19)	

It is interesting to notice that the differences in usefulness ratings between different strategies, although not significant,

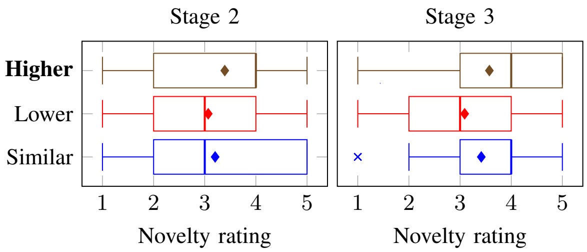

Fig. 7: Novelty ratings of ideas from workers exposed to others' ideas via three creative potential-based selection strategies.

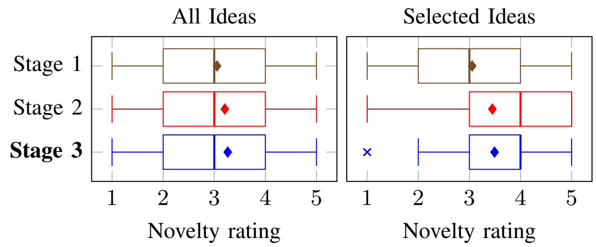

Fig. 8: Novelty ratings of ideas from Stages 1, 2, and 3.

follow similar patterns to those of novelty ratings. That is, the mixed strategy is the most effective personality-based strategy, and the higher creative potential strategy is the most effective creative potential-based strategy for producing useful ideas.

C. Comparing Ideas across Stages (H3)

Table VI and Figure 8 compare ideas from Stages 1, 2, and 3. First, we consider all ideas in each stage, and compare them across stages. Although we observe that both novelty and usefulness ratings increase as we progress from one stage to the next, the differences are not significant.

Next, instead of considering all ideas from Stage 2 and 3, we consider ideas produced only by workers exposed to mixed personality-based or higher creative potential-based idea selection strategies. With this modification, we observe that the novelty of ideas from Stages 2 and 3 is significantly higher than those from Stage 1. The difference between ideas from Stages 2 and 3 is not significant, though. However, we make an important observation about ideas going from Stage 2 to Stage 3: the first quartile shrinks considerably in Stage 3 (the difference in variance is not statistically significant, though, mainly because of the outlier in Stage 3). This suggests that there are fewer ideas of low novelty in Stage 3 than in Stage 2. Finally, we do not find significant differences for usefulness ratings across stages.

TABLE VI: Comparing ideas from the three stages.

Idea Selection Strategy	Stage	Novelty Mean (SD)	Novelty p	Usefulness Mean (SD)	Usefulness p
None	1	3.05 (1.37)		3.78 (1.21)	
	2	3.21 (1.34)	0.13	3.86 (1.15)	0.11
	3	3.26 (1.30)		3.96 (1.11)	
Mixed Personality or Higher Creative Potential	1	3.05 (1.37)		3.85 (1.16)	
	2	**3.45** (1.31)	0.01	3.85 (1.17)	0.45
	3	**3.49** (1.25)		3.96 (1.11)	

V. Discussion

We now answer the research questions we asked earlier based on the empirical insights we derive from our study.

Q_1. *How do a worker's personality traits and creative potential influence the creativity of the ideas the worker produces?*

Our multiple regression models indicate that a worker's personality traits and creative potential have significant influences on the creativity of the worker's ideas. Whereas the worker's creative potential, openness, and conscientiousness influence the novelty of his or her ideas, the worker's agreeableness, conscientiousness, and extraversion influence the usefulness of his or her ideas, in decreasing order of effect sizes.

Since none of the previous studies establish relationships between crowd workers' personality traits and creative performance, we compare our results to those from other settings. In previous studies about personality and creative performance, creative tasks are performed in groups [27], or individually but in a social environment, e.g., in a class [22]. Our result about openness is consistent with results in these studies.

A major disagreement between our results and those from previous studies is in regards to extraversion. Whereas previous studies find extraversion to be positively related to creativity, we find the relationship to be negative. We attribute this disagreement to the major difference in the two settings: social versus individual. An extrovert seeks excitement and stimulation [28]. These characteristics may yield benefits in social settings. Crowd work, though, is typically done by an individual with little overt social contact. As our results suggest, introverts, with a disposition to work independently, yield better creative performance in such settings.

The two previous studies, [27] and [22], do not agree on the influence of conscientiousness on creativity. Whereas Taggar [27] found evidence for a positive relationship, Sung and Choi [22] did not find the relationship to be significant. Our results agree with Taggar as we find conscientiousness to be positively related to both novelty and usefulness. We attribute this finding, too, to the crowd setting, where tasks are typically short lived. We imagine that producing creative ideas can be time consuming, and a worker's conscientiousness is important in influencing the individual to be creative.

Q_2. *What influence do the personality traits and creative potentials of previous workers, whose ideas a worker is exposed to, have on the ideas the worker produces?*

First, we find that exposing a worker to a mix of ideas—some from workers whose personalities are similar to the worker and some from workers whose personalities are dissimilar to the new worker—is the best personality-based strategy to stimulate the new worker to produce novel ideas. This result is similar to a recent result about collaborative crowd work that teams of balanced personalities yield better outcomes [12].

Since an individual's personality influences how he or she perceives an environment [11], we imagine that the personality has a bearing on the ideas the individual produces. Then, our result can be interpreted as follows. To facilitate a crowd worker to produce novel ideas, the worker must be exposed to some ideas to reinforce his or her thinking and some ideas the worker would not have thought about.

Second, we find that similar and higher creative potential-based strategies are more effective than the lower creative potential-based strategy in prompting a worker to produce novel ideas. We attribute this result to our incentive structure. Recall that we offered workers up to a bonus for producing more novel ideas than the ideas shown to them. Thus, showing ideas of lower novelty to a worker may not sufficiently induce the worker to exercise his or her creativity to the full extent.

Finally, we take a closer look at why we did not find significant differences for usefulness ratings. The distributions of our outcomes reveal a potential reason for this. As Figure 9 shows, there is greater variety in our data for novelty than for usefulness. The usefulness ratings are skewed toward high ratings, with more than 60% ratings as high or very high. We conjecture that this lack of variety could have lead to a Type II error (Kruskal-Wallis test failing to reject the false null hypothesis about differences in usefulness ratings).

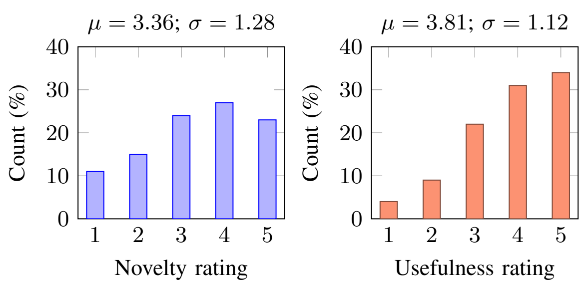

Fig. 9: Distributions of novelty and usefulness ratings.

Q_3. How effective is a sequential task design for acquiring creative ideas via Crowd RE?

We find that a sequential task design can be effective for acquiring novel ideas from the crowd, specifically, when the workers are exposed to carefully selected ideas. Although the creativity of ideas increase as the idea generation task progresses from one stage to another, we find that the increase is significant only when appropriate idea selection selection strategies are employed. This result demonstrates the practical utility of our findings. According to our results, if appropriate idea selection strategies are applied, novel ideas can be acquired in fewer stages than acquiring ideas via sequential process without applying any selection strategies.

Limitations and Threats to Validity

A key threat to validity in crowdsourced studies is data quality. To control quality, we required our workers to have completed at least 100 tasks and have an approval rate of at least 95%. Further, to control the quality of ideas produced (Phase 1), we (1) manually reviewed all ideas and filtered out ideas that we deemed as inappropriate or unrelated to the problem domain; and (2) asked members of the crowd to rate the clarity of the ideas and excluded ideas whose clarity was below a threshold. To control the quality of the idea ratings (Phase 2), we excluded unreliable ratings based on an interrater reliability measure. However, one aspect of the response quality threat that remains unaddressed is workers' responses to the personality and creativity surveys.

In addition to personality traits and creative potential, factors such as a worker's education, domain knowledge, and culture may influence the creativity of the worker's ideas and the worker's creativity ratings. Our study did not control for such factors and we defer their exploration to future work.

For logistical reasons, we limited the idea generation task (Phase 1) to three stages. Whether our findings on the effectiveness of the sequential task design (Q_3) generalize as more stages are included remains to be seen. We conjecture that the creativity of the ideas saturate after a certain number of stages.

VI. Related Work

Our work relates to research on creativity, personality, and crowdsourcing in requirements engineering.

A. Creativity in Requirements Engineering

Creativity is as an important aspect of RE and an increasing number of techniques seek to incorporate creativity in RE.

Bhowmik et al. [19] describe a creativity framework that mines ideas using topic modeling, and creates novel and innovative requirements by obtaining unfamiliar connections. Zachos and Maiden [29] describe AnTiQue, a tool to retrieve web services from domains analogical to a problem domain (in which requirements are sought). AnTiQue exploits natural language matching between requirements and service descriptions. Zachos and Maiden empirically evaluate AnTiQue, finding that it helps in generating novel requirements. Such tools can be employed within our Crowd RE process for facilitating effective creative thinking.

Svensson et al. [4] compare four creativity techniques (Hall of Fame, Constraint Removal, Brainstorming, and Idea Box) by conducting creativity workshops with students and practitioners. They find that (1) Brainstorming generates the most ideas, and (2) Hall of Fame generates the most creative requirements that are part of future release of products. Our study design does not ask workers to follow a specific creativity technique. An interesting extension to our study would be to investigate personality influences in the context of specific creativity techniques. However, an underlying challenge is to adapt some of these techniques from group to crowd settings.

Nguyen and Shanks [30] build a theoretical framework consisting of product, process, domain, people, and context to understand creativity in RE. They identify the need for investigating personality characteristics and traits of people involved in performing creative tasks. What we report here is precisely such an investigation.

Horkoff et al. [31] combine creativity techniques with goal modeling, and suggest starting with transformational technologies; following with exploratory creativity and combinatorial creativity; and ending with reflection and evaluation. However, whether such techniques can be adopted for untrained members of the crowd remains to be explored.

B. Personality in Requirements Engineering

Ample evidence (in domains such as psychology and management) suggests that personality traits impact creativity. Yet, none of the existing works relate personality and creativity in the context of an RE task.

Dallman et al. [32] study contextual factors including social and individual dimensions that influence creativity in RE. They identify that a risk-taking personality is an important individual dimension for creativity.

A few software engineering works study personality. Capretz and Ahmed [33] correlate software job requirements and soft skills with psychological traits, measured via the Myers-Briggs Type Indicator. They map software job roles to soft skills. Capretz and Ahmed's mapping of introversion with ability to work independently aligns with our finding that introversion positively influences the novelty and usefulness of the ideas crowd workers produce.

Wiesche and Krcmar [34] present a systematic literature review on software developers' personalities. They examine the effect of task-personality match on developers' satisfaction and performance. Cruz et al. [35] systematically review research on personality in software engineering. They identify that the role of personality in pair programming, education, team effectiveness, process allocation, individual performance, and leadership effectiveness is studied extensively. Thus, although personality is discussed in software engineering in general, there is no research in software engineering that relates personality and creativity. Our work fills this gap.

C. Crowdsourcing in Requirements Engineering

Hosseini et al. [36] understand the crowdsourcer, crowd, crowdsourced task, and crowdsourcing platform as four pillars of Crowd RE. They also analyze different features of the crowd and the crowdsourcer to see how these features impact the quality of elicited requirements [37]. Hosseini et al.'s findings are based on surveying focus groups of students and developers, and requirements engineering experts. Though broad, their list of features lacks personality and creativity of the crowd, which we find are valuable in Crowd RE.

Breaux and Schaub [3] describe a task decomposition workflow to scale the task of requirements acquisition from natural language sources to the crowd. Their results show that Crowd RE can both reduce the cost of requirements elicitation and increase coverage. Bhatia and Breaux [38] use crowdsourcing to construct a lexicon of information types for privacy policies. Whereas microtasks in these works require only basic cognitive abilities, our contribution ventures into exploiting crowd workers' creativity in requirements acquisition.

Lim and Finkelstein propose StakeRare [39] to facilitate requirements elicitation in large scale software projects. Their method (1) constructs a social network of stakeholders, (2) recommends requirements to stakeholders via collaborative filtering, and (3) prioritizes requirements based on stakeholders' project influences (computed via network measures). In Crowd RE, members of the crowd play stakeholder roles. An interesting direction is to construct a social network consisting of the crowd members contributing ideas to a product. Such a network would enable (1) rich interactions between members of the crowd, e.g., via arguments [40], yielding structured requirements, and (2) extending network-based techniques such as StakeRare to crowd settings.

Picazo-Vela et al. [41] describe a model based on the theory of planned behavior and Big-Five traits to study an individual's intention to provide online reviews. They empirically find that personality traits of neuroticism and conscientiousness have significant impact on an individual's intention to provide an online review. This result is similar to our finding that the conscientiousness of a crowd worker influences both the novelty and usefulness of his or her idea.

Maalej and Nabil [42] introduce techniques to classify application reviews as bug reports, feature requests, user experiences, or ratings. They test their techniques on AppStore and Play Store reviews, obtaining high precision and recall. An interesting opportunity would be to adapt such techniques to automatically classify ideas generated by crowd workers, albeit, potentially along different dimensions.

VII. Conclusions

We provide a new perspective on scaling Crowd RE by exploiting human factors. We describe a sequential task design for acquiring creative requirements from the crowd, considering that group work may not be viable in crowd settings. We show that a worker's personality traits and creative potential have significant influence on the creativity of the worker's ideas, which can eventually be expressed as requirements. Considering that a sequential design could lead to information overload on crowd workers, we develop strategies for selecting ideas from one stage to expose to workers in another. Further, we identify idea selection strategies that are most effective in stimulating crowd workers to generate creative ideas.

An important direction for future work is automating some parts of our sequential task design. Specifically, we envision that the need for crowd raters (e.g., our second phase workers) can be minimized by developing data-driven techniques, e.g., clustering the ideas based on the textual content may yield insights on the novelty of the idea. Our intuition is that novel ideas belong to smaller clusters, whereas mundane ideas belong to larger clusters. Another opportunity for automation in our process is to model a worker's personality and creative potential as functions of his or her ideas. Such models can eliminate the need for explicit personality and creativity questionnaires. However, such models and their validity with respect to established psychological instruments (questionnaires) remain to be explored.

Acknowledgments: Thanks to the NCSU Laboratory of Analytic Sciences for partial support, and to Özgür Kafalı and the anonymous reviewers for helpful comments.

References

[1] N. Maiden, S. Jones, K. Karlsen, R. Neill, K. Zachos, and A. Milne, "Requirements engineering as creative problem solving: A research agenda for idea finding," in *Proceedings of the 18th IEEE International Requirements Engineering Conference*, Sydney, Sep. 2010, pp. 57–66.

[2] A. Adepetu, K. A. Ahmed, Y. A. Abd, A. A. Zaabi, and D. Svetinovic, "CrowdREquire: A requirements engineering crowdsourcing platform." in *AAAI Spring Symposium: Wisdom of the Crowd*, ser. AAAI Technical Report, vol. SS-12-06, 2012.

[3] T. D. Breaux and F. Schaub, "Scaling requirements extraction to the crowd: Experiments with privacy policies," in *Proceedings of the 22nd IEEE International Requirements Engineering Conference*, Karlskrona, Sweden, Aug. 2014, pp. 163–172.

[4] R. B. Svensson and M. Taghavianfar, "Selecting creativity techniques for creative requirements: An evaluation of four techniques using creativity workshops," in *Proceedings of the 23rd IEEE International Requirements Engineering Conference*, Ottawa, Aug. 2015, pp. 66–75.

[5] A. Warr and E. O'Neill, "Understanding design as a social creative process," in *Proceedings of the 5th Conference on Creativity & Cognition*, London, 2005, pp. 118–127.

[6] P. B. Paulus and H.-C. Yang, "Idea generation in groups: A basis for creativity in organizations," *Organizational Behavior and Human Decision Processes*, 82(1):76–87, 2000.

[7] D. Retelny, S. Robaszkiewicz, A. To, W. S. Lasecki, J. Patel, N. Rahmati, T. Doshi, M. Valentine, and M. S. Bernstein, "Expert crowdsourcing with flash teams," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, Honolulu, 2014, pp. 75–85.

[8] P. André, R. E. Kraut, and A. Kittur, "Effects of simultaneous and sequential work structures on distributed collaborative interdependent tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, 2014, pp. 139–148.

[9] A. Fink, R. H. Grabner, D. Gebauer, G. Reishofer, K. Koschutnig, and F. Ebner, "Enhancing creativity by means of cognitive stimulation: Evidence from an fMRI study," *NeuroImage*, 52(4):1687–1695, 2010.

[10] L. Yu and J. V. Nickerson, "Cooks or cobblers? Crowd creativity through combination," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 1393–1402.

[11] T. Halfhill, E. Sundstrom, J. Lahner, W. Calderone, and T. M. Nielsen, "Group personality composition and group effectiveness: An integrative review of empirical research," *Small Group Research*, 36(1):83–105, 2005.

[12] I. Lykourentzou, A. Antoniou, Y. Naudet, and S. P. Dow, "Personality matters: Balancing for personality types leads to better outcomes for crowd teams," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing*, San Francisco, Feb. 2016, pp. 260–273.

[13] P. B. Paulus and V. R. Brown, "Toward more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming," *Social and Personality Psychology Compass*, 1(1):248–265, 2007.

[14] D. K. Simonton, "Teaching creativity: Current findings, trends, and controversies in the psychology of creativity," *Teaching of Psychology*, 39(3):217–222, 2012.

[15] J. Baer, "Domain specificity and the limits of creativity theory," *The Journal of Creative Behavior*, 46(1):16–29, 2012.

[16] W. Maalej, M. Nayebi, T. Johann, and G. Ruhe, "Toward data-driven requirements engineering," *IEEE Software*, 33(1):48–54, Jan. 2016.

[17] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality," *Psychological Assessment*, 18(2):192–203, Jun. 2006.

[18] H. G. Gough, "A creative personality scale for the adjective check list," *Journal of Personality and Social Psychology*, 37(8):1398–1405, 1979.

[19] T. Bhowmik, N. Niu, A. Mahmoud, and J. Savolainen, "Automated support for combinational creativity in requirements engineering," in *Proceedings of the 22nd IEEE International Requirements Engineering Conference*, Karlskrona, Sweden, Aug. 2014, pp. 243–252.

[20] M. A. Runco and G. J. Jaeger, "The standard definition of creativity," *Creativity Research Journal*, 24(1):92–96, 2012.

[21] K. A. Hallgren, "Computing inter-rater reliability for observational data: An overview and tutorial," *Tutorials in Quantitative Methods for Psychology*, 8(1):23–34, 2012.

[22] S. Y. Sung and J. N. Choi, "Do Big Five personality factors affect individual creativity? The moderating role of extrinsic motivation," *Social Behavior and Personality*, 37(7):941–956, 2009.

[23] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[24] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*. New York: Wiley, 1999.

[25] MATLAB, *Version 8.4.0 (R2014a)*. Natick, MA: The MathWorks Inc., 2014.

[26] A. J. Cooper, L. D. Smillie, and P. J. Corr, "A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale," *Personality and Individual Differences*, 48(5):688–691, 2010.

[27] S. Taggar, "Individual creativity and group ability to utilize individual creative resources: A multilevel model," *The Academy of Management Journal*, 45(2):315–330, 2002.

[28] H. Zhao and S. E. Seibert, "The Big Five personality dimensions and entrepreneurial status: A meta-analytical review," *Journal of Applied Psychology*, 91(2):259–271, 2006.

[29] K. Zachos and N. Maiden, "Inventing requirements from software: An empirical investigation with web services," in *Proceedings of the 16th IEEE International Requirements Engineering*, Barcelona, Sep. 2008, pp. 145–154.

[30] L. Nguyen and G. Shanks, "A framework for understanding creativity in requirements engineering," *Information and Software Technology*, 51(3):655–662, Mar. 2009.

[31] J. Horkoff, N. Maiden, and J. Lockerbie, "Creativity and goal modeling for software requirements engineering," in *Proceedings of the ACM SIGCHI Conference on Creativity and Cognition (C&C)*, Glasgow, Jun. 2015, pp. 165–168.

[32] S. Dallman, L. Nguyen, J. W. Lamp, and J. L. Cybulski, "Contextual factors which influence creativity in requirements engineering," in *Proceedings of the 13th European Conference on Information Systems (ECIS)*, Regensburg, Germany, May 2005, pp. 1734–1745.

[33] L. F. Capretz and F. Ahmed, "Making sense of software development and personality types," *IT Professional*, 12(1):6–13, Jan. 2010.

[34] M. Wiesche and H. Krcmar, "The relationship of personality models and development tasks in software engineering," in *Proceedings of the 52nd ACM Conference on Computers and People Research*. Singapore: ACM, 2014, pp. 149–161.

[35] S. Cruz, F. Q. da Silva, and L. F. Capretz, "Forty years of research on personality in software engineering: A mapping study," *Computers in Human Behavior*, 46(C):94–113, 2015.

[36] M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "The four pillars of crowdsourcing: A reference model," in *Proceedings of the 8th IEEE International Conference on Research Challenges in Information Science*, May 2014, pp. 1–12.

[37] M. Hosseini, A. Shahri, K. Phalp, J. Taylor, R. Ali, and F. Dalpiaz, "Configuring crowdsourcing for requirements elicitation," in *Proceedings of the 9th IEEE International Conference on Research Challenges in Information Science*, May 2015, pp. 133–138.

[38] J. Bhatia and T. D. Breaux, "Towards an information type lexicon for privacy policies," in *Proceedings of the 8th IEEE Workshop on Requirements Engineering and Law*, Ottawa, Aug. 2015, pp. 19–24.

[39] S. L. Lim and A. Finkelstein, "Stakerare: Using social networks and collaborative filtering for large-scale requirements elicitation," *IEEE Transactions on Software Engineering*, 38(3):707–735, May 2012.

[40] P. K. Murukannaiah, A. K. Kalia, P. R. Telang, and M. P. Singh, "Resolving goal conflicts via argumentation-based analysis of competing hypotheses," in *Proceedings of the 23rd IEEE International Requirements Engineering Conference*, Ottawa, Aug. 2015, pp. 156–165.

[41] S. Picazo-Vela, S. Y. Chou, A. J. Melcher, and J. M. Pearson, "Why provide an online review? An extended theory of planned behavior and the role of Big-Five personality traits," *Computers in Human Behavior*, 26(4):685–696, Jul. 2010.

[42] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? On automatically classifying app reviews," in *Proceedings of the 23rd IEEE International Requirements Engineering Conference*, Ottawa, Aug. 2015, pp. 116–125.