The Influence of Participants' Personality on Quantitative and Qualitative Metrics in Usability Testing

Thomas Schmidt
thomas.schmidt@ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Vera Wittmann
vera1.wittmann@stud.uni-regensburg.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Christian Wolff
christian.wolff@ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

ABSTRACT

We present the results of a usability study with 35 participants investigating the influence of personality on various metrics used in usability engineering. We conduct a task based usability test with a website integrating tasks of various difficulty and measure performance metrics like task completion rate and time on task. We also use standard questionnaire based usability metrics like the System Usability Scale. Furthermore, we gather qualitative data via open-ended questions and count the number of words as well as the mentions of positive and negative aspects. We measure personality using the well-known big five model, also often referred to as OCEAN model (openness, conscientiousness, extraversion, agreeableness, neuroticism) and three basic needs (need for influence and power, need for recognition and performance, need for security and tranquility). We analyze the relationship between personality and usability metrics via correlations and regression models. We identify multiple significant results and show that in our study the personality correlated with some of the usability metrics we inspected. Extraversion and the need for influence and power show the most and strongest correlations. Furthermore, we also show that regression models based on personality traits can explain up to 37% of the variance in usability metrics. The results have implications for the improvement of the selection process of usability test participants as well as for the interpretation of test results. We discuss these implications and give an outlook on further research in this area.

CCS CONCEPTS

• **Human-centered computing** → **Usability testing**; *HCI theory, concepts and models*; *Empirical studies in HCI*; • **Social and professional topics** → **User characteristics**.

KEYWORDS

Usability Engineering, Personality, Big Five, OCEAN, Usability, Usability Testing, Participant Selection, Extraversion

1 INTRODUCTION

Usability is one of the most important goals in software development and human computer interaction. There are various definitions of usability [19]. One of the most common ones is to describe usability as the degree to which users can achieve their goals effectively, efficiently and in a satisfying fashion when interacting with software or machines [25]. A high degree of usability has a positive effect on multiple important factors like the loyalty towards a software [20], trust [13] and the intention to buy something in online shops [27]. Because of this importance, the field of usability engineering has evolved strongly in the past years. Research in this area produced various methods and guidelines to evaluate and improve the usability of software [1]. One of the most important and common methods is the usability test. A traditional usability test is a session between a moderator and a test participant. The participant is presented with tasks to perform on the product that is tested. There are numerous variations concerning the details of this setting but some established metrics one can acquire during such tests are the *task completion rate*, the *time needed* as well as *self-reported data* by the participant via questionnaires or the "thinking aloud" method [1]. Usually, usability practitioners acquire a rather small sample of test persons for the usability test

(e.g. 5-10 participants [1, p. 53]) that is representative of the characteristics and demographics of the final end users [1].

To correctly interpret the results of usability tests, it is important for researchers and usability practitioners to know about factors that can influence and skew usability metrics and the behavior of participants. Various attributes concerning the setting like the influence of the laboratory set-up [47] or the prototype fidelity of the software [44] have been examined. Nevertheless, there is also reason to analyze the influence of the participants' characteristics on the performance and behavior in usability tests. They can be an important factor concerning the selection of test participants: since typical sample sizes are rather small, usability practitioners might want to recruit specifically test persons that are rather critical and give much and constructive feedback. Furthermore, keeping the influence of participants' characteristics in mind is important for the correct interpretation of usability metrics and might explain variance in performance behavior (e.g. if certain user groups in general tend to take longer for tasks or rate software better).

Most studies concerning this research question are focused on traditional demographic characteristics like the cultural background [14, 51] and age [48]. Another factor often examined is user expertise [39, 44]. However, there are only few studies regarding the influence of the participants' personality on performance metrics and behavior in usability tests. There is reason to examine this factor more deeply since research has shown that personality is an important predicting factor for holistically understanding human behavior in multiple other areas. Personality has an influence on cognitive performance [42], perceived well-being [23, 45], confidence [40], creativity [7] workplace behavior [37], job performance [24] and artistic preferences [22]. The influence of personality can also be observed in areas very close to human computer interaction like social media behavior [32], information privacy concerns [29], information seeking behavior on the web [46] or technology acceptance [50].

There are few exploratory studies that have indeed proven the influence of personality on various metrics of usability testing with the goal to find personality specific variances or to improve the participant selection process (see section 2). We want to contribute to this research field by using a more sophisticated personality questionnaire and quantitative as well as qualitative data as dependent variables to get a more comprehensive view of the interactions between personality dimensions, usability metrics and participants' test behavior. Therefore, we conducted a traditional task-based usability test with 35 participants operationalizing the personality via a questionnaire for the established *five-factor model* and analyze multiple qualitative and quantitative usability measures.

2 RELATED WORK

In the following section, we give a brief summary of the development of the five-factor personality model in psychology and describe existing work examining the influence of participants' personality on usability testing.

Five-Factor Personality Model

To operationalize personality we use a questionnaire based on the five-factor personality model. Therefore, we shortly describe the historical development of this model: As early as 1911, [49] presented a model in which personality was described as a collection of features and their values. A first listing of properties can be found in [33] who collected 20 psychogenic needs. Using factor analysis, [12] was able to define 16 bipolar dimensions which were later reduced to five by [36]. The five dimensions, also often referred to as *five-factor model, OCEAN model* or *big five* [26] have been thoroughly tested for stability and independence in countless studies [3, p. 107]. The dimensions are neuroticism, extraversion, conscientiousness, openness and social agreeableness. Finally, [16] developed *NEO – Personality Inventory* the most widely used and internationally accepted personality questionnaire based on the big five model which is also the foundation for the German questionnaire used in our study. This model differs from other by the way that people are not assigned a specific personality type like with another frequently used personality model: the *Myers-Briggs Type Indicator* [9]. Instead, personality is described by multiple values for every of the five personality dimensions. While the Myers-Briggs Type Indicator is predominantly used in the business world and training [21] its validity and reliability is often criticized in psychology [8, 38].In contrast, the five factor model is the established personality model in research and clinical psychology [26]. Therefore, we also refer to the five factor model for our study. We will describe the five personality dimensions in more detail in section 4.

Participants' Personality in Usability Testing

There are only few studies examining the influence of personality traits on performance and behavior in usability tests. [11] conducted a usability test for an e-commerce website (N=10). To operationalize personality they used the Myers-Briggs personality test [31] in a preselection process and invited 10 participants with strong forms of extraversion – five of them very introverted and five very extroverted participants. A traditional task based usability test was conducted. Participants were instructed to "think aloud" during the test. All mentioned usability problems were counted. It was shown that extroverts mentioned significantly more usability problems and took more time for all tasks than

introverts. There were no significant differences concerning ratings on the usability questionnaire *System Usability Scale.*

[5] focused for their research project on children as test participants and on how to identify children that will make effective participants in a user test. They regarded children as effective participants with respect to the number of identified usability problems and the ratio of verbally indicated problems among all possible problems. The test product was a computer game aimed at children and 26 preselected children participated. Since children cannot fill out complex questionnaires, personality traits of children were assessed during a longer preselection process with an observational method known in pedagogical psychology operationalizing multiple personality traits. Children could play with the game as they wanted and were instructed to "think aloud". The results of this study showed a significant correlation between the personality trait curiosity and the number of mentioned problems and for extraversion and friendliness on the proportion of problems indicated through self-initiated spoken output. These results lead to the conclusion that the selection of children as participants for usability tests can indeed be improved when taking personality traits into account.

[2] conducted a study similar to [11] and solely focused on the personality trait extraversion. They also employed the Myers-Briggs personality test [31] to operationalize personality and preselected overall 20 individuals with strong forms of extraversion. As test object they used an academic website and performed a traditional task-based usability test with the "thinking aloud" method. Furthermore, they measured multiple usability metrics like error rate, time on task or the number of identified usability issues. While they did not report on significance tests, they describe that extroverts were more successful in completing the task, found more usability problems, spend more time on the tasks and also had a higher error rate. Data gathered via the System Usability Scale did not show any relevant differences between extraverted and introverted concerning the rating.

[28] focused on the influence of personality on the established usability questionnaire System Usability Scale (SUS) and conducted a large-scale study (N=268). The study was carried out in form of an online survey. Participants rated multiple well-known IT products like *Microsoft Word, Google Search* or the *Nintendo Wii* with the SUS in case they reported experience with them. A small 20 question version for the five-factor model was used to operationalize the personality. Multiple significant correlations were found, most importantly *openness to experience* and *agreeableness* had the strongest positive correlations with the subjective usability assessment while *conscientiousness* was the weakest dimension and did not show any important interactions.

Although previous research was able to identify some interesting correlations, there are several limitations. Previous studies either focus on only one specific personality trait like extraversion [2, 11], on one type of usability metrics like survey data [28] or on a very specific user group like children [5]. We want to address those limitations with our study to advance the research area by analyzing multiple personality dimensions isolated and combined and their influence on various quantitative and qualitative usability metrics.

3 RESEARCH QUESTION

The overall research questions states as follows:

Is there a relationship between the personality of a test participant and usability metrics?

Since there is not a lot of research in this area, we formulate the research questions rather exploratory. Personality will be operationalized via the five-factor model and three basic needs. Multiple usability metrics like performance metrics, qualitative metrics and ratings of the SUS will be analyzed concerning relationships. First, we will examine correlations and use the five dimensions and three basic needs as separate independent variables with each of the usability metrics as dependent variables examining questions like e.g. *is there a correlation between neuroticism and the task completion rate?* Furthermore, to get a more comprehensive view on the relationship, we will analyze if the combined personality dimensions and basic needs can successfully explain variance concerning the usability metrics and in what way personality models can predict usability metrics via regression models.

4 METHODS

We conducted a task-based usability test similar to the frequently used *guerilla usability test* [34], gathered different established usability metrics and operationalized the personality via a psychometric instrument.

Test Object

We decided to use a website of a local adult education center as test object for the usability tests . For the majority of our sample, students of our university, the general usage and goals of this website are well known and understandable, as the main functionality of the website is to look for different types of courses, a task students are familiar with in the context of their own course of study at the university. Nevertheless, the website itself is rather unknown to students, since it is targeted towards employed adults. Furthermore, we also found numerous usability problems on this website. The website was fitting in confronting participants with different kinds of usability problems.

We conducted a heuristic expert evaluation of the website to systematically include usability issues in our tasks. We used *Nielsen's 10 Usability Heuristics* [35] and classified each issue according to these principles as well as ranked the

severity of each issue on a scale of 0 (not severe) to 4 (very severe). We identified 36 different usability issues in numerous parts of the website, e.g. minor issues like inconsistent names and symbols and more severe ones like important links that did not work.

Tasks

The process described in the previous subsection allowed us to design and select tasks with different types and levels of usability issues. We developed six tasks becoming more and more difficult regarding the number and severity of the included usability issues. Most of the tasks were information seeking tasks: Participants had to find specific information like the dates of courses, events or bus schedules. Several times participants had to interact with two tools the website offers: the calendar and the shopping cart.

We integrated multiple types of difficulty levels and usability issues so that we are sure participants are indeed confronted with usability problems like in a real usability test and we can analyze the variance concerning the reactions to these problems and tasks. Designing a usability test with no problems would likely lead to rather homogeneous reactions by the participants, but can be regarded in future research.

Personality

We use the *B5T personality questionnaire* [43] to operationalize the participants' personality. The B5T is a German questionnaire based on the five-factor personality model [43]. The five-factor personality model is one of the most well-known and established personality concepts in psychology [26] and less prone to critique than type based personality models (see section 2). It consists of the five dimensions: neuroticism, extraversion, conscientiousness, openness, agreeableness (see table 1).

The B5T questionnaire consists of 72 statements which participants have to judge concerning themselves. Participants can agree or disagree to a statement on a 4 point Likert scale. Examples for statements are "I am a loner", "I can put myself in others people's shoes" or "I make sure I'm always friendly". The majority of the questions relates to one of the personality dimensions of the five-factor model. The B5T also includes questions for three basic needs: *the need for recognition and performance* (in German: Bedürfnis nach Anerkennung und Leistung), *for influence and power* (in German: Bedürfnis nach Macht und Einfluss), *for security and tranquility* (in German: Bedürfnis nach Sicherheit und Ruhe). Note, that these basic needs are not part of the traditional five-factor model. We integrated these factors since [5] did indeed show that traits that are not part of basic personality dimensions can have an influence in usability testing. Via calculations the questionnaire produces a value between 1 (very low) and 9 (very high) for every personality dimension and basic need. The reliability and validity of the questionnaire was confirmed on a sample of approximately 5000 persons, with values for *Cronbach's Alpha* ranging between 0.76 and 0.90. The questionnaire is one of the most widely used personality questionnaires in German speaking countries [43]. Like most psychometric instruments, the questionnaire is based on subjective self-assessment and therefore dependent of the honesty of the participants. However, the questionnaire also includes question to measure the honesty of participants by asking unpleasant question every honest person should affirm. While there is never a guarantee for honest answers, this allows the filtering of rather suspicious participants.

Table 1 illustrates the five personality dimension and basic needs and describes characteristics of persons on the two extremes of those scales. All descriptions are based on [43].

Usability Metrics

We categorize the usability metrics in performance, survey and qualitative metrics.

Performance Metrics. As performance metrics, we collect the task completion rate and the time needed. Both metrics are established performance metrics in usability testing [1].

The tasks were designed in a way that the task completion per task is binary, meaning a task is either completed or failed. A task was successfully completed if the participants landed on a specific page (e.g. a page that contains the information the participant had to look for) or they orally informed the test moderator that the task was finished giving a correct answer concerning the task. The task was failed if the participants informed the test moderator that they cannot finish the task and gave up or assumed to have completed the task but informed the moderator with wrong information. We calculated a measure for the task completion rate by the ratio of successfully completed tasks divided by all tasks, e.g. a participant that succeeded in three tasks but failed for the other three had a task completion rate of 50%.

The time needed was measured in seconds needed to finish the task or to give up. The duration was counted with the first mouse movement per task and finished with the last click of a task performance. For the entire duration of the test, we summed up all task durations.

Survey metrics – System Usability Scale. We used the *System Usability Scale (SUS)* as standard questionnaire to measure perceived usability [10]. The System Usability Scale is one of the most used usability questionnaires for researchers and usability practitioners and is regarded as a highly robust and versatile tool [4]. It consists of 10 statements concerning different aspects of usability; the users can rate on a 5 point Likert scale to what degree they agree with the statement. Via recommended calculations a rated product receives a value

Table 1: Five Personality Dimensions and Three Basic Needs

Personality dimension	Low value	High value
Neuroticism	Emotional stable, calm, relaxed	Emotional unstable, tense, anxious, worried
Extraversion	Restrained, shy, pensive, independent	Sociable, talkative
Conscientiousness	Unorganized, undisciplined, spontaneous	Dutiful, diligent, accurate, structured
Openness (to experience)	Traditional, grounded, prefer the well-known	Open for new experiences, tolerant, curious
Agreeableness	Self-centered, direct, uncooperative	Friendly, cooperative, care for others, helpful
Need for recognition	Not attention seeking, performance focused	Focused on performance, success driven
Need for influence	No desire to have control, easy going, passive	Desire to have responsibilities and be in control
Need for security	Risky, open for novel experiences	Want calm, secure and peaceful conditions

between 0 (very low usability) and 100 (very high usability). We employed a German version of the SUS [41].

Qualitative metrics. Since we are interested in performance metrics we did not employ the "thinking aloud"-Method. The integration of this method can skew calculations for metrics like the time needed. Nevertheless, to gather qualitative data we integrated several open-ended questions in the final questionnaire. With two open-ended questions participants had to report in their own words about all positive and negative aspects of the website they noticed during the usability test. The participants could type as much as they wanted into two textboxes, one for positive and one for negative aspects.

To analyze this data we coded the entire produced text per participant. We marked all negative and all positive mentions in the text. Repetitions of aspects were ignored. Overall, we use the number of positive aspects and the number of negative aspects as dependent variables. Furthermore, we counted the number of words to measure the length of the text and the overall productivity of the text production. Table 2 summarizes all dependent variables.

Familiarity with the Website

It is important for our study design that the participants are unfamiliar with the website. To control this confounding variable we integrated two questions into our questionnaires asking if the website was familiar for the participants (yes/no) and how often they used the website on a 4 point scale with the items: never, rarely, occasional and frequent.

Procedure

The entire usability test was performed in a laboratory room on a laptop. After a short introduction, participants were presented with the tasks one by one. The test moderator briefly explained each task. The tasks were also presented for the participant on a sheet of paper. Participants were instructed to inform the moderator if they feel they have finished the task or if they wanted to give up. During the task performance, the moderator did not distract or influence the participants in any way.

After the usability test, participants completed the questionnaire about the familiarity with the website, the SUS and the open-ended questions. Finally, participants completed the B5T and a standard demographic questionnaire. We put the personality questionnaire on purpose in the end so participants were not influenced by the overall idea of the study and under the impression that this was a normal usability test. This is a common approach in studies examining the influence of personality [46]. Furthermore we informed the participants that all data collected was anonymized and that they should answer the questions as honest as possible.

Sample

35 persons participated in the study. 13 of them males and 22 females. Almost all of the participants were students of media informatics or other degree programs (n=31). The majority of participants were not familiar with the website (n=33) and reported that they never used it (n=34). One participant stated to rarely use the website. Examining the data of this person, no particularities were found. The participants' ages ranged between 20 and 66 years but the majority of participants (n=29) was between 20 and 29 years old (M=28.77, Sd=10.94). The results concerning the honesty scale of the B5T suggest that all participants did indeed answer the personality questions honestly. However, please be cautious since this scale is also prone to subjectivity.

5   RESULTS

We examine the influence of personality on usability metrics via correlations and linear regression models. As level of significance we employ $p < 0.05$.

Descriptive Statistics

Personality. In table 3, we report all important metrics concerning the distribution of the personality dimensions among the test participants. Analyzing the histograms of the personality dimensions they show to be normally distributed among the test participants with most manifestations among values close to the middle (e.g. 5 and 6). However, note that

Table 2: Description of Usability Metrics

Usability metric	Description
Task completion rate (TCR)	Proportion of successfully completed tasks to all 6 tasks
Time (in seconds)	The entire time needed to complete all 6 tasks; sum of every time on task
SUS rating	Final rating considering the SUS questionnaire
Number of words	Number of words used in the open-ended questions
Positive mentions	Number of mentioned positive aspects in the open-ended questions
Negative mentions	Number of mentioned negative aspects in the open-ended questions

Table 3: Descriptive Statistics - Personality Dimensions

Dimension	Min	Max	M	Med	Sd
Neuroticism	3	7	5.46	6	1.25
Extraversion	1	7	4.86	5	1.31
Conscientiousness	2	8	4.94	5	1.51
Openness	1	8	4.49	4	1.85
Agreeableness	1	8	5.37	5	1.85
Need for recognition	2	8	5.26	5	1.44
Need for influence	1	9	4.43	5	1.91
Need for security	1	9	5.71	6	1.79

Table 4: Descriptive Statistics - Usability Metrics

Usability metric	Min	Max	M	Sd
TCR	0.67	1.0	0.92	0.12
Time	424.09	1781.94	837.33	313.36
SUS rating	25.0	92.5	56.29	21.13
Number of words	4	107	33.94	26.63
Positive mentions	0	2	0.83	0.79
Negative mentions	0	6	2.49	1.61

the chosen test for correlations in section 5 is not dependent of normal distribution.

Usability Metrics. Table 4 illustrates the descriptive statistics concerning the usability metrics. All of the participants completed at least 4 of the 6 tasks successfully, the task completion rate solely ranges concerning the completion of two tasks. The entire test took on average about 14 minutes with the fastest participant finishing in about 7 minutes, while the highest duration was 30 minutes. The website achieved an average SUS of 56.29%. According to [4] this can be considered as "OK" overall rating but is very close to poor usability. Note that the reason for this low ratings might very well be that we included on purpose tasks with usability problems. Regarding qualitative data, the data shows that participants varied strongly concerning the length of their answers and remarks, on average around 34 words were used. Participants rarely mentioned positive aspects and rather mentioned negative aspects which is in line with the rather low SUS ratings. Overall, the low ratings and negative mentions validated our choice for this website and the tasks since we needed a software that indeed has usability problems so that metrics are not solely perfect and variance can be found.

Correlations

We analyze the influence of personality on usability metrics via Spearman's rank correlation coefficient *Rho*. All preconditions to use this correlation coefficient are met by our variables. We examine correlations concerning all personality

dimension and the dependent variables. We report significant results ($p < .05$) and other results pointing to general tendencies via the p- and r-value (strength and direction of the correlation). Please note to interpret the general tendencies with caution. To interpret the strength of the correlation we refer to the recommendations of [17]. We made no significant or otherwise important findings concerning the factors *conscientiousness* and the *need for security and tranquility*, therefore we do not report on any results concerning these factors.

Neuroticism. Neuroticism correlates significantly and positively with the task completion rate. It is a correlation (rs(33) = .40, p = .018) of moderate strength according to [17]. The higher the neuroticism value the more likely participants succeeded in completing the tasks. No other variable showed significant correlations.

Extraversion. We identified a significant, moderate and positive correlation between the overall time needed and extraversion (rs(33) = .441, p = .008). However, on the other hand, the correlation between task completion rate and extraversion is negative (rs(33) = -.298, p = .082) but rather weak. Extraverted persons were rather unsuccessful in completing tasks and took more time while introverted were more successful in less time.

Openness. We found a significant correlation between the openness dimension of a participant and the task completion rate (rs(33) = -.342, p = .045). The correlation is rather weak and negative.

The Influence of Participants' Personality in Usability Testing

Agreeableness. Agreeableness did not show any significant correlations. Although not significant, we identified a tendency concerning the open-ended questions. The higher the agreeableness value the higher the number of positive aspects mentioned (rs(33) = .286, *p* = .095).

Need for recognition and performance. A general tendency was found examining the correlation for the SUS value and the need for recognition and performance (rs(33) = -.295, *p* = .086). Persons with higher values on this personality scale tended to rate the website lower concerning the SUS items.

Need for influence and power. Multiple significant correlations were found for the need for influence and power: The scale correlates significantly, moderately and negatively with the task completion rate (rs(33) = -.434, *p* = .009) and positively and moderately with the overall time needed to complete the tasks (rs(33) = .380, *p* = .025). Furthermore, concerning the open-ended questions this personality dimension correlates negatively, moderately and significantly with the number of positive mentions in the produced text (rs(33) = -.370, *p* = .029). On a minor note, there is a tendency concerning the SUS value. We identified a weak and negative correlation (rs(33) = -.295, *p* = .085). Overall, test persons with high values on this scale tend to perform rather bad concerning task completion, take longer and report more negative aspects as well as rate the page lower on the SUS scores. Persons with low values on this scale tend to behave the other way around.

Regression Models

We performed linear regression analysis to examine if and how multiple personality dimensions combined influence usability metrics and if one can combine those dimensions to models that can successfully predict usability metrics and therefore participant performance and behavior.

Similar to [5], we performed exploratory linear regression analysis of all subsets of personality traits. However we only report on models consisting of subsets of variables that showed significant results or general tendencies concerning the correlations (see the previous section). Our analysis showed that those models perform best. We are testing hypothesis like *Does the model consisting of extraversion and the need for influence and power significantly predict the time needed for tasks?* We did not make any significant findings concerning the number of words and the negative mentions in the open-ended questions. We conducted exploratory analysis with multiple linear regression models on those metrics but did indeed not find any significant prediction models. Thus, we will not report on results concerning those usability metrics.

In the following, we report the significance value (*p*) and the F-value (F) of the regression model. We also report the

Table 5: Regression Coefficients (Task Completion Rate)

Variable	Regression coefficient	T	p
(Constant)	1.036	8.815	.000
Neuroticism	.022	1.629	.114
Extraversion	-.027	-2.066	.048
Openness	-.011	-1.079	.289
Need for influence	-.013	-1.378	.178

Table 6: Regression Coefficients (Time)

Variable	Regression coefficient	T	p
(Constant)	317.789	1.546	.132
Extraversion	81.118	2.039	.050
Need for influence	28.348	1.042	.305

R^2-value by describing what proportion of variance in the usability metric can be explained by the model as well as the effect size and interpretation of this size by [15] speaking of a strong effect with f-values higher than 0.4. Next, we report the results concerning all regression coefficients via a table. Besides regression metrics, the tables consist of the results concerning a t-test on the regression coefficients. We report the T- and *p*-value, thus one can inspect if a variable contributes significantly to a model.

Task completion rate. The personality model consisting of neuroticism, extraversion, openness and the need for influence and power shows a significant linear regression on the usability metric task completion rate (F(4, 30) = 4.314, *p* = .007). The model explains 37% of the variance of the task completion rate. According to [15], this is a strong effect (f = .76). Table 5 shows the individual directions and strengths of the personality dimensions concerning the effect on the task completion rate. The strongest predictors are extraversion and neuroticism with extraversion being a negative (T = -2.066, *p* = .048) and neuroticism a positive and rather weak predictor (T = 1.629, *p* = .114). We also tested models with fewer and more variables but the reported one does indeed perform the best.

Time. Concerning the time needed, a significant linear regression with the predictors extraversion and the need for influence and power (F(4,30) = 3.402, *p* = .046) proved to be significant. The model explains 18% percent of the variance which is still a strong effect (f = .47). The detailed analysis (see table 6) of the regression coefficients shows that both variables have a positive correlation with the metric extraversion being the strongest predictor (T = 2.039, *p* = .050).

SUS rating. We conducted a regression analysis with two basic needs: the need for influence and power and the need

Table 7: Regression Coefficients (Positive Mentions)

Variable	Regression coefficient	T	p
(Constant)	.957	1.733	.093
Agreeableness	.087	1.241	.224
Need for influence	-.135	-1.987	.056

for recognition and performance. The linear regression did not show a significant influence of this model on SUS ratings (F(4,30) = 2.085, *p* = .141).

Positive mentions. With the model of agreeableness and the need for power and influence we identified a significant linear regression concerning the positive mentions in the open-ended questions (F(4,30) = 3.602, *p* = .039). The model explains 18% of the variance. This is considered a strong effect (f = .47). Table 7 illustrates the interactions of all influencing factors. Note that both factors do not have a significant influence on their own but only together as regression model, with the need for influence and power having a stronger and negative effect (T = -1.987, p = .056) and agreeableness a weaker and positive one (T = 1.241, p = .224).

6 DISCUSSION

In the following section, we discuss the major findings of our study in the context of previous research, formulate implications for the research area and usability practitioners and furthermore describe ideas for future research. Please note the limitations of our study we outline in section 7 when interpreting the data, most notable the limited sample size.

Extraversion

We identified two personality dimensions as having the most influences on usability metrics according to our analysis, one of them being the factor extraversion. Rather extraverted test persons took more time to complete the test and also had more problems with completing the tasks while introverted persons were more successful. Although, these findings point to extraverted persons having much more problems, there were no significant findings concerning the rating or the statements in the open-ended questions. Concerning the regression models, extraversion is also the strongest predictor for the time needed as well as for the task completion rate. Those findings validate the fact that previous research focused on extraversion as especially important factor in the context of usability tests. The absent influence on the SUS ratings is in line with research by [2, 5, 11, 28]. In contrast with [2] extraverted persons were not more successful in completing the tasks. We were not able to validate the findings of [11], [5] and [2] that extraverted persons report more usability problems and generally share more information via

our open-ended questions. Note that those studies employ "thinking aloud" while we used open-ended questions, so the reporting behavior of extraverted persons might be influenced by the chosen method. More research is necessary to examine the sophisticated influence of extraversion on usability testing in more detail.

The Need for Influence and Power

We also identified another factor that has strong correlations on multiple usability metrics, which is not part of the traditional five-factor model but a specific dimension employed by the questionnaire we used: the need for influence and power [43]. The higher this factor the lower the task completion rate and the higher the time needed to complete all tasks. Furthermore, persons with higher manifestations on this scale mentioned fewer positive aspects in the open-ended questions and rate the website lower on the SUS scale while persons with low levels of this need behaved the opposite way. One explanation for this behavior might be that persons with a higher need for influence and power want to control their surroundings and, when failing during the test, they feel not in control anymore and therefore tend to be more critical concerning their ratings. In regards to the regression models, this factor is part of all of the best predicting regression models, in the case of the positive mentions also as strongest predictor. Therefore, we propose this factor next to extraversion as a very important influencing factor in usability testing. Since both factors show strong influences but are part of differing models we also examined if these factors correlate with each other and we actually measure the same concept. However, there are no significant correlations between those two factors.

Our findings concerning the three basic needs also show that the extended analysis of personality traits beyond the well-known and established models (like the five-factor model) can be very helpful in explaining the variance in participants' behavior. [5] proved the success of this extension by regarding factors like curiosity and intelligence. We recommend exploring various factors in future research to get a more comprehensive view of the possible influencing factors.

Agreeableness

There are several correlations between personality factors and usability metrics, which, contrary to our assumption did not show strong significant effects. We assumed that the factors extraversion and agreeableness might have an influence on the quantity of input produced in the open-ended questions since persons with high values on this dimensions are supposed to be talkative, friendly and helpful. However, we only found a weak positive correlation between agreeableness and mentions of positive aspects in the open-ended questions. Regarding the regression model for this usability

metric, the need for influence and power has much stronger influence than agreeableness. There is not a lot of research about the factor agreeableness in usability testing. We recommend analyzing this factor with methods like "thinking aloud" or interviews to identify more differences. Our assumption as well as research by [28] that more agreeable persons tend to rate the SUS better (since those persons are more friendly) cannot be validated by our data.

Conscientiousness

The trait conscientiousness did not show any significant correlations. This finding is rather surprising to us, since this factor is regarded as one of the most influencing ones in other important areas [6] . Since test persons with higher conscientiousness are more dutiful and accurate, we assumed they might have a higher task completion rate and are therefore less critical concerning the tool. This was not the case and is in line with research on a larger sample size concerning SUS scores by [28]. We do not want to speculate about the reasons for this absence of influence and recommend future research with other usability metrics and methods but as it stands for now, this factor seems to be not important for interpreting participants' behavior during usability tests.

Personality and the System Usability Scale

Considering the results about the System Usability Scale in general, our findings suggest that personality for the most part, does not have an important influence. We identified several rather weak correlations but our regression model consisting of the factors agreeableness and the need for influence and power does not significantly predict the SUS values. This is in line with research by [2, 11] for the factor extraversion. We validated those findings concerning most of the other personality dimensions as well. For a larger sample, [28] were able to make significant findings for openness and agreeableness. However they did not employ a traditional usability tests but rather let the participants rate well-known products by their memory, which is a very different approach. Overall, our data suggests that the SUS is a rather robust instrument not influenced by personality and therefore very recommended for usage. For further research, we propose to also analyze questionnaires that are not solely usability focused like e.g. the *User Experience Questionnaire* [30] since factors describing user experience are often rather subjective and more dependent of the opinion of the user (e.g. aesthetics; [13]). For example, highly agreeable users might tend to give out better ratings for subjective metrics.

Regression models

The majority of research in this area only regarded one or very few personality dimension like extraversion. The integration of multiple dimensions and more sophisticated

personality manifestations was not only helpful since many of those showed significant correlations when analyzed in isolation. But also since we were able to analyze prediction models consisting of multiple dimensions via linear regression analysis and compare the influence of those dimensions to each other. We were indeed capable to define significant regression models that explain the variance of the variables: task completion rate, time needed and the positive mentions, thus confirming that in our study personality does indeed predict partially usability metrics. Furthermore, we found that the combination of multiple dimensions does predict usability metrics better than isolated dimensions which is to be expected since humans do not only consist of one personality trait but multiple ones with different manifestations. This interaction and combination of multiple personality dimensions is often ignored in research. We recommend future work to integrate regression models with multiple dimensions and not focus on the extremes of singular traits.

The best model was found concerning the task completion rate and consisted of neuroticism, extraversion, openness and the need for influence and power explaining 37% of the variance of this metric. This is a rather fascinating insight for us since we assumed that the task completion rate is a rather objective metric highly dependent of the task and the evaluated tool. On the other hand, for the more subjective qualitative metrics like the number of words no significant models could be found except for the positive mentions. Overall, our findings suggest that quantitative as well as qualitative metrics can be predicted by personality models.

Implications for Research and Usability Engineering

Our results have implications for research employing methods like usability tests and usability practitioners as well.

On the one hand, since personality does show correlations with usability metrics in our study and in other research, we recommend usability practitioners and researchers to keep the possibility of this influence in mind and pay attention to possible variances. Just like researchers currently examine the results by persons with different ages or cultural backgrounds more in depth, researchers also might regard important influencing personality factors like extraversion and the need for influence and power.

While the acquisition of demographic variables is rather normal in usability tests, the acquisition of personality is not. Certainly, it is not possible to integrate large questionnaires with over 70 questions like the questionnaire used in our study. Nevertheless, for most personality models, very small questionnaires consisting only of up to 20 questions exist e.g. the *Mini-IPIP* [18] for the five-factor model. When integrating those small questionnaires usability professionals and researchers can also check among extreme manifestations of

certain personality dimension and analyze if they did skew data which improves the validity of the interpretation.

Another important implication relies in the process of participant selection. For many usability practitioners, it is important to acquire talkative test persons that identify many usability problems and give constructive critique. When targeting this specific group, one can make the argument that usability practitioners need fewer participants since this specific group is more effective in identifying and communicating problems which is of course an important advantage in the context of time and cost resources. We propose that based on our study and on the results of other research the identification of test persons can be facilitated by using personality models. Although we did not employ "thinking aloud" and also did not directly measure the identified usability problems we can derive fitting personality dimensions based on our usability metrics. We heuristically assume that persons with lower task completion rates, longer time needed and lower SUS ratings do indeed encounter more problems and are more critical participants. Based on our results, usability professionals should especially target test persons with high levels of extraversion, openness, the need for influence and power and low levels concerning neuroticism. The acquisition of these test persons can be done via purposive and snowball sampling techniques [2] . Via a short online personality questionnaire, usability professionals and researchers might only invite test persons that fit the required personality manifestations. Although, especially in the context of qualitative data, it would be important to also acquire test persons that are talkative and formulate precise information about usability problems, we were not able to identify personality traits that strongly influence the input produced concerning open-ended questions. However other research suggest that certain personality manifestations like high levels of extraversion [2, 11] do influence qualitative data produced with the thinking aloud method.

Overall, we want to highlight that the current research suggests that different personality dimensions influence different types of usability engineering methods. We recommend to also keep the chosen method in mind when employing purposive sampling based on the personality.

7 LIMITATIONS

While our sample size is larger than in the majority of similar studies [2, 11], [28] showed that one can make more significant and valid findings with larger sample sizes. Note that [28] were able to acquire a larger sample size by employing a test design consisting only of online questionnaires. Furthermore, our sample is very homogeneous in the way that the majority of participants were university students. It is rather challenging to gather those number of heterogeneous test persons with lab based usability testing and no monetary incentives. However, we propose to explore remote usability tests to gather larger and more heterogeneous samples.

As mentioned before, we did not employ "thinking aloud" as technique in our usability test to not skew the performance metrics. However, as previous research has shown, the behavior in "thinking aloud"-settings can be influenced by the personality, so we want to integrate more qualitative methods in our future research. We did use open-ended questions but were not able to find strong effects. Furthermore, we do not want to limit our research on thinking aloud but assume that methods like interviews or focus groups are also prone to be influenced by the participants' personality. We also propose to improve the analysis of open-ended questions or the output in thinking aloud settings by not only coding usability problems and positive or negative mentions but also more advanced factors like the mentions of possible solutions and improvements. By this, research can further identify the most advantageous and constructive participants for usability tests. In the context of more quantitative metrics, we did also not count several other quantitative metrics like the error rate or the click behavior. Paying attention to those metrics might uncover other important correlations.

Furthermore, the usage of a questionnaire to operationalize personality is always dependent of the subjectivity and honesty of the participants. The questionnaire we used also tests the honesty of participants via specific questions and we did not find dishonest persons based on these questions. However, this is still not a guarantee for honesty and furthermore, the self-image of persons can often differ strongly from the reality.

8 CONCLUSION

We addressed several limitations and aspects we want to include in future work in section 6 and 7. We are currently planning similar studies like this one but with varying usability engineering methods beyond the simple usability test, e.g. interviews and focus groups and also with other individual factors and personality traits to get a more holistic view on the influence of personality in usability engineering. Note that we only regarded personality as influencing factor. However, other research showed that there are other influencing factors concerning test participants like age [48], cultural background [14] or intelligence and curiosity [5]. We want to analyze the interaction of those factors, personality traits and other variables to get a more detailed understanding on the influence of individual factors in usability engineering so researchers and practitioners can improve the test participant selection process and the interpretation of variance in result data. We contributed to this research goal with the presented study for the factor personality and plan to continue our research towards this and other factors.

REFERENCES

[1] William Albert and Thomas Tullis. 2013. *Measuring the user experience: collecting, analyzing, and presenting usability metrics.* Newnes.

[2] Ali Alnashri, Obead Alhadreti, and Pam J Mayhew. 2016. The Influence of Participant Personality in Usability Tests. *International Journal of Human Computer Interaction (IJHCI)* 7, 1 (2016), 1.

[3] Jens B Asendorpf and Franz J Neyer. 2012. *Psychologie der Persönlichkeit.* Springer-Verlag.

[4] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[5] Wolmet Barendregt, Mathilde M Bekker, Don G Bouwhuis, and Esther Baauw. 2007. Predicting effectiveness of children participants in user testing based on personality characteristics. *Behaviour & Information Technology* 26, 2 (2007), 133–147.

[6] Murray R Barrick and Michael K Mount. 1991. The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology* 44, 1 (1991), 1–26.

[7] Mark Batey, Adrian Furnham, and Xeniya Safiullina. 2010. Intelligence, general knowledge and personality as predictors of creativity. *Learning and individual differences* 20, 5 (2010), 532–535.

[8] Gregory J Boyle. 1995. Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations. *Australian Psychologist* 30, 1 (1995), 71–74.

[9] Katharine C Briggs. 1976. *Myers-Briggs type indicator.* Consulting Psychologists Press Palo Alto, CA.

[10] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[11] GE Burnett and D Ditsikas. 2006. Personality as a criterion for selecting usability testing participants. In *Proc. int. conf. on information and communications technologies.* Citeseer, 599–604.

[12] Raymond B Cattell. 1965. *The scientific analysis of personality.* Penguin Books, Oxford, England.

[13] Marie Christine Roy, Olivier Dewit, and Benoit A Aubert. 2001. The impact of interface usability on trust in web retailers. *Internet research* 11, 5 (2001), 388–398.

[14] Torkil Clemmensen, Morten Hertzum, Kasper Hornbæk, Qingzin Shi, and Pradeep Yammiyavar. 2008. Cultural cognition in the thinking-aloud method for usability evaluation. *Icis 2008 Proceedings* (2008), 189.

[15] Jacob Cohen. 1992. A power primer. *Psychological bulletin* 112, 1 (1992), 155–159.

[16] Paul T Costa and Robert R McCrae. 1992. Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment* 4, 1 (1992), 5.

[17] Christine P Dancey and John Reidy. 2007. *Statistics without maths for psychology.* Pearson Education.

[18] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment* 18, 2 (2006), 192.

[19] Sanjay Kumar Dubey and Ajay Rana. 2010. Analytical roadmap to usability definitions and decompositions. *International Journal of Engineering Science and Technology* 2, 9 (2010), 4723–4729.

[20] Carlos Flavián, Miguel Guinalíu, and Raquel Gurrea. 2006. The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & management* 43, 1 (2006), 1–14.

[21] Adrian Furnham. 1996. The big five versus the big four: the relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences* 21, 2 (1996), 303–307.

[22] Adrian Furnham and John Walker. 2001. The influence of personality traits, previous experience of art, and demographic variables on artistic preference. *Personality and Individual Differences* 31, 6 (2001), 997–1017.

[23] Veronica Gomez, Franciska Krings, Adrian Bangerter, and Alexander Grob. 2009. The influence of personality and life events on subjective well-being from a life span perspective. *Journal of Research in Personality* 43, 3 (2009), 345–354.

[24] Gregory M Hurtz and John J Donovan. 2000. Personality and job performance: The Big Five revisited. *Journal of applied psychology* 85, 6 (2000), 869.

[25] W Iso. 1998. 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs). *The international organization for standardization* 45, 9 (1998).

[26] Oliver P John, Laura P Naumann, and Christopher J Soto. 2008. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research* 3, 2 (2008), 114–158.

[27] Udo Konradt, Hartmut Wandke, Björn Balazs, and Timo Christophersen. 2003. Usability in online shops: scale construction, validation and the influence on the buyers' intention and decision. *Behaviour & Information Technology* 22, 3 (2003), 165–174.

[28] Philip Kortum and Frederick L Oswald. 2018. The impact of personality on the subjective assessment of usability. *International Journal of Human–Computer Interaction* 34, 2 (2018), 177–186.

[29] Melinda L Korzaan and Katherine T Boswell. 2008. The influence of personality traits and information privacy concerns on behavioral intentions. *Journal of Computer Information Systems* 48, 4 (2008), 15–24.

[30] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group.* Springer, 63–76.

[31] Maurice Lorr. 1991. An empirical evaluation of the MBTI typology. *personality and Individual Differences* 12, 11 (1991), 1141–1145.

[32] Kelly Moore and James C McElroy. 2012. The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior* 28, 1 (2012), 267–274.

[33] Henry Alexander Murray. 1938. Explorations in personality: A clinical and experimental study of fifty men of college age. (1938).

[34] Jakob Nielsen. 1994. Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. *Cost-justifying usability* (1994), 245–272.

[35] Jakob Nielsen. 1995. 10 usability heuristics for user interface design. *Nielsen Norman Group* (1995). https://www.nngroup.com/articles/ten-usability-heuristics/

[36] Warren T Norman. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology* 66, 6 (1963), 574.

[37] Irina Sumajin Parkins, Harold D Fishbein, and P Neal Ritchey. 2006. The influence of personality on workplace bullying and discrimination. *Journal of Applied Social Psychology* 36, 10 (2006), 2554–2577.

[38] David J Pittenger. 2005. Cautionary comments regarding the Myers-Briggs Type Indicator. *Consulting Psychology Journal: Practice and Research* 57, 3 (2005), 210.

[39] Jochen Prümper, Michael Frese, Dieter Zapf, and Felix C Brodbeck. 1991. Errors in computerized office work: differences between novice and expert users. *ACM SIGCHI Bulletin* 23, 2 (1991), 63–66.

[40] Briony D Pulford and Harjit Sohal. 2006. The influence of personality on HE students' confidence in their academic abilities. *Personality and Individual Differences* 41, 8 (2006), 1409–1419.

[41] M Rauer. 2011. Quantitative Usability-Analysen mit der System Usability Scale (SUS). (2011). https://blog.seibert-media.net/blog/2011/

04/11/usablility-analysen-system-usability-scale-sus/

[42] Heiner Rindermann and Aljoscha C Neubauer. 2001. The influence of personality on three aspects of cognitive performance: Processing speed, intelligence and school performance. *Personality and individual differences* 30, 5 (2001), 829–842.

[43] L Satow. 2012. Big-Five-Persönlichkeitstest (B5T): Testmanual und Normen. *Available online: URL: http://www. drsatow. de* (2012).

[44] Jürgen Sauer, Katrin Seibel, and Bruno Rüttinger. 2010. The influence of user expertise and prototype fidelity in usability tests. *Applied ergonomics* 41, 1 (2010), 130–140.

[45] Ulrich Schimmack, Jürgen Schupp, and Gert G Wagner. 2008. The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social indicators research* 89, 1 (2008), 41–60.

[46] Thomas Schmidt and Christian Wolff. 2016. Personality and information behavior in web search. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology.* American Society for Information Science, 121.

[47] Andreas Sonderegger and Juergen Sauer. 2009. The influence of laboratory set-up in usability tests: effects on user performance, subjective ratings and physiological measures. *Ergonomics* 52, 11 (2009), 1350–1361.

[48] Andreas Sonderegger, Sven Schmutz, and Juergen Sauer. 2016. The influence of age in usability testing. *Applied Ergonomics* 52 (2016), 291–300.

[49] William Stern. 1911. *Die differentielle Psychologie in ihren methodischen Grundlagen [Differential psychology in its methodological foundations].* Barth, Leipzig, Germany.

[50] Gunnvald B Svendsen, Jan-Are K Johnsen, Live Almås-Sørensen, and Joar Vittersø. 2013. Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology* 32, 4 (2013), 323–334.

[51] Ravi Vatrapu and Manuel A Pérez-Quiñones. 2006. Culture and usability evaluation: The effects of culture in structured interviews. *Journal of usability studies* 1, 4 (2006), 156–170.