Acquiring Creative Requirements from the Crowd

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

Conducted on Amazon MTurk, our study involves (1) acquiring requirements for smart home applications from 300 participants, and (2) rating the creativity (usefulness and novelty) of the acquired requirements from an additional 300 participants. We acquire requirements from the crowd in a sequential process, where requirements acquired in one stage are exposed to workers in the next stage to acquire more requirements. To reduce the potential for information overload in a sequential Crowd-RE process, we develop strategies for selecting requirements from one stage to expose to workers in a later stage.

We find that (1) a crowd worker's personality traits and creativity potential have significant influences on the creativity of the worker's ideas; (2) exposing a new worker to ideas from previous workers cognitively stimulate the new worker to produce creative ideas; and (3) personality traits and creative potential are an effective means of selecting a few requirements from a pool of previous requirements to expose a new worker to. Our findings offer insights on how to carry out Crowd RE effectively.

***Index Terms*—Requirements; 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 is still far from the reach of the artificial. Crowd RE [2] is an emerging avenue for soliciting human intelligence for various RE tasks from the members of the public (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 [4].

Accordingly, our **first objective** is to facilitate crowdsourcing of RE tasks that require humans to exercise creativity, an advanced 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 [5]. As Maiden et al. [1] note, many existing RE techniques are designed to explore a search space of known requirements, but not novel requirements; thus, idea finding is an area that promises a high potential for importing established creativity techniques into the RE domain. Further, we focus on acquiring requirements for products for which a crowd member can naturally play the stakeholder role during idea generation—potentially all consumer products fit the bill.

Idea generation is often understood as a social process [6]. 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*) [7]. Despite recent efforts [8], [9], 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) [6]. Although such threats can be reduced by recruiting groups carefully, doing so is is more difficult in crowdsourcing than in traditional settings.

Our **second 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 [10]. A potential solution to facilitate 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 [11], and may even perform better than simultaneous task designs in some cases [9]. A challenge with sequential design, though, is selecting ideas from previous stages to expose to a new worker in the current stage. Expecting a new worker to process all previous ideas (potentially thousands) is not viable both economically and because of the associated cognitive overload.

Our **third objective** is to understand how to select a few ideas from a pool of previously acquired ideas to cognitively stimulate a new 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 influences how one perceives their environment and interacts within it. There is evidence in both traditional [12] and crowdsourced [13] groups 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 [14]. Thus, it is conceivable that a worker's ability to make such connections depends on his or her creative potential.

The foregoing intuitions prompted us to systematically investigate how personality traits and creative potential of workers influence crowdsourced creative tasks. To do so, we conducted a study on Amazon Mechanical Turk (MTurk), requiring 300 participants to come up with ideas (in a user story format) for a future smart home in three sequential stages. We measured these participants personality traits and creative potential before collecting the ideas. We then employed an additional 300 participants to rate the acquired ideas for their clarity and creativity (novelty and usefulness).

We seek to answer the following research questions.

RQ1. How do a worker's personality traits and creative potential influence the creativity of the ideas he produces?
RQ2. What influence do matches or mismatches in personality traits and creative potential of (1) a new worker and (2) the workers whose ideas the new worker is exposed to have on the ideas produced by the new worker?
RQ3. How effective is a sequential task design for acquiring creative ideas via Crowd RE?

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 potential 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 will 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 [15], an increasing amount of evidence suggests that creativity is domain-specific [16]. 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 (RQ1). Second, our study is the first, across domains, to understand how exposing a user to others' ideas based on the personalities and creative potentials of the user and the others influence the creativity of the user's ideas (RQ2 and 3).

Significance: There is an increasing recognition to the crowd's role in RE. 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 [17]. In contrast to these, we offer a fresh perspective on scaling Crowd RE by exploiting the human factors or personality traits and creative potential. We show that these factors can be inexpensive to measure, yet 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. Section IV presents our results. Section V discusses our results. 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 generate ideas for smart home applications. In the second phase, an additional 300 participants 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 design a three-stage sequential process for acquiring ideas from the crowd. Participants in the first stage generate an initial set of ideas. Participants in the second stage are exposed to ideas from the first stage and asked to generate more ideas. Similarly, participants in the third stage are exposed to ideas from the second stage and asked to generate more ideas.

Figure 1 shows an overview of the two phases and three stages of the first phase of our study.

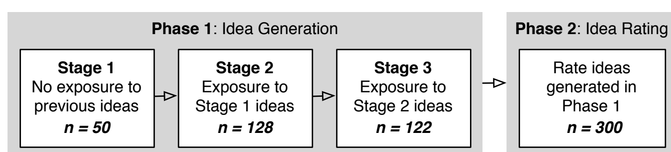

Fig. 1: An overview of our two-phase study design.

As Figure 2 shows, each participant in the idea generation phase answers a personality survey and a creativity survey before completing the idea generation task. The idea selector chooses a few ideas from the pool of previous ideas based on one of the six strategies we specify (II-A2). Since the first stage in this process does not have any previous ideas, the idea selector applies only to the second and third stages.

1) Personality and Creativity Surveys: Before acquiring ideas, we assess each participant's personality and creative potential. First, we employ the Mini-IPIP (International Personality Item Pool) [18] scales to measure a participant's Big Five personality traits of Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to experience (O). The Mini-IPIP scales consists of 20 items

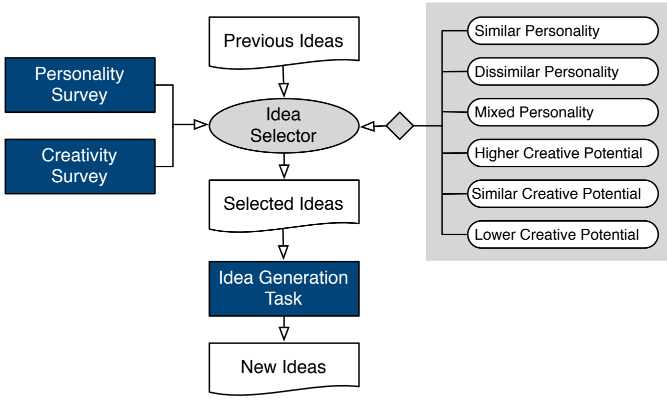

Fig. 2: An overview of our task design the idea generation phase.

(11 negative items)—four items for each Big Five trait. A participant responds to each of the 20 items on a Likert scale of 1 (strongly disagree) to 5 (strongly agree). A score for each Big Five trait is then computed as the mean of the positive and reverse-scored negative items corresponding to the trait.

Second, we employ the Creative Personality Scale (CPS) [19] for assessing a participant'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). A participant answers whether each of those items describes them on a Likert scale of 1 (strongly disagree) to 5 (strongly agree). A single creative potential score is then computed 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 increase the likelihood that research participants would drop out before getting to the main task [18].

2) Idea Generation Task: The main task for participants in this phase is to come up with ideas for smart home applications. We provided basic information about smart homes and smart home requirements, and encouraged participants to conduct additional research on smart homes. We described the task and incentivized participants 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 of your requirement that is more creative than the ones shown to you (as judged by us), you will receive a 20 cents bonus (up to a maximum of USD 1).

We required each participant to produce at least ten ideas, each in a user story format, as shown in Figure 3. We asked participants to produce ideas distinct from the samples shown to them. However, we encouraged combinational creativity [20] by telling participants that they can make creative connections among the samples shown to them to come up with new ideas. Finally, we asked participants to choose an application category (health, safety, energy, entertainment, or other) and, optionally, provide a few descriptive tags for each idea.

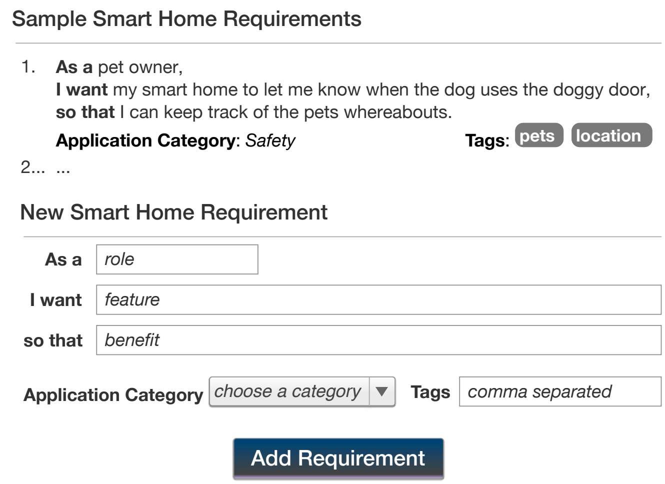

Fig. 3: A screen mockup of the ideas survey (user story format).

For participants in the second and third stages of our sequential process, we showed ten sample requirements from the previous stage (for the first stage, we showed one example requirement from us). We choose the sample requirements via one of the three personality-based and three creative potential-based strategies (shown in Figure 2) described next.

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

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

Next, for each new participant, we find a few previous participants based on the personality distances, and expose the new participant to ten ideas randomly selected from those of the chosen participants. We choose the previous participants, depending on the selection strategy, as follows.

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

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

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

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

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

Higher creative potential: Three previous participants from one quantile above the quantile to which the new participant's creative potential belongs ($n_2 = 23$; $n_3 = 20$).

Similar creative potential: Three previous participants from one same quantile as the one to which the new participant's creative potential belongs ($n_2 = 25$; $n_3 = 22$).

Lower creative potential: Three previous participants from the quantile below the one to which the new participant's creative potential belongs ($n_2 = 19$; $n_3 = 21$).

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

B. Phase 2: Idea Rating

Participants in the first phase generated a total of 2,966 ideas, after we excluded a participant's idea if it is unrelated to smart homes or identical to a sample idea shown to the participant. However, to answer our RQs, we must compare the creativity of the generated ideas. Thus, in the second phase of our study, we ask 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 participants (distinct from the first phase participants), such that at least three raters rated each idea bundle.

Just as we did in the first phase, we provided background information about smart homes and smart home requirements to the second phase participants. We also required the second phase participants to generate at least three smart home ideas before proceeding to the rating task. We did so to encourage participants incubate on 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 its broadly accepted definition [21], entails *novelty* and *usefulness*. Thus, we ask participants to rate each idea on these two criteria. Further, we realize that a participant may not clearly understand some ideas. Considering creativity ratings for such ideas may lead to false conclusions. Thus, as shown in Figure 4, we also ask participants to rate the *clarity* of each idea. We described these criteria to the participants 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 as original and unexpected, i.e., something that is not commonplace, mundane, or conventional.

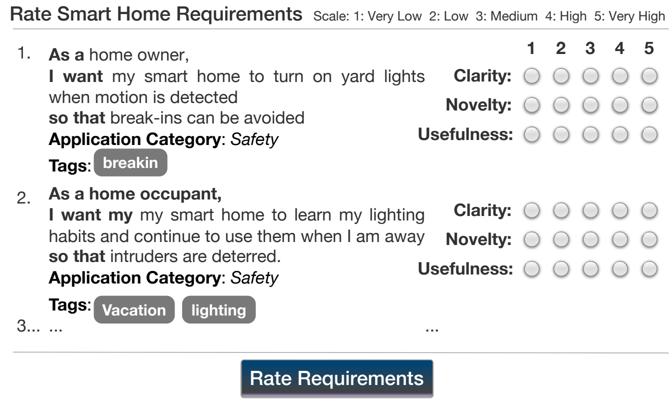

Fig. 4: A screen mockup of an idea rating screen.

We asked participants 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 participant—one for each bundle of ten ideas assigned to the participant—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 our Phase 1 and Phase 2 participants. We provide USD 3 as base for each participant, who completed all assigned tasks. As mentioned earlier, Phase 1 participants could earn an extra dollar as creativity bonus. Considering the difficulty of Phase 1 task, we provided the bonus to more than half of Phase 1 participants.

TABLE I: Time spent by our participants and the payment we provide

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

III. EVALUATION

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

A. Hypotheses

To answer R1, we evaluate the following hypotheses.

$H1A_{null}$ (Null hypothesis): A worker's personality trait has no influence on the creativity of the ideas the worker produces.

$H1A_{alternate}$ (Alternate 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 also evaluate a similar pair of hypotheses for creative potential.

$H1B_{null}$ A worker's creative potential has no influence on the creativity of the ideas the worker produces.

$H1B_{alternate}$ A worker's creative potential influences the creativity of the ideas the worker produces.

To answer R2, we evaluate the following hypotheses.

$\mathbf{H2A_{null}}$ 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 strategies.

$\mathbf{H2A_{alternate}}$ 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 strategies.

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

$\mathbf{H2B_{null}}$ 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 strategies.

$\mathbf{H2B_{alternate}}$ 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 strategies.

To answer R3, we evaluate the following hypotheses.

$\mathbf{H3_{null}}$ The creativities of the ideas produced in Stages 1, 2, and 3 of our sequential Crowd-RE process are the same.

$\mathbf{H3_{alternate}}$ 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 alternate 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 inter-rater reliability (IRR) for ratings of each idea via the intra-class correlation (ICC). The ICC is a commonly-used statistic for assessing IRR for ordinal data [22] 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 analyses.

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*, based on 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 [23].

We employ multiple regression (MR) models [24] 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 only indicates whether a relationship exists at all. We reject the null hypothesis about a factor influence at 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 higher the value of a factor's coefficient, the higher 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 [24]. That is, considering the outcome as Y and the $\hat{Y}$ as the outcome predicted from optimally weighted factors, $R^2 = \mathrm{sd}^2_{\hat{Y}}/\mathrm{sd}^2_{Y}$. 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 [25]. The Kruskal-Wallis test is an extension of the Wilcoxon ranksum test and a non-parametric version of the one-way ANOVA (and thus, it does not require 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) on Matlab [26].

IV. RESULTS

Figure 5 shows the distributions of our participants' personality traits and creative potentials. In these and other boxplots we show, the diamond dots represent the mean of the distribution, and the x marks outside boxes indicate outliers.

The distributions of individual traits in Figure 5 shows that our data represents a variety of personalities. Also, the mean and standard deviation of the traits in our data follow a pattern similar to those reported in previous studies [18], [27].

A. H1A and H1B: Factor Influences on the Outcomes

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

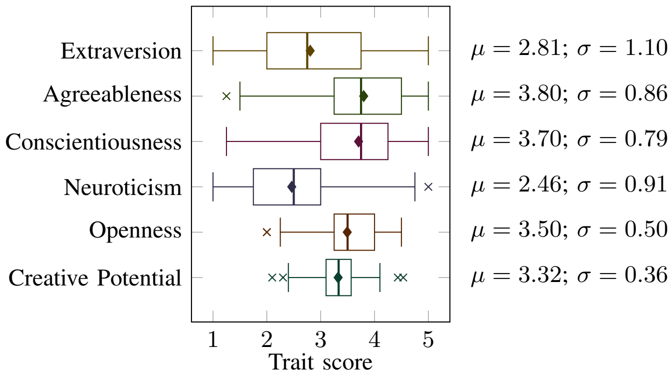

Fig. 5: Distributions of the factors of our interest (personality trait and creative potential scores) in our data.

TABLE II: 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	.012		.020	

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

Second, we find that a participant's agreeableness has the highest influence on the usefulness of the participant's ideas, closely followed by conscientiousness. Further, we find that extraversion has a significant, but negative influence on the usefulness of a participant'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 bad in predicting the outcomes from factors in our data. Nonetheless, as we alluded to before (III-B2), R^2 values do not affect our conclusions about the trends in the data, specifically, the significance of factor influences.

B. H2A and H2B: Idea Selection Strategies

Table III and Figure 6 compare the three personality-based idea selection strategies for Stages 2 and 3, separately. These results suggest that participants exposed to others' ideas via the mixed personality-based strategy produce more novel idea than those exposed via the other personality-based strategies. We also find that this result is consistent for both stages. However, we did not find sufficient evidence to reject the null hypothesis that the usefulness of a participant's ideas are the same irrespective of the personality-based strategies employed.

Table IV 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, though, 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. Next, we fail to reject the null hypothesis about the usefulness ratings for creative potential-based strategies, too.

TABLE III: Comparing personality-based idea selection strategies

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

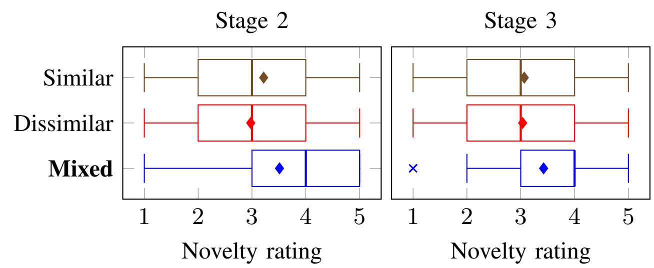

Fig. 6: Novelty ratings of ideas generated by participants exposed to others' ideas via different personality-based selection strategies

TABLE IV: Comparing creative potential-based selection strategies

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

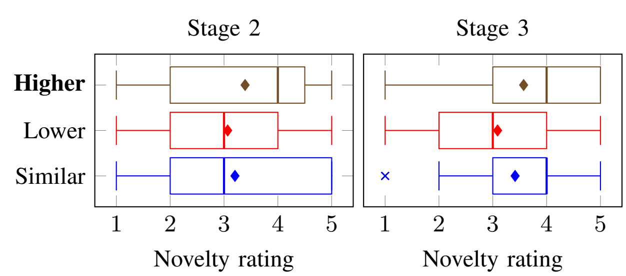

Fig. 7: Novelty ratings of ideas from participants exposed to others' ideas via different creative potential-based selection strategies

Although not significant, it is still important to notice that the differences in usefulness ratings between different strategies follow similar patterns to those of novelty ratings. That is, the mixed personality-based strategy and the higher creative potential-based strategy are the most effective strategies, of their kind, in producing useful ideas.

C. H3: Comparing Ideas across Stages

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

Next, instead for considering all ideas from Stage 2 and 3, we consider ideas produced by participants exposed to mixed personality-based and higher creative potential-based idea selection strategies, only. 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 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 compared to Stage 2. Finally, we do not find significant differences for usefulness ratings across stages.

TABLE V: 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)	0.13	3.78 (1.21)	0.11
	2	3.21 (1.34)		3.86 (1.15)	
	3	3.26 (1.30)		3.96 (1.11)	
Mixed Personality or Higher Creative Potential	1	3.05 (1.37)	0.01	3.85 (1.16)	0.45
	2	**3.45** (1.31)		3.85 (1.17)	
	3	**3.49** (1.25)		3.96 (1.11)	

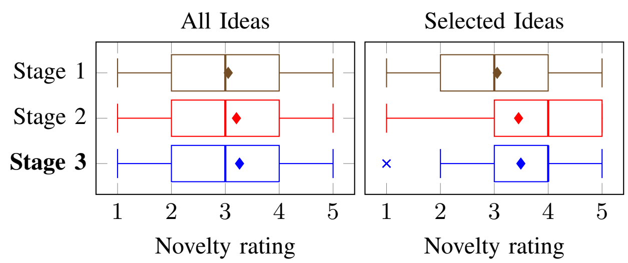

Fig. 8: Novelty ratings of ideas from Stages 1, 2, and 3

D. Demographics

Table VI summarizes the demographics information about participants we collected in a presurvey. We compare if idea ratings across different levels (e.g., male, female, or other for gender) of a demographic factor. We do not find significant differences in novelty and usefulness ratings based on the demographic factors of gender, age, and education.

TABLE VI: Demographics of our study participants

Gender	Male: 48.1%, Female: 51.3%, Other: 0.6%
Age	18–24: 15%, 25–34: 55%, 35–45: 20%, 45–54: 6.5%, 55 or older: 3.5%
Education	Graduate degree: 15.8%, Bachelor degree: 41.6%, Some college but no degree: 28.2%, High school: 14.4%

V. Discussion

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

RQ1. 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 can have significant influences on the creativity of the ideas he produces. Whereas a 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 the 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 [28], or individually but in a social environment, e.g., in a class [23]. Our result about openness is consistent with results in these studies.

A major disagreement between our results and those from previous studies is about 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 extravert seeks excitement and stimulation [29]. 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, [28] and [23], do not agree on the influence of conscientiousness on creativity. Whereas Taggar [28] found evidence for a positive relationship, Sung and Choi [23] 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 plays an important role in influencing the individual to be creative.

RQ2. What influence do matches (or mismatches) in personality traits and creative potential of a new worker and the workers whose ideas the new worker is exposed to have on the ideas produced by the new worker?

First, we find that exposing a new worker to a mix of ideas—some from previous workers whose personalities are similar to the new worker and some from previous 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 [13]. Since an individual's personality influences how he or she perceives an environment [12], we imagine that the personality also 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 exposing a new worker to ideas from previous workers of similar or higher creative potential than the new worker prompts the worker to produce ideas of greater novelty than exposing the new worker to ideas from workers of lower creative potential. We attribute this result to our incentive structure. Recall that we offered workers up to a one dollar bonus for producing more novel ideas than the ideas shown to them. Thus, showing less novel ideas to a worker, to start with, may not force 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 more variety in our data for novelty ratings than for usefulness ratings. The usefulness ratings are skewed toward high ratings, with more than 60% ratings as high or very high. It is possible that, this amount of variety is not sufficient for our tests (Kruskal-Wallis) to detect significance, even if there was underlying significance (a potential false negative or Type II error).

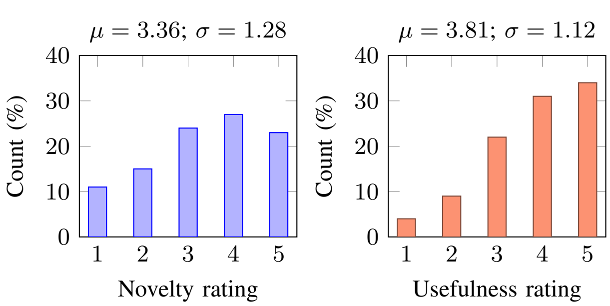

Fig. 9: Distributions of our outcomes (novelty and usefulness ratings).

RQ3. 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 ideas to expose new workers to are selected carefully. Although the creativity or 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.

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. [20] describe a creativity framework that mines ideas using topic modeling, and creates novel and innovative requirements by obtaining unfamiliar connections. Zachos and Maiden [30] describe AnTiQue, a tool based on Structured Mapping Technique to retrieve web services analogical to a requirement problem. They empirically evaluate the tool and find that it can help in generating novel requirements. Such tools are complementary to our Crowd RE process. Such tools can can benefit our process by helping to reduce mundane tasks in our process, e.g., ratings ideas.

Svensson et al. [5] compare four creativity techniques (Hall of Fame, Constraint Removal, Brainstorming, and Idea Box) by conducting creativity workshops with students and practitioners, finding that (1) Brainstorming generates the most number of ideas, and (2) Hall of Fame generates maximum number of creative requirements that are part of future release of products. Our study does not require participants to follow a specific creativity technique. An interesting extension to our study would be to investigate personality influences in context of specific creativity techniques. However, a challenge is to adapt some of these techniques from group to crowd settings.

Nguyen and Shanks [31] build a theoretical framework consisting of product, process, domain, people and context to understand creativity in RE. They acknowledge the need for investigating personality characteristics and traits of people involved in performing creative tasks. That is what we do.

Horkoff et al. [32] combine creativity techniques with goal modeling, and suggest starting with transformatory technologies, followed by exploratory creativity, 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

Dallman et al. [33] study contextual factors including social and individual dimensions that influence creativity in RE. They identify that risk taking personality characteristic is an important individual dimension for creativity.

Despite ample evidence that personality traits impact creativity in domains such as psychology and management, to the best of our knowledge, none of the existing works relate personality and creativity in the context of an RE task.

Broadly in Software Engineering: There are a few software engineering works that study personality. Capretz and Ahmed [34] correlate software job requirements and soft skills with psychological traits, measured via Myers-Briggs Type Indicator. They map software job roles and soft skills. Capretz and Ahmed's mapping of introversion with ability to work independently and analyse business requirements aligns with our finding that introvert crowd workers working independently perform better on creative tasks.

Wiesche and Krcmar [35] present a systematic literature review on software developers' personalities, and examine the effect of the fit of task characteristics and developer's personality on satisfaction and performance. Although they found certain review results to be contradicting, their review and other works within, are consistent with our finding that openness to experience influences creativity.

Cruz et al. [36] also systematically review research on personality in software engineering. Topics such as team composition in pair programming, education, team effectiveness, process allocation, individual performance, and leadership effectiveness have been studied extensively. 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. [37] understand the crowdsourcer, crowd, crowdsourced task, and crowdsourcing platform as four pillars of Crowd RE. They also analyse different features of the crowd and the crowdsourcer to see how these features impact the quality of elicited requirements [38]. Their 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 as valuable to Crowd RE. as their features.

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 extractions and increase the coverage. Bhatia and Breaux [39] use crowdsourcing to construct a lexicon of information types for privacy policies. Whereas microtasks in these works only require basic human abilities, our ventures into exploiting crowd workers creativity in a requirements acquisition task.

Lim and Finkelstein propose StakeRare [40], a collaborative filtering based method to support requirements elicication in social networks. They empirically evaluate StakeRare on a large-scale software project. StakeRare identifies and prioritize stakeholders, and it recommends requirements to a stakeholder based on his or her initial set of ratings. Crowd workers often work independently. However, crowdsourcing in social networks is an emerging topic. Future work can explore the role personality and creative potential play in tasks crowdsourced on social networks.

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 evaluate their model and 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 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 can have significant influence on the novelty 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. We show that some of the strategies we develop are quite effective in stimulating crowd workers to generate novel 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 Phase 2 participants) 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, the effectiveness of such models against carefully designed questionnaires remains to be seen.

Acknowledgments: Thanks to the US Department of Defense for support through the NCSU Laboratory of Analytic Sciences and the Science of Security Lablet.

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 IEEE 22nd International Requirements Engineering Conference*, Karlskrona, Sweden, Aug. 2014, pp. 163–172.

[4] A. J. Quinn and B. B. Bederson, "Human computation: A survey and taxonomy of a growing field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, 2011, pp. 1403–1412.

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

[6] 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.

[7] P. B. Paulus and H.-C. Yang, "Idea generation in groups: A basis for creativity in organizations," *Organizational Behavior and Human Decision Processes*, vol. 82, no. 1, pp. 76–87, 2000.

[8] 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.

[9] 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.

[10] 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*, vol. 52, no. 4, pp. 1687–1695, 2010.

[11] 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.

[12] 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*, vol. 36, no. 1, pp. 83–105, 2005.

[13] 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 Fransisco, Feb. 2016, pp. 260–273.

[14] 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*, vol. 1, no. 1, pp. 248–265, 2007.

[15] D. K. Simonton, "Teaching creativity: Current findings, trends, and controversies in the psychology of creativity," *Teaching of Psychology*, vol. 39, no. 3, pp. 217–222, 2012.

[16] J. Baer, "Domain specificity and the limits of creativity theory," *The Journal of Creative Behavior*, vol. 46, no. 1, pp. 16–29, 2012.

[17] W. Maalej, M. Nayebi, T. Johann, and G. Ruhe, "Toward data-driven requirements engineering," *IEEE Software*, vol. 33, no. 1, pp. 48–54, Jan 2016.

[18] 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*, vol. 18, no. 2, pp. 192–203, Jun. 2006.

[19] H. G. Gough, "A creative personality scale for the adjective check list," *Journal of Personality and Social Psychology*, vol. 37, no. 8, pp. 1398–1405, 1979.

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

[21] M. A. Runco and G. J. Jaeger, "The standard definition of creativity," *Creativity Research Journal*, vol. 24, no. 1, pp. 92–96, 2012.

[22] K. A. Hallgren, "Computing inter-rater reliability for observational data: An overview and tutorial," *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, pp. 23–34, 2012.

[23] 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: An International Journal*, vol. 37, no. 7, pp. 941–956, 2009.

[24] 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.

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

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

[27] 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*, vol. 48, no. 5, pp. 688–691, 2010.

[28] S. Taggar, "Individual creativity and group ability to utilize individual creative resources: A multilevel model," *The Academy of Management Journal*, vol. 45, no. 2, pp. 315–330, 2002.

[29] H. Zhao and S. E. Seibert, "The Big Five personality dimensions and entrepreneurial status: A meta-analytical review," *Journal of Applied Psychology*, vol. 91, no. 2, pp. 259–271, 2006.

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

[31] L. Nguyen and G. Shanks, "A framework for understanding creativity in requirements engineering," *Information and Software Technology*, vol. 51, no. 3, pp. 655 – 662, 2009.

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

[33] S. Dallman, L. Nguyen, J. Lamp, and J. Cybulski, "Contextual factors which influence creativity in requirements engineering," in *Proceedings of the 13th European Conference on Information Systems*, Regensburg, May 2005, p. 107.

[34] L. F. Capretz and F. Ahmed, "Making sense of software development and personality types," *IT Professional*, vol. 12, no. 1, pp. 6–13, Jan 2010.

[35] 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.

[36] 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*, vol. 46, pp. 94–113, 2015.

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

[38] 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 (RCIS)*, May 2015, pp. 133–138.

[39] J. Bhatia and T. D. Breaux, "Towards an information type lexicon for privacy policies," in *Eighth IEEE International Workshop on Requirements Engineering and Law (RELAW)*, Ottawa, Aug 2015, pp. 19–24.

[40] S. L. Lim and A. Finkelstein, "Stakerare: Using social networks and collaborative filtering for large-scale requirements elicitation," *IEEE Transactions on Software Engineering*, vol. 38, no. 3, pp. 707–735, May 2012.

[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 tole of Big-Five personality traits," *Computers in Human Behavior*, vol. 26, no. 4, pp. 685 – 696, 2010, emerging and Scripted Roles in Computer-supported Collaborative Learning.

[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 (RE)*, Aug 2015, pp. 116–125.