**Full citation**: Farhangian, M, Purvis, M.K., Purvis, M.A., and Savarimuthu, B.T.R. 2016.Personalities and software development team performance – a psycholinguistic study, in proceedings of the 24th European Conference on Information Systems (ECIS 2016) (Istanbul, Turkey, June 12-15, 2016).

# PERSONALITIES AND SOFTWARE DEVELOPMENT TEAM PERFORMANCE, A PSYCHOLINGUISTIC STUDY

*Completed Research*

Farhangian, Mehdi, University of Otago, Dunedin, New Zealand, mehdi.farhangian@otago.ac.nz
Purvis, Martin, University of Otago, Dunedin, New Zealand, martin.purvis@otago.ac.nz
Purvis, Maryam, University of Otago, Dunedin, New Zealand, maryam.purvis@otago.ac.nz
Savarimuthu, Bastin Tony Roy, University of Otago, Dunedin, New Zealand, tony.savarimuthu@otago.ac.nz

## Abstract

*Successful software development often depends on the ability of specifically skilled individuals to cooperate effectively in order to achieve the project goals. This cooperative ability goes beyond the individual technical skills of the team participants and concerns how easy it is for people to work together as a team. The primary goal of this paper is to present a decision-support model that can assist software team managers to form teams that have likely appropriate personality combinations of the team members. To achieve this goal, there needs to be an efficient and unobtrusive method of measuring employee personality profiles and a methodological approach to establish useful rules for team formation that uses these personality profiles. We present how we have achieved these goals in this paper, and then we demonstrate the effectiveness of our approach by deriving rules for team formation in connection with software development of the Python programming language.*

*Keywords: Software development projects, Personality, Team performance, Psycholinguistics*

# INTRODUCTION

How to compose effective teams for project work is a difficult but important problem for large organizations. This is particularly true in the area of software engineering, where groups of people must work on specific technically-specified modules and make them work together efficiently and effectively. Often teams are established purely on the basis of personal familiarity and experience, but this approach is less likely to work in today's more dynamic and distributed work environment. What is needed is a more systematic approach, and the present paper seeks to make progress in that area. Our goal in this paper is to provide insights into the effect of personality on overall team effectiveness in software development projects.

Previous work has investigated factors that may affect team performance, including age, gender, personal traits, knowledge, and skills, and among these personality traits have been singled out as particularly important (LaPine et al., 2011; Cruz et al., 2014; André et al., 2011; Salleh et al., 2012). However, with respect to personality, most of the studies in the software engineering domain have focussed on the relationship between individual personality and individuals' behaviour. Thus group factors such as cohesion, conflict, and coordination have not been fully covered.

There have been a few previous general studies with respect to the significance of personality, and the degree of personality diversity on team performance, but the results have been mixed and with conflicting conclusions (Bowers et al., 2000; Day and Bedeian, 1995; Aamodt and Kimbrough, 1982; Barr et al., 2011). One study did find that high aggregate values of conscientiousness and openness did contribute to the success of research teams (McGrath, 1986). Narrowing the focus to studies concerning personality and its diversity in the specific area of software engineering again reveals conflicting results (Andre et al., 2011; Bradley and Hebert, 1997; Miller and Yin, 2004; Peslak et al., 2006; Rutherfoord, 2001; Lewis and Smith, 2008; Karn and Cowling, 2006). Some researchers suggest that a homogeneous team results in more effective performance e.g.(Bowers, Pharmer, & Salas, 2000), other researchers suggest that a heterogeneous team results in more efficient performance e.g. (Aamodt & Kimbrough, 1982).

Some of these conflicting results may be attributable to difficulties in measuring the personalities of team members. So in connection with the overall goal of our work, there are actually two problems that must be addressed: (a) how to measure personality in a manner that is useful for our purposes and (b) how to measure performance of software teams. We don't pretend to offer general, global solutions to these problems, but we do believe that progress can be made within the scope of specific domains. Thus there are two research questions that we address in this paper:

RQ1: How can we measure personalities in a reliable and unobtrusive manner?

RQ2: What is the relationship between combined group personalities of software teams and team performance?

In the following, we describe how we have developed procedures to assist team composition that can be applied to a specific software development context, and we demonstrate our approach with respect to the Python Enhancement Proposal (PEP) process, which is used in the development and maintenance of the software programming language Python (Warsaw et al., 2000).

Our overall goal is to answer the second research question (RQ2) which is on estimating team performance based on combined team personalities. In order to answer the second question, we will first need to infer the individual personalities (RQ1). .

The rest of the paper is organised as follows. Section 2 reviews the personality measures for Team Performance. Section 3 explains the relationship between personality and linguistic style. The data extraction and methodology are described in section 4. In section 5, we present our findings about the relationship between personality and texts and also the effect of team personality on their performance. Section 6 contains conclusion.

# PERSONALITY MEASURES FOR TEAM PERFORMANCE

Measuring and assessing personalities is problematic for two primary reasons. For one thing people often resist personality measures and may fake answers on questionnaires (Martin et al., 2002). Moreover, there is no universally recognized measure of what constitutes personality. With respect to the first problem, a possible approach is to use indirect measures -- previous researchers have shown that people inadvertently reveal their personality traits by their textual communications (e.g. Pennebaker and King 1999). Since software teams use text-based communication tools (e.g. wikis, mailing lists, blogs, and instant messengers), the idea of extracting personality knowledge from these texts is a feasible strategy for software managers. However, only a few studies on collaborative software development teams have analysed personality from shared texts (Bazelli et al., 2013; Rigby and Hassan, 2007; Licorish and MacDonell, 2014a). These studies have mostly focussed on the personalities of top software developers and contributors and have not offered recommendations about team compositions and group factors. So our approach is to use textual analysis to the whole software engineering team.

With respect to the second above-mentioned problem, that of using an objective personality, there are two main proposed measurement schemes in this area.

• The Five Factor Model (FFM, also known as “Big Five”) (Costa and McCrae, 1992) is popular in the social sciences.

• The Myers-Briggs Type Indicator (MBTI) psychometric scheme (Myers, 1962) is based on the work of Carl Jung (Jung, 1921) and is widely used in the business world.

Although the FFM is popular in the social science community, there are three reasons why we have chosen to use the MBTI for our study.

1. The MBTI is the most prominent psychometric scheme that has been applied to software engineering studies (Wiesche and Kremar, 2014). One research article showed that MBTI was used in 48% of all software engineering studies, and the closely-related Keirsey Temperament Sorter (Keirsey, 1998) was used for an additional 9%, while the FFM model was used for 19% of such studies (Cruz et al., 2015). Thus MBTI and related metrics accounted for 57% of the studies compared to FFM’s 19%.

2. The FFM does not introduce a standardized cluster to compare groups of people, while the MBTI does.

3. The FFM’s “factors” all identify positive characteristics, which makes self-assessing participants less likely to score themselves low on any of these scales. The MBTI dimensions, on the other hand, are bipolar – each end of each scale, such as “thinking” and “feeling” can have a positive interpretation. This makes self-assessors more likely to provide accurate information about themselves with the MBTI scale than with the FFM.

The MBTI, which is based on a psychological type scheme originally developed by Carl Jung (Jung 1921) and then modified by Myers and Briggs (Myers et al., 1985), has four personality dimensions:

- Extraversion vs. Introversion– An extraverted person prefers to focus on the outer world, while an introverted person keeps more to him or herself.

- iNtuition vs. Sensing– An intuitive type is more abstract and understands according to his or her inner compass, while a sensor gathers information that is in concrete, objective form.

- Thinking vs. Feeling– A thinker makes decisions based on logic and demonstrable rationality, while a feeler is more empathetic and attempts to see things from given perspectives.

- Judging vs. Perceiving– A judger wants things settled and organized, and a perceiver is flexible and open to new developments.

Each of these four dimensions represents a continuum, with the labels identifying the extreme ends of each continuum. Thus one could be considered to be very Extraverted or, if he or she is a bit closer to the middle of the spectrum, perhaps only slightly more Extraverted than Introverted. In both cases, though, the person whose personality is closer to the Extraverted end of that particular dimension would be given the label "E" for that dimension. This means that an individual could be given four MBTI labels that represent one of 16 possible MBTI types. For instance an MBTI label of ISFJ would mean that the person's personality is closer to the Introversion (I), Sensing (S), Feeling (F), and Judgemental (J) ends of the respective dimensions.

For our work, we calibrate each dimension with number between 0 and 100. So, for example, for the E-I (Extraverted-Introverted) dimension, a value between 0 and 50 would mean that the person is extraverted, and a value between 50 and 100 would mean that that person is introverted. This scheme is summarized as follows:

- E-I (Extraversion/Introversion): (range 0-50 → *Extraverted*; 50-100 →*Introverted*)
- N-S (Intuitive/Sensing): (range 0-50 → *Intuitive;* 50-100 → S*ensor*)
- T-F (Thinking/Feeling): (range 0-50 → *Thinker;* 50-100 → *Feeler*)
- J-P (Judging/ Perceiving): (range 0-50 → *Judger;* 50-100 → *Perceiver*)

Since we are aggregating these values in connection with teamwork, we introduce to additional indicators that are used in connection with personality profiles (Neuman et al., 1999):

- Team Personality Elevation (TPE): a team's mean level for a particular personality trait.
- Team Personality Diversity (TPD): the variance with respect to a particular personality trait among team members.

# PERSONALITY AND TEXT ANALYSIS

People express themselves in their own unique styles, both verbally and textually. Thus forms of "linguistic fingerprinting" have been employed to identify styles of political leaders (Hart, 1984), to determine the anonymous authors of books (Foster, 1984), and to distinguish the behaviour of software developers (Rigby and Hassan, 2007; Licorish and MacDonell, 2014b). To facilitate and help automate this operation, Pennebaker developed the Linguistic Inquiry and Word Count (LIWC) textual analysis tool (Pennebaker, 2001).

The LIWC tool counts over 4500 functional words and word particles, and with a psychometrically-based dictionary classifies them into 80 meaningful categories or dimensions. These categories encompass linguistic processes (e.g. personal pronouns, adverbs, prepositions, etc.), psychological processes (e.g. social processes, positive and negative emotions, etc.), personal concerns (e.ge. work, achievement, leisure, etc.), and spoken categories (e.g. assents, nonfluencies, fillers, etc.).

Text-mining tools such as this can be used not only to uniquely identify individuals, but also to match individuals with certain behaviour patterns. Indeed, Pennebaker and King (1999) attempted to match some people who had written essays with their FFM profiles. Their results were suggestive, but the correlations were not strong. We believe that we can make further progress in this area. For one thing, we think that written assignments and essays, which are the main textual sources for Pennebaker and King (1999) may not always reflect the characteristic natural feelings, morals, and values of the author. If the subject matter of the written assignment is shifted to other topics, then other textual expression patterns may emerge. In other words, behaviour is discriminative and not always consistent across the range of situations – the assignments and scientific articles that Pennebaker and King used may not cover a sufficient variety of behavioural dimensions. In addition, formal writing may be more a reflection of verbal ability than of personality.

A possibly more fertile arena for the authentic expression of personality is that of social networking websites. On those sites people are more likely to feel free from arbitrary constraints, since they voluntarily choose and contribute to the discussion topics and are not trying to conform to the constraints of an “assignment”. The users on those sites establish the “rules of the game” in terms of expression content, which can be in the form of posts, comments, and blogs.

So in this article we use the LIWC tool to operate on text from social networking Websites two reveal the relationships between these texts and the self-identified personalities (MBTI) of their authors.

In order to address the limitations of the previous studies in this domain we develop a model to relate linguistic styles and personality. These limitations are summarized as follows:

1. There is no study to focus on the relationship between LIWC and MBTI. However, Lee et al. (Lee, Kim, Seo, & Chung, 2007)introduce correlations between the Korean version of Linguistic Inquiry and Word Count KLIWC and Myers-Briggs types.

2. Previous studies (e.g. (Pennebaker & King, 1999)) collected texts under laboratory settings. People may not express their real feelings, morals and values when they did not choose a topic to write about.

3. In the previous studies the size of samples is not considerable and each writing sample is less than a few thousand words. Moreover, these data are gathered from a small number of participants that limit the results.

# METHOD AND DATA EXTRACTION

Our method for extracting insights about the relationship between Python developers and their personality consisted of 3 main steps:

1. Gathering the data

2. Finding relationships between personality and text usage

3. Finding relationship between personality of teams’ compositions and their effectiveness.

The first step involved extracting data from social networking website where people report their MBTI personality profiles. Note that in most social networking websites, people are often careless about their spelling and grammar, which can lead to ambiguities in interpretation in the writing and it leads to various types of ambiguities. To help avoid this issue, we used a popular social networking website, Quora (http://www.quora.com), in which people cannot be anonymous or use fake names. We believe this tended to increase accuracy in terms of spelling and properly grammatical sentence construction. In addition we explored two other popular social networking websites to validate our findings – Reddit (http://www.reddit.com) and College Confidential(http://www.collegeconfidential.com).

On the Quora website discussing and reporting MBTI personality profiles is a popular activity. We identified users who reported their MBTI personalities, and we extracted their posted textual information. In addition we posted several questions and asked users to report their personalities to enrich our data. We took similar steps with the two other sites, College Confidential and Reddit.

In the second step, we extracted the text of the Quora users from their responses to other questions. Users' texts were analysed with the LIWC tool and the value for all the 80 LIWC dimensions were identified for each user. After generating the value of all the variables in our Quora samples, we measured correlations between personality types and these variables. These correlations were then used to develop a computational model that can be used to determine the personality of people from their texts. To validate this computational model, our formula was cross-tested with our Reddit and College Confidential data.

In the third step, we conducted an analysis of teams that developed Python Enhancement Proposals (PEP). PEPs are documents that contain information about the important decisions taken by the Python development team. The performance of these teams (i.e. success and failure of PEP projects) is available to the public. Our proposed computational model was then employed to determine the personality of each team member involved in the PEP teams based on written text of these members made available in the social media (mostly their own blogs). After determining the personality of each PEP group using the process discussed in the previous section, we analysed the collected dataset in order to generate rules showing the relationship between team personality and their performances.

# 5 RESULTS AND DISCUSSIONS

For the Quora platform, we identified 393 users who reported their MBTI personality profiles. Since the distribution of personalities among users differed from some previous reports of personality distributions (e.g. "The Myers & Briggs Foundation", 2015), we compared our derived Quora personality distribution with those of the two other social networking Websites (Reddit and College Confidential). This comparison allowed us to assess whether the Quora data was typical and similar to the trends at other social networking Websites. Thus we gathered MBTI reports from 185 users of College Confidential ("Collegeconfidential", 2015) and from 39 uses of Reddit ("Reddit", 2015). These data and the distribution of personality among the users of these websites are shown in Table 1.

| Personality | Quora | | College Confidential | | Reddit | |
|---|---|---|---|---|---|---|
| | Number | Percentage | Number | Percentage | Number | Percentage |
| ESTJ | 8 | 2.03% | 10 | 5.40% | 1 | 2.60% |
| ISTJ | 12 | 3.05% | 11 | 5.95% | 2 | 5.13% |
| ISFJ | 4 | 1.02% | 6 | 3.24% | 1 | 2.56% |
| ESFJ | 5 | 1.30% | 5 | 2.70% | 0 | 0% |
| ESFP | 9 | 2.30% | 6 | 3.24% | 0 | 0% |
| ISFP | 5 | 1.30% | 1 | 0.54% | 0 | 0% |
| INFP | 52 | 13.30 | 5 | 2.70% | 3 | 7.69% |
| INFJ | 44 | 11.20% | 12 | 6.49% | 2 | 5.13% |
| ENFJ | 18 | 4.60% | 12 | 6.49% | 5 | 12.82% |
| ENTJ | 31 | 7.90% | 12 | 6.49% | 4 | 10.26% |
| ISTP | 8 | 2.03% | 4 | 2.16% | 2 | 5.13% |
| INTP | 60 | 15.30% | 27 | 14.59% | 7 | 17.94% |
| INTJ | 59 | 15.01% | 43 | 23.24% | 5 | 12.82% |
| ESTP | 6 | 1.53% | 4 | 2.16% | 1 | 2.56% |
| ENFP | 30 | 7.63% | 14 | 7.57& | 1 | 2.56% |
| ENTP | 42 | 10.69% | 13 | 7.03% | 5 | 12.82% |
| Total | 393 | 100% | 185 | 100% | 39 | 100% |

*Table 1. Correlations between MBTI personality and LIWC*

Table 2 illustrates the similarity across the three social networking Websites. For all the Websites, the introverted types were slightly more in abundance than extraverts. Intuitive types were far more than prevalent than sensing types. Thinkers were somewhat more prominent than feelers, while judgers and perceiver's were roughly equal.

| | E-I | N-S | T-F | J-P |
|---|---|---|---|---|
| College Confidential | 59% | 25% | 33% | 40% |
| Reddit | 56% | 18% | 31% | 49% |
| Quora | 62% | 15% | 42% | 54% |

*Table 2. Comparing the distribution of personality in three social networking websites*

These results differ from some results posted about the general population. For example, "The Myers & Briggs Foundation" (2015) report that, along the intuition-sensing (N-S) scale, sensors (S) make up 73% of the population; whereas for the social networking Websites that we examined, intuition (N) was dominant and made up 75%, 82%, and 85% of the users on College Confidential, Reddit, and Quora, respectively.

With these results, we went on to address the second of our two research questions (RQ1), how to measure the relationship between writing style and personality.

## 5.1 Measuring the relationship between writing style and personality

After identifying Quora users who reported their personalities, we extracted the text of these users from their responses to other questions. We were only interested in the English texts written by users, so their non-English characters and quotes were removed. Additionally, we removed those users who did not contribute to enough answers to provide enough text material. That left a total of 228 users who were included in the final analysis, and their texts were analysed with the LIWC tool.

With LIWC each word in the input file is associated with a word category, and the output data provides the percentage of words for that category. After generating the values of all the LIWC dimensions in our Quora samples, we used Pearson correlation to find the relationship between personality and these dimensions. These correlations can be used to develop a computational model that characterizes a person's personality from the texts that he or she writes.

After analysing the text of the users by the LIWC tool, we considered the LIWC dimensions for which their correlations were significant at the 0.05 level, and these variables with their correlations are presented in the Table 3.

| **Personality** | **Correlation** | | |
|---|---|---|---|
| Introversion | words > 6 letters (-0.167)<br>dictionary words (0,153)<br>unique words (- 0.238)<br><br>negate (0.241)<br>number (0.147)<br>negative emotion (0.204)<br>anxiety (0.147)<br>sad (0.177) | cognitive processes (0.194)<br>cause (0.189)<br>insight (0.138)<br>discrepancy (0.161)<br>tentative (0.2)<br>hear (0152)<br>social (0.132)<br>common verbs (0.135) | humans (0.133)<br>present tense (0.178)<br>inclusive (-0.166)<br>occupation (-0.2)<br>school (-0.151)<br>job (-0.219)<br>music (- 0.168)<br>body (0.146) |
| Sensing | we (-0.158)<br>optimism (0.139) | cause (-0182)<br>occupation (0.145) | sports (0.141)<br>colon (0.173) |
| Feeling | words > 6 letters (- 0.191)<br>unique (0.165)<br>I (0.147)<br>self (0.144)<br>effect (0.150)<br>positive emotion (0.185)<br>positive feeling (0.244) | certain (0.166)<br>time (0.201)<br>job (- 0.175)<br>money (- 0.181)<br>physical states and functions (0.244) | feel (0.136)<br>body (0.192)<br>sexual (0.215)<br>sleep (- 0.246)<br>exclamation points (0.181)<br>other P (-0.132) |
| Perceiving | assent (-0.134) | common verb (-0.180) | family (-0.145)<br>humans (-0.15) |

*Table 3. Correlations between MBTI personality and LIWC*

As the correlations indicate in the table, people of each personality type tend to express certain kinds of words. For each personality, we develop a formula by adding or subtracting the independent, correlated LIWC dimensions. The formula expressed as follows:

$$.\mathrm{Rp} = \mathrm{B} \sum_{\mathrm{i}=1}^{80} \mathrm{LIWC_i} * \mathrm{Correlation_{LIWC_i}} \quad (1)$$

where $Rp$ indicates the relative personality, and indicates whether the correlation between the LIWC dimension and the personality is significant or not. If this correlation is significant $B = 1$, otherwise $B = 0$. For example, Perceiving is equal to - *assent**0.134 – *common_verb**0.180 - *family**0.145 -

$humans$*0.15. If after analysing one of the users' texts, it turns out that $assent$, $common\_verb$, $family$, and $humans$ are equal to 0.16, 1.2, 0.07 and 0.52 respectively, then $Rp = -0.16 * 0.134 - 1.2 * 0.180 - 0.07 * 0.145 - 0.52 * 0.15 = -0.32559$. After calculating the relative values of the personality of each user in the four dimensions (Introversion, Sensing, Feeling, Perceiving), we scaled these relative values to an absolute value that is between 0 and 100.

To validate our formula for measuring personality, our formula was cross-tested with Reddit and College Confidential data. We gathered public texts from 35 Reddit users and 135 College Confidential users who reported their MBTI personalities. As a result, we used the LIWC tool to generate all the 80 dimensions from the texts of each user. Then, by using the proposed formula, we predicted the personality of each user in each dimension. After predicting the personality of the users, we labelled each dimension with Y or N depending on whether it matched with the real (self-identified) personality or not. Dividing number of Y's by the total number of participants yields the accuracy of the proposed model, which was 65% for Reddit and 73% for College Confidential as presented in Table 4.

| | **College Confidential** | **Reddit** |
|---|---|---|
| Accuracy | 73% | 65% |

*Table 4: Validation of the proposed formula in Reddit and College Confidential*

Thus with our capability of predicting personality form text usage in hand, we proceeded to address our second research question on determining the relationship between combined group personalities of software teams and team performance.

## 5.2 Measuring the relationship between group personalities of software teams and performance

Using the above-discussed methods, we proceeded to investigate the effect of teams' personalities in connection with Python software development teams and their effectiveness in achieving their goals.

Python is largely developed through the Python Enhancement Proposal (PEP) process (Warsaw et al., 2000). A PEP is a design document that describes a new feature for Python or its processes or environment. The developers use mailing lists as the primary forum for discussing PEPs and the language's development. Out of 363 PEPS that we examined, 83 of them were developed by more than one person. We removed the individual works in order to study teams' behaviour.

We gathered texts of the member of these 83 teams from their public activities on the Internet (mostly their blogs). We could have simply analysed the "commits" (saving the results to a common database) of these users, but we believe that commits do not provide a rich source of texts and mainly contain only short and technical messages that do not reveal the personality of the developers. Consequently we extracted other online texts, such as from blogs and other public activities, to achieve more reliable results. We only analysed teams that produced enough text from all the team members, and this accounted for 78 of the teams. Then, we calculated the relative personality of each member based on our proposed method that is presented in Formula 1. These relative values were converted to an absolute value that is a number between 0 and 100.

Based on these values that represent the personality of users in four dimensions, Extraversion-Introversion (E_I), iNtuition-Sensing (N_S), Thinking-Feeling (T_F), Judging- Perceiving (J_P), we calculated 8 new variables: the TPE's of the four dimensions (TPE_EI, TPE_NS, TPE_TF, and

TPE_JP) and the TPD's of the four dimensions (TPD_EI, TPD_NS, TPD_TF, and TPD_JP. Recall that the TPE in each personality dimension is team's average personality values for that dimension, and the TPD in each personality dimension is the team's variance in that personality dimension. The values less than the median of each category labelled as "Low" and others labelled as "High".

For the work outcome, we referred to the status attribute in the PEP document that represents the state of the proposal, and it is labelled with one of these categories: draft, accepted, rejected, withdrawn, active, deferred, replaced, and final. Some PEPs are never finalized, so we labelled them as failed projects, if they are rejected or withdrawn or deferred or replaced. Otherwise the projects are labelled as successful projects.

We applied Apriori algorithm (Agrawal & Srikant, 1994) for association rules mining to analyse the collected dataset with an aim to generate the rules to show the relationship between personality of the teams and their performances. The analysis was conducted with the support threshold setting at 0.1 and 0.7 as the confidence level. This means each candidate with support greater than 0.1 and with a confidence level over 0.7 is considered a candidate with strong association rules. Measuring lift of these rules determines the ability of the rules for predicting the future behaviour. The results which are presented below have a lift higher than 1. A lift ratio larger than 1 implies that the relationship between the antecedent and the consequent is more. The larger the lift ratio, the more significant the association.

Table 5 shows a group of useful rules to managers so they can make decisions about which compositions would affect the team performance. As an example, the first recommended rule is as follows:

TPE_NS=LOW & TPE_JP=LOW ➔Success

It indicates that teams with high intuition personality as well as high judging personality are successful. The support, confidence and lift of these rules are 0.173, 0.87, and 1.809, respectively. In general, our findings are summarized as follows:

Some rules such as 4, 5, and 6 suggest that we need a heterogeneous team in intuitive-sensing dimension to improve the efficiency of teams. Heterogeneity in intuition-sensing personality results in successful group performance, since sensing types bring facts and details and Intuitive types provide new possibilities and ideas. This fining is confirmed by Kyungsub S. Choi et al.( 2008). They found diverse sensing and intuition preferences would challenge each other and offer wider array and solutions. Also, they considered different make up of pair programmers. In their studies the most successful teams were diverse teams who are not totally opposite (e.g. TN-FS) or alike (e.g. TS-TS). One of the successful pairs were ST-NT and based on that they concluded that similarities in thinking-feeling dimension provide common ground for reconciling differences and diversity in intuition-sensing dimension helps them to generate new ideas. In their study, in comparison to opposite teams and alike teams, opposite teams were more successful.

Some rules such as 5 add another suggestion to the previous rules and suggest in conjunction with heterogeneity in terms of intuition-sensing, having generally high intuition personality in team is beneficial. In this connection, Cheng et al showed that diverse pairs in intuition and sensing performed significantly better than homogeneous sensing type pairs, but not better than intuition type pairs (Cheng, Luckett, & Schulz, 2003). Intuition in MBTI is related to openness to experience (McCrae & Costa, 1989) and several studies confirmed the positive role of openness in team (e.g. (Neuman et al., 1999) and B. H. Bradley, Klotz, Postlethwaite, & Brown, 2013)).

Some other rules such as 1 suggest low perceiving leads to better team performance, in the other words, having judging personality has a positive influence on the team effectiveness. This finding is consistent with another study about the relationship between MBTI personality and software development teams that suggests judging personality is involved in dealing with external world and meeting deadlines (Gorla & Lam, 2004). In FFM, judging is correlated to conscientiousness (McCrae & Costa, 1989), however, this correlation is not that strong. The positive relationship between conscientiousness and team performance is shown by several researchers (e.g. (English, Griffith, & Steelman, 2004).

Some other rules such as 6 suggest having a heterogeneous team in terms of perceiving and judging personality improves the efficiency of teams, since having perceivers in teams help them to consider alternatives in decision making and judging people help the team to be on schedule. These findings are confirmed by (Bradley & Hebert, 1997) where their experiments showed a successful team had a better balance of judging type and perceiving type (70% J, 30% P) than less successful team with (100% J). However, if we consider Judging in MBTI similar to contentiousness in FFM the results are conflicting, while some studies (e.g. Humphrey & Hollenbeck, 2007) show homogeneity in terms of conscientiousness negatively affects the performance some others showed heterogeneity on conscientiousness was not significantly related to oral performance e.g. (Mohammed & Angell, 2003) and (Prewett & Walvoord, 2009).

| Rules' Number | Antecedent | Consequent | Support | Confidence | Lift |
|---|---|---|---|---|---|
| 1 | TPE_NS=LOW & TPE_JP=LOW | Success | 0.173 | 0.87 | 1.809 |
| 2 | TPE_TF=HIGH &TPE_JP=LOW | Success | 0.16 | 0.86 | 1.788 |
| 3 | TPE_JP=LOW | Success | 0.213 | 0.84 | 1.747 |
| 4 | TPE_EI=HIGH & TPD_NS=HIGH | Success | 0.2 | 0.83 | 1.726 |
| 5 | TPE_NS=LOW &TPD_NS=HIGH | Success | 0.226 | 0.77 | 1.601 |
| 6 | TPD_NS=HIGH &TPD_JP=HIGH | Success | 0.173 | 0.76 | 1.58 |
| 7 | TPE_TF=HIGH &TPD_NS=HIGH | Success | 0.24 | 0.75 | 1.56 |
| 8 | TPE_TF=HIGH & TPD_JP=HIGH | Success | 0.2 | 0.75 | 1.56 |
| 9 | TPE_NS=LOW &TPE_TF=HIGH & TPD_NS=HIGH | Success | 0.186 | 0.74 | 1.539 |
| 10 | TPE_NS=LOW & TPE_TF=HIGH &TPD_JP=HIGH | Success | 0.173 | 0.72 | 1.497 |

*Table 5. Rules for relating personality and team performance*

Some rules such as 7 add another suggestion that feeling personality improves the efficiency of the teams. Intuitively we can say F types may focus on the harmony of teams, while T type may prefer focus on getting the job done which might frustrate each other. Bradley et al. (J. H. Bradley & Hebert, 1997) studied and compared two teams of software developers. The successful team had a larger percentage of F types. They conclude feeling personality types help team to focus more on group harmony and consequently result in successful performance. These findings are in line with our findings that High TPE in feeling can help teams to be more successful. Also, if we consider feeling similar to agreeableness (McCrae & Costa, 1989), more evidences support this finding (e.g. Barrick & Stewart, 1998).

Some rules such as rule 4 suggest that the most successful teams are introverted. In contrast to our findings, several studies suggest that extroversion improves the efficiency of software teams e.g. (Barrick & Mount, 1991) and (Barry & Stewart, 1997a). An explanation for our different results can be due to the different communication type in our study comparing to the relevant studies in literatures. These researches study organizations and teams who had oral communications, but in our case, teams use text-based tools. Behaviours of extraverted and introverted people are highly related to the type of communication and it seems that introverted overcome the issue of lack of communication when the communication mode is verbal.

In general our rules suggest that managers may achieve enhanced team outcomes by choosing teams that are (a) generally introverted, judging, feeling and intuition-oriented, (b) are heterogeneous in terms of cognitive style (iNtuition-Sensing dimension) and life approach (Judging-Perceiving) dimension.

# 6 CONCLUSIONS

This work's primary goal is to provide a methodological construct for studying and managing the relationship between team personalities and performance. However, to achieve that it provides three further contributions which are (a) insight about the personality distribution among social network users, (b) a computational model to determine the relationship between MBTI personality and the LIWC dimensions, and (c) information in connection with Python software development projects concerning the relationships between teams' 'performances and their MBTI personality profiles. Software project managers may find these insights useful for managing their own software development projects, but we want to stress that our general approach can be used to derive a new rule set that is customized for a specific software engineering context. Most of our findings are supported by some empirical studies in the literature as discussed in Section 5.

Some people might object that the MBTI psychometric measure is not an objective measure of individual personality. In the context of our work, however, this is not really an issue. As long as people can be relatively objectively identified with these different MBTI profiles and then have those profiles correlated with software development performance, then our methodology can be useful, irrespective of any larger implications about the MBTI metric.

Our approach does have some recognized limitations. The LIWC tool has a bias against individuals whose first language is not English, and we did not separate non-English users and developers. Social roles, gender, age and other demographic factors which are not covered in this study might be involved in the linguistic styles. Moreover, our proposed rules about the linkage of personality and team performance may not be applicable to other development contexts. We had a limited number of developers' teams in a particular domain. Further experiments and validations must be performed before our correlations and results can be generalized.

## References

Aamodt, Michael G., And Wilson W. Kimbrough. 1982. "Effect Of Group Heterogeneity On Quality Of Task Solutions." *Psychological Reports* 50(1): 171–74.

Agrawal, R, and R Srikant. 1994. "Fast Algorithms for Mining Association Rules." *Proc. 20th int. conf. very large data bases, VLDB*.

André, Margarita, María G. Baldoquín, and Silvia T. Acuña. 2011. "Formal Model for Assigning Human Resources to Teams in Software Projects." *Information and Software Technology* 53(3): 259–75.

Barr, Stewart, Andrew Gilg, and Gareth Shaw. 2011. "'Helping People Make Better Choices': Exploring the Behaviour Change Agenda for Environmental Sustainability." *Applied Geography* 31(2): 712–20.

Barrick, MR, and GL Stewart. 1998. "Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness." *Journal of applied psychology 83.3*.

Barrick, Murray R., And Michael K. Mount. 1991. "The Big Five Personality Dimensions And Job Performance: A Meta-Analysis." *Personnel Psychology* 44(1): 1–26.
Barry, B, and GL Stewart. 1997. "Composition, Process, and Performance in Self-Managed Groups: The Role of Personality." *Journal of Applied psychology*.
Bazelli, Blerina, Abram Hindle, and Eleni Stroulia. 2013. "On the Personality Traits of StackOverflow Users." In *2013 IEEE International Conference on Software Maintenance*, IEEE, 460–63.
Bowers, C. A., J. A. Pharmer, and E. Salas. 2000. "When Member Homogeneity Is Needed in Work Teams: A Meta-Analysis." *Small Group Research* 31(3): 305–27.
Bradley, Bret H, Anthony C Klotz, Bennett E Postlethwaite, and Kenneth G Brown. 2013. "Ready to Rumble: How Team Personality Composition and Task Conflict Interact to Improve Performance." *The Journal of applied psychology* 98(2): 385–92.
Bradley, John H., and Frederic J. Hebert. 1997. "The Effect of Personality Type on Team Performance." *Journal of Management Development* 16(5): 337–53.
Cheng, Mandy M., Peter F. Luckett, and Axel K-D. Schulz. 2003. "The Effects of Cognitive Style Diversity on Decision-Making Dyads: An Empirical Analysis in the Context of a Complex Task." *Behavioral Research in Accounting* 15(1): 39–62.
Choi, Kyungsub S., Fadi P. Deek, and Il Im. 2008. "Exploring the Underlying Aspects of Pair Programming: The Impact of Personality." *Information and Software Technology* 50(11): 1114–26.
"Collegeconfidential." 2015. http://talk.collegeconfidential.com/college-confidential-cafe/151904-whats-your-myers-briggs-personality-type-6.html.
Costa, Paul T., and Robert R. McCrae. 1992. "Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)." *Odessa, FL: Psychological Assessment Resources*.
Cruz, Shirley, Fabio Q.B. da Silva, and Luiz Fernando Capretz. 2015. "Forty Years of Research on Personality in Software Engineering: A Mapping Study." *Computers in Human Behavior* 46: 94–113.
English, A, RL Griffith, and LA Steelman. 2004. "Team Performance The Effect of Team Conscientiousness and Task Type." *Small Group Research*. Foster, Don. 2014. *Author Unknown: On the Trail of Anonymous*. Henry Holt and Company.
Gorla, Narasimhaiah, and Yan Wah Lam. 2004. "Who Should Work with Whom?" *Communications of the ACM* 47(6): 79–82.
Hart, RP. 1984. "Verbal Style and the Presidency: A Computer-Based Analysis." Academic Pr.
Humphrey, SE, and JR Hollenbeck. 2007. "Trait Configurations in Self-Managed Teams: A Conceptual Examination of the Use of Seeding for Maximizing and Minimizing Trait Variance in Teams." *Journal of applied psychology 83.3*.
Jung, C.G. 1921. *Psychological Types: Or the Psychology of Individuation.* Harcourt, Brace.
Karn, J, and T Cowling. 2006. "A Follow up Study of the Effect of Personality on the Performance of Software Engineering Teams." *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering. ACM.*
Keirsey, D. 1998. *Please Understand Me II: Temperament, Character, Intelligence*. Prometheus Nemesis Book Co. ISBN 1-885705-02-6.
LePine, Jeffery a., Brooke R. Buckman, Eean R. Crawford, and Jessica R. Methot. 2011. "A Review of Research on Personality in Teams: Accounting for Pathways Spanning Levels of Theory and Analysis." *Human Resource Management Review* 21(4): 311–30.
Lewis, TL, and WJ Smith. 2008. "Creating High Performing Software Engineering Teams: The Impact of Problem Solving Style Dominance on Group Conflict and Performance." *Journal of Computing Sciences in Colleges*.
Licorish, Sherlock A., and Stephen G. MacDonell. 2014a. "Personality Profiles of Global Software Developers." In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, New York, New York, USA: ACM Press, 1–10.

Licorish, Sherlock A., and Stephen G. MacDonell.. 2014b. “Understanding the Attitudes, Knowledge Sharing Behaviors and Task Performance of Core Developers: A Longitudinal Study.” *Information and Software Technology* 56(12): 1578–96.

Martin, B.A, C.-C Bowen, and S.T Hunt. 2002. “How Effective Are People at Faking on Personality Questionnaires?” *Personality and Individual Differences* 32(2): 247–56.

McCrae, R R, and P T Costa. 1989. “Reinterpreting the Myers-Briggs Type Indicator from the Perspective of the Five-Factor Model of Personality.” *Journal of personality* 57(1): 17–40.

McGrath, JE. 1986. “Studying Groups at Work: Ten Critical Needs for Theory and Practice.” *Designing effective work groups*.

Miller, J, and Z Yin. 2004. “A Cognitive-Based Mechanism for Constructing Software Inspection Teams.” *Software Engineering, IEEE Transactions on*.

Mohammed, Susan, and Linda C. Angell. 2003. “Personality Heterogeneity in Teams: Which Differences Make a Difference for Team Performance?” *Small Group Research* 34(6): 651–77.

Myers, I. B. 1962. “The Myers-Briggs Type Indicator.” *Consulting Psychologists Press*.

Myers, Isabel Briggs, Mary H. McCaulley, and Robert Most. 1985. “Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator.” *Consulting Psychologists Press*.

Neuman, G. A., S. H. Wagner, and N. D. Christiansen. 1999. “The Relationship between Work-Team Personality Composition and the Job Performance of Teams.” *Group & Organization Management* 24(1)

Pennebaker, JW. 2001. “Linguistic Inquiry and Word Count: LIWC 2001.” *Mahway: Lawrence Erlbaum Associates71 (2001)*

Pennebaker, JW, and LA King. 1999. “Linguistic Styles: Language Use as an Individual Difference.” *Journal of personality and social psychology 77.6*

Peslak, AR. 2006. “The Impact of Personality on Information Technology Team Projects.” *Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research:. ACM*.

Pieterse, V, DG Kourie, and IP Sonnekus. 2006. “Software Engineering Team Diversity and Performance.” *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries. South African Institute for Computer Scientists and Information Technologists,.*

"Reddit". 2015. http://www.reddit.com/r/AskReddit/comments/cj6fi/whats_your_myersbriggs_personality_type_reddit/.

Rigby, Peter C., and Ahmed E. Hassan. 2007. “What Can OSS Mailing Lists Tell Us? A Preliminary Psychometric Text Analysis of the Apache Developer Mailing List.” In *Fourth International Workshop on Mining Software Repositories (MSR’07:ICSE Workshops 2007)*, IEEE, 23–23.

Rutherfoord, Rebecca H. 2001. “Using Personality Inventories to Help Form Teams for Software Engineering Class Projects.” *ACM SIGCSE Bulletin* 33(3): 73–76.

Salleh, Norsaremah, Emilia Mendes, and John Grundy. 2012. “Investigating the Effects of Personality Traits on Pair Programming in a Higher Education Setting through a Family of Experiments.” *Empirical Software Engineering* 19(3): 714–52.

The Myers & Briggs Foundation. 2015."How Frequent is My Type", *The Myers & Briggs Foundation*. http://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm (June 9, 2015).

Warsaw, Barry; Hylton, Jeremy; Goodger, David. 2000. “PEP 0001 -- PEP Purpose and Guidelines” *Python Enhancement Proposals. Python Software Foundation*.

Wiesche, M, and H Krcmar. 2014. “The Relationship of Personality Models and Development Tasks in Software Engineering.” *Proceedings of the 52nd ACM conference on Computers and people research. ACM.*