On the influences of personality traits on employees engagement with gamified enterprise tools

Master's Thesis

Lei Yen Cheung

On the influences of personality traits on employees engagement with gamified enterprise tools

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE
TRACK INFORMATION ARCHITECTURE

by

Lei Yen Cheung
born in Voorburg, The Netherlands

Web Information Systems
Department of Software Technology
Faculty EEMCS,
Delft University of Technology
Delft, the Netherlands
http://wis.ewi.tudelft.nl

IBM
Center for Advanced Studies
Johan Huizingalaan 765
1066 VH Amsterdam, the Netherlands

http://www.ibm.com

On the influences of personality traits on employees engagement with gamified enterprise tools

Author: Lei Yen Cheung
Student id: 1358219
Email: lieyen@gmail.com

Abstract

Gamification techniques are used in enterprises to support employees' engagement with computer-mediated business processes. The potential effectiveness of the incentives brought by gamification techniques are, however, not equally appealing to individuals. To better understand when gamification can be an effective engagement aid, it is important to study how individual differences (personal or character-related) of employees relate with the effectiveness of game mechanics applied to enterprise-class computer tools. *Personality* is a property of an individual that is known to influence, among others, task performance, learning styles, and gaming preferences. Despite the existence of an abundant body of research, the relationship between the effectiveness of game mechanics in an enterprise setting and the personality of employees is yet to be fully understood. This thesis contributes new knowledge on the matter, by studying the influence of personality traits and gender stereotypes on the behavior of 177 IBM employees that participated in an experiment on gamified learning and socialness experience. We engaged with the employees of the IBM Netherlands in the Netherlands and performed a personality trait and gender stereotype inventory by means of a questionnaire. The results of the questionnaire supported our investigation on the descriptive power of personality traits in explaining the differences in participation and engagement in the targeted population. Finally, we validated the effectiveness of state-of-the-art techniques for automated personality assessment, to assess the possibility of developing large-scale experiments on the effect personality traits without the need for questionnaires.

Thesis Committee:

Chair:	Prof. Dr. Ir. Geert-Jan Houben, Faculty EEMCS, TUDelft
University supervisors:	Dr. Alessandro Bozzon, Dr. Judith Redi, Faculty EEMCS, TUDelft
Company supervisor:	Drs. Robert-Jan Sips, Center for Advanced Studies, IBM Netherlands
Committee Member:	Dr. Pablo Cesar, Faculty EEMCS, TUDelft

Preface

In November 2014 I started this research as a graduation project for my masters, it turned out to be a long battle against the time. This document is the report of this long process. It cannot express the long days working on this research, traveling frequently from The Hague to Amsterdam, working together with my fellow colleagues inside and outside IBM, the hope for good results and the sadness and tiredness with each failed attempt. Nevertheless, I learned many things during this project and I am proud of what I have accomplished. And now I am about to earn my Master Degree in Computer Science from Delft University of Technology. This research is conducted in the Web Information Systems (WIS) group at Delft University of Technology in collaboration with The Center for Advanced Studies (CAS) department at IBM, Amsterdam.

Several persons have contributed academically, practically and with support to this master thesis. I would therefore first like to thank my direct supervisors Alessandro Bozzon and Robert-Jan Sips for providing me the opportunity to work in a professional environment to conduct my research and guiding me throughout this journey. Alessandro Bozzon supervised me step-by-step to make sure that my work was solid and helped me with setting up my research. Robert-Jan Sips also helped me to make sure that I was on the right track. I am very grateful to my second university supervisor Judith Redi for her patience and support to help me with the statistical analysis, providing me feedback and guiding me through the process. A great deal of appreciation goes to Professor Johnson of providing me a reference dataset and a detailed description for each personality trait. In addition, I would like to thank Professor Geert-Jan Houben for his feedback on this work. Likewise, I would like to thank Pablo Cesar for his participation in the thesis committee. Furthermore, I also wish to thank all of the respondents, without their cooperation I would not have been able to conduct this research. Last but not least, I am very obliged to my parents, brother and some dearest friends for their support through the whole journey of my masters, without them I wouldn't have been able to achieve this.

Lei Yen Cheung
Delft, the Netherlands
February 29, 2016

Contents

List of Figures

Chapter 1

Introduction

Modern enterprises make an increasing use of computer tools to support the daily activities of their employees. The issue of understanding how these tools can actually help companies and employees in reaching their business goals, and how they affect the engagement of employees with the enterprise has been subject of study for a rather long time.

Recent years saw an increasing research effort in the investigation of the potential of *gamification* – i.e. the use of game mechanics in non-game context – as a means of supporting user engagement and fostering user activity, social interaction and quality and productivity of actions[33].

Businesses are turning to gamification both to engage and motivate their employees and to meet business needs. Studies showed that when gamification is implemented successfully, it can give an edge to enterprises by increasing user motivation and achievement of business goals[70].

While potentially beneficial, gamification techniques are not appealing to all individuals [19, 69, 70]. Several studies investigated the effectiveness of different game mechanics in engaging employees with social network sites[23], information systems [37] and e-commerce system[35]. For example Farzan et al. indicated that gamification only has a positive effect on engaging some users with social network sites for a short time [23].

Others studies suggested that the context of the service might be an essential antecedent for engaging people with gamification [80, 33]. In a recent experiment performed in IBM[78] it has been shown that gamification can be used to support user engagement and drive the online social behavior of employees in an enterprise environment. However, variations in the level of engagement were observed in groups of the experiment. Assuming that engagement and motivations are personal, we advocate the need to study the influence of the so called individual differences on the effectiveness of game mechanics applied to enterprise-class computer tools.

1.1 Motivation

Computer tools are often created with a one-size-fits-all philosophy, where a single design (of functionalities, interaction, user interfaces, etc.) suffices for the needs of all the intended users. While the trend towards personalisation and adaption in Web applications demonstrate the need to cope with this rather short-sighted mindset, enterprise-class applications are, to some extent, falling behind. Recognizing and valorizing individual differences can be as important within the enterprise as it is on the Web: the acquisition of knowledge and understanding about an employee's individual properties is beneficial for both the company and the employee. On the one hand, the company could use these insights to adapt applications to the employee's needs, thus implicitly achieving greater engagement; on the other hand, the employee can feel more empowered by the availability of tools that actually support her in achieving business goals.

These individual differences can be seen as something "personal" that stirs and drives the user in the usage (and the modalities of usage) with a computer application. In this work, we characterise the "personal" dimension by using the *Theories of personality*[32], which defines personality as: "*The characteristics or blend of characteristics that make a person unique*". To describe the personality of an individual we use the *Trait Theory of Personality*, which defines a trait as follows: "*a trait can be thought of as a relatively stable characteristic that causes individuals to behave in certain ways"*. The trait theory suggests that individual personalities are composed of these broad dispositions. Psychologists who support the trait theory of personality believe that there are five basic traits (called the "*Big Five*"): extraversion, openness to experience, conscientiousness, neuroticism and agreeableness[60].

Studies performed in the social sciences highlighted how personality can influence task performance, learning styles, and gaming preferences. The relationship between the effectiveness of game mechanics in an enterprise setting and the personality of employees is yet to be fully understood.

1.2 Research objectives

Previous work conducted in IBM Netherlands[78] has shown that gamification can help the enterprise to engage people, although with different effectiveness across the targeted population. This work inspired the activities of this thesis, and drove the formulation of the main research question addressed by this thesis:

To which extent do personality traits affect engagement strategies in enterprise gamification?

To answer this question, we defined a set of objectives that guided the thesis activities:

Objective 1 *Understand the construct of personality, and its role in an enterprise environment.*

The goal is to understand what personality is from a scientific perspective, and how can it be measured in enterprise. A literature study guides the investigation of which are the most relevant individual traits that may impact gamification, and the identification of techniques to measure such traits. This objective links to the literature study presented in Chapter 2 and the measurement of personality traits described in Chapter 4. In addition, we also analyze the specificity of our population in terms of personality traits and characteristics, which is presented in Chapter 5.

Objective 2 *Define an experimental methodology.*

The goal is to design the methodology used to conduct our study, according to the findings produced by the literature study. This objective links to Chapter 3 where we describe which decisions we made and sketch the methodology applied in this work.

Objective 3 *Measure and understand how personality traits can be related to the interaction with gamified application.*

This objective relates to the investigation of how the outcome of the gamified application developed in [78] (and described in Chapter 7) can be used to study how personality traits affect the gameplay and the employee engagement.

Objective 4 *Identify a reliable technique to automatically assess the personality of employees without recurring to questionnaires or other invasive testing techniques.*

This investigates the existence and effectiveness of automatic methods that could be used to infer employees' personality. The availability of such method would enable large-scale studies. Our findings are presented in Chapter 6.

Objective 5 *Understand to what extent the findings obtained from the activities of Objective 3 can be generalized to a broader population.*

This objective refers to the generalization study explained in Chapter 8, where we attempt to generalize the findings produced in Objective 3 to the general IBM population and a general population which is not specific to IBM.

1.3 Contribution

This thesis provides the following original contributions:

1. **A data set: Participants' game behavior and personality measurements.**

   We extend a previous data set from [78] with information about players' personality traits as calculated from a standard personality questionnaire. A complete description of the dataset is provided in Chapter 4. This dataset contains rich information about users and their scores for each measured personality trait. This information is utilized in our work to analyze the personalities of employees in enterprise (described in Chapter 5), the relationship between personality traits and game play, described in Chapter 7.

2. **Novel insights on the relationship between personality traits and game mechanics applied on enterprise-class applications.**

   The second contribution of this thesis is a thorough analysis carried out on the gathered data. The aim of this analysis is to investigate the descriptive power of personality traits in the interpretation of the behaviour of employees interacting with a gamified enterprise-class application. Through this analysis we were able to prove that personality traits could be used to explain behaviour difference in the game (see Table D.1), as discussed in Chapter 7.

3. **Validation of an automated tool for personality measurement based on social media data.**

   By capitalising on the questionnaire data described in the previous point, we performed a study aimed at validating the effectiveness of an automated tool for personality measurement (described in Chapter 6), fed with data coming from 1) enterprise social media (i.e. IBM Connections), and 2) public social media (i.e. Twitter). The goal was to understand the capability of this tool, so to pave the way to future studies performed at a larger scale.

1.4 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 introduces the scientific background of this thesis and discusses related work on gamification, gameplay and personality measurement. The methodology of this work is described in Chapter 3. Chapter 4 focuses on explaining the sub-studies and processes to collect the data we need for this study. We summarize the analysis of personality traits in Chapter 5. The study on automatic personality assessment is described in Chapter 6. In the same chapter, we also present a study on the reliability of this alternative method to measure personality automatically. Personality and Gameplay, Chapter 7, presents the investigation on the predictive power of personality traits in the prediction of employees interaction with a gamified enterprise-class tool. In Chapter 8 we investigate whether the observed outcomes can be generalised to a broader population. Each chapter concludes the work with a chapter conclusion. Afterward, we summarize the results and observations from our study, along with concluding remarks and proposals for future work in Chapter 9.

Chapter 2

Background and Related work

This chapter discusses several topics which put this thesis into context: user engagement, gamification, personality, the relationship between personality and gameplay, and measuring personality. In Section 2.1 we define what user engagement is based on literature and discuss some related work. In Section 2.2 we describe enterprise gamification, together with previous work in this field. We explain the concept of personality and personality traits in Section 2.3. Subsequently, we describe the Big Five model in Section 2.4. Afterwards, we explain how personality traits can be measured in Section 2.5. In Section 2.6 we discuss work related to the relationship between personality and gameplay based on previous work in literature. Finally, we conclude the chapter with a chapter conclusion in Section 2.7.

2.1 User Engagement

According to O'Brien and Toms[66], user engagement is "*a quality of user experiences with technology that depends on the aesthetic appeal, novelty, and usability of the system, the ability of the user to attend to and become involved in the experience, and the user's overall evaluation of the experience*". Another definition of user engagement is "*the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource*" [3].

Both definitions indicate that user engagement can be defined as a connection between the user and a *property*. In the first definition, this property is linked to the user experience, while in the second definition the property is linked to a resource. This property is dependent on the application. In addition, engagement depends on the depth of participation the user is able to achieve with respect to each experiential property. Moreover, there are large individual differences with respect to which elements could trigger engagement and maintain it.

Users in an enterprise environment often use computer tools to reach business values, objectives and support their job. Due to individual differences, these computer tools are not supportive to all users within the enterprise as they wish. A common problem that researchers try to solve in the last decades is to engage users within the enterprise with the means of these computer tools[23]. Within the enterprise, user (employee) engagement is particularly important, because engaged employees are more committed to their organization's goals and values, motivated to contribute to organizational success, and are able at the same time to enhance their own sense of well-being.

Literature shows that there are many ways to measure engagement, subjectively or/and using metrics. The most common ones are self-reported measurements, like the 7-item questionnaire from Webster and Ho's[82]. Another way which has been used in previous research is the use performance indicators not as measures of engagement, but as correlates of engagement.

Many researches have shown that engagement is related to personality[39] and that personality can be used to predict engagement. For example, in a study[54] it has been shown that work engagement is characterized by high scores on extroversion in combination with low scores on neuroticism. However, there is also a need to develop an understanding of engagement on an individual level. People who are more engaged or less engaged are likely to differ in certain personality traits as well as in the nature of their jobs [54]. Driving engagement is related to personal aspects which means that a proper understanding of users is required. One way to better understand users is to define a user model. User modeling can be done either manually or automatically.

Having defined and explained the concept of user engagement, we move on to gamification which presents a way to trigger and support user engagement.

2.2 Gamification

Gamification is the application of game design principles and mechanics to non-game environment[52]. It represents a means of supporting user engagement and positive patterns in service use, but these effects depend heavily on the context in which it is implemented as well as on the users it targets [35]. Businesses are turning to gamification both to engage and motivate their employees and supporting them to meet their business needs.

In the last couple of years, research efforts have been devoted to investigating the potential of gamification as a means of supporting user engagement and enhancing positive patterns in service use, for example increasing social interaction and user activity. Hamari et al. provided a literature review of empirical studies on gamification in [35]. They found out that points, leaderboards, and badges were clearly the most commonly used motivational affordances used in literature [34]. According to their study, it seems that gamification does produce positive effects and benefits. An example which indicates that gamification has a positive effect on some users for a short time is the study done in [23]. While another study[80] done by Tom et al. shows that removing gamification might have detrimental effects on those users who are still engaged by gamification.

When enterprise gamification is implemented successfully it can increase user motivation and achievement of goals. Moreover, it also helps enterprise to engage employees and meet business needs [70]. There are a few enterprise gamification successes that have spawned from carefully applying game mechanics to fit the unique needs of each organization. Most enterprise gamification applications focus on points, leaderboards and badges. For example, in a previous study[78, 79] it has been shown that gamification can be used to support user engagement and drive the online social behavior of employees in an enterprise environment. In this study, the authors used points, badges and leaderboard as game mechanics to engage employees with an enterprise class tool. Employees could earn points and badges by interacting with the tool (i.e. answering quiz questions, inviting colleagues to play the game and sharing news). Another example of gamification applied within the enterprise is *Badgeville*, which is a software service that has many customizable options for companies to configure any type of goal ranging from task-related goals such as completing expense reports to learning goals such as leveling up a key industry skill. Employees are rewarded with points for participating in training programs as well as badges. A leaderboard is used to visualize the scores among colleagues, which is updated weekly with the highest scores. The three top scorers every month could be given the opportunity to participate as one of the faces of the enterprise's annual marketing campaign[14].

Having defined and explained the concept of enterprise gamification, we now turn to personality, which is another important aspect of this study.

2.3 Personality

A well-accepted definition of personality is defined as follows: "*Personality is that pattern of characteristic thoughts, feelings, and behaviors that distinguish one person from another and that persists over time and situations*" [68].

Throughout the last decades, many theories have been developed to characterize a person based on the personality they have, by analyzing and studying the action, attitude and behavior of a person. To exploit available features that might enhance systems' performances, researchers are investigating different fields including psychology.

The study of personality has a broad and varied history with an abundance of theoretical traditions. Psychologists use different approaches when describing personality, and there are at least 18 theories to define it. However these can be scaled down to four types of personality theories[75]: Psychodynamic, Humanistic, Trait and Social Cognitive approaches to personality.

Since we are not psychologists, we will focus on the *Trait Theory*. This theory is an approach to studying human personality that identifies and measures the degree to which certain personality *traits*, a set of relatively stable characteristic that causes individuals to behave in certain ways, exist from individual to individual[49]. Experts use the personality traits theory to find and measure the psychological differences among people. Computer systems can easily deal with personality when it is described with personality traits [58, 65] instead of the broader term *personality*. Therefore, we focus on personality traits in the remaining part of this study.

2.3.1 Personality Traits

As mentioned before personality traits can be defined as a set of relatively stable characteristics that causes individuals to behave in certain ways. Personality traits often refer to enduring personal characteristics that are revealed in a particular pattern of behavior in a variety of situations. It describes actions, attitudes, behaviors, thoughts and emotion that an individual possess. Psychologists who support the trait theory of personality believe that there are five basic traits (called the "*Big Five*"): extraversion, openness to experience, conscientiousness, neuroticism and agreeableness [60].

In the last decades research efforts have been devoted to studying personality traits of people. For example, Komarraju et al. investigated the relationship between the Big Five personality traits and academic motivation[50]. While in another work the associations between the Big Five personality traits and environmental engagement is examined[61].

Previous research[17, 53, 73, 73, 40] has shown that personality traits correlate with a number of real-world behaviors and that a strong correlation exists between personality traits and how people behave. For example, in [17] the relationship between people's personality traits and social media use is explored. Another work investigated the relationship between Big Five personality traits and Internet usage [53]. Moreover, the authors in [40] investigated the relationship between personality and motivation for playing online games.

Personality traits are relatively stable over time, they differ across individuals and influence behavior. In some theories and models, personality traits are properties a person could have, but in other models personality traits are dimensions in which a person can be rated. In the next section, we describe several models which are frequently used and commonly accepted in literature.

2.4 The Model of Traits

There are two models, which describe personality along different traits. For each model personality traits can be measured by means of self-reported measurements (i.e. questionnaires or tests). The first model is the *Three Factor Model of Traits* created by Eysenck which describes personality into three major traits: neuroticism, extraversion and psychoticism [21]. The second model is *Five Factor Model of Traits*, also know as the *Big Five personality model*, which describes personality in five dimensions: extraversion, openness to experience, agreeableness, and conscientiousness [18].

Both models use self-reports and construct hierarchical taxonomies by using factor analysis. However, the trait *psychoticism* marks the two models apart, because the Big Five personality model does not contain such a trait. Research[59] has shown that this trait does not fit into a normal distribution curve, like the other trait in both models. This causes scores to be rarely high and considerable overlap with some psychiatric conditions. Therefore, we choose to use the Big Five Model as a general approach to describe human personality. Moreover, this model is the most widely used and accepted model of personality in psychology. Another reason to use this model is because in previous studies [12, 48] it have been shown that an unique combination of each of the Big Five traits influences an individual's success at work. It was shown that the combination of the Big Five determines why and how a person is motivated to achieve certain goals.

2.4.1 The Big Five Model

The Big Five model defines human personality series of five dimensional traits [60]. The dimensions represent the main differences among personalities and explain why individuals behave in certain ways. The five dimension or factors are Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A) and Neuroticism (N).

Each factor addresses many different personality characteristics which we describe later.

The first trait, Openness to experience, represents an individual's willingness to consider alternative approaches and describes a dimension of cognitive style that distinguishes imaginative, creative people from down-to-earth, conventional people. Conscientiousness, the second trait, reflects the degree to which an individual is organized, diligent and scrupulous. The third trait, Extraversion, is marked by pronounced engagement with the external world and the tendency to be sociable and able to experience positive emotions. Agreeableness, the fourth trait, reflects individual differences in concern with cooperation and social harmony. The last dimension is Neuroticism, which reflects a person's tendency to experience psychological distress and emotional suffering. The traits of the Big Five model are also known as *OCEAN*, which is an abbreviation of the first letter of each trait.

These five traits describe personalities across five broad domain. However, it is also possible to use the six associated subordinate traits to describe personality more specifically. The six subordinate traits belonging to a Big Five dimension are negatively correlated with the opposite of the other dimension, meaning that they are positively correlated with the Big Five dimension. These subordinate traits are known as facets. The Big Five model and the 30 facets are shown in Table 2.1. A detailed explanation[1] about each trait and facet can be found in Appendix A.

2.4.2 Relationship between Big Five traits and work engagement

An unique combination of each of the Big Five traits influences an individual's success at work in three ways. First, the combination of the Big Five determines why and how a person is motivated to achieve certain goals. For example people who score high on extraversion are more motivated to achieve a goal if a reward is involved. Second, an individual's mood is affected by personality which reflects in the way a person responds to situations at work[12, 48]. In several studies[47, 46] it has been proven that conscientiousness and agreeableness are related to job satisfaction, because these traits indirectly affect performance behavior. In other words, if a person is happy with his job than it is more likely that this person has better *performances* at work. Third, a person's interpersonal relationship is affected by a person's personality profile, which makes it an important determinant of work success when the work involves getting along with other people.

2.5 Measuring personality traits

Two common ways of assessing the personality of an individual are doing a variety of tests (1) or doing a self-report inventory (2). Self-report inventories categorize people based on the scores for the personality dimensions and provide much more mature understanding of ways in which attitudes and personality traits affect behavior.

In the remaining part of this section, we describe two different approaches of how personality can be measured in the remaining part of this section. We consider the ap-

[1] The description is provided by John Johnson, Professor Emeritus of Psychology at the Pennsylvania State University

The Big Five personality traits	Facets and correlated trait adjective
Openness	Ideas (curious) Fantasy (imaginative) Aesthetics (artistic) Actions (wide interests) Feelings (excitable) Values (unconventional)
Conscientiousness	Competence (efficient) Order (organized) Dutifulness (not careless) Achievement striving (thorough) Self-discipline (not lazy) Deliberation (not impulsive)
Extraversion	Gregariousness (sociable) Assertiveness (forceful) Activity (energetic) Excitement-seeking (adventurous) Positive emotions (enthusiastic) Warmth (outgoing)
Agreeableness	Trust (forgiving) Straightforwardness (not demanding) Altruism (warm) Compliance (not stubborn) Modesty (not show-off) Tender-mindedness (sympathetic)
Neuroticism	Anxiety (tense) Angry hostility (irritable) Depression (not contented) Self-consciousness (shy) Impulsiveness (moody) Vulnerability (not self-confident)

Table 2.1: The Big Five traits and their corresponding facets. Chart recreated from [41]

proaches described in Section 2.5.1 as *subjective measurements* of personality traits. With subjective measurement, we mean that in the process of determining personality traits questionnaires are involved or directly asking people to provide information. However, in Section 2.5.2 we elaborate on studies in which personality traits can be "recognized" by using available data created by users on Social Media, these approaches are considered as *objective measurements* of personality traits.

2.5.1 Subjective approach to measure personality traits

In this section, we describe several methods to measure different personality traits on a subjective way (i.e. using personality tests or questionnaires).

International Personality Item Pool NEO Inventory

The International Personality Item Pool NEO inventory (IPIP-NEO) is a personality inventory created by Lewis Goldberg[30]. The inventory was designed to measure constructs similar to those assessed by the 30 facets scales in the NEO Personality Inventory[18]. A decade later professor Johnson created a version of Goldberg's inventory that could be assessed on the World Wide Web, this version has been referred as the IPIP-NEO 300-item inventory. The IPIP-NEO inventory calculates scores for both the five domains of the Five Factor model and also six facets of each main domain[18].

Myer-Briggs Type Indicator

This personality type indicator is perhaps the most common one used by career counselors and in the corporate world to help crystallize people's understanding of themselves. The questionnaire is based on the work of psychologist Carl Jung [63] which sorts people into categories based on four areas: sensation, intuition, feeling and thinking, as well as extraversion/introversion. It measures psychological preferences in how people perceive the world, make decisions and describe how people choose to interact with the world [62].

The validity of the MBTI has been the subject of many criticisms. In [27] it has been shown that many studies which supported MBTI are methodologically weak or unscientific. Moreover, the MBTI does not use validity scales to assess exaggerated or socially desirable responses. A study was done by Francis et al. showed that some MBTI dimensions are correlates weakly with the Eysenck Personality Questionnaire lie scale [26].

Rosenberg Self-esteem Scale

This personality test, also known as RSES, has been developed by Rosenberg[72] in the 1960s for a study of adolescent self-image, the RSES has become the most widely used general purpose measure of self-esteem in psychological research. This test contains ten items which should be answered with a four point Likert-type scale. Five out of 10 items are stated using positive words while the other five statements have negative words in their statements. The scales of this test measure state self-esteem by asking the respondents to reflect on their current feelings.

The reliability and validity of RSES have been tested in many settings. In a number of studies, it appears to have low reliability because of the appearance of negative-worded items in the statement.

Minnesota Multiphasic Personality Inventory

This inventory, abbreviated as MMPI, is the most widely used and researched standardized psychometric test of adult personality and psychopathology [9]. MMPI is originally developed by Hathaway et al. but is copyrighted by the University of Minnesota, because this university first published it in 1943. From then on the MMPI has been revisioned several times and the current version of the test (MMPI-2) can be completed on the computer which provides a range of profile choices.

Over the years the use of MMPI has been greatly expanded. The MMPI was designed to help identify personal, social, and behavioral problems in psychiatric patients. Due to the different validity scales, it makes it difficult for people without expertise in social psychology to understand and interpret the results.

16 Personality Factor Questionnaire

The sixteen personality factor questionnaire(16PF) consist of 164 statements that need to be ranked using the 5-likert scale and is created by the British psychologist Raymond Cattell. This psychologist assumes that variations in human personality could be best explained by a model that has sixteen variables. These sixteen variables can be interpreted as personality traits. This questionnaire provides clinicians with a normal-range measurement of anxiety, adjustment, emotional stability and behavioral problems. Besides the 16 primary traits, the 16PF also constructs a version of the Big Five secondary traits which are Introversion/Extraversion, low Anxiety/high Anxiety, Receptivity/Tough-Mindedness, Accommodation/Independence, and Lack of Restraint/Self-Control. The 16PF has also been translated into over 30 languages and dialects and is widely used internationally. One study shows that the five global factors of 16PF seem to correspond closely to the *Big Five personality traits* [11].

Bem Sex Role Inventory

The Bem Sex Role Inventory(BSRI) is a self-administered questionnaire and a widely used instrument in measuring gender role perceptions [6]. BSRI measures masculine and feminine gender roles separately and is able to yield a measure of androgyny(i.e. the combination of masculine and feminine characteristics), and has adequate psychometric properties.

Research has supported the idea that androgyny correlates with a number of other positive attributes, such as higher levels of identity formation in college students. In addition, androgynous individuals have been demonstrated to have more reasons for living than gender-typed individuals[20]. These findings suggest that androgynous individuals tend to be more psychologically healthy and function more adaptively in modern living. In contrast, research suggests that individuals who are undifferentiated in terms of gender role (low on both masculinity and femininity) tend to be less adaptable. Moreover, in the past numerous studies have found that gender categorizations are correlated with many stereotypical gendered behaviors [38].

This test consists of 60 items. Each item represents a different personality trait. For each item, participants need to rate themselves based on a 7 point Likert scale. The 60 different personality traits are evenly dispersed, 20 masculine, 20 feminine, and 20 filler traits thought to be gender neutral. This results in four possible categorizations: *masculine*, *feminine*, *androgynous* and *undifferentiated*.

Masculine represents people who score high on the masculine traits and low on the feminine trait. People who score high on feminine traits and low on masculine traits are being categorized as *Feminine*. People who score equally on masculine and feminine traits will result in the category *Androgynous*. The fourth type of score, *Undifferentiated*, was seen as the result of scoring below the median on both masculine and feminine traits.

2.5.2 Automatic approaches

The methods to measure personality described in Section 2.5.1 all use a self-reporting way to determine the personality of an individual, coming from the field of Psychology. In this research, we want to take a computer science angle, by exploring the possibilities and approaches to determine personality traits automatically, without asking people to complete questionnaires. In this section, we, therefore, focus on describing alternative approaches which do not use questionnaires or a self-reporting way to measure an individual's personality.

One of the most well known examples of recognizing personality traits without intrusively asking people to fill a questionnaire is the *MyPersonality Project* [51], in which the authors show that *Facebook Likes* can be used to automatically and accurately predict a range of personal characteristics. In last decades many studies have been conducted in the field of predicting individual traits and characteristics based on various inputs like written text [24], using psychometric test [18], and the habitat of individuals[31]. Another study has examined the accuracy of personality judgment based on physical appearance [64] in which the authors provide the first evidence that even without nonverbal expressive aspects of appearance more traits can be detected and judgment are generally more accurate than observing nonverbal expressive behavior. Farnadi et al. tried to infer user's personality traits from their Facebook *status* updates using machine learning techniques [22].

The studies mentioned previously are just a begin towards a new generation of recognizing personality traits automatically without intrusively bothering users to fill in questionnaires which are boring and require quite some time to complete. User generated content in online social networking sites form a potentially rich source of information for researches, business application to explore and analyze data that leverage content for personalization. Sips et al. presents a vision of an "*Inclusive Enterprise*" [77], where the authors aim to create an automated system that senses and influences the working environment of an employee within the enterprise. The idea is to first collect data from existing infrastructure and Enterprise Social Media by sensing from the background without bothering employees to provide additional information. Then use the gathered data supported by platforms like Bluemix and Watson to enable personalized interaction and working experiences.

Watson Personality Insights[2], a tool developed by IBM which was officially released in March 2015, has the capability to infer personality traits automatically based on an input text of minimum 100 words. The initial version of this tool had been evaluated and discussed in [4]. Moreover, in another work this service was used to examine the accuracy of three types of personality traits derived from Twitter[3].

2.6 Personality and Gameplay

The ability to predict personality has implications in many areas. Existing research has shown connections between personality traits and success in both professional and personal relationships [1][2]. In this section, we focus on studies related to personality and gameplay behavior, because one of our objectives is to measure and understand how personality traits can be related to interactions with gamified applications within the enterprise.

[2] `https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/personality-insights.html`

[3] `http://www.twitter.com`

Personality types refer to the psychological classification of individuals based on their personality. In gamification, this is often referred to player types. In general, a game needs to be motivating, addictive and encouraging with very short term goals so that users can fail and try again until they succeed. This stimulates users to continue playing the game. However since we have seen that due to individual differences users are not attracted in the same way. Player types are therefore used to help classify game mechanics and game features. Player types of users playing the game are often taken into consideration to see if these users can be given ways to all have fun in the same experience, or if the personality type of a specific demographic is skewed heavily towards a certain personality type to tailor the gamification experience accordingly[25].

In a study Ferro et al. investigated the relationship between player types, and personality types and traits [25]. They identified possible relationships between player typologies and personality types and traits with that of game elements and mechanics. Five categories of player types have been identified: Dominant user, Objectivist, Humanist, Inquisitive User, and Creative Individuals. Each of these player types is matched with traits similar across trait theories in psychology, such as the Big Five, Eysenk and Myer-Briggs. Each of these theories has been described in Section 2.5.1. In another work, Codish et al. [15] studied the differences between how introverts and extroverts perceived playfulness in a gamified educational setting. The authors showed that the enjoyment from leaderboards had a negative effect on the playfulness for extroverts. Moreover, extroverts enjoyed the presents of badges more than introverts and rewards were perceived as more enjoyable by extroverts than introverts.

Additionally, a set of studies has been performed by IBM to understand whether personality characteristics inferred from social media data can predict people's behavior and preferences. One of those studies found that people who score high on excitement-seeking are more likely to respond to an information request on social media. People who score high on cautiousness are less likely to respond to an information request on social media[57]. Another study showed that people who score high on modesty, openness, and friendliness are more likely to spread information[55].

2.7 Chapter conclusions

This chapter focuses on presenting the basic concept of user engagement, gamification and personality. Additionally, we have described different methods in this chapter to determine the personality of individuals subjectively and subjectively. This chapter also stresses the importance that a strong correlation exists between personality traits and how people behave. Some of these findings are taken into account in Chapter 3 when we describe design decisions of our experimental study in our methodology.

Chapter 3

Methodology

In a previous work performed in IBM Netherlands, an enterprise computer tool (named the **IBM game**) has been developed to study gamification within the enterprise with the aim to support employee engagement and drive online social behavior of employees in a community. The study demonstrated that the implementation of game mechanics in the IBM game helped to engage employees in enterprise. However, the level of engagement was dependent on the game mechanics the population received.

In this work, we used the IBM game to investigate the relationship between the effectiveness of game mechanics in an enterprise setting and the personality of employees. To study this problem we defined a methodology consisting of three parts: (1) a literature study about engagement, gamification and personality; (2) a study on personality traits in enterprise, in which we measure and analyze personality traits of IBM employees, and investigate if these traits affected the performance of the game mechanics that were observed in the IBM game; and (3) a generalization study to investigate whether what we have observed from (2), can be generalized by taking a look at the IBM population and broader population which is not specific to IBM.

In this work we take a computer science angle, by exploring the possibilities and approaches to infer personality automatically, without asking people to complete personality questionnaires.

The findings from the literature study, presented in Chapter 2 are used to define the experimental methodology (Objective 2) described in this chapter. In the remaining part of this chapter we describe each step of our methodology in more detail and specify how we meet our objectives described in Section 1.2. An overview of our methodology is presented in Figure 3.1.

Part two: Studying personality traits

The goal of this part is to understand the population we have by measuring and analyzing their personality and investigate if the employees' personality affected the performance of the game mechanics that were observed in the IBM game. This part of the study composed of different components which we describe one by one. Each element of this part of the study contributes to meeting objective 1, 3 and 4.

- **Personality Survey**. We inject our understanding about personality combined with the findings from literature review to create a survey with the purpose to

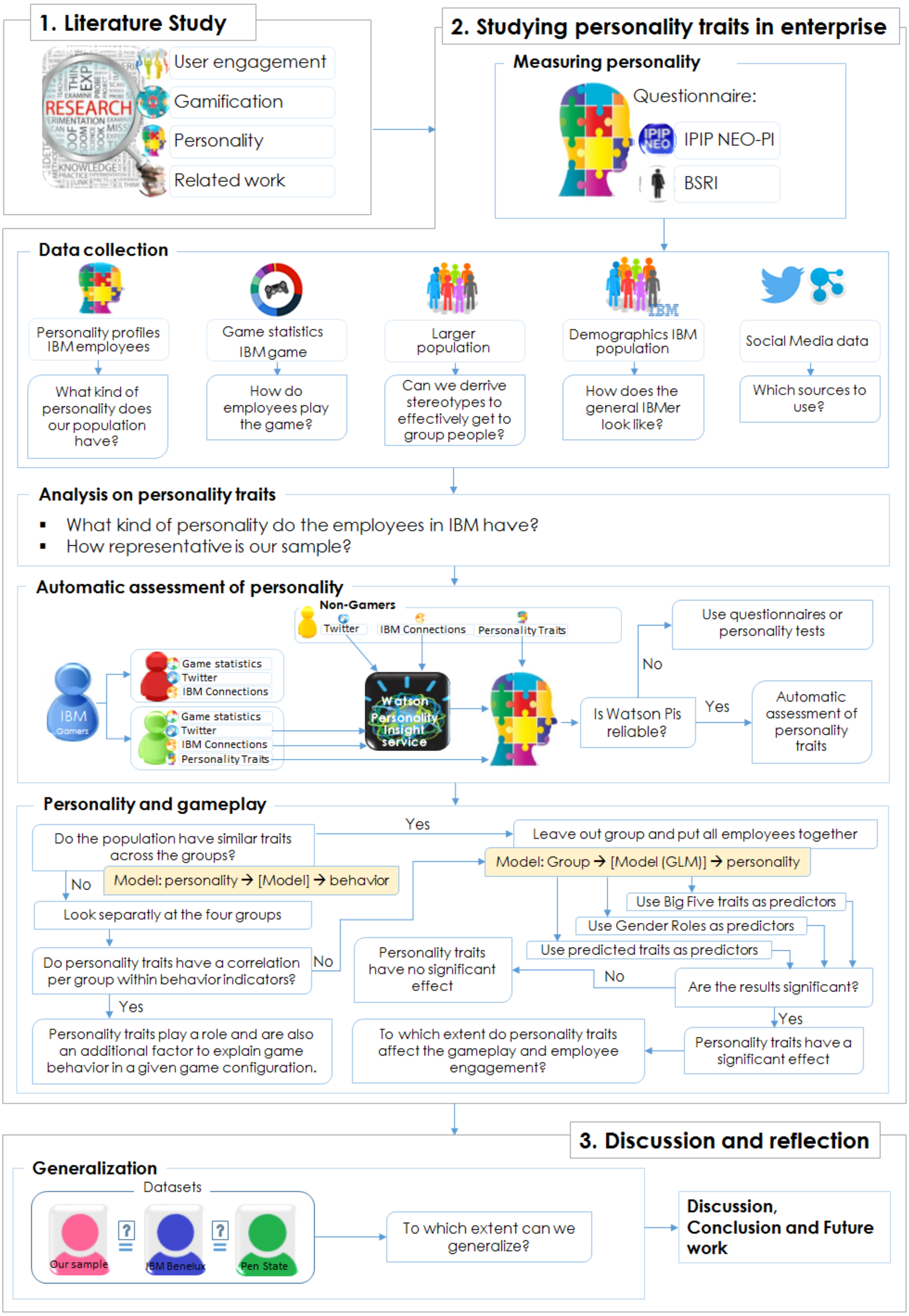


Figure 3.1: Overview of our methodology.

measure personality of employees in an enterprise environment.

Based on literature findings of personality measurement described in Section 2.5, we decided to use two subjective personality questionnaires and one objective approach to measure the personality of employees. The first personality test that we decided to use is the 120-item version of the *International Personality Item Pool NEO Inventory* (IPIP-NEO), described in Section 2.5.1. This public-domain tool for personality describes personality in terms of the Big Five Model which is the most widely accepted model. In addition, this inventory is free of charge and has been addressed in many studies as discussed in Sections 2.3.1, 2.4.2 and 2.6. Moreover, IPIP-NEO is one of the most widely used and well-validated commercial inventories in the world[42]. It provides a more detailed description of the personality of an individual, not only in terms of the Big Five dimensions but also on a lower scale of 30 facets. For the interested reader, a detailed description of each of the Big Five traits and 30 facets is presented in Appendix B. The second personality test that we decided to use is the 60-item Bem Sex Role Inventory(BSRI), described in Section 2.5.1. This inventory describes human personality in terms of gender roles instead of the Big Five Model. Gender role self-perception is one particularly promising variable that may help in further explaining within-sex and between-sex variations in empathy and affective responses. The objective approach we decided to use is Watson Personality Insights service on which we will elaborate more at a later stage.

The idea of this survey is to measure the personality of as many employees as possible to use as ground-truth for the following parts of our study. We recruited employees by means of an e-mail. We choose to implement an additional functionality into our survey which instantly measures the personality of the employee and shows a detailed report on the observed personality explained using Big Five traits and gender traits, once an employee completes the personality survey. We assumed that this functionality would attract more employees to do the survey because they get a direct reflection of the measured personality instead of receiving results afterward via email. Furthermore, to get more awareness we created a little advertisement which was published in the newsletter of our department and we posted messages on the enterprise's internal social media platform. More details on the experimental settings for the survey to determine employees' personality is described in details in Section 4.1.

- **Data Collection** To examine the objectives, described in Section 1.2 we need to collect different types of data. The required data are described below.

  - *Personality profiles of IBM employees*: the personality data that is required (1) to understand the population we have; (2) to investigate the differences in participation and engagement in the IBM game; and (3) to identify the reliability of a technique to automatically assess the personality of employees without recurring to questionnaires or other invasive testing techniques. This data is collected via the personality survey described in Section 4.1.
  - *Game statistics of the IBM Game*: this dataset contains the game statistics from the employees who played the IBM game in the period of May 2014

until August 2014. The game statistics during this period have been gathered in a dataset which has been given to us for our research. This dataset is required to meet objective 3, which is to measure and understand how personality traits can be related to the interaction with a gamified application. This dataset is described in Section 4.2

- *Dataset of a larger population*: this dataset contains personality scores of the IPIP-NEO for a larger population. Standalone scores cannot be interpreted in isolation to explain personality profiles. After reviewing previous work related to the IPIP-NEO, described in Section 2.5.1, we decided to contact Professor John Johnson, Professor of Psychology at Penn State University [45], who created the 120 item version of the IPIP-NEO, to get a reference group to whom we can compare the measured personality of people participated in our survey. We also use this data to group people with similar personality traits. In addition, this data is also required for our generalization study. This dataset is described in Section 4.3
- *Dataset with demographics of IBM population*: this dataset describes the demographics of the general IBM population in terms of age and gender. This dataset is required to investigate the representativeness of our sample to the IBM population. This dataset is retrieved by contacting the *Human Resource* Department of IBM. Due to privacy concern, the content of this dataset was not allowed to be distributed externally.
- *Social media data*: in our literature finding described in Section 2.5.2, we found a tool which is able to derive personality insights of users automatically. But in order to use this tool, data from users are needed that contains users' expressions, personal thoughts or opinions. This data are fed to the service for automatically personality assessment. We considered using data from social media, because the user often shares their feelings, thoughts and opinions on these platforms. In this study we use social media data from Twitter[1] and IBM Connections[2]. These data are used for testing the reliability of the service on which we will elaborate more in the next section. The experimental settings to retrieve these data are presented in Section 4.5.

- **Analyzing personality traits and characteristics**. To understand the population we have it is required to analyze the specificity of our population in terms of personality traits and characteristics. A descriptive analysis has been performed on our population to see what kind of personality our IBM sample has. Then we perform an exploratory data analysis to get more insights into the personality characteristics of our population by comparing it to a larger sample not specific to the IBM population. In this study, we reached out to a domain expert to help us with the analysis to interpret the calculated scores of the personality traits. This analysis contributes to meet objective 1 and 3. The results of this study are presented in Chapter 5.

[1] `http://www.twitter.com`
[2] `http://www.ibm.com/software/products/nl/conn`

- **Automatic assessment of personality**. In this research, we also take a computer science angle, by exploring the possibilities and approaches to infer personality automatically, without asking people to complete personality questionnaires. We examine the possibilities of IBM's Watson Personality Insights service(PIs) [3], which was one of the alternative methods to measure personality discussed in Section 2.5.2 of our literature study. This tool is able to infer personality traits automatically based on a given set of input text. We analyze how reliable this tool is for potential usage in a later phase of our study. The idea is to compare personality insights determined by our personality survey with the personality insights generated by the Watson Personality Insights service. We do this by observing the differences between the two sources and perform statistical analysis to see to which extent the tool is reliable. This assesses the possibility of developing large-scale experiments on the effect of personality traits without the need for questionnaires. This analysis is required to meet objective 4. Results of this investigation are presented in Chapter 6.

- **Personality and gameplay**. One of our objectives is to measure and understand how personality traits can be related to the interaction with gamified application (objective 3). In this study, we use the IBM game as a mean to investigate the predictive power of personality traits in the prediction of employees game behavior and engagement. In our literature study described in Section 2.6, we explained what previous work has done to find possible relationships between personality and gameplay. In our study, we performed various statistical analysis to find possible relationships between personality traits and game behavior. The results of this analysis is presented in Chapter 7.

Part three: Discussion and reflection

To meet objective 5, we reflect and discuss our observations and findings related to part two of this methodology. We examine if we can generalize these results to the general IBM population and to a broader population not specific to IBM. To achieve this, we need a larger data set which represents the general IBM population and a general population which is not specific to IBM. Then we conduct statistical tests between the two population to observe the generalizability of our sample. The experimental settings and results of this part are described in Chapter 8.

Concluding remarks

So, in order to proof this study in a real word content, we need to conduct several sub-studies and implement several crawlers to retrieve information and data we need to answer our research question and meet our objectives. The studies and experiments and the required data are summarized in Table 3.1.
To answer the main objective of this work, we summarize our findings and provide concluding remarks about this foundational study in Chapter 9.

[3] https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/doc/personality-insights/

[4] http://www.ibm.com/cloud-computing/bluemix/

The studies/experiments	Required data and preliminaries
1. Understanding personality and how it can be measured	- Literature study
2. Measuring employees' personality traits in enterprise	- Personality questionnaire - Tools to create questionnaire - Understanding IBM Watson Personality Insights service
3. Understanding our population	- Personality profiles and characteristics of IBM sample
4. Understanding and usage of Watson Personality Insight Service	- IBM Watson Personality Insights service - IBM Bluemix[4] - Integrated Development Environment (IDE)
5. Reliability of Watson Personality Insight Service	- IBM Watson Personality Insights service - Employees' Twitter data - Employees' IBM Connections data - Employees' personality based on survey
6. Investigating the relationship between personality and gameplay	- Game statistics IBM Gamers - Personality profiles of IBM Gamers
7. Explore generalizability of observations	- Data describing IBM Gamers - Data describing a broader population not specific to the IBM population - Data describing general IBM population

Table 3.1: The studies and experiments in our methodology.

3.1 Chapter conclusions

This chapter focused on explaining the methodology used in this research to conduct the study about studying personality in an enterprise context, observe the influences of personality traits on game behavior and an approach to generalize the observed results for further use. We described the decisions we made based on literature findings from Chapter 2. In addition, we described the studies we need to achieve our objectives, including the required data for the studies. Finally, we defined the path of our research and where the results of each part of the study can be found.

Next chapter, Chapter 4 describes how we collected data to examine the objectives and discuss our observations on the gathered data.

Chapter 4

Data collection and analysis

This chapter describes how we collected the data we needed for the experiments described in Table 3.1. For each of these datasets, we explain how the data has been retrieved along with the experimental settings and our observation and analysis of the separated data sets.

4.1 Measuring personality traits within the enterprise

The most common technique for personality measurement is asking people to complete a personality questionnaire in which they rate themselves for each trait adjectives. To measure the personality of employees within the enterprise we decided to create a personality survey and recruit employees to fill the survey. However, we also want to explore whether it is possible to measure employees' personality traits in an automated fashion. In the remaining part of this section, we describe how we created an online survey to assess personality traits of the IBM gamers via questionnaires, how we deal with legal aspects and implemented additional features. Afterward, we analyze and discuss the results of the survey. The analysis on the scores of the personality traits of the employees within IBM is described and discussed in the next Chapter.

4.1.1 Personality Survey

We created an online personality survey which is only accessible via the intranet of IBM to measure personality traits of IBM gamers subjectively (i.e. using a questionnaire). For the interested reader, the survey used in our presented in Appendix A. This survey requires about 15$\sim$30 minutes in total and consist four parts. First, we ask participants to provide some personal general information like their name, age and email address. Second, participants are asked to answers 120 statements from the IPIP NEO Inventory. Each question has a 5-point Likert Scale[1], varying from *Very Inaccurate* to *Very Accurate*. Participants are asked to rate themselves for each statement based on this scale. Third, the participants are asked to fill in the BSRI questionnaire consisting of 60 statements, each of the has a 7-point Likert Scale varying from *never or almost never true* to *always or almost always true*. In the last part of the personality survey some general questions related to the survey are asked.

[1] `http://en.wikipedia.org/wiki/Likert_scale`

4.1.2 The participants

The *core participants* were recruited via an invitation through email. With core participants, we mean the people who have also participated in the IBM Game. We assigned a token to each of them to prevent that individuals complete the survey using the name of another person. Moreover, this allows us to see which participants were willing to complete the survey and which opted out.

A reminder has been sent three times(i.e. after the first, second and third week of the invitation). If the participant has not responded after the reminder, we approached them via the social network on the intranet or contacted them personally. In the invitation and reminder, we described the initial purpose of the survey and provided a personalized link for each participant to do the survey. The participants had the option to opt out if they did not want to participate in the survey.

Besides this group of people, we also decided to create the same version of the personality survey which is accessible for people who have not played the IBM game. These people were recruited via email, personal contact, posts on the forum and online communities and the internal CAS newsletter. We define this group as the *open group*. The personality insight from the open group is gathered to validate the alternative tool to measure personality traits automatically, which we will elaborate in Chapter 6.

4.1.3 Legal aspect

Since we are dealing with personality traits which are part of personal data, we needed to make sure this work accounts for the best practices concerning transparency and privacy and is compliant with the Data Protection Act[2]. In the Informed Consent Statement (ICS) presented as part of the survey in Appendix A, we specified the purpose of the personality survey, what kind of data is requested and how this data is used in our research. Additionally, we recorded a video in which we explain the purpose of our study and the information described in the ICS in person. Both ICS and video recording are included at the begin of the survey. All participants of the survey must read through the statement and accept it before they can actually participate in the survey. Moreover, the survey is only accessible via the intranet of IBM. The gathered data is stored on the server within our research department and is only accessible by our research team to prevent data from leaking out unwanted.

4.1.4 Additional features

Several additional features have been added to the online questionnaire to convince people to participate and enhance the user experience (e.g. the functionality to pause a session and continue later on).

One feature is the video recording presented to the participant at the start of the personality survey, which main purpose is to specify the ICS, but also specify that participation is crucial for our research. A second feature is a functionality to save the current session of the questionnaire and continue later on. The reason for this implementation is because people within the enterprise use to have limited *spare time* to fill in these type of questionnaire, which usually need a minimum of 15 minutes to

[2] `http://ec.europa.eu/social/BlobServlet?docId=2507&langId=en`

complete. Instead of emailing the results afterward to the participant, we have chosen to implement a feature which generates a detailed feedback report about the measured personality traits as soon as all the necessary fields are filled in. The participant can choose to read through the report once it is generated, but they can also choose to save to report as a PDF version and read it afterward. This feature is included in both personality questionnaires of the survey and an example of this report is presented in Appendix B for the interested reader.

4.1.5 Internal consistency

The internal consistency of IPIP-NEO in our survey has been measured using Cronbach's alpha[28] which provides an overall reliability coefficient for our set of questions. In our sample, the values for Cronbach's alpha are ≥ 0.746 for the traits, which indicates an acceptable level of internal consistency for our scales with this specific sample. For the interested reader, the alpha coefficients for each of the facets are presented in Table C.4 in Appendix C.

We also computed the Cronbach alpha for BSRI which resulted in $(\alpha = 0.83)$ for the masculine scale and $(\alpha = 0.79)$ for the feminine scale. These are lower though still comparable to Bem's internal reliabilities for these scale [5]. This shows that the BSRI part in our survey provides high internal consistency.

4.1.6 Results

The personality survey which consists of the IPIP and BSRI has been completed by 177 employees of IBM in the period of December 28th, 2014 to February 16th, 2015. The age of the employees varies between 19 and 61 years. The 177 employees spent in the survey a total number of 85 hours, with an average of 29 minutes and standard deviation(SD) of 14 minutes. The individual completion time was between 10 minutes and 81 minutes. In this population, 42 people are females and 135 are males.

As mentioned in section 4.1.2 we have two groups within our sample. The *core people* who also played the game and the *open group* who have not participated in the game. A total number of 212 people participated in the IBM game from whom we also have the game statistics. However only 112 people have completed the survey, 42 people opted out, 14 people never responded, 13 people would take a look at it if they had time[3], 18 people were not working at IBM anymore, 7 people only filled in partially of the survey and 6 people were out of office for a long period of time. On the other hand, we managed to reach 95 people in the *open group* (i.e people who only completed the personality survey and have not participated in the IBM game), but only 65 of them completed the whole personality survey. In the remaining part of this work, we define this data set as the ***survey data set***. This data set contains the personality profiles of IBM employees who completed the personality survey.

Demographics of the participant

To get a better overview of all the participants we divide them into three groups. Group A is the group of people who have not completed the survey but have played the IBM

[3] this has been established by contacting participants personally or via intern social media.

		Number of employees			
		Group A	Group B	Group C	Total
		n=100	n=112	n=65	n=277
Gender	Female	30	16	26	72
	Male	70	96	39	205
AgeGroup	17-24	12	1	19	32
	25-39	70	33	34	137
	40-65	18	78	12	108
Country	Argentina	0	1	0	1
	Belgium	8	9	0	17
	China	0	0	1	1
	Croatia	0	0	2	2
	Denmark	1	0	0	1
	France	1	1	1	3
	India	0	0	1	1
	Netherlands	80	87	55	222
	Peru	1	0	0	1
	Poland	0	1	0	1
	Romania	7	8	0	15
	United Kingdom	2	1	0	3
	United States	0	4	5	9
Participant is a manager	False	78	88	65	231
	True	22	24	0	46

Table 4.1: Employee demographics across the three groups.

game. Group B contains the *core group*, those who have played the IBM game and completed the personality survey. Group C represents the *open group*; the group of IBM employees who completed the personality survey but who have not played the IBM game. In the remaining part of this section we are only interested in Group B and Group c, because in both groups people have completed the personality survey.

Table 4.1 shows the gender, country and the number of managers/non-managers within the three groups. It is noticeable that 24% of the employees in these groups are female and 76% are male. However, this ratio is expected since in the Benelux the International Business Machines(IBM) employs 25% female and 75% male, and worldwide 24% of the IBM employees are females and 76% are males.

The people in these groups are spread across 12 countries and three continents. The majority of the people work in the Netherlands (80,22%) , Belgium (5,08%) , Romania (4,52%) and less than 4% from other countries in Europe and America. This is not unexpected since the IBM game and personality survey are advertised in the IBM HQ in the Netherlands. A small group of other countries like the US participated in the survey mostly due to the advertisement on internal social communities. The distribution of manager along the participants is as follow: 13,56% of the participants have a manager flag and 86,44% do not have a manager flag.

Further participation

In the last part of our survey, we asked the participants whether they would like to participate in further research. Of the 177 people 120 stated to be open and reachable to participate in further research, 55 people clearly mentioned not willing to participate in any further research, 2 people did not answer this question. A number of 113 people indicated that they liked the survey and are willing to complete the same type of surveys, while 62 people strongly indicated that they do not want to complete a similar type of surveys.

The feedback of the participants had four similarities. The first point refers to the length of the survey. Some people though that the survey was too long and took much time to complete. The second point refers to the difficulty to understand and interpret some terms used in the statements. The third point is that the survey was very interesting and provided a clear and detailed explanation[4]. The fourth common point is that the results reflected on recognizable personality characteristics in people's real life.

We understand that completing this type of personality survey might take some time, especially when people are asked to complete two inventories in a row. But we have already decided to use a shorter form of the IPIP-NEO. Instead of the 300-items one we used the 120-item IPIP-NEO which has been proven to be favorable to the properties of the longer form [44]. Other shorter forms of personality inventories provide less dimension for reliable analysis because according to literature study in Chapter 2 it would be difficult to explain the observed differences between the main traits of the Big Five Inventory. About the difficulties of words in the statements, we did not thought this population would have many problems understanding the terms, unfortunately, this appeared not to be the case. We provided two types descriptions of stereotyping, one based on personality traits and one based on gender. This is perceived well by the participants.

4.2 Analyzing gaming behaviors of IBM gamers

The dataset which represents the employees of IBM who have played the IBM game is gathered in the study done by Stanculescu et al. [78], which was given to us at the begin of our study. In that study, the dataset has been used to research how game elements can be used to support user engagement and drive online social behavior of employees in an enterprise environment. In the remaining part of this chapter, we call this data set the ***gamers data set***.

In this section, we describe this gamers dataset. First we explain the IBM game in more detail. Then we describe the gamers who played the IBM game. Afterward, we analyze how the gamers played the game and investigate whether gamers who completed the personality survey played the game differently than gamers who did not completed the personality survey. This information could be useful to explain differences in game behavior afterward.

[4]one might argue that this point is in contrast with point one, but different people participated in the personality survey. Some people liked and enjoyed the survey while other did not like it.

4.2.1 The IBM Game

This tool is an online website, accessible by all IBM employees on the intranet of the company. This tool provides four core features which can be summarized as:

1. a quiz game with questions from three categories: *IBM Facts*, *World Wide Technology* and *You & Your Network*.

   IBM Facts questions are related to IBM history, technology and facts. *World-Wide Technology* questions touch upon knowledge about information technology and technology in general. *You & Your Network* questions revolve around user's connections headlines and skills.

   The users score points by answering questions and can receive badges for reaching different milestones.

2. a social hub, allowing users to link and see their connections from social networks as *Facebook*, *LinkedIn* and *IBM Connections*. The list of connections is combined with information from the game such as score, earned badges or current position on the leaderboard.

3. a mechanism allowing users to invite their connections from social network to play the game.

4. a mechanism allowing users to share trending news articles about IBM on their social network.

However depending on the experimental group users may also have to option to view leaderboards and/or badges. This tool has been deployed for 2 months in the period of May 12th, 2014 to July 11th, 2014 to experiment with different engagement mechanics into various groups with the aim to observe which gamification mechanism are useful for engaging employees in IBM.

Different game elements were measured in the IBM game, for example the number of times a user has played the game, the number of LinkedIn connections a user has and the number IBM question a user has answered correctly. For the interested reader the complete list of game elements is presented in Table D.1.

4.2.2 The participants

The IBM game has been played by 212 people in the period of May 12th, 2014 to July 11th, 2014. They spent a total of 66 hours to play the game while individual play time varied between 1 minute and 8.32 hours. The average age of this group is 39.14 years with a standard deviation of 10.19 years. In this sample, 46 candidates are females and 166 are males, as shown in Table 4.1. Each of the participants was assigned into one of the four treatment groups. The game had three game mechanics: badges, points, and leaderboards. In each of the four treatment group a different set of game mechanics were presented. For the interested reader, the groups are described in Table D.2.

4.2.3 Game statistics

The IBM game has been played 632 times in the period of May 12th, 2014 to July 11th, 2014. The number of games per user ranged between 0 and 71. The 212 employees spent in the game a total number of 66 hours while individual play time was between 1 minute and 9.12 hours. The gamers answered a total of 2256 questions in IBM Facts, 1827 questions in World Wide Technology and 4199 social questions, invited 552 colleagues to play and shared company related news 248 times.

To get an overview and better understanding how gamers played the IBM game, we conducted a quantitative analysis of the gamers data set. We have split this data set into two groups: group 1 ($n = 100$) contains the game behaviors of gamers who did not completed the personality survey and group 2 ($n = 112$) contains the observations of those who completed the survey. For the interested reader the descriptive statistics are shown in Table D.3.

Since not every gamer has completed the personality survey, we would like to study whether group 2 is representative for the whole population of gamers. For each game element we used the Mann-Whitney Test to determine whether the two groups were significantly different from each other. However, for none of the dependent variables (i.e. game elements) significant differences were found. This indicates that group 2, gamers who completed the personality survey, maintained the characteristics of the whole gamers population. For the interested reader, a detailed test statistics are presented in Table D.4.

4.3 The Penn State population

In a study done by Johnson in [44], a large dataset of personality traits has been collected using an online personality questionnaire. This personality questionnaire is an online version of the 300-item IPIP representation of the NEO-PI-R [30]. The validity of the dataset has been tested in [43]. The research in this article addressed the validity of 23,994 cases of personality records by analyzing linguistic incompetence, inattentiveness and international misrepresentation in web-based versus paper and pencil personality measures. Since we conduct the same type of personality questionnaire (i.e. IPIP), we contacted Professor Johnson to retrieve the dataset used in the article to see if we could derive *stereotypes*, groups of people who show similar personality characteristics, to effectively get to group people. In the remaining part of this thesis, we define this dataset as the ***Penn State dataset***.

4.3.1 The participants

This dataset contains 20992 protocols produced by individuals who have completed a web version of the 300-item IPIP-NEO. Reported gender of the participants is 7743 males and 13249 females. The age varied from 10 to 99 years, with a mean age of 26.2 and SD of 10.8 years. The participants discovered the website at their own without actively recruiting.

This population is considered as a broader population in which we would like to find stereotypes(i.e. people with a similar set of personality characteristics). The idea is to find a way to effectively grouping people with similar personality characteristics.

4.3.2 Facets determination

Since we are dealing with a 5 dimension and 30 facets of personality characteristic it would be pleasant to first investigate which facets have the most variability, less skewness and highest difference between male and female participants. The main reason for checking this is to reduce the number of traits that we would like to investigate. This could be useful to find personality traits which are able to explain more about the personality of the population in a later stage of our research. Moreover, it does not make sense to look for correlations or for a significant effect on variables that have very little variance.

First, we do an exploratory analysis on the Penn State data set to get more insights about how people in this population score for each personality trait. Based on this analysis we take the top 10 traits that have the highest standard deviation, the highest variance which are: *Cautious*, *Sympathy*, *Liberalism*, *Cheerful*, *Vulnerability*, *Self-discipline*, *Modesty*, *Intellect*, *Excitement-seeking*, *Self-Conscientiousness* and *Immoderation*.

Second, we analyze which of the above mentioned 10 traits correlate with the sex of people, because this could provide insights which personality traits can explain more differences between female and male. In addition, we also checked which traits has the highest skewness for females and males. These traits are not very useful to analyze because these traits do not have much variability and are almost a *constant value*. So, we removed the traits with the highest skewness from the top 10 list mentioned above. Based on this analysis, we picked the top five traits with the highest variability and lowest skewness out of ten mentioned above.

The results are presented in Table 4.2. The first column presents the top five traits for males that are most skewed. The second column presented the most skewed traits for females. The last column presents the top five traits along which male and female in the Penn State data set vary the most. In the Pen state population is seems that the trait *Openness* does not show much variance, compared to the other four traits. Two facets with the highest variability and less skewness between male and female belong to the trait *Neuroticism*. This information could be meaningful in a later stage when we analyze personalities traits of people who completed our survey or when we want to explain how gender is related to these traits.

	Most skewness for Male	Most skewness for Female	Highest variability[a]
1	Intellect(O)	Cheerful(E)	Modesty(A)
2	Cheerful(E)	Achievement-Striving(C)	Self-Consciousness(N)
3	Vulnerability(N)	Intellect(O)	Immoderation(N)
4	Achievement-Striving(C)	Sympathy(A)	Excitement-seeking(E)
5	Adventurousness(O)	Cooperativeness(A)	Activity-level(E)

[a]: the top five traits with most variability and less skewness between male and female.
facet(X) means that this facet belong to one of the Big Five dimensions: **O**penness, **C**onscientiousness, **E**xtroversion, **A**greeableness or **N**euroticism.

Table 4.2: Facets determination of Penn State dataset

4.3.3 Finding stereotypes for clustering

In our work stereotypes can be defined groups of people who show similar personality characteristics. We use clustering to see which dimensions of personality show more variability in the Penn State population and associate those dimensions with stereotypes of people like in demographic types. In this section, we provide five different methods to cluster people of the Penn State data set. Given the Penn State data set, then a record can be defined as an individual case in the dataset, for instance, a participant with the calculated scores for each personality trait can be seen as one record.

Kohonen/Self-Organizing Map

This technique is based on neural networks that apply a specific type of competitive learning to cluster data.

The idea behind this clustering method is that there are basic units, called *neurons* which are organized into two layers: the *input layer* and the *output layer*. All of the input neurons are connected to all of the output neurons, and each of these connections is associated with *strengths* or *weights*. The output neuron with the strongest response is said to be the *winner* and is the answer for that input. When an output neuron wins, its weights are adjusted along its neighborhood and form a two-dimensional map of the clusters. This process is repeated many times until for each input node the winning output node is found. At the end when the whole network is fully trained, records that are similar will be close together on the output map whereas records which differ will be far apart.

Kohonen networks attempt to find relationships and overall structure in the data. The output from a Kohonen network is a set of (X,Y) coordinates, which can be used to visualize groupings of records and can be combined to create a cluster membership code.

Applying this technique to the Penn State dataset provided us poor clustering quality. Unfortunately, this indicates that the findings are not very meaningful due to the number of clusters. Figure 4.1 shows the results of clustering. The left part shows a pie chart that contains 35 possible clusters, each color stands for a different cluster. X and Y are coordinates pointing to the output from the Kohonen Network. The right part of the figure presents the model summary view that shows a snapshot of the cluster model, including the quality of the cluster. In this case figure 4.1 shows that the cluster results are of poor quality. This means that no strong evidence is found for the 35 cluster structures. The records are not distinct from each other and are hardly distinguishable in term of personality explanation.

TwoStep

The next methods we used for clustering is TwoStep clustering. TwoStep clustering consists of two steps. In the first step of the procedure, it scans the records one by one and decides if the current record should merge with the previously formed clusters or start a new cluster based on the distance criterion. In the second step, it clusters the sub-clusters from the pre-cluster step using a hierarchical clustering method which does not require a number of clusters to be selected ahead of time. Unfortunately, this

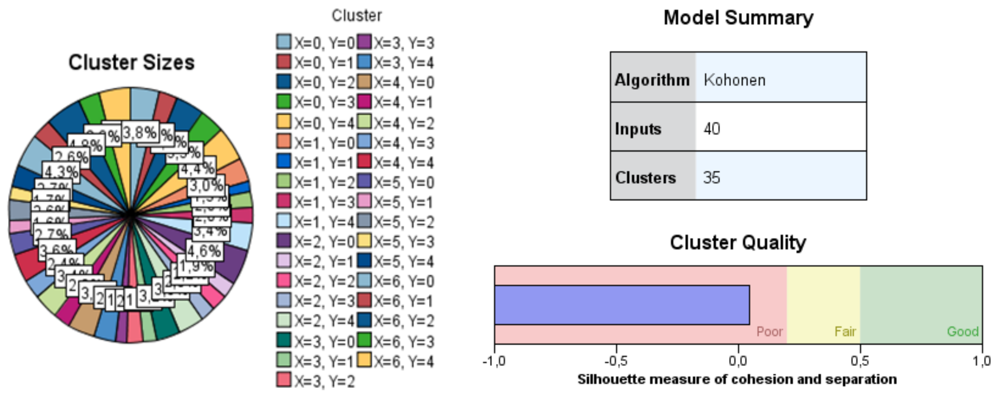


Figure 4.1: Results of Kohonen clustering.

technique also provided clusters of poor quality. The proposed clusters are difficult to distinguish from each other in terms of personality differences.

K-means

This technique aims to partition n observations into k clusters in which each record is assigned to the cluster with the nearest mean, serving as a prototype of the cluster. After all, cases have been assigned, the clusters are updated to reflect the new set of records assigned to each cluster. The records are then checked again to see whether they should be reassigned to a different cluster. This process iterates continue until either the maximum number of iterations is reached, or the change between one iteration and the next fails to exceed a specified threshold. Using this technique different personality profiles were created for each cluster but the differences between clusters were not significant. Unfortunately, the cluster results were also of poor quality, similar to the *Kohonen/Self-Organizing Map*. This indicates that the created clusters are very similar and thus makes it difficult to have a clear distinction between the groups. Hence, it is not very useful.

Density-based spatial clustering of applications with noise

The Density-based spatial clustering of applications with noise(DBSCAN) algorithm can identify clusters in large spatial data sets by looking at the local density of database elements. It captures the insight that clusters are dense groups of points. The idea behind this technique is that if a particular point belongs to a cluster, it should be near to lots of other points in that cluster. Applying this technique to the Penn State dataset provided us the results that a the majority of the population belongs to the random noise instead of to a clear cluster, which is also shown in Figure 4.2.

Agglomerative hierarchical clustering

In this approach, each record starts with its own cluster and pairs of clusters are merged as one moves up the hierarchy. A measure of dissimilarity between a set of personality traits decides whether a cluster should be combined or split. Using N clusters, and from

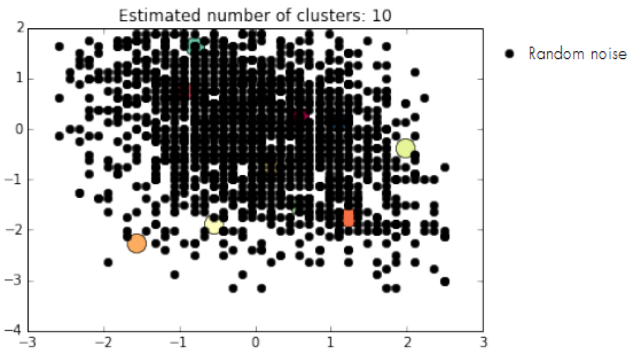
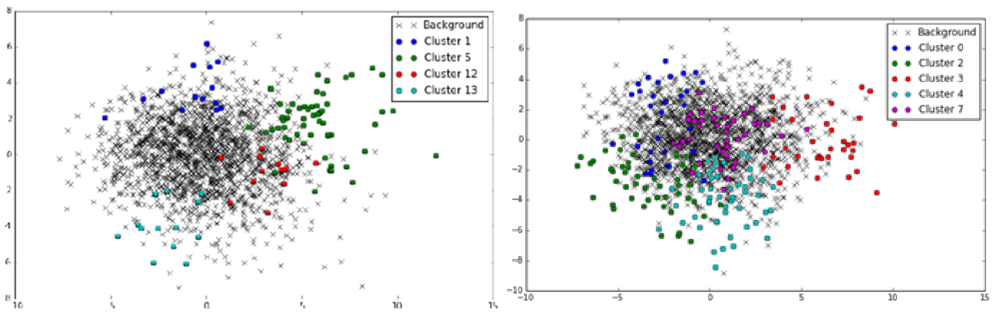


Figure 4.2: Results of DBSCAN.

those clusters, manually select clusters which had more than 10 elements to present one group of participants. The plot of the clusters is projecting the agglomerated clusters on the first two components, which are the two largest components of the Principal Component Analysis (PCA) on the dataset.

We tested this approach on a random sample size of 2000 in the Penn State dataset to see if the results were more promising than previous ones. In our first try, about 12.4% of the data points were clustered, result shown in Figure 4.3a.



(a) First attempt (b) Second attempt

Figure 4.3: PCA view of group clustering.

In a second attempt, we used another random sample of 2000 records to see if we could get more promising results. The result was that in this sample only 9.8% of the records were clustered, which is less promising (see Figure 4.3b). So we tried it four more times on another random sample of 2000 data points to see what the results were. During this third attempt on average 14.3 % of the different samples were clustered, which provided more promising results than previous times.

The results of this attempt are shown in Figure 4.4. After several attempts of clustering smaller samples of the whole population we may conclude that there is some kind of consistency in the output, therefore we would like to run this clustering method one more time.

We decided to go for a final run to apply this clustering method to the whole population of the Pen state dataset to observe whether stereotypes could be found in the Penn State population. The result shows that using this approach, it is possible to clus-

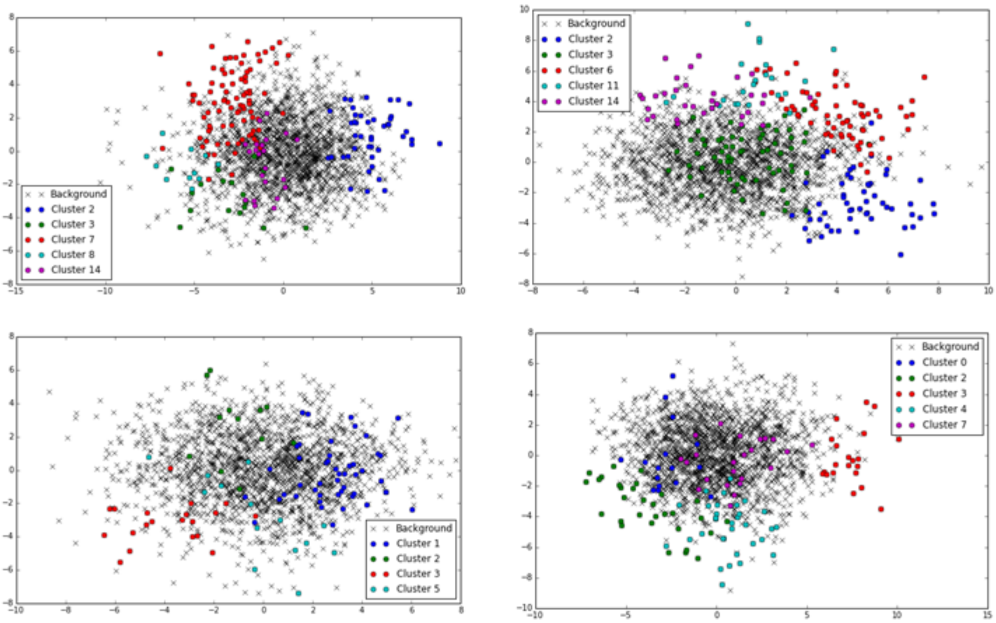


Figure 4.4: PCA view of group clustering, attempt 3.

ter on average 14.6% of the population. Around 10% upper bound to be classified into the cluster, even after different iterations.

So we may conclude that currently we cannot provide evidence to derive clear stereotypes from the Penn State dataset to effectively group people together. Whereas more research is needed to find effective ways to cluster people based on similar personality characteristics.

4.4 The IBM Benelux population

This dataset represents the general population of IBM Benelux. It contains several descriptive information from all employees working in IBM Benelux, like gender, age, department, salary scale, and years of service. This dataset is retrieved via Human Resource of IBM. In addition, this dataset is used to observe the general characteristics of an IBM employee representing IBM Benelux in the generalization study, presented in Chapter 8. Due to privacy concern, the content of this dataset was not allowed to be distributed externally. In the remaining part of this work we call this data set ***IBM Benelux***, representing the general IBM population.

4.5 Collecting Social Media data

In this section we describe how we collected data from social media platform Twitter[5] and IBM Connections[6]. These data are needed to test the automatic personality assessment tool in a later stage, described in Chapter 6. We decided to crawl content

[5] `http://www.twitter.com`
[6] `http://www.ibm.com/software/products/nl/conn`

of social media because this is a place where users present themselves to the world, revealing personal details and insights into their lives. Users may use social media to express their feelings, thoughts and share their interest.

Twitter[7] was our first option because most tweets are publicly available. Moreover, it has been shown that tweets can be used to predict a user's personality can be accurately[29]. A tweet is a very short message (i.e. maximal 140 characters) posted on the Twitter website. The message may include text, keywords, mentions of specific users, links to websites, and links to images or videos on a website.

Our second option is IBM Connections[8], this is a social network platform accessible via the intranet of IBM. Since this study is done in IBM and related to the IBM game, we thought it would be a good opportunity to explore IBM Connections.

4.5.1 Twitter crawling

In order to crawl tweets from the IBM gamers, we first need to find them on Twitter and see if they have a Twitter account. Unfortunately, this needs to be done manually by searching for a gamers' name and comparing the profile pictures and profile description. After finding a user account which matches with an IBM gamer, we can map these findings with each other manually.

We crawled all available tweets from an IBM gamer using the Twitter API[9]. We specifically crawl tweets from a user's profile timeline, because we are only interested in tweets which may show insights into their personality. Tweets that are only reposted are not that interesting because it is not phrased by the user itself.

After mapping a Twitter user account to an IBM gamer, we implemented a script which crawls all the tweets for a given username. Once we collected all the names of the Twitter accounts, we wrote another script to crawl tweets in a batch based on the usernames, instead of crawling tweets from a single user at the time.

Subsequently, we collected all the tweets for a given user, aggregated the tweets and applied a basic filtering to cleanup the data. This filtered data serves as input for determining a user's personality profile using Watson Personality Insight service.

The idea is to get insights about personality traits based on the tweets input and compare the personality profile with the profile from the user's personality survey to see if differences in personality traits scoring occur. If the observed difference is significant we may conclude that the Watson Personality Insight service does not provide reliable insights into personality traits of users based on a given set of tweets.

The Twitter dataset

In total we retrieved 149 user profiles of Twitter, 38 of them were from people who played the IBM game but did not complete the survey, 68 profiles are from people who played the game and also completed the survey, 43 profiles are from people who completed the survey but did not play the game. Results presented in figure 4.5.

Using a wordcounter and text analyzer we did some analysis on the gathered tweets. The size of a collection of tweets varies between 2 KB and 29KB. The reported

[7] http://www.twitter.com

[8] http://www-03.ibm.com/software/products/nl/conn

[9] https://dev.twitter.com/cards/getting-started#crawling

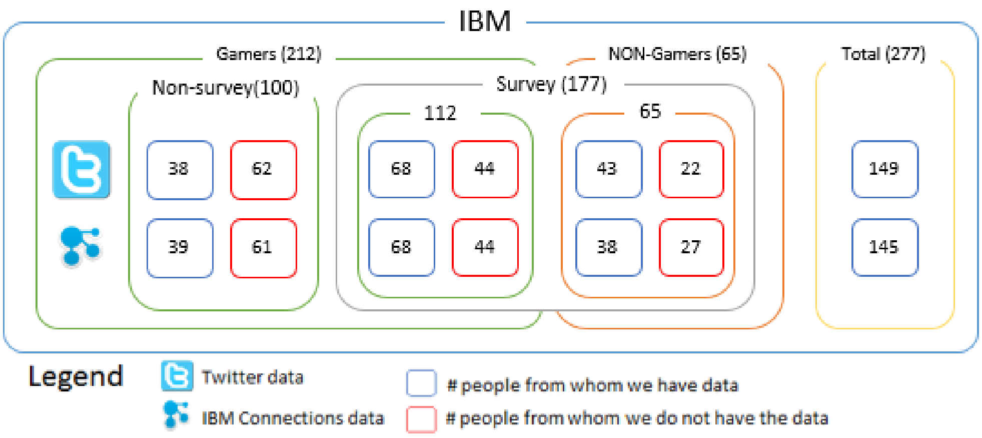


Figure 4.5: Overview of Twitter and IBM Connections datasets.

word count range is 171 word till 3166 words. The top 10 frequent used words[10] are: *"social", "connect", "happy", "data", "ibm", "cloud", "future", "service", "smarter", "project"*.

4.5.2 IBM Connections crawling

Crawling content from the internal social media platform IBM Connections is more difficult than crawling from Twitter. The main reason is because the content of IBM Connections is only available via the intranet and due to privacy reason there are a limited amount of data we can collect. Fortunately, several methods exist to crawl social content from IBM Connections. One of the options is to use the SaND Streams[11] application, which is a novel application that uses a faceted search approach to provide employees with advanced capabilities of search, navigation, attention management, and other types of analytics on top of an enterprise activity stream. Another option is to use IBM Connections API[12].

The idea is to use an IBM gamers' email address as a unique identifier to map an IBM Connections account to an IBM gamer. We crawl activities, blog post and status updates from a user if they exist. Because these content may provide insights into their emotion expressions.

Afterward, the gathered content can be inserted into the Watson PI service to get insight about personality traits based on the input crawled from IBM Connections, and compare the personality profile with the profile from our personality survey to see if differences in personality traits scoring occur. If the observed difference is not significant we may conclude that the Watson Personality Insight service is reliable given the set of social content crawled from Connections.

[10] with words we mean words in the tweets as well as used in hashtags (i.e #) in a tweet.

[11] `https://www.research.ibm.com/haifa/projects/imt/social/sand_streams.shtml`

[12] `http://www-10.lotus.com/ldd/lcwiki.nsf/xpAPIViewer.xsplookupName=IBM+Connections+5.0+API+Documentation#action=openDocument&res_title=IBM_Connections_APIs_ic50&content=apicontent`

The IBM Connections dataset

A total number of 145 user profiles where retrieved from IBM Connections, 39 user profiles come from users who have played the IBM game but have not completed the survey, 68 profiles are from users who played the game and also completed the survey and 38 profiles are from participants of the survey who have not played the IBM game. This is also shown in figure 4.5.

Using a wordcounter and text analyzer we analyzed this dataset. The size of a user's Connections data varies between 3 KB and 1655KB. The reported word count range is 182 word till 4325 words. The top 10 frequent used words are: *"connect", "partner", "cloud", "ibm", "project", "young", "network", "woman", "smarter", "community"*.

4.6 Chapter conclusions

This chapter focused on describing and analyzing the datasets which we need to conduct our study. An overview of all the gathered dataset is shown in Figure 4.6. To summarize, we have three major data sets; ***IBM***, ***IBM Benelux*** and ***Penn State***. However, the IBM data set contains the sub data sets: *Survey*, *Gamers*, *Twitter* and *IBM Connections*.
The experiments which need the data described in this chapter will come along in the remaining part of this thesis.

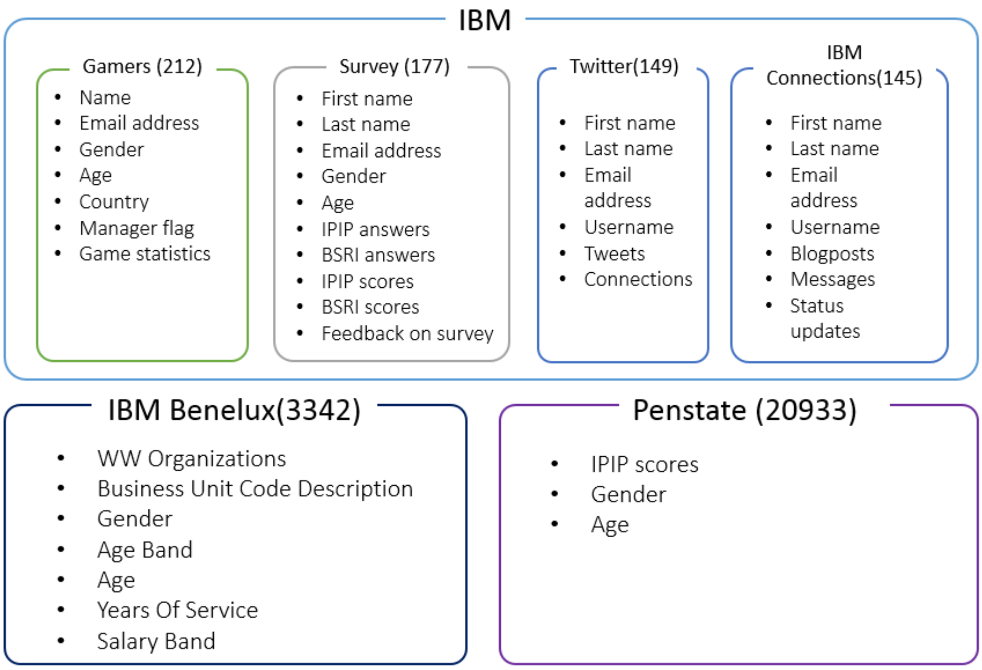


Figure 4.6: Overview of all gathered datasets.

Chapter 5

Analysis of personality traits and characteristics

This chapter describes the analysis of the personality traits of the people who have completed our personality survey. The purpose is to have a proper understanding of different personality traits of people who have played the game and to get insight about how specific our sample is compared to the Penn State population. In Section 5.1 we analyze the scores of the personality traits in the Penn State data set. In Section 5.2 we first determine which facets have the most variability and are less skewed, after which we analyze the scores determined by the IPIP questionnaire and then we do the same for the scores of the BSRI. In addition, in Section 5.3, we do not only look at personality traits, but we also take gender and age into account in order to get more insights in the specificness of our sample.

Since the assumptions of normality and homogeneity in our data were violated, we decided to use non-parametric tests throughout our analysis, unless specified otherwise. For the interested reader, the results of assumption tests for the IBM population are presented in Table C.1 and for the Penn State population are presented in Table C.2.

5.1 Personality traits of Penn State population

According to professor Johnson[45], personality scores cannot be interpreted in isolation. The scores from personality tests are only if it is compared them to a reference group, which usually contains people of the same gender and approximately the same age. In our study, we used the Penn State dataset (described in Section 4.3) as a reference group to compare to our IBM sample. We analyze the scores of the personality traits of the Penn State population and compare them with our sample. For each Big Five personality trait, we analyzed the personality score (see Table 5.1). For the interested reader the scores of the related individual facets are presented in Table C.3 of Appendix C.

As can be seen in Table 5.1) shows that women score average higher on the Big Five personality traits than men, and this is especially the case for *Neuroticism* and *Agreeableness*. This results replicated findings of previous research[13, 83, 74].

Trait	Total Mean n = 20993	TotalSD	MaleMean n=7743	MaleSD	FemaleMean n=13249	FemaleSD
Openness	3.64	0.43	3.59	0.42	3.70	0.45
Conscientiousness	3.56	0.43	3.52	0.41	3.60	0.44
Extraversion	3.36	0.26	3.33	0.29	3.39	0.26
Agreeableness	3.69	0.39	3.55	0.37	3.85	0.41
Neuroticism	2.60	0.24	2.48	0.28	2.73	0.24

Table 5.1: The personality scores of the Big Five personality traits

5.2 Personality traits of IBM sample

This section describes the personality traits scores of the group who have participated in our personality survey. First, we determined which facets in our sample have the most variability and less skewness. Second, we explain the observations in terms of the IPIP-NEO questionnaire. Third, we discuss the results on the BSRI scores of this sample. Additionally, we compare our observations with findings from previous research to get a better understanding of our sample.

5.2.1 Facet determination of the IBM population

Similar to Section 4.3.2, we investigate which facets have to most variability, less skewness and largest difference between male and female participants. This information provides insight in which personality traits are useful to explain personality differences between males and females. Moreover, we also investigate which traits are the most skewed and therefore less useful to explain differences in personality between men and women.

The 10 facets with the highest variability in our IBM sample are, *Anxiety, Orderliness, Anger, Gregariousness, Self-Consciousness, Deliberation, Depression, Modesty, Self Discipline* and *Impulsiveness*. None of these facets (with the highest variability) are components of the trait *Openness*, what indicates that this trait might not be very useful to analyze at a later stage. Only one facet (*modesty*) belongs to the trait *Agreeableness*, as well as *Gregariousness* to the trait *Extroversion*. Three facets (*Orderliness, Deliberation and Self Discipline*) are components of to the trait *Conscientiousness* and five facets (*Anxiety, Anger, Self-Consciousness, Depression and Impulsiveness*) are components of to the trait *Neuroticism*. This indicates that the dimensions *Conscientiousness* and *Neuroticism* are useful to analyze their relationship with gameplay because these traits have the most variance.

The five traits with the highest skewness for males and females are presented in the first and the second column of Table 5.2, while the five traits with the highest variability are presented in the last column of Table 5.2.
Comparing these results with the facet determination described in Section 4.3.2, we observe that the facet *Self-Consciousness* which is a component of the trait *Neuroticism* shows the most variability in both populations while facets belonging to the trait *Agreeableness* seem to be very skewed for both males and females.

Self-Consciousness reflects how sensitive individuals are to what others think of them. Their concerns about rejection and ridicule cause them to feel shy and uncom-

	Most skewness for male	Most skewness for female	Highest variability
1.	Achievement Striving(C)	Straightforwardness(A)	Anxiety(N)
2.	Depression(N)	Dutifulness(C)	Self-Consciousness(N)
3.	Modesty(A)	Compliance(A)	Impulsiveness(N)
4.	Trust(A)	Assertiveness(E)	Gregariousness(E)
5.	Anger(N)	Deliberation(C)	Orderliness(C)

Note. the five traits with most variability and less skewness between male and female.
Note. facet(X) means that this facets belong to one of the Big Five dimensions: **O**penness, **C**onscientiousness, **E**xtroversion, **A**greeableness or **N**euroticism.

Table 5.2: Facet determination of the IBM sample.

fortable around others. They are easily embarrassed and often feel ashamed. Their fear that others will criticize them or make fun of them are exaggerated and unrealistic, and their awkwardness and discomfort may make these fears a self-fulfilling prophecy. Low scorers, in contrast, do not suffer from the impression that everyone is watching and judging them. They do not feel nervous in social situations.

Agreeableness reflects individual differences in concern with cooperation and social harmony. In our sample, people are more negatively skewed indicating that people are more agreeable. Agreeable individuals value getting along with others. They are therefore considerate, friendly, generous, helpful, and willing to compromise their interests with others. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy. This might be the reason why the IBM population completed the personality survey, but more research is needed to investigate this.

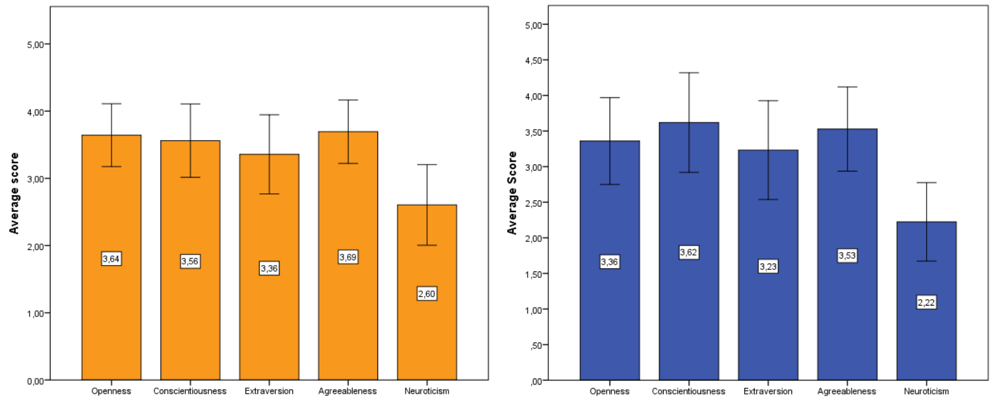
5.2.2 Big Five personality scores

As mentioned before the IPIP-NEO measures a user's personality traits in terms of the Big Five personality traits and the facets belonging to each trait.

Taking a closer look at the scores from our sample of the IBM population, we observe that this population scores around the average scores of the Penn State population for all traits, except for *Neuroticism* which is shown in Figure 5.1. To verify this observation we came up with the following hypothesis:

Hypothesis 1. *Our IBM sample is identical to the Penn State population in terms of the Big Five traits.*

We used the Mann-Whitney-Wilcoxon test to examine this hypothesis. We conducted five pairwise non-parametric tests on the scores of one trait for our IBM sample and Penn State population. The results from the Mann-Whitney-Wilcoxon tests showed that no significant differences were found between traits in the two populations, except for the trait *Neuroticism* which had a p-value of .0134; therefore we rejected the null hypothesis that the distributions of both groups are identical. This means that the our IBM sample scores significantly lower on *Neuroticism* ($p = 0.0134$), compared to the Penn State population. Hypothesis 1 is therefore rejected. This signifies that people in IBM are exceptionally calm, composed and unflappable. Moreover, they are less likely to react with intense emotions, even to situations that most people would describe as stressful.



(a) For Penstate data set.

(b) For our sample population of IBM.

Figure 5.1: Average scores on each personality factor shown with standard deviation bars.

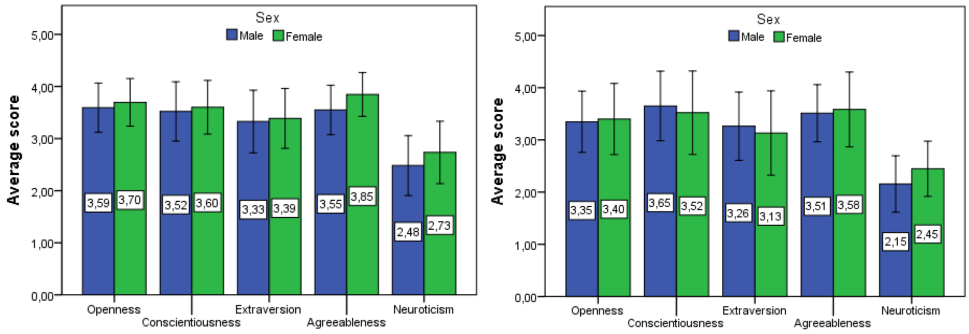
In a previous research, it is found that *Neuroticism* is related to different uses of the Internet[36]. The fact that the IBM sample scores low on *Neuroticism* could provide insights in how they played the IBM game. In addition, research[73] has also shown that individuals who score high in *Neuroticism* are more likely than emotionally stable individuals to use the Facebook Wall[1]. This might be the reason why people do not share information in the IBM game. Further research is needed to confirm this.

Sex differences in the Big Five personality traits

Sex differences in personality traits are often characterized in terms of which sex has higher scores on average on which trait. Previous studies[8, 13, 83] have shown that women reported higher scores for *Extraversion, Agreeableness*, and *Neuroticism* than men. Therefore, we analyzed the differences in personality scores between females and males in our sample.

Figure 5.2b shows the distribution of personality scores across sexes. We observed that in our sample females score higher on *Agreeableness* and *Neuroticism*. However, the results of Mann-Whitney U tests (presented in Table 5.3) have proven that only the scores for *Neuroticism* were significantly higher for females $(M = 2.45, SD = .53)$ than for males$(M = 2,15, SD = .54)$, $U = 1870.500, p = .001$. The p-values indicate no significant differences depending on sex in the other four traits, which is unsurprising because previous researches have only found that woman score significantly higher than men on the traits: *Extraversion, Agreeableness,* and *Neuroticism.*
The average scores of male and female on the level of facets of the Big Five personality traits are presented in Table C.4 in Appendix C. Significant differences between females and males were found on the following facets: *Aesthetics, Feelings, Actions, Assertiveness, Altruism, Compliance, Modesty, Sympathy, Anxiety, Depression, Self Conscientiousness, Impulsiveness* and *Vulnerability*. Remarkable is that for the facet

[1]Each Facebook user has a Wall that their friends can use to write messages or post links for the user to see. Communication on the Wall is asynchronous, and the posted information is generally viewable for other Facebook users



(a) Penn State population. (b) Sample population IBM.

Figure 5.2: Average scores on each personality factor shown with standard deviation bars clustered by sex.

	Mann-Whitney-U result	
Openness	$U = 2675.00$	$p = .583$
Conscientiousness	$U = 2667.00$	$p = .562$
Extroversion	$U = 2621.50$	$p = .462$
Agreeableness	$U = 2469.00$	$p = .072$
Neuroticism	$U = 1870.50$	$p = .001$

Table 5.3: Mann-Whitney-U significance test for Big 5 traits for the IBM sample.

cheerfulness no significant differences were found between males and females in our sample, because research[76, 56] have shown that woman are more cheerful than men. This indicates that men and women in our sample are experiencing the same range of positive feelings, including happiness, enthusiasm, optimism, and joy.

In summary, we found that our population scores significantly lower on *Neuroticism* than the Penn State population. Furthermore, we identified that the biological sex differences had a significant impact on the scores for some of the personality traits and facets which replicated the findings of previous studies. These findings implicated that in our sample, people are more calm, composed and capable of handling stressful situations compared to other people from the same age group and nationality. Moreover, the women in our sample are more nurturing, and to a greater extent tender-minded, and altruistic than men.

5.2.3 BSRI scores

The Bem Sex-Role Inventory (BSRI) is a measure of masculinity-femininity and gender roles. It assesses how people identify themselves psychologically. In our personality survey, we used the original 60-item version of the BSRI.

In the literature, there are multiple methods for classifying people into gender roles. The median split method was used here to classify the gender roles of these participants, because this method is most commonly used and it avoids methodological issues that occur when other approaches are used [7].

First, we summed the scores of the questions in the questionnaire referring to the masculine scale and then those referring to the feminine scale. Then we took the median (masculinity = 4.90; femininity = 4.70) of our IBM sample to split between *Androgynous*, *Undifferentiated*, *Feminine* and *Masculine*. the scores of each participant were compared to the median. Mean scores that fell on the median were classified as "*high*" rather than "*low*" scores. If the participant's mean score was below the median in both the feminine and masculine scale, this person was classified as *Undifferentiated*. If the participant's mean scores on both the masculine and feminine scale were equal to or above the median that participant was classified as *Androgynous*. Those people who scored equal to or higher than the median on the feminine scale and lower on the masculine scale were classified as *Feminine*. Finally, those who scored equal to or higher than the median on the masculine scale and lower on the feminine scale were classified as *Masculine*. The distribution of participants into these four groups is presented in Table 5.4.

Gender roles	**Female sex (%)**	**Male sex (%)**	**Total (%)**
Undifferentiated	7 (17)	15 (11)	22 (12)
Masculine	4 (9)	43 (32)	47 (27)
Feminine	11 (26)	17 (13)	28 (16)
Androgynous	20 (48)	60 (44)	80 (45)
Total	42(100)	135 (100)	177 (100)

Table 5.4: Gender roles across biological sexes.

As can be seen in Table 5.4, $27\%(n = 11)$ were classified as *Feminine*, compared to $32\%(n = 43)$ of the men being categorized as *Masculine*. Interestingly, 48% of the females (n = 20) and 44% of the males (n = 60) were classified as *Androgynous*, which means that individual score high on both the feminine and masculine scales. In addition, more males (11%, n = 15) than females(17%, n = 7) were classified as *Undifferentiated*. Furthermore, it is interesting that a small group (12.6%, n = 17) of male participants are categorized as *Feminine*. This indicates that these men see themselves possessing more feminine traits than masculine traits. The other way around, 9.5% (n = 4) of the females were categorized as *Masculine*.

People who are classified as *Androgynous* have an adjusted character that incorporates the characteristics of both sexual orientations. They disregard what traits are culturally constructed specifically for males and females within a specific society, and rather focus on what behavior is most effective within the situational circumstance.

In contrast to our findings, research showed that in an older Brazilian population[10], the majority(34%) of the people were classified as *Undifferentiated*. Indicating that the people in the older Brazilian population show behaviors or characteristics that society associates with the male gender role or with the female gender role. Only 32% of the people in the Brazilian population were classified as *Androgynous*. The contrast shown here might occur because of cultural differences.

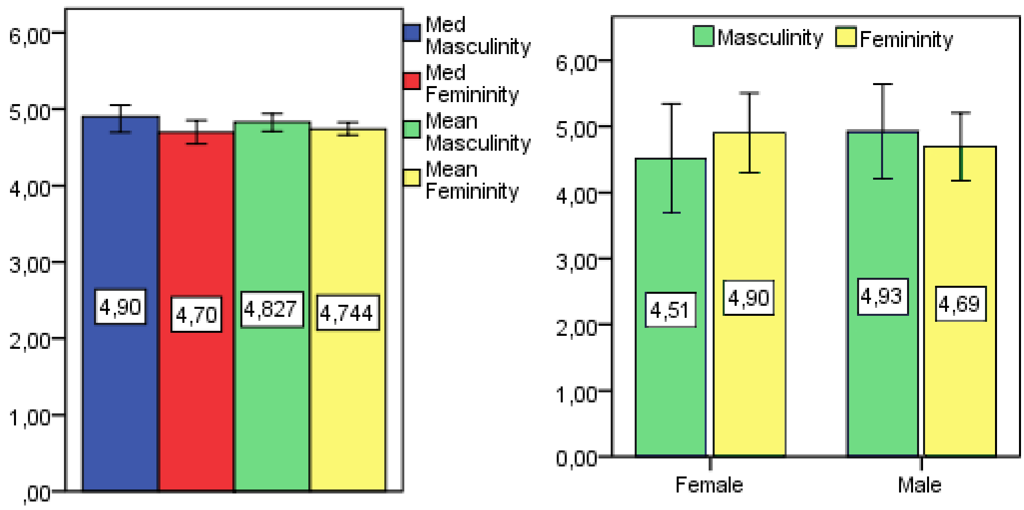
Gender differences in BSRI scores

To get a better understanding whether males and females differ in their self-perception of gender role, we conducted a statistical test to examine these differences. The group statistics of our population are shown in Table 5.5.

Gender role	Bio. Sex	N	Mean	Std. Deviation	Std. Error Mean
Masculinity	female	42	4.51	0.83	0.13
	male	135	4.93	0.72	0.06
Femininity	female	42	4.90	0.60	0.09
	male	135	4.69	0.51	0.04

Table 5.5: Group statistics BSRI for females and males.

When analyzing the participants' scores based on gender, the mean masculinity score for the male participants (on a Likert-scale of 1 to 7 with 7 being the highest degree of masculine possible) is 4.93 with a standard deviation of 0.71 (Table 5.5). The mean femininity score for the male participants (with 7 being the highest degree of femininity possible) is 4.51 with a standard deviation of 0.83. This can be contrasted with the scores for the female participant. The mean masculinity score for the female participants is 4.69 with a standard deviation of 0.51. The mean femininity score for the female participants is 4.90 with a standard deviation of 0.60.



(a) Average scores en median values.

(b) Average scores across sexes.

Figure 5.3: BSRI scores of IBM population.

We defined the following two hypothesis to test the self-perception of gender roles in our sample:

Hypothesis 2. *Females score higher on the feminine scale than males.*

Hypothesis 3. *Males score higher on the masculine scale than females.*

		Sex		Total
		female	male	
masculinity level	low	27	32	59
	high	15	103	118
Total		42	135	177

Table 5.6: Masculinity level across sexes

		Sex		Total
		female	male	
femininity level	low	11	72	86
	high	31	63	91
Total		42	135	177

Table 5.7: Femininity level across sexes

Since the assumption of normality and homogeneity of the BSRI data were not violated (presented in Table C.1), we used parametric-tests in this analysis. Two independent samples t-test were conducted to examine whether there was a significant difference between females and males on the *femininity* and *masculinity* scores. Comparisons between men and women showed a statistically significant difference on the masculinity scale of the BSRI, $t(175) = 3.15, p < .05$, and a significant difference on the femininity scale, $t(175) = -2.21, p < .05$. This means that women (M = 4.90, SD = .60) scored significantly higher on femininity than men $(M = 4.70, SD = .51)$. Similarly, men $(M = 4.93, SD = .72)$ scored significantly higher on masculinity than females $(M = 4.51, SD = .83)$. Based on these results we accepted hypothesis 3 and 4.

Furthermore, Table 5.6 shows that social stereotypes can still be seen for men: of the 135 men, 103 (76.30%) are in the *high-masculinity* (i.e. category *Masculine* and *Androgynous*) category, compared to only 32 (23.70%) in the *low-masculinity* (i.e. category *Feminine* and *Undifferentiated*) category. The same holds for women possessing feminine traits presented in Table 5.7. A total number of 31 females (73.80%) were in the *high-feminine category*, while 11 (26%) females are in the *low-feminine* category.

To evaluate the relationship between the level of masculinity/femininity and biological sex, the data was analyzed using the *Chi* square goodness of fit test with a significance level of 0.05.

Hypothesis 4. *Male participants score high on Masculinity while females score low on Masculinity.*

For this hypothesis, the null hypothesis that the relationship exists between biological sex and level of masculinity was rejected, $\chi^2(1) = 5.432, p < .05$. Hypothesis 5 is accepted. So, there is a statistically significant relationship between sex and masculinity. This means that males and females did not score similarly on Masculinity, indicating that both sexes had a different way to identify themselves in a masculine role.

Hypothesis 5. *Female participants score high on Femininity while males score low on Femininity.*

For this hypothesis, the null hypothesis that the relationship between biological sex and level of femininity is due to chance was not rejected, $\chi^2(1) = 0.156, p > .05$. No statistically significant relationship between sex and *Femininity* were found for our IBM sample. So, females and males scored similarly on *Femininity*, indicating that both sexes had a similar way to identify themselves in a feminine role.

In summary, gender role classification categorized our population into four types: *Masculine, Feminine, Androgynous* and *Undifferentiated*. The majority of our IBM sample were classified as *Androgynous*. Additionally, we have shown that male participants are described as in scoring *high* on masculine traits. Contrary to our predictions, gender role classification did not reflect biological sexes. In fact, we observed that a higher percentage of both males and females were classified as either *Androgynous* or *Undifferentiated* instead of the traditional gender roles of *Masculine* and *Feminine*. This indicates that the people in our sample are more likely to focus on what behavior is most effective within the situational circumstances rather than adhere to culturally constructed behavior for men and women. Furthermore, we found no differences in the way the two sexes associate themselves with feminine traits. This indicates that females and gender roles are different entities in this population.

5.3 Representativeness of our sample

We would like to know how well our sample reflects the population with regard to participants distribution of age and gender. Using the frequencies of our sample set a One-Sample Chi-Squared Test was executed. The analysis involves the weighted cases method to determine the extent to which our sample ($n = 277$) generalizes to the general IBM population ($n = 3342$) in terms of the distribution of participants across the age categories (1) and sex (2). The same method is applied to explore the generalization of our sample to the Penn State population ($n = 20933$).

5.3.1 IBM population

Data reported by IBM indicated that among the general IBM population 5.02% of the employees are in the age of 17-24 years (category 1), 40.21% are in the age of 25-39 years (category 2), 54.78% are in the age of 40-65 years (category 3). To determine whether our sample differed significantly from those of the general IBM population, we analyzed the age categories of our sample. In our sample of 277 employees, 32 (11.55%) fall into age category 1, 137 (49.46%) fall into age category 2, and 108 (38.99%) fall into category 3. A Chi-squared goodness-of-fit test based on the values presented in Table 5.8 shows that the age category in our sample differed significantly from the general IBM population, $\chi^2(2, N = 277) = 30.903, p < .001$. Thus, the distribution across age categories in our sample is not the same as in the general IBM population. This means that our sample is not representative for the general IBM population in term of age categories.

Age category	Age range	Observed N	Expected N	Residual
1	17 - 24	32	2,8	29,2
2	25 - 39	137	47,7	89,3
3	40 - 65	108	226,5	-118,5
Total		277		

Table 5.8: Calculated frequency for age category

To investigate whether our sample is representative of the general IBM population in term of sex distribution, we conducted a One-Sample-Chi-Squared-Test. Result shows that our test statistic is not statistically significant: $\chi^2(1) = 1.401, p = 0.237$. This indicates that the null hypothesis that the proportions in our sample are equal to the proportions in the IBM population is accepted. Thus, we may conclude that our sample proportions are not significantly different from the IBM population proportions across sex. The distribution across sex in our sample is the same as in the IBM population. Thus, our sample is representative of the general IBM population in terms of sex.

In summary, our population is not a representative sample of the IBM population in terms of age. However, no statistical differences were found between the average age of the two groups.On the other hand, our sample is a representative sample of the IBM population in terms of sex distribution.

5.3.2 The Penn State population

The same approach has been applied to determine whether our sample is representative of the Penn State population. The calculated frequencies are presented in Table 5.9.

Age category	Age range	Observed N	Expected N	Residual
2	17 - 24	32	141.3	-109.3
3	25 - 39	137	96.1	40.9
4	40 - 65	108	39.6	68.4
Total		277		

Table 5.9: Calculated frequency for age category.

A Chi-Square test showed that the age class distribution is different for our IBM sample and the Penn State population ($\chi^2(39) = 220.307, p = .001$). In fact, a ranksum test showed that the age of the Penn State population is significantly lower than our sample ($p = 0.001$). However, the distribution of sex is not significantly different, as proven by a Chi-square test ($\chi^2(1) = 72.068, p = .201$).

In summary, our population is a representative sample of the Penn State population in term of biological sex but not in terms of age distribution.

5.4 Chapter conclusions

In this chapter, we analyzed and discussed the personality traits and characteristics of our sample in term of the Big Five dimensions, BSRI, age, and gender. This provides some insights in the personality of our IBM sample. These insights can be used at a later stage to explain if what we observed is related to the personality of our sample.

In the first section, we found that women score higher on *Neuroticism* and *Agreeableness* than men in the Penn State population. In the second section, it was shown that our IBM sample scores lower on *Neuroticism* compared to the Penn State population, which means that people in our IBM sample are calm, composed and unflappable. This might also be a reason why people in the IBM game did not show many sharing activities. But more research is needed to investigate this. Furthermore, it has also been shown that women and men in our IBM sample differed significantly on personality traits. Women were more nurturing and to a greater extent tender-minded, and altruistic than men. This could mean that women could play the IBM game differently because of the different personality.

Our hypothesis that a larger proportion of males would be classified in the *Masculine* category and a larger proportion of females would be classified in the category *Feminine* was partly accepted. We found some overlap between biological sex and gender roles. However, based on the BSRI scores the majority of our IBM sample were classified as *Androgynous*. This means that the people in our sample are more likely to focus on what behavior is most effective within the situational circumstances rather than ad- here to culturally constructed behavior for men and women.In addition, no evidence was found that indicates that females are more likely to be sex-typed as *Feminine*. This means that the women in our IBM sample show masculine traits and deviate from the sex-typed female. This might also indicate that these women play the game differently than sex-typed women. But more research is needed to verify this.

In the third section, we identified that our sample is not a representative sample of the general IBM population in term of age, but is in terms of sex distribution. Moreover, if we compare our IBM sample with the Penn State population in terms of age distribution is it not a representative sample of the Penn State population, but in terms of biological sex, it is.

Now that we understand the personality traits and characteristics of our population we continue with the study to describe and analyze a tool which is able to determine personality traits automatically based on a given input text.

Chapter 6

Automatic assessment of personality

In this chapter, we want to take a computer science angle, by exploring the possibilities and approaches to determine personality traits automatically, without intrusively asking people to complete questionnaires. The tool we address in this chapter is *Watson Personality Insight Service*. This tool plays a central role in this thesis because it supports the research of automatic personality assessment. In Section 6.1 we give an overview of the tool with an emphasis on how it work, what it does and how we use it in our study. In Section 6.2 we conduct an experiment to test the reliability of this tool to meet objective 4, which is to investigate the use and reliability of an enterprise tool that is able to assess personality automatically without recurring to questionnaires or other invasive testing techniques. The results are presented in Section 6.3. Afterwards, we discuss our observations and results in Section 6.4 and conclude the work of this chapter in Section 6.5.

6.1 Watson Personality Insights Service

A way to measure personality traits objectively is using IBM Watson™ Personality Insights service[1]. This service provides an Application Programming Interface (API) that enables applications to derive insight about cognitive and social characteristics, including *Big Five*, *Values*, and *Intrinsic Needs* based on a given set of input text. The Personality Insights service is a generally available service by the end of February 2015 that was formerly known as the User Modeling service while in beta.

In this experiment, we want to understand the science behind the service and find out whether this service can be used to determine personality automatically. If this is the case then the personality traits of the remaining IBM gamers who did not complete the personality survey can be calculated.

Personality Insights by Watson

The idea behind this service is based on the fundamental premise that the words one uses in daily life reflect one's personality [67]. However, the text to be analyzed by

[1] `https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/doc/personality-insights/`

the service needs to be reflective. This means that the text should expose personal experiences, responses and thoughts of a person. This service provides three kinds of personality insights.

1. Personality characteristics are defined in the terms of the Five Factor model: *Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism*. Not only the five dimensions are computed, but also thirty facets of the Big Five Model are determined that are described in Table 2.1. This model that PIs uses was trained by learning the personality characteristics from blogs in a research performed by Yarkoni[84].

2. The intrinsic needs are resonated in 12 terms: *Excitement, Harmony, Curiosity, Ideal, Closeness, Self-expression, Liberty, Love, Practicality, Stability, Challenge, and Structure*. This model was trained with Twitter data.

3. The values are identified across five dimensions: *Self-transcendence / Helping others, Conservation / Tradition, Hedonism / Taking pleasure in life, Self-enhancement / Achieving success, and Open to change / Excitement*. For the sake of this research, we are only interested in the personality characteristics(i.e Big Five dimensions) provided by this service. Forum post was used to train this model for PIs.

6.1.1 The science behind Watson PIS

This service uses linguistic analytics based on the Linguistic Inquiry and Word Count (LIWC) dictionary on a given set of input text to calculate personal scores for the personality characteristics, values, and intrinsic needs. The given input text can be all types of communications such as social media, forum/blog posts, email, text messages.

The Personality Insight service computes the personality scores from textual information by tokenizing the input and matching the created tokens with the LIWC psycholinguistics dictionary. A score for each dictionary category is computed. Afterward, a weighted combination approach is used to derive personality characteristics (i.e Big Five, Needs and Values) from the computed LIWC category scores.

In order to compute statistically significant results a minimum of 3500 words and ideally 6000 words or more are preferred because this could reduce the sampling error.

6.2 Reliability of Watson PI service

In the previous section, we described how IBM Watson Personality Insight service works and what the preliminaries are in order to use it. With this experiment we want to investigate whether this service is reliable and provide accurate insight into personality traits, therefore we set up an experiment to test the reliability of this service before we decide to use this tool in our research to substitute measuring personality traits using a personality survey. Moreover, we would also like to get a better understanding of how the service works according to different types of data. In order to use this tool, we gathered two data sets, described in Section 4.5, which serves as input for the service to derive insights about personality traits of the IBM population.

The idea is to compare our survey-based scores with the ones derived by Watson PIs in which we put the content of social media to assess personality traits. To establish ground truth the participants took the psychometric test described in Section 4.1. Then we analyze the differences with statistical tests to see if this tool is reliable to use in our research.

The reason why we ask ourselves this is because the number of people that actually answered the personality questionnaire are a subset of the previous IBM game experiment. In general, we would like to see if this approach by which personality can be assessed automatically, could be used at a scalable level at the company instead of using personality questionnaires.

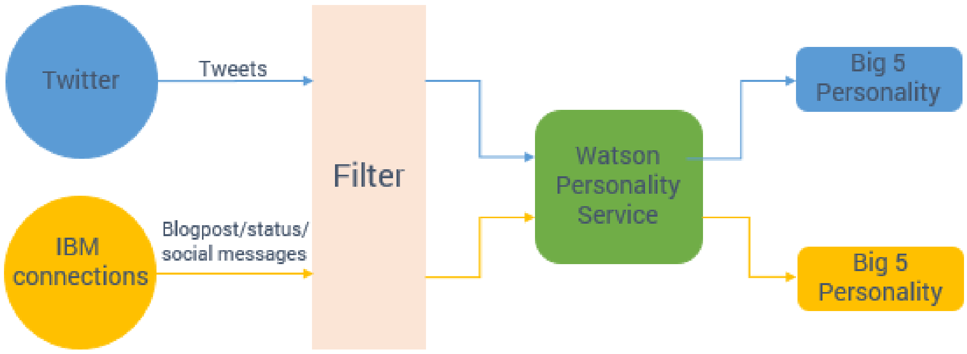


Figure 6.1: Automatic personality assessment.

6.3 Observations and Results

As described in Section 4.5 we retrieved two datasets(i.e Twitter and IBM Connections) from participants of our population. In this study, we only used the datasets separately, because it has not been validated yet whether analyzing text combined with different media platform produce reliable results and is therefore not recommended by IBM.

Figure 6.2 provides the average calculated scores for the Big Five dimensions derived by the service for the two datasets. It is observable calculated Big Five trait scores deviate from the scores of the personality survey.

The scores derived by Watson PIs for the Big Five dimensions based on Twitter dataset seems to be different than the actual scores computed by our survey, especially for the *Conscientiousness* dimension. Also, the scores derived by Watson PIs for the Big Five dimensions based on IBM Connections dataset seems to be different on all dimensions.

Correlational analyzes were performed to compare the derived traits and the psychometric measures. One straightforward approach is to calculate the correlation coefficients for all individual factors of all the traits. We did this analysis twice, ones with the psychometric scores from our survey and *Twitter* dataset and one with the survey-based scores and the scores derived with *Connection* dataset. The correlation coefficients were calculated per trait across all people.

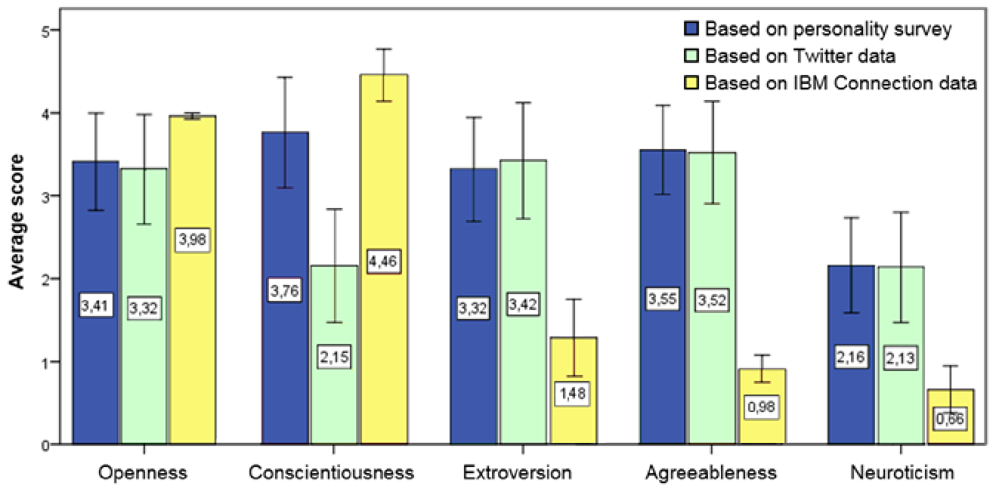


Figure 6.2: Overview of personality scores across difference sources.

The resulting coefficients for the first analysis were in the range of $-0.17 < r < 0.91$ and for the second analysis the coefficients were in the range of $-0.07 < r < 0.15$. Both of the results indicate that some personality traits are highly correlated while other are not correlated at all. The results of the findings are presented in Table 6.1 and 6.2. Only the ones that are marked blue are of our interested.

Statistically significant positive correlations were found for the traits: *Openness* and *T_Openness*, *Extroversion* and *T_Extroversion*, *Agreeableness* and *T_Agreeableness*, *Neuroticism* and *T_Neuroticism*. It was predictable that no significant correlation was found between the scores derived from the survey for *Conscientiousness* and the score derived from the PIs based on the Twitter data.

No statistically significant correlation were found between the measures main traits and the predicted traits using data from IBM Connections. This means that the increases or decreases computed by our survey do not significantly relate to increases or decreases in the scores of the derived dimensions by PIs.So we may conclude that increases or decreases in the dimensions *Openness, Extroversion, Agreeableness, and Neuroticism* only significantly relate to increases or decreases in scores derived by PIs (based on Twitter data) of the same personality dimension.

But since our personality traits are multiple dimensional constructions, we would like to use a measure that takes this into account. Therefore the RV-coefficient[71] was used to examine the overall correlation between derived traits and the corresponding psychometric scores. We did this for both of the datasets separately. RV coefficient is a multivariate generalization of squared Pearson correlation coefficient, and it measures the closeness of two sets of points in a multiple-dimensional space.

Our analysis showed that the RV-coefficient test was not significantly strong for both cases. The scores were significant for Big 5 $(rv = 0.1160, p = 0.0461)$ computed from Twitter data, while scored were not significant for Big 5 $(rv = 0.0217, p = 0.6081)$ computed from Connections data.

Twitter data	T_Open.	T_Cons.	T_Extr.	T_Agre.	T_Neur.
	Corr.	Corr.	Corr.	Corr.	Corr.
Openness	,832**	,229	,676**	,594**	,097
Conscientiousness	,526**	-,017	,623**	,621**	-,070
Extroversion	,624**	-,004	,871**	,540**	-,105
Agreeableness	,628**	,142	,555**	,799**	,111
Neuroticism	,102	,847**	-,043	,067	,910**

The top row presents the predicted Big Five personality traits using data from Twitter. For example, T_Open. means predicted trait *Openness* using Twitter data.
**. Correlation is significant at the 0.01 level (2-tailed).

Table 6.1: Pearson correlation values between derived score based on Twitter data and personality scores.

Connections Data	C_Open.	C_Cons.	C_Extr.	C_Agree.	C_Neur.
	Corr.	Corr.	Corr.	Corr.	Corr.
Openness	-,072	-,004	,228	,141	-,084
Conscientiousness	-,015	,079	,088	,048	-,134
Extroversion	-,034	,004	,153	,011	-,084
Agreeableness	-,048	,056	,152	,141	-,108
Neuroticism	-,115	-,053	-,048	,122	,074

The top row presents the predicted Big Five personality traits using data from IBM connections. For example, C_Open. means predicted trait *Openness* using IBM connections.

Table 6.2: Pearson correlation values between derived score based on Connections data and personality scores.

On a normalized 0-1 scale, the Mean Absolute Percentage Error(MAPE)[2] for each personality dimension determined by Watson PI using Twitter data was roughly 11%, except for the dimension *Conscientiousness*. The MAPE for the dimension scores derived using data from IBM Connections varies from 67% to 99%. Results are shown in Table 6.3. This means that Watson PIs predicts a user's score for a personality trait based on Twitter data to within just more than one-tenth of its actual value for all personality dimension except for *Conscientiousness*. Personality scored computed by Watson PIs on IBM Connection data are not accurate at all compared to the actual scores of the personality dimensions.

	Openness	**Conscientiousness**	**Agreeableness**	**Extroversion**	**Neuroticism**
Watson PIs (Twitter)	10.92%	69.54%	10.41%	12.53%	11.32%
Watson PIs (IBM Connections)	67.41%	76.24%	99.24%	95.31%	93.12%

Table 6.3: Mean Absolute Percentage Error for each test and personality trait.

[2]This measures the size of the error in percentage terms. Calculated by $\left(\frac{1}{n}\sum\frac{|Actualvalue-Predictedvalue|}{Actualvalue}\right)*100$

Hence, we may conclude that scores derived from the Watson PIs based on Twitter data provide significant comparable scores compared to the scores computed via our survey. Thus, the PI service provides reliable scores for the following personality dimension: *Agreeableness, Openness, Extroversion, and Neuroticism*. However, this does not hold for personality scores derived from Watson PIs using data from IBM Connections.

6.4 Discussion

The results provide support for the fact that Twitter data could be useful to derive insight in four of five personality traits. In contrast data from IBM Connections does not seem to work as good as we hoped. In this section, we discuss how Watson personality Insight service works according to different types of data.

According to the documents about IBM's Personality Insights service, a personality Insights portrait can be created only where sufficient data of suitable quantity and quality is provided. Because language use varies naturally from document to document and from time to time, a small sample of text might not be representative of an individual's overall language patterns. The service computes the percentile and sampling error that describe the extent to which the input text exhibits a characteristic and the possible range of deviation.

The main reason why tweets were a useful source to derive personality insight from is because tweets are often composed of people to express their own opinion and experience. The average word count per tweet data per person is 2902 words. The number of words improves the sampling error from the service, results in more accurate insights. Therefore, four of the five traits were successful predictable. Only Conscientiousness was difficult to predict, but in literature, it has been shown that Conscientiousness is difficult to describe in words. This could be the reason why the predicted Conscientiousness is not reflective to the user's conscientiousness score.

On the other hand, from the previous section, it was shown that Extroversion and Agreeableness were not predicted very well based on the data from IBM Connections. Agreeableness is a person's tendency to be compassionate and cooperative toward others. Extroversion is a person's tendency to seek stimulation in the company of others. Data from IBM Connections has a bit different structure compare to Twitter data. The input text could exist of corporate text messages pre-written by IBM, news articles or announcement of the company, messages promoting a special event. these messages do not have the intention to reflect the personality of someone who shares or publishes such messages. These are just a few examples that are included in the IBM Connection data which might not be very useful to derive personality insight from. These might be a possible reason why the predicted trait is not reflective to the user's personality profile. Another reason that could lead to less accurate results is the number of words used in the text. The average number of words used in IBM data for a person is 562 words. The documentation of the personality insights service already stated that using fewer than 2000 words can result in a sampling error that is greater than 30 percent, this might also be the reason why the derived personality scores are not very accurate.

6.5 Chapter conclusions

The results presented in this chapter suggest some interesting observations. The feasibility and validity of automatically deriving personality traits of an individual from Tweets in a real world are shown in the first place. Four out of five personality traits were successfully derived from Twitter data. Although data from IBM Connections seems to be less useful to derive personality insights from, this might due to the length and the context of the input. Less input text results in less accurate personality insights. In this study, none of the predicted personality traits based on data from IBM connections were reflective to the user's personality traits.

Now that we have validated the use and reliability of the tool to automatic assess personality traits we move on to the analysis to see if personality traits affect the way people played the IBM game. We will use the predicted personality traits which seems to be provided accurate personality insights in Openness, Extroversion, Agreeableness and Neuroticism in the next section to investigate the influences of these predicted traits on the game behavior.

Chapter 7

Personality and Gameplay

In this chapter, we present the study to determine whether personality traits also play a role in the way people played the IBM game. The insights from this study will help us to answer the main research question of this work which we defined as: "***To which extent do personality traits affect engagement strategies in enterprise gamification?***".

In the previous study about the IBM game, it was shown that the differences in game behavior between the treatment groups could have been explained by the fact that each group had a given game mechanic. In this work we add a new dimension *personality* to the previous study, to see if personality traits(i.e. Big Five dimension and gender roles) can be used as an additional dimension to explain the difference in the way IBM employees played the IBM game(see figure 7.1).

We first verified whether the populations of four treatment groups involved in the IBM game could be differentiated in terms of personality. A Kruskal-Wallis H test revealed that the distribution of the scores of each personality traits was similar (i.e. not significantly different) across the four groups. Therefore, we did not analyze the interaction effects between the treatment groups. Instead, we analyzed the main effect of personality traits on the game behavior. Test details are reported in Table 7.1.

In the remaining part of this chapter we first explain in Section 7.1 how we measure the gameplay in the IBM Game. In Section 7.2 we conduct a series of regression analysis to investigate the relationship between personality and gameplay. Section 7.3 presents a discussion on the results and observations. Concluding remarks of this chapter are described in Section 7.4.

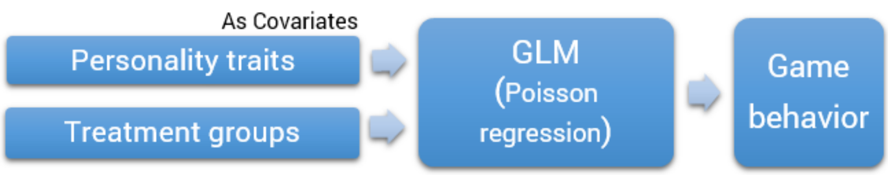


Figure 7.1: The prediction model.

	Neur.	Agree.	Extr.	Open.	Consc.	Masculinity	Femininity
Chi-Square	2.602	.239	1.715	2.443	.834	3.341	1.130
df	3	3	3	3	3	3	3
Asymp. Sig.	.457	.971	.634	.486	.841	.342	.770

Table 7.1: Test statistics of personality across treatment groups.

7.1 Measuring gameplay

As mentioned in Section 4.2.1 different game elements were measured in the IBM game. These game elements were used to measure a specific behavior. To specify the gaming behaviors, we structured the game elements of the IBM game and grouped them based on their purpose. In the current context we measure the following gaming behaviors:

- **Engagement**. The level of engagement is measured by the average session length, a number of game sessions, total time spent, users' game score and the total number of questions a user has answered.
- **Social behavior**. The social behavior of a user is measured by the number of invites a user sent and accepts, and the number of news shared.
- **Popularity**. The popularity of a user is measured by the number of connections across social media (LinkedIn, Facebook and in the game).
- **Expertise**. The level of expertise is measured by the number of correct answers given on the different type of questions.
- **Curiosity**. The curiosity behavior is measured by the number of times a user is looking at his peers and news.
- **Controlled behavior**. This behavior is not measured for each user because not every user received the same game mechanic. Only the given game mechanic was measured in this behavior. Game elements that are measured in this behavior are for example the number of badges a user has, the number of times a user looks at his achievements and the leaderboard.

Since the assumption of normality and homogeneity in our data was violated, tested using the *Shapiro Wilk Test* and *Levene's Test for Equality of Variances* presented in Table 7.2, we preferred to use a generalized linear model instead of an ANCOVA.

7.2 Main effects using generalized linear model

A generalized linear model(GLM) is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. Each of these components is described below.

Game Elements	Shapiro-Wilk			Levene's Test of Equality of Variances	
	Statistic	df	Sig.[a]	F	Sig.[b]
Total time	,268	109	,000	,089	,766
Average session	,890	109	,000	6,120	,014
Number of sessions	,208	109	,000	,001	,982
Total IBM Questions	,490	109	,000	,008	,928
Total wwt questions	,370	109	,000	,496	,482
Total social questions	,184	109	,000	,045	,832
Max. Score	,176	109	,000	,042	,837
News shared on LinkedIn	,301	109	,000	,704	,403
Invites sent	,202	109	,000	,002	,962
Invites sent pm	,158	109	,000	,946	,332
Invites accepted	,225	109	,000	1,976	,161
Invites accepted pm	,155	109	,000	,150	,699
Correct IBM questions	,480	109	,000	,138	,710
Correct wwt questions	,362	109	,000	,412	,522
Correct social questions	,170	109	,000	,007	,936
LinkedIn connections	,778	109	,000	2,527	,113
Facebook connections	,376	109	,000	10,896	,001
In game connections	,856	109	,000	,250	,618
Peers impression	,301	109	,000	,946	,332
News impression	,304	109	,000	,218	,641
Achievement impressions	,204	109	,000	5,767	,017
Leaderboard impressions	,226	109	,000	2,198	,140
Badges earned	,728	109	,000	2,286	,132

[a]. Variable is normal distributed if Sig. value is > 0.05.
[b]. Variable has equal variance between gamers who completed the survey and those who did not complete the survey if Sig. value is > 0.05.

Table 7.2: Results of tests for Normality and Homogeneity of the game elements.

- *Random Component*: this is known as the error model which explains how random error is added to the prediction that comes out of the link function. So, the random component is the dependent variable, on which the normality constraint is now relaxed.

- *Systematic Components*: this is known as the linear predictors in the linear regression model which specifies the explanatory variables. In this case, the components are the predictors, for example, the personality scores, that need to be combined to predict the random component (and their combination is linear).

- *Link Function*: specifies the link between random and systematic components. It explains how the expected value of the response variables relates to the linear predictor of explanatory variables. In a GLM different link functions can be used that denotes a different relationship between the linear model and the response variable(e.g. log, inverse, etc.). This function maps the outcome of the linear combination into the domain of the response variable.

In this study, we perform a series of Poisson regression analysis to examine the game behaviors of people who played the IBM game based on the treatment groups in the game, while we are controlling for different personality traits in each analysis. The series of Poisson regression analysis consists of investigating four models. In the first two models, we use the whole questionnaire dataset of the gamers, while we are controlling for the Big Five dimension in Model 1 and controlling for gender roles in Model 2. In Model 3 and 4, we apply the Poisson regression analysis on a subset of the whole questionnaire dataset of the gamers from whom we also have the personality insight derived via Twitter data. In Model 3, we use the Big Five dimensions calculated from the survey as predictors, while in Model 4, we use the derived Big Five dimension from Twitter as predictors.

The Poisson regression analysis is a special case of the Generalized Linear Model where the data follows a Poisson distribution that is frequently encountered with count data. This method is applicable to our data because our dependent variables (e.g. the game elements) consists of "count data".

7.2.1 Results Model 1

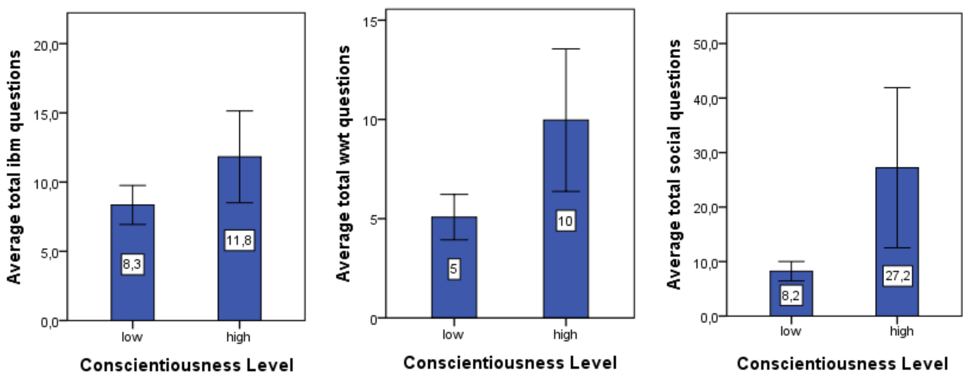
The results of first Poisson regression analysis is presented in Table 7.3. It shows that the Big Five dimensions have a significant influence in the prediction of the game behavior.

		Model 1: With Big Five dimensions as covariates ($n = 112$)									
Game behavior	**Dependent variables**	Openness		Conscientiousness		Extroversion		Agreeableness		Neuroticism	
		Sig.	$Exp(\beta)^a$	Sig.	$Exp(\beta)^a$	Sig.	$Exp(\beta)^a$	Sig.	$Exp(\beta)^a$	Sig.	$Exp(\beta)^a$
Engagement	Total time	**.000**	1.288	**.000**	6.053	**.000**	.263	**.000**	.407	**.000**	.713
	Average. session	.783	.965	.032	1.248	.006	.733	.024	.746	**.001**	1.293
	Number of sessions	.758	.941	**.000**	4.474	**.000**	.438	.124	.720	**.000**	.646
	Total IBM questions	**.000**	1.661	**.000**	2.033	**.000**	.483	**.000**	.594	**.000**	.877
	Total social questions	**.000**	2.424	**.000**	29.234	**.000**	.091	**.000**	.195	**.000**	.475
	Total wwt questions	**.000**	1.661	**.000**	5.960	**.000**	.301	**.000**	.356	.187	.915
	Max. score	**.000**	1.472	**.000**	15.339	**.000**	.177	**.000**	.307	**.000**	.544
Social behavior	News shared on LinkedIn	**.010**	1.042	**.000**	2.907	.098	1.661	.110	.601	.060	.712
	Invites sent	.731	.919	**.000**	9.751	**.000**	.167	.090	.630	**.000**	.628
	Invites sent pm	**.008**	1.043	**.000**	22.478	**.000**	.172	.426	.789	**.000**	.435
Popularity	LinkedIn connections	**.000**	1.385	**.000**	.782	**.000**	1.128	**.000**	.854	**.000**	1.004
	Facebook connections	**.002**	.689	**.000**	.656	**.000**	2.574	.637	.947	**.000**	.574
	In game connections	.064	1.154	.655	.973	.255	1.086	.262	1.092	**.000**	.812
Expertise	Correct IBM questions	**.000**	1.789	**.000**	2.218	**.000**	.419	**.000**	.575	**.000**	.732
	Correct social questions	**.000**	2.484	**.000**	39.423	**.000**	.071	**.000**	.176	**.000**	.451
	Correct wwt questions	**.002**	1.575	**.000**	8.441	**.000**	.228	**.000**	.296	.028	.823
Curiosity	Peers impression	.312	1.182	**.000**	4.242	**.000**	.313	.036	.676	**.000**	.680
	News impression	.668	.891	**.000**	3.901	**.000**	.378	.355	.747	.198	.803
	Invites accepted	.928	1.060	**.000**	10.626	.065	.313	.028	.161	.395	1.380
	Invites accepted pm	.886	1.125	**.000**	26.817	.095	.265	.020	.081	.939	1.039
Controlled behavior	Leaderboard impression	.300	1.537	**.000**	17.342	**.000**	.057	.307	.689	.020	.557
	Achievement impression	.784	1.081	**.000**	16.893	**.000**	.142	**.006**	.384	.097	.766
	Badges earned	.636	1.100	.048	1.668	.033	.652	.290	.794	.556	1.072

* the significant values $p < .01$ are marked bold. [a] the exponentiated values of the coefficients.

Table 7.3: Results of Poisson Regression using Big Five dimensions calculated from the survey (Model 1).

The results of the present investigation indicate that the individual characterization of the Big Five dimensions can be used to explain variance in game behavior. One of the key components to engage gamers with gamification are the use of badges. Badges provide players with a sense of achievement and thus encourage players to think carefully about their in-game behavior[81]. Given that engagement is something personal, we would expect that personality would affect the number of badges a gamer earns. However, we found no significant associations between the Big Five dimensions and the number of badges earned in the IBM game. This means that none of the personality dimensions affected the number of badges a gamer has in the game. This is also the only game element which does not has a relationship with any of the Big Five dimensions. On the other hand, we found strong evidence that a relationship exists between the Big Five dimensions and several game behaviors. For example, the more *Conscientious* a gamer is the more questions this gamer answered. The exponential beta values ($Exp(\beta)$) for the total number of questions are 2.033 (IBM questions), 5.960 (World wide technology (wwt) questions) and 29.234 (Social questions). These results can be interpreted as someone who with high *Conscientious* scored 2.033 more *IBM questions* than someone with less *Conscientiousness*. Someone with high *Conscientious* scored 5.960 more world wide technology questions than someone with low *Conscientiousness*. The other results of the Poisson regression analysis can be interpreted in the a similar way. For the purpose of visualization we are going to take *Conscientiousness* and divide the gamers in *low* and *high Conscientiousness* depending on the median. In figure 7.4 it shows that the more Conscientious a gamer is, the more questions are answered per category.

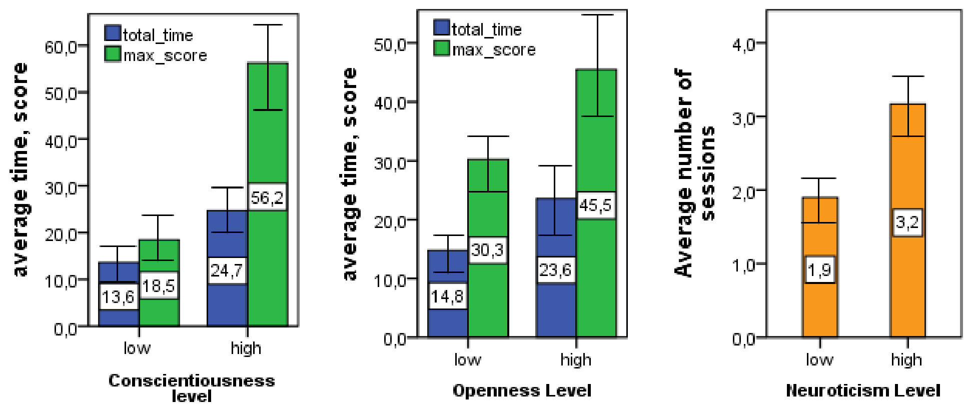


(a) IBM questions (b) wwt questions (c) Social questions

Figure 7.2: Relationship between Conscientious and number of questions answered.

In summary, the following list highlights the potential moderating effects of the Big Five personality traits affecting the different game behaviors. For each behavior, we visualized the effects of the personality dimensions with the largest exponentiated value (see Table 7.3). Depending on the median we divided the people into *low* scores and *high* scores for each dimension.

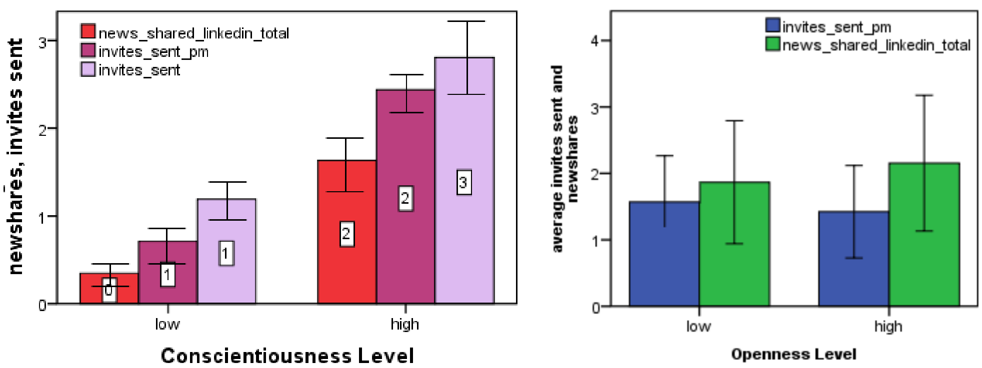
Outcome 1: People who score high on each of the Big Five dimensions will spend more time in the game than people who score less on these dimensions. Moreover, these people also tend to answer more IBM questions in general. In addition, this also holds for answering social questions. Furthermore, people with these characteristics also have a higher score in the game and people who are more conscientious, extrovert and neurotic play the game more often and send more invites than people who score low on these dimensions. Figure 7.3 presents the influences of Conscientiousness, Openness, Neuroticism on several engagement aspects like total time spent, game score and the number of sessions.



(a) total time spent (b) game score (c) number of sessions

Figure 7.3: The influence of Conscientiousness, Openness, Neuroticism on engagement aspects.

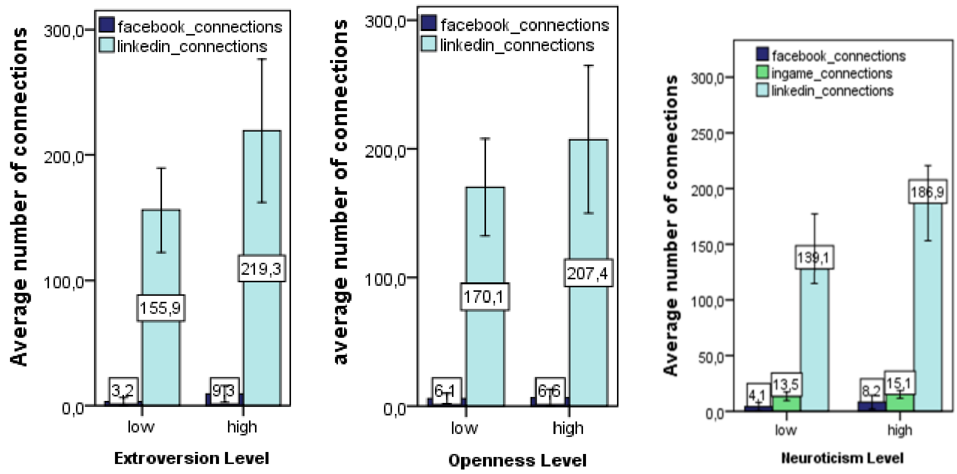
Outcome 2: The traits *Openness, Conscientiousness, Extroversion and Neuroticism* influence the social behavior of the gamers. Gamers who are more open en conscientious share more news in the IBM game than those who score low on these dimensions. In addition, gamers who score high on *Conscientiousness, Extroversion and Neuroticism* have sent more invites than gamers scoring low on the mentioned dimensions. And if gamers are more open then they also sent more invites via personal messages. In short, *Agreeableness* is the only dimension which does not influence social behavior in the game. Similar to the previous visualization, in figure 7.4a we only visualize *Conscientiousness* and *Openness* because these traits have the biggest impact on all the aspects of social behavior.



(a) Influences of Conscientiousness (b) Influences of Openness

Figure 7.4: The influence of Conscientiousness and Openness on Social behavior.

Outcome 3: People who are more open and extrovert are more popular in the game. This means that these people have more connections on social media LinkedIn, Facebook and in the game. People who are more open also have more connec-

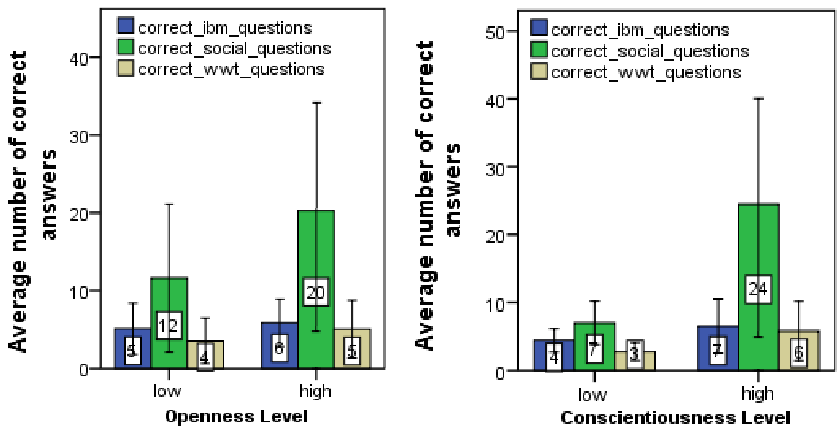


(a) influence of Extroversion (b) influence of Openness (c) influence of Neuroticism

Figure 7.5: The influences of Extroversion, Openness and Neuroticism on popularity.

tions in the game. Figure 7.5 presents the influences of *Openness, Extroversion,* and *Neuroticism* on popularity.

Outcome 4: The expertise of gamers are influenced by all five dimensions. The number of correct answers is mostly influenced by a higher score in *Conscientiousness* and *Openness*. In other words, the more agreeable and conscientious a person is the more likely this person answers questions correctly. Figure 7.6 presents the influences of *Openness* and *Conscientiousness* on Expertise.



(a) Influences of Openness (b) Influences of Conscientiousness

Figure 7.6: The influences of Openness and Conscientiousness on Expertise.

Outcome 5: The curiosity of peoples is mostly influenced by the level of *Conscientiousness*. People who score higher on *Conscientiousness* have more accepted invites and look more often at their peers and the news. Figure 7.7 presents the influence of *Conscientiousness* on Curiosity.

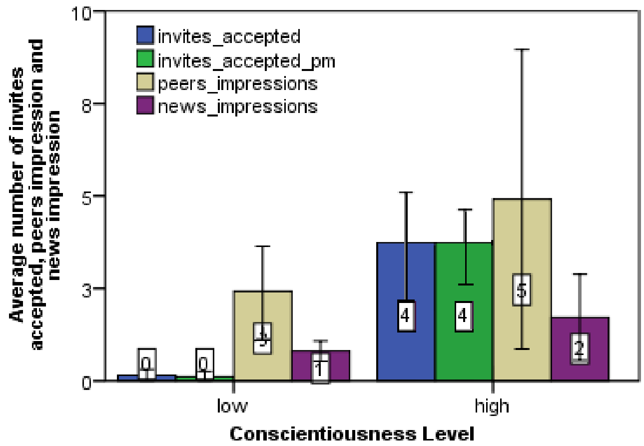


Figure 7.7: The influence of Conscientiousness on Curiosity.

Outcome 6: The controlled behavior is mostly affected by the level of conscientiousness. People who are more conscientious look more often at the leaderboard and their achievements. Figure 7.8 presents the influence of *Conscientiousness* on Controlled behavior.

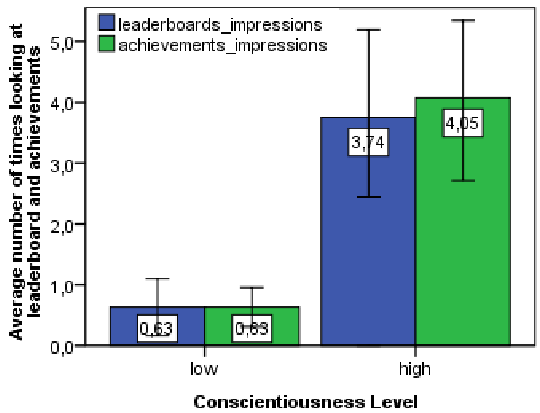


Figure 7.8: The influences of Conscientiousness on Controlled behavior.

7.2.2 Results Model 2

In this regression analysis, the calculated gender roles *Masculinity* and *Femininity* were used as predictors to predict game behavior. Table 7.4 presents the results of this regression analysis. It shows that *Masculinity* as well as *Femininity* have a significant effect on several game behaviors.

The following list highlights the potential moderating effects of the gender role traits affecting the different game behaviors. For each result, we visualized the influences of the gender roles on the game behaviors. Depending on the median of the gender roles we divided the people into *low* scores and *high* scores for *Masculinity* and *Femininity*. This is only used for visualization purposes.

Game behavior	**Dependent variables**	Model 2: With gender roles as covariates ($n = 112$)			
		Masculinity		Femininity	
		Sig.	Exp(B)[a]	Sig.	Exp(B)[a]
Engagement	Total time	**.000**	.734	**.000**	1.463
	Average session	.468	.758	.475	.822
	Number of sessions	**.003**	.763	**.000**	1.552
	Total IBM Questions	**.003**	1.155	.946	.996
	Total wwt questions	**.001**	1.271	**.006**	1.135
	Total social questions	**.000**	.970	**.000**	1.860
	Game score	.346	.724	**.000**	1.578
Social behavior	Invites sent	.321	.339	**.000**	1.830
	Invites sent pm	.432	.442	**.000**	2.218
	News shared on LinkedIn	**.000**	2.830	**.019**	1.546
Popularity	LinkedIn connections	**.000**	1.425	**.000**	1.530
	Facebook connections	.448	.775	.164	.665
	In game connections	.503	.887	.443	.737
Expertise	Correct IBM questions	.192	1.090	.946	.995
	Correct wwt questions	.201	1.100	.293	.914
	Correct social questions	**.000**	.535	**.000**	1.977
Curiosity	Peers impression	.325	.549	**.006**	1.292
	News impression	.043	.559	.047	1.377
	Invites accepted	.032	.448	.254	1.488
	Invites accepted pm	.092	.334	.187	1.730
Controlled behavior	Achievement impressions	**.000**	.556	**.000**	1.735
	Leaderboard impressions	.039	.213	**.000**	2.719
	Badges earned	.253	.893	.712	1.044

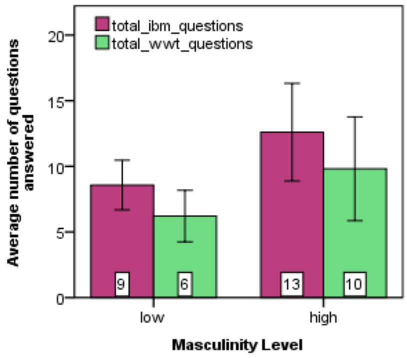
* the significant values $p < .025$ are marked bold.
[a] the exponentiated values of the coefficients.

Table 7.4: Results of Poisson Regression using BSRI scores calculated from the survey (Model 2).

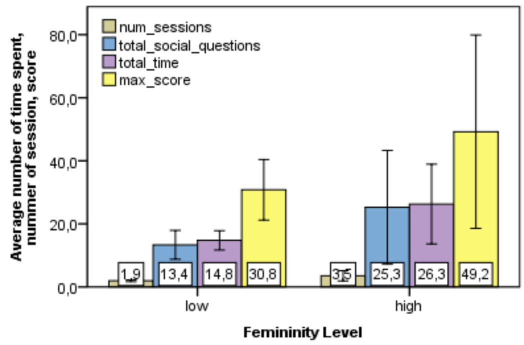
Outcome 1: People who are more masculine answer more questions than people who have low scores on the masculine trait. Someone who is scoring high on *Femininity* has a higher game score than someone scoring low on *Femininity*. Figure 7.9 presents the influences of *Masculinity* and *Femininity* on several engagement aspects like total time spent and the number of answered questions.

Outcome 2: Both gender roles influence the social behavior. The more feminine people are the more invites people sent. While the more masculine people are the more news articles they share on LinkedIn. Figure 7.10 presents the influences of *Masculinity* and *Femininity* on social behavior.

Outcome 3: Both gender roles influence the number of LinkedIn connections of a person. The more feminine/masculine people are the more LinkedIn connections they have. Figure 7.11 presents the influences of *Masculinity* and *Femininity* on popularity aspects in the game.

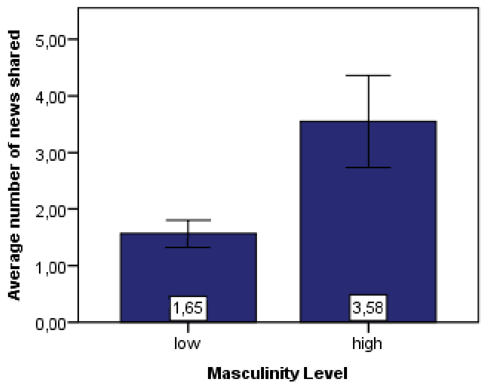


(a) Masculinity influences

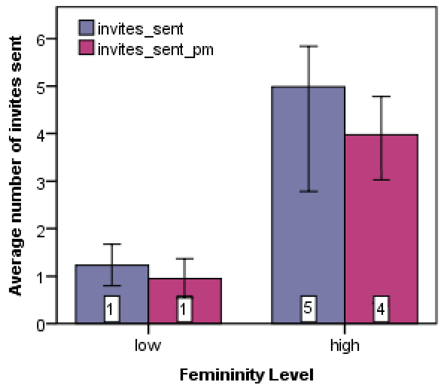


(b) Femininity influences

Figure 7.9: The influence of Masculinity and Femininity on engagement aspects.

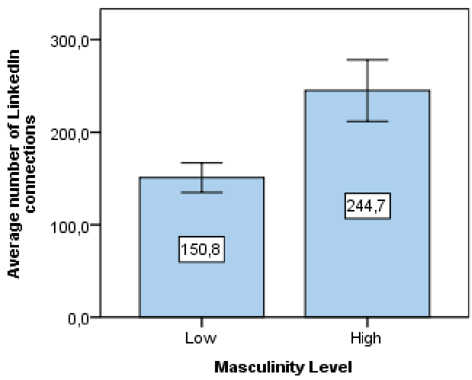


(a) Masculinity influences

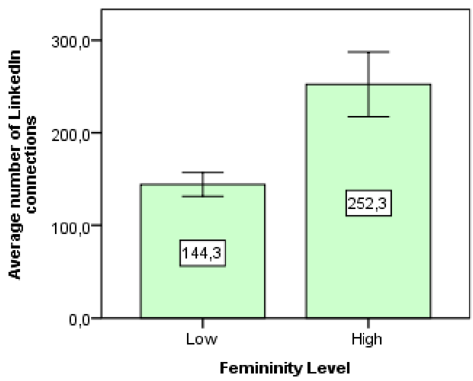


(b) Femininity influences

Figure 7.10: The influence of Masculinity and Femininity on social behavior.



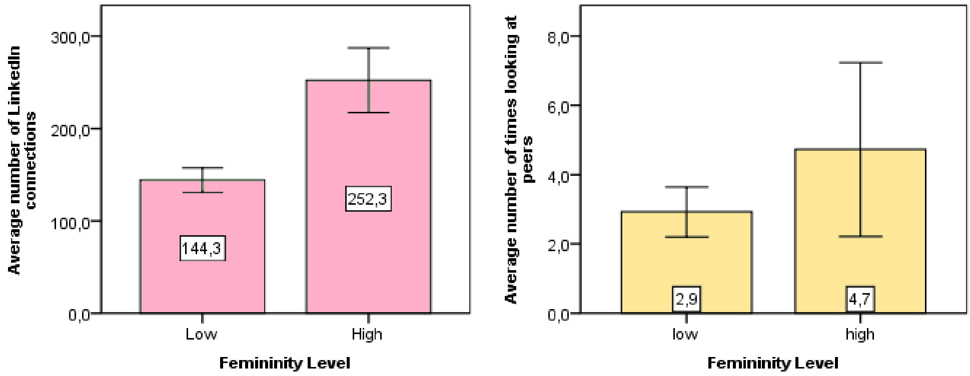
(a) Masculinity influences



(b) Femininity influences

Figure 7.11: The influence of Masculinity and Femininity on the number of Linkedin connections.

Outcome 4: The more feminine people are the more correct answers on social questions are given. Figure 7.12a presents the result as described.



(a) influences on number of correct answers (b) influences on peer impression

Figure 7.12: The influence of Femininity on popularity and expertise.

Outcome 5: Only femininity has limited influences in the curiosity behavior. The more feminine people are the more often they will take a look at their peers. Figure 7.12b visualizes this result.

Outcome 6: *Femininity* seem to have the largest influence on the controlled behavior. The more feminine people are the more they look at the leaderboard and their achievements. The results are visualized in Figure 7.13.

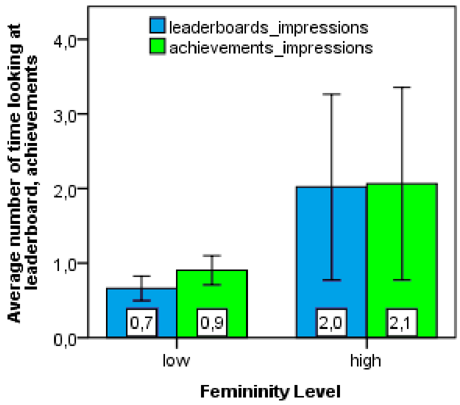


Figure 7.13: The influence of Femininity on controlled behavior.

7.2.3 Results Model 3

The purpose of this analysis is to observe whether similar results (as presented in Model 1) can be achieved by using self-reported and predicted personality information. To achieve this, we recomputed Model 1 using the self-reported data, but only the population for which we also have collected Twitter data.

Note that in this case we only used four of the five traits, because in the previous chapter we observed that the prediction of *Conscientiousness* by Watson PIs were bad. Therefore, it would not make sense to use *Conscientiousness* in model 4; to achieve comparability of model 3 and 4, we use four traits(*Openness, Extroversion, Agreeableness* and *Neuroticism*) in model 3 as well. The result of this analysis is presented on the left side of Table 7.5.

Similar outcomes were found in Model 3. There were no cases that were significant in Model 1 but not significant in Model 3 and vice-versa. However, we do observe differences in the exponential beta values (i.e. $Exp(\beta)$). Some of these values increased and some of the beta values decreased. But the observed difference were all below 0.500. These results were expected to be similar since Model 3 included a subset of the population involved in Model 1.

7.2.4 Results Model 4

This section presents the results for the fourth Poisson regression analysis which has also been applied to a subset of the whole questionnaire dataset of gamers from whom we also have the Twitter information. The predicted personality traits were used to as predictors to predict the game behavior. So, in this analysis we used the traits: *Openness, Extroversion, Agreeableness* and *Neuroticism*, predicted from Watson PIs as covariates in the GLM. The result of this analysis is presented on the right side of Table 7.5.

Similar results were found compared to model 3 described in Section 7.2.3. When we compare the significance values and the exponentiated values of the coefficients for each row(marked purple in Table 7.5), we observe that in all cases similar significance values and exponentiated values are found for each personality trait and game element. This indicates that the predicted personality traits were not only reliable in predicting the employees' personality scores, but it also seems that similar results (as described in Section 7.2.1) can be achieved using the predicted personality information.

7.3 Discussion

The objective of this study was to highlight the potential moderating effects that personality traits affected the gameplay and the employee engagement. Our research question was to investigate if there are differences between how people with different personality traits perceive employee engagement in a gamified enterprise setting. Significant differences in engagement were found in the previous study and in the current study personality traits were added to the perceived engagement and game behavior. The results of the current study provide support for personality traits in the prediction of game behavior in a sample from a large international enterprise. Specifically, results provide support for the Big Five personality traits and the gender roles *Masculinity* and *Femininity*. Each of these traits offered significant prediction of game elements in game behavior.

Across the treatment groups is was shown that the distribution of the scores of each personality trait was similar. This means that no significant differences in personality could be found between the groups. The main reason for not finding significance

Game behavior	Dependent variables	Model 3: With four main dimensions from personality survey as covariates ($n = 68$)								Model 4: With four main dimensions predicted from Twitter data as covariates ($n = 68$)							
		Openness		Extroversion		Agreeableness		Neuroticism		Openness		Extroversion		Agreeableness		Neuroticism	
		Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$	Sig.	$Exp(\beta)$
Engagement	Total time	**.000**	1.535	**.000**	.621	**.007**	.476	**.000**	.542	**.000**	1.579	**.000**	.520	**.002**	.421	**.000**	.538
	Average session	.423	1.157	.822	.967	.017	.629	**.000**	1.430	.245	.886	.802	1.024	.112	1.189	**.009**	1.406
	Number of sessions	.519	1.147	.386	.838	.077	.1.428	**.000**	.493	.023	1.363	.044	.703	.015	1.424	**.000**	.540
	Total IBM Questions	**.000**	1.718	.837	.980	**.000**	.546	**.000**	.961	**.002**	1.631	.890	.771	**.000**	.538	**.000**	.837
	Total social questions	**.000**	2.349	**.000**	.084	**.001**	1.146	**.000**	.276	**.000**	2.231	**.000**	.075	**.001**	1.114	**.000**	.311
	Total wwt questions	**.000**	1.902	**.000**	.369	**.005**	.710	.122	.889	.016	1.202	**.003**	.378	**.001**	.725	.019	.861
	Max. Score	**.000**	1.262	**.000**	.149	**.000**	.210	**.000**	.367	**.000**	1.284	**.000**	.193	**.000**	.254	**.000**	.422
Social behavior	News shared on LinkedIn	**.000**	1.048	**.000**	1.623	.987	1.005	.588	.889	**.000**	1.001	**.000**	1.728	.140	1.330	.176	.810
	Invites sent	.268	1.308	**.001**	.460	.083	1.930	**.000**	.308	.241	2.416	**.000**	.488	.118	1.227	**.000**	.410
	Invites sent pm	.783	.927	.018	.506	.090	2.987	**.000**	.237	.753	2.574	.981	.487	.091	2.633	**.000**	.256
Popularity	LinkedIn connections	**.008**	1.045	**.000**	1.380	**.000**	.800	**.000**	1.087	**.000**	1.300	**.000**	1.344	**.000**	.895	**.000**	1.035
	Facebook connections	**.000**	.541	**.000**	2.624	.854	.415	**.000**	.613	**.000**	.564	**.000**	2.693	.163	1.176	**.000**	.693
	In game connections	.047	.833	.250	1.368	.800	1.023	**.000**	.742	.243	1.068	.663	1.176	**.000**	.818	**.000**	.746
Expertise	Correct IBM questions	**.000**	1.708	**.001**	.475	**.000**	.580	**.000**	.771	**.000**	1.741	**.001**	.457	.001	1.068	**.000**	.699
	Correct social questions	**.000**	2.423	**.000**	.089	**.001**	.123	**.000**	.464	**.000**	2.419	**.000**	.451	**.000**	.125	**.000**	.412
	Correct wwt questions	**.000**	1.974	**.001**	.279	**.000**	.347	.030	.772	**.006**	1.912	**.000**	.330	**.002**	.334	.041	.778
Curiosity	Peers impression	.112	1.539	**.000**	.325	.186	1.248	**.000**	.696	.124	1.745	**.000**	.379	.822	1.023	**.000**	.699
	News impression	.976	.991	**.001**	.753	.327	1.420	.018	.649	.507	1.137	**.000**	.789	.095	1.351	.021	.613
	Invites accepted	.532	1.545	.924	.939	.454	.591	.304	.635	.021	5.248	.012	.346	.598	1.228	.075	.516
	Invites accepted pm	.393	1.944	.666	.728	.534	.628	.060	.390	.032	6.033	.015	.318	.736	1.153	.016	.339
Controlled behavior	Leaderboard impressions	.297	1.393	**.000**	.064	.619	2.179	.332	.314	.035	1.513	**.000**	.061	.312	1.866	.000	.308
	Achievement impressions	.219	.1.403	**.002**	.159	**.000**	.402	.100	.471	.050	1.945	**.000**	.364	**.000**	.404	.230	.474
	Badges earned	.214	1.327	.457	.850	.457	.840	.216	.838	.496	1.104	.028	.744	.387	1.126	.081	.825

* the significant values $p < .0125$ are marked bold.
.XXX highlight the results that are significant for both Model 3 and Model 4.

Table 7.5: Results of Poisson regressions on subset of our IBM sample (Model 3 and 4).

between the groups is probably due to the fact that the people were randomly assigned to the treatment groups, no control for personality traits was taken into account.

When we are controlling for the Big Five personality traits, it indicates that *Conscientiousness* has the biggest impact of on how people played the game. We found that engagement was mostly influenced by the trait *Conscientiousness*. The more conscientious a person is the more this person is engaged in the game. The engagement has been measured by the total time spent, the average session length, the number of sessions, the total answered questions and the game score. *Conscientiousness* was significantly related to each of these game elements. Moreover, people who are more open spent more time on the game and answers more questions which also resulted in a higher game score. The social behavior in the game is to a large extent influenced by the level of *Conscientiousness* of a person. The more conscientious a person is the more invites this person would send and the more news this person shares. The popularity of the game, measures by the number of connections on social networks, is mostly affected by *Openness* and *Extroversion*. The more open and extrovert a person is the more connections this person has on his social network compared to people scoring low on these dimensions. The expertise, the number of correctly answered questions were also depending on the level of *Conscientiousness*. People who are more conscientious tend to answer more questions correctly than people who score low on this trait. Also, the curiosity behavior and controlled behavior are impacted the most by *Conscientiousness*. The current findings indicate that a higher neurotic personality may have more impact on the engagement, social behavior, popularity and expertise in the IBM game. Specifically, individuals in the highly neurotic group were found to spent more time per game session on the game than individuals in the less neurotic group. However, in the previous chapter, we have shown that people in our IBM sample are less neurotic. This might be the reason why significant influences of neuroticism on game behavior was found, but the reported beta values of the coefficients were rather low.

When we control for gender roles as personality traits. The result shows that *Masculinity*, as well as *Femininity*, have influences on game behaviors. The biggest impacts of the two gender roles are visible in engagement and social behavior. In the other game behaviors, *Masculinity* and *Femininity* do have a significant influence but the impact is not huge. The finding that people who score higher on masculine traits and feminine traits are more engaged might be specific for our IBM sample. In previous research, we identified that the majority of our IBM sample is classified as *Androgynous*, which describes people that score high on both masculine and feminine traits, but none of the traits dominate. This indicates that the females in our IBM sample are different than the sex-typed females and therefore they may also show more masculine behaviors. For example, women are often expected to be passive and submissive, while men are usually expected to be active and competitive. But the analysis shows that the more feminine a person is the more this person engages in the game and therefore showing an active behavior (i.e. answering more social questions, spent more time on the game and having a higher game score).

When we only control for four dimensions of personality (i.e. *Openness, Extroversion, Agreeableness*, and *Neuroticism*), the results show that *Openness* has the biggest influence on engagement and expertise. While *Extroversion* strongly affected the popularity of the game. *Agreeableness* is an important factor that influenced the social behavior and several game elements of engagement. *Neuroticism* also influenced several game elements in the game, but the impact was not huge. Similar results were found when we used the four predicted personality dimensions based Twitter data. This indicates that the predicted personality traits not only provide accurate insights of employees' personality but the predicted personality insights could also be used to achieve the same observations as using the self-reported personality information. The findings in this chapter supported the fact that personality traits influence the game behavior.

7.4 Chapter conclusions

In this chapter, we have conducted a series of regression analysis using the self-reported personality traits as well as predicted personality information as covariates, to observe the relationship between personality and gameplay. It turns out that strong evidence was found for the fact that the new information, personality traits, is useful to explain differences in game behavior.

Even though we did not find strong evidence for significant interactions effects between the treatment groups, we were still able to find many interesting main effects. The effects of including the Big Five dimensions and gender roles into the prediction of the game behavior have been confirmed by the Poisson regression analysis. This research suggests that both the Big Five dimensions(*Openness, Conscientiousness, Extroversion, Agreeableness,* and *Neuroticism*) and gender roles(*Masculinity* and *Femininity*) are important predictors of game behavior. Results supported that *Conscientiousness* has the strongest influence on engagement, social behavior, expertise, curiosity and controlled behavior. *Openness* was found to be the second strongest predictor of the game behaviors. *Masculinity* and *femininity* were found to have an impact on the engagement, social behavior and part of other game elements.

Now that we have completed the analysis to find a possible relationship between personality traits and gameplay, we would like to explore the generalizability of these results which is presented in the next chapter.

Chapter 8

Generalization

In this chapter, we discuss an approach which calculates the probability that our sample is being taken from a population we want to generalize to. This method is called randomization. First we explain what this randomization is. Then we apply this method to the IBM population as well as to the Penn State population. The work of this chapter will be concluded with a chapter conclusion.

8.1 Randomization

In this section, we explore a the *randomization* method described in [16], which can tell us whether two samples are related without any reference to population parameters.

The idea behind this method is to combine two samples with each other and randomly sample it into two pseudo samples. Then calculate the probability that this random sample is taken out of the sample we would like to compare with. Based on the observation we may determine how reasonable it is that the two samples are related to each other. The complete procedure of this *randomization* method is presented in figure 8.1.

First we want to determine the representativeness of our sample for the general IBM population. Second we want to compute the probability that our sample is representative of the Penn State population.

IBM population

The null hypothesis that our sample is drawn from the IBM population will be rejected if the sample statistic Θ[1] fall in the upper or lower 2.5% of the distribution of Θ^*[2].

Let d_{IQR} denote the difference between the interquartile range of our sample and the IBM population. In this case the interquartile differences is: $d_{IQR} = 12.5 - 12 = 0.5$. To determine the probability that this difference arose by chance, we determine the distribution of the values of d_{IQR} that could arise by chance using the *randomization* procedure.

Figure 8.2. shows a distribution of 10000 values of d^*_{IQR}, generated by the *approximate randomization* procedure. It shows the 10000 differences of the interquartile

[1] Let $\Theta = f(S_A, S_B)$ the difference of the interquartile ranges of the samples A and B.

[2] $\Theta^* = f(A^*_i + B^*_i)$, with A^*_i the first randomized pseudo sample and B^*_i the remaining pseudo sample.

Procedure 5.7 Approximate Randomization to Test Whether Two Samples Are Drawn from the Same Population

i. Let S_A and S_B be two samples of sizes N_A and N_B, respectively. Let $\theta = f(S_A, S_B)$ be a statistic calculated from the two samples, such as the difference of the interquartile ranges of the samples. Let $S_{A+B} = S_A + S_B$, that is, the merge of S_A and S_B.
ii. Do K times:
   a. Shuffle the elements of S_{A+B} thoroughly.
   b. Assign the first N_A elements of S_{A+B} to a randomized pseudosample A_i^* and the remaining N_B elements to B_i^*.
   c. Calculate $\theta_i^* = f(A_i^*, B_i^*)$ and record the result.
iii. The distribution of θ_i^* can now be used to find the probability of the sample result θ under the null hypothesis that the samples are drawn from the same population.

Figure 8.1: The approximate randomization procedure, recreated from [16].

ranges of two randomized samples A_i^* and B_i^*, of sizes 277 and 3342, respectively, drawn without replacement from a sample that contains all 3619 age samples from our population and the general IBM population. The mean of the distribution is 1.132 with a stand deviation of 0.887. From this results, we can see that our original sample result, $d_{IQR} = -0.5$, is certainly not unusual because roughly 44% of the values of the differences of interquartile ranges lie below it. Therefore, we may accept the null hypothesis and conclude that the samples are drawn from the same population without drawing any conclusions about the populations.

Penn State population

We apply the same procedure to our sample in combination with the Penn State population.

Let d_{IQR} denote the difference between the interquartile range of our sample and the Penn State population. In this case the interquartile differences is: $d_{IQR} = 11 - 12 = -1$. To determine the probability that this difference arose by chance, the distribution of the values of d_{IQR} that could arise by chance is needed.

Figure 8.3. shows a distribution of 10000 values of d_{IQR}^*, generated by this procedure. It shows the 10000 differences of the interquartile ranges of two randomized samples A_i^* and B_i^*, of sizes 277 and 2847, respectively, drawn without replacement from a sample that contains all 3123 age samples from our population and the Penn State population. The mean of the distribution is 1.896 with a stand deviation of 1.325. From this results, we can see that our original sample result, $d_{IQR} = -1$, is certainly not unusual because roughly 11% of the values of the differences of interquartile ranges lie below it. Therefore, we may accept the null hypothesis and conclude that the samples are drawn from the same population which means that our sample could be a representative sample taken from the Penn State population.

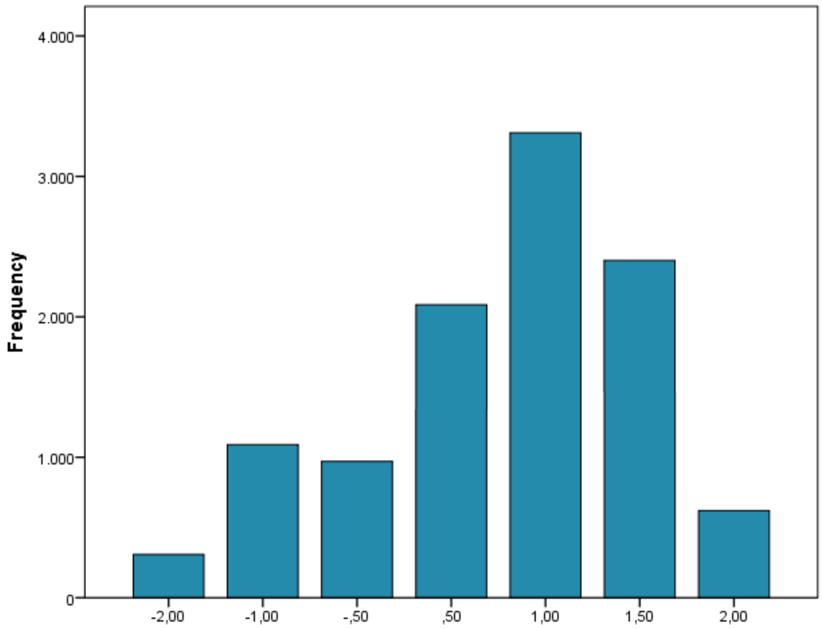


Figure 8.2: A distribution of d^*_{IQR} generated by approximate randomization procedure using our sample and the IBM population.

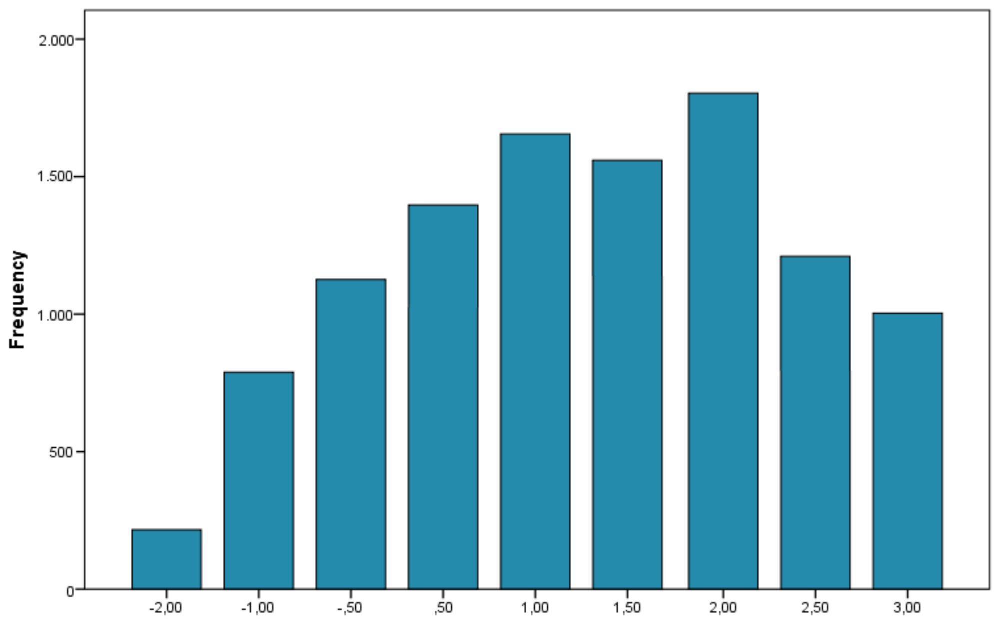


Figure 8.3: A distribution of d^*_{IQR} generated by approximate randomization procedure using our sample and the Penn State population.

8.2 Chapter conclusions

This chapter provided a way of determining whether our sample is representative for the general IBM population as well as for the Penn State population. Using the approximate randomization procedure, where we do not make any assumptions about the population parameters, we may conclude that our sample is representative for the general IBM population as well as the Penn State population.

Chapter 9

Conclusions and Future Work

This chapter gives an overview of the project's contributions, the concluding remarks of our study, and proposals for future work.

9.1 Contributions

Driven by the need to better understand the benefits of gamification techniques in an enterprise context, we performed a study in IBM Netherlands that involved 117 employees. We aimed at increase the understanding of personality traits affect engagement strategies within enterprise gamification. In the pursuit of this goal, we provided three original contributions:
(1) A novel dataset, an extension of a previous effort [78], that contains: a) information about employees' personality traits – as calculated using standard personality questionnaires; b) information about the interaction of such employees with a gamified enterprise tool developed in [78]; and c) personality trait estimates, as derived from state-of-the-art automated techniques implemented by IBM Watson. A complete description of the dataset is provided in Chapter 4. Given the sensible nature of the data contained in the dataset, its content cannot be publicly released; however, it will be available to support future research efforts.
(2) Based on (1), we contribute **an analysis of the personalities of employees in enterprise** (described in Chapter 5), and **an analysis of the relationship between personality traits and game play**, described in Chapter 7. The analysis aimed at understanding if and how personality traits could be used to interpret the behaviour of employees interacting with a gamified enterprise-class application.

We observed the population under study to score significantly lower on *Neuroticism* compared to a reference group not specific to IBM. We could not find strong evidence that this also holds for *Agreeableness*. These findings suggests that the IBM population contains people who are more calm, composed and capable of handling stressful situations compared to people from the same age group. Moreover, the women in our sample are more nurturing, tender-minded, and altruistic more often and to a greater extent than men. In addition, contrary to our predictions, it has also been shown that gender role classification was not reflective for their biological sexes. No differences in the way the two sexes associate themselves with feminine traits were found. However, is has been shown that men score higher in masculine traits.

We then pursued a better understanding of how personality traits played a role in the way people played the IBM Game. Results show that the Big Five personality traits, as well as the gender roles, may use to explain the observed game behaviors of people who have played the IBM game if we do not make a distinction between treatment groups (i.e. different configurations of game mechanics). Otherwise, and coherently with a previous study [78], differences among groups can only be explained by the presence and absence of the game mechanics (leaderboard and badges).

On the other hand, we were able to find many interesting main effects. The effects of including the Big Five dimension and gender roles into the prediction of the game behavior have been confirmed by the Poisson regression analysis. This research suggests that both the Big Five dimensions(*Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism*) and gender roles(*Masculinity* and *Femininity*) are important predictors of game behavior. Results supported that *Conscientiousness* has the strongest influence on engagement, social behavior, expertise, curiosity and controlled behavior. *Openness* was found to be second strongest predictor for the game behaviors. *Masculinity* and *Femininity* were found to have impact on the engagement, social behavior and part of other game elements.

Finally, we explored whether what we observed can be generalized to a wider population. Based on the results we may conclude that using the randomization method for generalization, our sample is representative of the general IBM population and the Penn State population.

(3) We validated the effectiveness of the *Watson Personality Insights service* for personality measurement (described in Chapter 6). We compared the results of questionnaires with the output of the automated tool fed with data coming from 1) enterprise social media (i.e. IBM Connections), and 2) public social media (i.e. Twitter). Subsequently, we studied the possibilities and reliability of Watson Personality Insights service, which is able to derive personality traits automatically based on a given set of input text. This study shows that Twitter data from users provided promising and reliable results compared to the scores determined by the personality survey. Unfortunately, this did not hold when data from IBM Connections was used as input. Using Twitter data four traits (*Openness, Extroversion, Agreeableness,* and *Neuroticism*) were successfully derived, but none of the traits were reflective to the user's personality profile if data from IBM Connections was used. Moreover, we found that similar results related to the relationship between personality and gameplay can be achieved by using the predicted personality information.

Altogether, these contributions improve on the state-of-the-art by offering a comprehensive outlook on the effectiveness of personality traits as a construct to describe and predict employees engagement with gamified applications.

We found strong significant evidence for the fact that personality traits affected the game behaviors. In additional, the results of this work indicated that personality traits did have a significant effect on predicting game behavior. With other words, a strong evidence was found that a relationship exists between personality traits and different game elements, which are part of different game behaviors. The significant insights were presented in Section 7.2.1.

To our best knowledge, this study was the first one that explored the possible influences of personality traits on different game elements and mechanics in enterprise.

Moreover, contrary to other studies, individuals from an international organization with different backgrounds were involved. Our findings carry significant, theoretical and practical implications. The most important one is that it has proven that personality traits did affect the engagement of employees. If more is known about the psychological factors why people like gamification, organizations are better able to influence their employee's engagement, for example, to build implementation that will not disengage (part of) the employees.

Limitations

This work has some methodological limitations that should be taken into account. First, in this study we worked with different samples, due to the fact that not everyone who participated in the IBM game completed our survey. So, to get a better understanding of the different samples we performed a quantity analysis on the different samples. We found that the sample who contributed to the previous study about the IBM game is not significantly different from our sample. However, we could still be missing people that acted differently which could result in different observations.

Second, our sample was relatively small and the participants were randomly assigned to one of the four treatment groups in the IBM game without considering for any personality characteristics. In fact, it might also be one of the reasons why no significant differences in game behavior between the groups were found when controlling for personality traits in the present study. Further research is needed to verify this.

Third, the personality data were based solely on employees' self-reports using two personality questionnaires. It is possible that some people rated themselves differently by selecting a different scale for a given statement which resulted in a deviated reflection of their personality. However, the respondents completed the questionnaires voluntarily and could not gain anything by giving biased responses.

Fourth, Watson Personality Insights service has been studied and used in this research to assess personality traits of people automatically. This tool seems to provide promising and reliable outcomes for the personality scores based on Twitter data. However, this service has several limitation. The models used by the service to determine the personality traits were trained on specific online media, so the computed scores might deviate 2 to 16%[1] when used with input from different media. Unfortunately, the personality scores derived by analyzing data from IBM Connections was not as promising as we hoped. This might due to the fact the service it not trained properly to handle this type of input and therefore deviated more than it should. It might also due to the fact the we did not have enough data(e.g. more than 2000 words) per person for Watson Personality Insight Service to produce reliable inferences for personality scores. This service requires a sufficient number of words in the input text to compute reliable inferences for personality traits. So, in several cases, we did not have enough data crawled to compute the personality scores, so we were forced to exclude several people from our experiment. Another limitation using this tool is that IBM does not recommend combining and analyzing input text from multiple media sources which we actually had in mind since we crawled data from Twitter as well as from IBM Connections.

[1] `https://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/doc/personality-insights/science.shtml#researchLimitations`

9.2 Future work

Our study, being of an exploratory and interpretive nature, raises a number of opportunities for future research, both in terms of methodology and concept validation.

In this study, we highlighted the potential moderating effects that Big Five personality traits and gender roles have on the different game behaviors within enterprise setting. Identifying these differences can assist developers in designing solutions to engage people with different personalities using gamification. In future research, it is recommended to take the underlying facets of each personality dimension into account when performing the regression analysis. The individual facets of the personality factors could provide more insights to explain differences in game behavior. Moreover, the current work on the relationship between personality and employee engagement and game play can be expanded by including personality characteristics such as age, sex and nationality. A main point of interest here is to explore whether these these personality characteristics influence game behavior or elements given a set of game mechanics.

Another important finding is the validity of assessing personality based on Twitter data. This opens a lot of research possibilities. One potential approach could be to build a recommendation system for recommending tweets, which takes the personality of the user into account. This could improve the performance of recommender systems due to the personalized recommendations.

Furthermore, the automatic assessment of personality also opens a variety of opportunities for researchers to study and investigate personality traits of individuals without the need of using self-reports or personality questionnaires. The availability of this tool would enable large-scale studies.

Finally, we recognize the importance of conducting similar studies in other companies to see if the conclusion drawn by the current work are generalizable beyond the IBM use case.

Bibliography

[1] Y. Amichai-Hamburger, H. Kaplan, and N. Dorpatcheon. Click to the past: The impact of extroversion by users of nostalgic websites on the use of internet social services. *Computers in Human Behavior*, 24(5):1907–1912, 2008.

[2] Y. Amichai-Hamburger and G. Vinitzky. Social network use and personality. *Computers in human behavior*, 26(6):1289–1295, 2010.

[3] S. Attfield, G. Kazai, M. Lalmas, and B. Piwowarski. Towards a science of user engagement (position paper). In *WSDM Workshop on User Modelling for Web Applications*, 2011.

[4] H. Badenes, M. N. Bengualid, J. Chen, L. Gou, E. Haber, J. Mahmud, J. W. Nichols, A. Pal, J. Schoudt, B. A. Smith, Y. Xuan, H. Yang, and M. X. Zhou. System u: Automatically deriving personality traits from social media for people recommendation. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 373–374, New York, NY, USA, 2014. ACM.

[5] S. L. Bem. On the utility of alternative procedures for assessing psychological androgyny. *Journal of consulting and clinical psychology*, 45(2):196, 1977.

[6] S. L. Bem. *Bem sex-role inventory: Professional manual*. Consulting Psychologists Press, 1981.

[7] S. Blackman. Comments on three methods of scoring androgyny as a continuous variable. *Psychological Reports*, 51(3f):1100–1102, 1982.

[8] L. A. Burton, J. Hafetz, and D. Henninger. Gender differences in relational and physical aggression. *Social Behavior and Personality: an international journal*, 35(1):41–50, 2007.

[9] W. J. Camara, J. S. Nathan, and A. E. Puente. Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice*, 31(2):141, 2000.

[10] L. F. Carver, A. Vafaei, R. Guerra, A. Freire, and S. P. Phillips. Gender differences: Examination of the 12-item bem sex role inventory (bsri-12) in an older brazilian population. *PloS one*, 8(10):e76356, 2013.

[11] H. E. Cattell. The original big five: A historical perspective. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée*, 1996.

[12] T. Chamorro-Premuzic and A. Furnham. Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, 37(4):319–338, 2003.

[13] B. P. Chapman, P. R. Duberstein, S. Sörensen, and J. M. Lyness. Gender differences in five factor model personality traits in an elderly cohort. *Personality and individual differences*, 43(6):1594–1603, 2007.

[14] Y.-k. Chou. Top 10 enterprise gamification cases employees productive. *Gamification Examples*, 2014.

[15] D. Codish and G. Ravid. Personality based gamification-educational gamification for extroverts and introverts. In *Proc. 9 th Chais Conf. for the Study of Innovation and Learning Technologies: Learning in the Technological Era*, 2014.

[16] P. R. Cohen. Empirical methods for artificial intelligence. *IEEE Intelligent Systems*, (6):88, 1996.

[17] T. Correa, A. W. Hinsley, and H. G. De Zuniga. Who interacts on the web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2):247–253, 2010.

[18] P. T. Costa and R. R. MacCrae. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO FFI): Professional Manual*. Psychological Assessment Resources, 1992.

[19] S. Deterding. Gamification: designing for motivation. *interactions*, 19(4):14–17, 2012.

[20] J. B. Ellis and L. M. Range. Femininity and reasons for living. *Educational & Psychological Research*, 1988.

[21] H. J. Eysenck and S. B. G. Eysenck. *Manual of the Eysenck Personality Questionnaire (junior and adult)*. Hodder and Stoughton, 1975.

[22] G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock. Recognising personality traits using facebook status updates. In *Proceedings of the workshop on computational personality recognition (WCPR13) at the 7th international AAAI conference on weblogs and social media (ICWSM13)*, 2013.

[23] R. Farzan, J. M. DiMicco, D. R. Millen, C. Dugan, W. Geyer, and E. A. Brownholtz. Results from deploying a participation incentive mechanism within the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 563–572. ACM, 2008.

[24] L. A. Fast and D. C. Funder. Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. *Journal of personality and social psychology*, 94(2):334, 2008.

[25] L. S. Ferro, S. P. Walz, and S. Greuter. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, page 7. ACM, 2013.

[26] L. J. Francis and S. H. Jones. The relationship between the myers-briggs type indicator and the eysenck personality questionnaire among adult churchgoers. *Pastoral Psychology*, 48(5):377–386, 2000.

[27] W. L. Gardner and M. J. Martinko. Using the myers-briggs type indicator to study managers: A literature review and research agenda. *Journal of Management*, 22(1):45–83, 1996.

[28] J. A. Gliem and R. R. Gliem. Calculating, interpreting, and reporting cronbachâs alpha reliability coefficient for likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, 2003.

[29] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting personality from twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 149–156. IEEE, 2011.

[30] L. R. Goldberg. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7:7–28, 1999.

[31] S. D. Gosling, S. J. Ko, T. Mannarelli, and M. E. Morris. A room with a cue: personality judgments based on offices and bedrooms. *Journal of personality and social psychology*, 82(3):379, 2002.

[32] C. S. Hall and G. Lindzey. Theories of personality. 1957.

[33] J. Hamari. Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic commerce research and applications*, 12(4):236–245, 2013.

[34] J. Hamari and V. Eranti. Framework for designing and evaluating game achievements. *Proc. DiGRA 2011: Think Design Play*, 115:122–134, 2011.

[35] J. Hamari, J. Koivisto, and H. Sarsa. Does gamification work?–a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, pages 3025–3034. IEEE, 2014.

[36] Y. A. Hamburger and E. Ben-Artzi. The relationship between extraversion and neuroticism and the different uses of the internet. *Computers in human behavior*, 16(4):441–449, 2000.

[37] P. Herzig, M. Ameling, and A. Schill. A generic platform for enterprise gamification. In *Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference on*, pages 219–223. IEEE, 2012.

[38] R. M. Hoffman and L. D. Borders. Twenty-five years after the bem sex-role inventory: A reassessment and new issues regarding classification variability. *Measurement and Evaluation in Counseling and Development*, 34(1):39–55, 2001.

[39] I. Inceoglu and P. Warr. Personality and job engagement. *Journal of Personnel Psychology*, 2015.

[40] S.-P. Jeng and C.-I. Teng. Personality and motivations for playing online games. *Social Behavior and Personality: an international journal*, 36(8):1053–1060, 2008.

[41] O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

[42] J. A. Johnson. Predicting observers' ratings of the big five from the cpi, hpi, and neo–pi–r: a comparative validity study. *European Journal of Personality*, 14(1):1–19, 2000.

[43] J. A. Johnson. Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of research in personality*, 39(1):103–129, 2005.

[44] J. A. Johnson. Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120. *Journal of Research in Personality*, 51:78–89, 2014.

[45] J. A. Johnson. Personal Communication, 2014-2015.

[46] T. A. Judge, J. E. Bono, and E. A. Locke. Personality and job satisfaction: the mediating role of job characteristics. *Journal of applied psychology*, 85(2):237, 2000.

[47] T. A. Judge, D. Heller, and M. K. Mount. Five-factor model of personality and job satisfaction: a meta-analysis. *Journal of applied psychology*, 87(3):530, 2002.

[48] T. A. Judge, C. A. Higgins, C. J. Thoresen, and M. R. Barrick. The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3):621–652, 1999.

[49] S. Kassin. *Psychology*. Pearson/Prentice Hall, 2004.

[50] M. Komarraju and S. J. Karau. The relationship between the big five personality traits and academic motivation. *Personality and individual differences*, 39(3):557–567, 2005.

[51] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[52] J. Kumar. *Gamification at work: designing engaging business software*. Springer, 2013.

[53] R. N. Landers and J. W. Lounsbury. An investigation of big five and narrow personality traits in relation to internet usage. *Computers in Human Behavior*, 22(2):283–293, 2006.

[54] S. Langelaan, A. B. Bakker, L. J. Van Doornen, and W. B. Schaufeli. Burnout and work engagement: Do individual differences make a difference? *Personality and individual differences*, 40(3):521–532, 2006.

[55] K. Lee, J. Mahmud, J. Chen, M. Zhou, and J. Nichols. Who will retweet this?: Automatically identifying and engaging strangers on twitter to spread information. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 247–256. ACM, 2014.

[56] A. Locksley, E. Borgida, N. Brekke, and C. Hepburn. Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, 39(5):821, 1980.

[57] J. Mahmud, M. X. Zhou, N. Megiddo, J. Nichols, and C. Drews. Recommending targeted strangers from whom to solicit information on social media. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 37–48. ACM, 2013.

[58] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*, pages 457–500, 2007.

[59] G. Matthews, I. J. Deary, and M. C. Whiteman. *Personality traits*. Cambridge University Press, 2003.

[60] R. R. McCrae and P. T. Costa Jr. A five-factor theory of personality. *Handbook of personality: Theory and research*, 2:139–153, 1999.

[61] T. L. Milfont and C. G. Sibley. The big five personality traits and environmental engagement: Associations at the individual and societal level. *Journal of Environmental Psychology*, 32(2):187–195, 2012.

[62] I. B. Myers, M. H. McCaulley, and R. Most. *Manual, a guide to the development and use of the Myers-Briggs type indicator*. Consulting Psychologists Press, 1985.

[63] I. B. Myers, M. H. McCaulley, N. L. Quenk, and A. L. Hammer. *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*, volume 3. Consulting Psychologists Press Palo Alto, CA, 1998.

[64] L. P. Naumann, S. Vazire, P. J. Rentfrow, and S. D. Gosling. Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin*, 2009.

[65] M. A. S. N. Nunes, J. S. Bezerra, and A. A. de Oliveira. Personalityml: a markup language to standardize the user personality in recommender systems. *GEINTEC-Gestão, Inovação e Tecnologias*, 2(3):255–273, 2012.

[66] H. L. O'Brien and E. G. Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955, 2008.

[67] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. The development and psychometric properties of liwc2007, 2007.

[68] E. J. Phares. *Introduction to personality*. Good Year Books, 1991.

[69] E. C. Prakash and M. Rao. Introduction to gamification in enterprises. In *Transforming Learning and IT Management through Gamification*, pages 47–72. Springer, 2015.

[70] M. Rauch. Best practices for using enterprise gamification to engage employees and customers. In *Human-Computer Interaction. Applications and Services*, pages 276–283. Springer, 2013.

[71] P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods: the rv-coefficient. *Applied statistics*, pages 257–265, 1976.

[72] M. Rosenberg. Society and the adolescent self-image. 1965.

[73] C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr. Personality and motivations associated with facebook use. *Computers in human behavior*, 25(2):578–586, 2009.

[74] G. Rubinstein. The big five among male and female students of different faculties. *Personality and Individual Differences*, 38(7):1495–1503, 2005.

[75] R. Ryckman. *Theories of personality*. Cengage Learning, 2012.

[76] V. E. Schein. Relationships between sex role stereotypes and requisite management characteristics among female managers. *Journal of applied psychology*, 60(3):340, 1975.

[77] R.-J. Sips, A. Bozzon, G. Smit, and G.-J. Houben. The inclusive enterprise: Vision and roadmap. In *Engineering the Web in the Big Data Era*, pages 621–624. Springer, 2015.

[78] A. B. Stanculescu, Laurentiu Catalin and R.-J. Sips. Driving engagement and online social behavior of employees in an enterprise environment, 2014.

[79] L. C. Stanculescu, A. Bozzon, R.-J. Sips, and G.-J. Houben. Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 346–358, New York, NY, USA, 2016. ACM.

[80] J. Thom, D. Millen, and J. DiMicco. Removing gamification from an enterprise sns. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1067–1070. ACM, 2012.

[81] H. Wang and C.-T. Sun. Game reward systems: gaming experiences and social meanings. In *Proceedings of DiGRA 2011 Conference: Think Design Play*, pages 1–12, 2011.

[82] J. Webster and H. Ho. Audience engagement in multimedia presentations. *ACM SIGMIS Database*, 28(2):63–77, 1997.

[83] Y. J. Weisberg, C. G. DeYoung, and J. B. Hirsh. Gender differences in personality across the ten aspects of the big five. 2011.

[84] T. Yarkoni. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373, 2010.

Appendix A

The survey

In this appendix we provide the two personality questionnaires which we have combined and used in our survey to determine the personality traits of the IBM employees.

General Information

My Personality Inventory

This is a survey which determines your personality based on two questionnaires.

The first one is the International Personality Item Pool Representation of the NEO PI-R™ (IPIP-NEO).

The second one is The Bem Sex Role Inventory (BSRI).

The personality inventories

The 120-item NEO-FFI is a shortened version of the NEO PI-R that provides a quick, reliable, and accurate measure of the five domains of adult personality.
More information about the IPIP can be found on: http://ipip.ori.org/

The BSRI included 60 dichotomous items divided into 3 subscales (Masculinity, Femininity, and Neutral) of 20 items each.

More information about the BSRI can be found on http://psycnet.apa.org/journals/psp/31/4/634/

Directions

This survey requires about **15~30 minutes** in total and consist of the following **four parts**:

First, you will be asked to provide some general information.

Second, you will be asked to answers some of the questions from the IPIP NEO Inventory.

Each question has a 5 point scale, varying from *Very Inaccurate* to *Very Accurate*. Based on this scale you are asked to rate yourself for each question.

Third, you will be asked to answers the BSRI questionnaire.

For each of the questions you have to rate yourself on a scale from *never or almost never true* to *always or almost always true.*

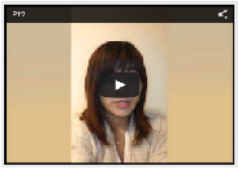
Fourth, you will be asked some general questions about this survey.

If you have any questions please contact me.

Informed consent statement

Hi all,

My name is Lie Yen Cheung. I am a Computer Science student from Delft and currently I am doing an internship for my master thesis here at IBM in the Netherlands. In my thesis I would like to investigate whether personality traits and skills are related with somehow people played the game "How much of an IMBer are you?". For example, people who are extroverts and competitive play the game more often to get a higher score. Based on this type of analysis I would like to research if there is a relation between a given user activity or activeness and given a personality trait. The results will be used to optimize and improve the social game "How much of an IMBer are you?". In order to achieve this, I need your help by filling in the survey, so I can collect data for my research. The potential benefits to you includes feedback about your personality and learning about psychology.



Informed consent statement:

- Your participation is voluntary.
- Your participation in this study will be kept confidential
- The questionnaires does not reveal hidden, secret information about you nor does it assess serious psychological disorders.
- Your data will be safely stored and used only and exclusively for academic research purposes, and will be accessible **exclusively** by the members of the research team.
- The gathered data will be not be distributed to any third a party and they are never used for commercial purposes.
- Please answer the questions as honestly and accurately as possible.
- Please make sure you respond to all the items and do not leave any blanks
- It is assumed that you are taking the test purely for interest: you should never use the information given here for any serious "real life" purposes. It is not psychological advice or a diagnostic of any kind.
- If you do not feel comfortable after submitting the survey you could contact me (lieyen.cheung@nl.ibm.com) and I will remove your data from the database and will not be using it in my study.

***Acknowledgment that you understand the limitations of the test results and agree with the conditions mentioned above.**

Before you can access the test, you must endorse the statement below by clicking the checkbox.

☐ Yes, I understand that the primary purpose of this survey is to educate the respondent about the five factor model of personality and the Bex Sex Role Inventory (BSRI). Secondarily, this survey is used to estimate the respondent's standing within the five factor model and the BSRI. These are two different inventories to characterize personality. The five factor model characterizes a personality within five factors (openness, conscientiousness, extraversion, agreeableness, and neuroticism), while the BSRI characterize the personality of the respondent as masculine, feminine, androgynous, or undifferentiated. Moreover, I understand my responses may be anonymously collected for academic research purposes as outlined in the informed consent statement. Furthermore, I understand that the program that generates the report is designed to produce estimates that are as accurate as possible, but that measurement error or improper responding can produce inaccurate results.

General questions

* **Please enter your name**

First name

Last name

The reason for your to enter your name is because I would like to personalize the personality report by calling you by your name.

* **Select your gender**

Female Male

* **Please enter your email adress**

* **How old are you?**

Only numbers may be entered in this field.

The International Personality Item Pool

These are the questions for NEO IPIP 120 items which is used to determine your personality.
For each of the following items, please select the option which describes you the best.

Directions: The following statements concern your perception about yourself in a variety of situations. Your task is to indicate the strength of how accurately each item describes your, utilizing a 5 point scale varying from *Very Inaccurate* to *Very Accurate*. Very Inaccurate denotes strong disagreement, Very Accurate denotes strong agreement, and the other options represent intermediate judgments. In the boxes after each statement, click option which describes you the best.

There are no "right" or "wrong" answers, so select the option that most closely reflects you on each statement.

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Worry about things					
Make friends easily					
Have a vivid imagination					
Trust others					
Complete tasks successfully					
Get angry easily					
Love large parties					
Believe in the importance of art					
Use others for my own ends					
Like to tidy up					
Often feel blue					
Take charge					
Experience my emotions intensely					
Love to help others					
Keep my promises					
Find it difficult to approach others					
Am always busy					
Prefer variety to routine					
Love a good fight					
Work hard					
Go on binges					
Love excitement					
Love to read challenging material					
Believe that I am better than others					
Am always prepared					

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Remain calm under pressure					
Radiate joy					
Tend to vote for liberal political candidates					
Sympathize with the homeless					
Jump into things without thinking					
Fear for the worst					
Feel comfortable around people					
Enjoy wild flights of fantasy					
Believe that others have good intentions					
Excel in what I do					
Get irritated easily					
Talk to a lot of different people at parties					
See beauty in things that others might not notice					
Cheat to get ahead					
Often forget to put things back in their proper place					
Dislike myself					
Try to lead others					
Feel others' emotions					
Am concerned about others					
Tell the truth					
Am afraid to draw attention to myself					
Am always on the go					
Prefer to stick with things that I know					
Yell at people					
Do more than what's expected of me					

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Rarely overindulge					
Seek adventure					
Avoid philosophical discussions					
Think highly of myself					
Carry out my plans					
Am calm even in tense situations					
Have a lot of fun					
Believe that there is no absolute right or wrong					
Feel sympathy for those who are worse off than myself					
Make rash decisions					
Am afraid of many things					
Avoid contacts with others					
Love to daydream					
Trust what people say					
Handle tasks smoothly					
Lose my temper					
Prefer to be alone					
Do not like poetry					
Take advantage of others					
Leave a mess in my room					
Am often down in the dumps					
Take control of things					
Rarely notice my emotional reactions					
Am indifferent to the feelings of others					
Break rules					

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Only feel comfortable with friends					
Do a lot in my spare time					
Dislike changes					
Insult people					
Do just enough work to get by					
Often eat too much					
Enjoy being reckless					
Have difficulty understanding abstract ideas					
Have a high opinion of myself					
Waste my time					
Feel that I'm unable to deal with things					
Love life					
Tend to vote for conservative political candidates					
Am not interested in other people's problems					
Rush into things					
Get stressed out easily					
Keep others at a distance					
Like to get lost in thought					
Distrust people					
Know how to get things done					
Am not easily annoyed					
Avoid crowds					
Do not enjoy going to art museums					
Obstruct others' plans					
Leave my belongings around					

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Feel comfortable with myself					
Wait for others to lead the way					
Don't understand people who get emotional					
Take no time for others					
Break my promises					
Am not bothered by difficult social situations					
Like to take it easy					
Am attached to conventional ways					
Get back at others					
Put little time and effort into my work					
Am able to control my cravings					
Act wild and crazy					
Am not interested in theoretical discussions					
Boast about my virtues					
Have difficulty starting tasks					
Know how to cope					
Look at the bright side of life					
Believe that we should be tough on crime					
Try not to think about the needy					
Act without thinking					

The Bem Sex Role Inventory

* These are the questions for BSRI 60 items which is used to determine your personality.
For each of the following items, please select the option which describes you the best.

Directions:
The following statements concern your perception about yourself in a variety of situations. Your task is to rate yourself on each item, utilizing a 7 point scale varying from *Never or almost never true* to *always or almost always true.*

Never or almost never true denotes strong disagreement, *always or almost always true* denotes strong agreement, and other options represent intermediate judgments. In the boxes after each statement, click option which describes you the best.

There are no "right" or "wrong" answers, so select the option that most closely reflects you on each statement.

	NEVER OR ALMO! NEVER TRUE	USUALLY NOT TRUE	SOMETIMES, BUT INFREQUENTLY TRUE	OCCASIONALLY TRUE	OFTEN TRUE	USUALLY TRUE	ALWAYS OR ALMOST ALWAYS TRUE
DEFEND MY OWN BELIEFS	○	○	○	○	○	○	○
AFFECTIONATE	○	○	○	○	○	○	○
CONSCIENTIOUS	○	○	○	○	○	○	○
INDEPENDENT	○	○	○	○	○	○	○
SYMPATHETIC	○	○	○	○	○	○	○
MOODY	○	○	○	○	○	○	○
ASSERTIVE	○	○	○	○	○	○	○
SENSITIVE TO OTHERS' NEEDS	○	○	○	○	○	○	○
RELIABLE	○	○	○	○	○	○	○
STRONG PERSONALITY	○	○	○	○	○	○	○
UNDERSTANDING	○	○	○	○	○	○	○
JEALOUS	○	○	○	○	○	○	○
FORCEFUL	○	○	○	○	○	○	○
COMPASSIONATE	○	○	○	○	○	○	○
TRUTHFUL	○	○	○	○	○	○	○
HAVE LEADERSHIP ABILITIES	○	○	○	○	○	○	○
EAGER TO SOOTHE FEELINGS	○	○	○	○	○	○	○
SECRETIVE	○	○	○	○	○	○	○
WILLING TO TAKE RISKS	○	○	○	○	○	○	○
WARM	○	○	○	○	○	○	○
ADAPTABLE	○	○	○	○	○	○	○
DOMINANT	○	○	○	○	○	○	○
TENDER	○	○	○	○	○	○	○
CONCEITED	○	○	○	○	○	○	○
WILLING TO TAKE A STAND	○	○	○	○	○	○	○

	NEVER OR ALMO! NEVER TRUE	USUALLY NOT TRUE	SOMETIMES, BUT INFREQUENTLY TRUE	OCCASIONALLY TRUE	OFTEN TRUE	USUALLY TRUE	ALWAYS OR ALMOST ALWAYS TRUE
LOVE CHILDREN	○	○	○	○	○	○	○
TACTFUL	○	○	○	○	○	○	○
AGGRESSIVE	○	○	○	○	○	○	○
GENTLE	○	○	○	○	○	○	○
CONVENTIONAL	○	○	○	○	○	○	○
SELF-RELIANT	○	○	○	○	○	○	○
YIELDING	○	○	○	○	○	○	○
HELPFUL	○	○	○	○	○	○	○
ATHLETIC	○	○	○	○	○	○	○
CHEERFUL	○	○	○	○	○	○	○
UNSYSTEMATIC	○	○	○	○	○	○	○
ANALYTICAL	○	○	○	○	○	○	○
SHY	○	○	○	○	○	○	○
INEFFICIENT	○	○	○	○	○	○	○
MAKES DECISIONS EASILY	○	○	○	○	○	○	○
FLATTERABLE	○	○	○	○	○	○	○
THEATRICAL	○	○	○	○	○	○	○
SELF-SUFFICIENT	○	○	○	○	○	○	○
LOYAL	○	○	○	○	○	○	○
HAPPY	○	○	○	○	○	○	○
INDIVIDUALISTIC	○	○	○	○	○	○	○
SOFT-SPOKEN	○	○	○	○	○	○	○
UNPREDICTABLE	○	○	○	○	○	○	○
MASCULINE	○	○	○	○	○	○	○
GULLIBLE	○	○	○	○	○	○	○

	NEVER OR ALMO! NEVER TRUE	USUALLY NOT TRUE	SOMETIMES, BUT INFREQUENTLY TRUE	OCCASIONALLY TRUE	OFTEN TRUE	USUALLY TRUE	ALWAYS OR ALMOST ALWAYS TRUE
SOLEMN	○	○	○	○	○	○	○
COMPETITIVE	○	○	○	○	○	○	○
CHILDLIKE	○	○	○	○	○	○	○
LIKABLE	○	○	○	○	○	○	○
AMBITIOUS	○	○	○	○	○	○	○
DO NOT USE HARSH LANGUAGE	○	○	○	○	○	○	○
SINCERE	○	○	○	○	○	○	○
ACT AS A LEADER	○	○	○	○	○	○	○
FEMININE	○	○	○	○	○	○	○
FRIENDLY	○	○	○	○	○	○	○

Optional questions

Thank you! The survey is completed. While you can now submit, we ask another minute of your time to go through the last part of this survey.

As mentioned before I am a master student from Delft University of Technology, and currently I am working on my master thesis to study the behaviour of employees in enterprise with the ultimate goal of improving the quality and enjoyability of work and optimizing and improving systems used in enterprise, while guaranteeing high performance. This survey is part of my research activities.

I will be working on more experiments in the future, so if you are interested to participate in future survey/experiment please fill in the questions below.

Would you like to perform more surveys of this type?

○ Yes ○ No

Do you want to participate in futher research?

○ Yes ○ No

Do you have some feedback to share with me about the survey you just completed?

Appendix B

The detailed report about personality traits

In this appendix we provide an example of a detailed report[1] which is generated instantly after the participants have entered their ratings for each of the inventories.

The first report describes how a person fits into the IPIP questionnaire. For each of the single main trait a detailed description is provided along with a description for each of the subtraits based on the individual scores.

The second report describes how a person can be classified based on their given answers to the BSRI. The report states whether a person is Androgynous, undifferentiated, Masculine or Feminine.

[1]These example reports do not represent who I am, I generated these just to give an indication what kind of feedback the participant would receive after filling in the personality survey. The reports are generated based on randomly answers provided to the items in both personality inventories.

Report for the IPIP

Dear LieYen, you just completely filled in the IPIP Questionnaire. This is a detailed report of your results based on your given answers.

This report estimates the individual's level on each of the five broad personality domains of the Five-Factor Model. The description of each one of the five broad domains is followed by a more detailed description* of personality according to the six subdomains that comprise each domain.

Please keep in mind that "low," "average," and "high" scores on a personality test are neither absolutely good nor bad. A particular level on any trait will probably be neutral or irrelevant for a great many activities, be helpful for accomplishing some things, and detrimental for accomplishing other things. As with any personality inventory, scores and descriptions can only approximate an individual's actual personality. High and low score descriptions are usually accurate, but average scores close to the low or high boundaries might misclassify you as only average.

On each set of six subdomain scales it is somewhat uncommon but certainly possible to score high in some of the subdomains and low in the others. In such cases more attention should be paid to the subdomain scores than to the broad domain score. Questions about the accuracy of your results are best resolved by showing your report to people who know you well.

** The description of each of the five domains and thirty subdomains is provided by Dr. John A. Johnson, Professor of Psychology, Pennsylvania State University.*

These are your scores for the IPIP NEO Inventory, based on the answers you have given in the IPIP NEO questionnaire.

There are 5 tables which stand for each of the five factors, the columns are the facets described in each factor.

OPENNESS TO EXPERIENCE	**Imagination**	**Artistic Interests**	**Emotionality**	**Adventurousness**	**Intellect**	**Liberalism**
3	0	3	3	4.5	4.5	3
average	low	high	high	high	high	high

CONSCIENTIOUSNESS	**Self-Efficacy**	**Orderliness**	**Dutifulness**	**Achievement-Striving**	**Self-Discipline**	**Cautiousness**
3.25	0	4.5	3	3	3	6
average	low	high	high	high	high	high

EXTRAVERSION	**Friendliness**	**Gregariousness**	**Assertiveness**	**Activity Level**	**Excitement-Seeking**	**Cheerfulness**
1.25	3	3	1.5	0	0	0
low	high	high	low	low	low	low

AGREEABLENESS	**Trust**	**Morality**	**Altruism**	**Cooperation**	**Modesty**	**Sympathy**
4.25	1.5	6	3	6	6	3
high	low	high	high	high	high	high

NEUROTICISM	**Anxiety**	**Anger**	**Depression**	**Self-Consciousness**	**Immoderation**	**Vulnerability**
2	0	1.5	1.5	1.5	3	4.5
average	low	low	low	low	high	high

Detailed description per factor and per scale

Openness to Experience

Openness to Experience describes a dimension of cognitive style that distinguishes imaginative, creative people from down-to-earth, conventional people. Open people are intellectually curious, appreciative of art, and sensitive to beauty. They tend to be, compared to closed people, more aware of their feelings. They tend to think and act in individualistic and nonconforming ways. Intellectuals typically score high on Openness to Experience; consequently, this factor has also been called Culture or Intellect. Nonetheless, Intellect is probably best regarded as one aspect of openness to experience. Scores on Openness to Experience are only modestly related to years of education and scores on standard intelligent tests.

Another characteristic of the open cognitive style is a facility for thinking in symbols and abstractions far removed from concrete experience. Depending on the individual's specific intellectual abilities, this symbolic cognition may take the form of mathematical, logical, or geometric thinking, artistic and metaphorical use of language, music composition or performance, or one of the many visual or performing arts. People with low scores on openness to experience tend to have narrow, common interests. They prefer the plain, straightforward, and obvious over the complex, ambiguous, and subtle. They may regard the arts and sciences with suspicion, regarding these endeavors as abstruse or of no practical use. Closed people prefer familiarity over novelty; they are conservative and resistant to change.

Openness is often presented as healthier or more mature by psychologists, who are often themselves open to experience. However, open and closed styles of thinking are useful in different environments. The intellectual style of the open person may serve a professor well, but research has shown that closed thinking is related to superior job performance in police work, sales, and a number of service occupations.

Your score is: ***average***

- *If your score on Openness to Experience is low, indicating you like to think in plain and simple terms. Others describe you as down-to-earth, practical, and conservative.*
- *If your score on Openness to Experience is average, indicating you enjoy tradition but are willing to try new things. Your thinking is neither simple nor complex. To others you appear to be a well-educated person but not an intellectual.*
- *If your score on Openness to Experience is high, indicating you enjoy novelty, variety, and change. You are curious, imaginative, and creative.*

Openness Facets

- Imagination. *To imaginative individuals, the real world is often too plain and ordinary. High scorers on this scale use fantasy as a way of creating a richer, more interesting world. Low scorers are on this scale are more oriented to facts than fantasy. Your level of imagination is* ***low***.
- Artistic Interests. *High scorers on this scale love beauty, both in art and in nature. They become easily involved and absorbed in artistic and natural events. They are not necessarily artistically trained nor talented, although many will be. The defining features of this scale are interest in, and appreciation of natural and artificial beauty. Low scorers lack aesthetic sensitivity and interest in the arts. Your level of artistic interests is* ***high***.
- Emotionality. *Persons high on Emotionality have good access to and awareness of their own feelings. Low scorers are less aware of their feelings and tend not to express their emotions openly. Your level of emotionality is* ***high.***
- Adventurousness. *High scorers on adventurousness are eager to try new activities, travel to foreign lands, and experience different things. They find familiarity and routine boring, and will take a new route home just because it is different. Low scorers tend to feel uncomfortable with change and prefer familiar routines. Your level of adventurousness is* ***high***.
- Intellect. *Intellect and artistic interests are the two most important, central aspects of openness to experience. High scorers on Intellect love to play with ideas. They are open-minded to new and unusual ideas, and like to debate intellectual issues. They enjoy riddles, puzzles, and brain teasers. Low scorers on Intellect prefer dealing with either people or things rather than ideas. They regard intellectual exercises as a waste of time. Intellect should not be equated with intelligence. Intellect is an intellectual style, not an intellectual ability, although high scorers on Intellect score slightly higher than low-Intellect individuals on standardized intelligence tests. Your level of intellect is* ***high***.
- Liberalism. *Psychological liberalism refers to a readiness to challenge authority, convention, and traditional values. In its most extreme form, psychological liberalism can even represent outright hostility toward rules, sympathy for law-breakers, and love of ambiguity, chaos, and disorder. Psychological conservatives prefer the security and stability brought by conformity to tradition. Psychological liberalism and conservatism are not identical to political affiliation, but certainly incline individuals toward certain political parties. Your level of liberalism is* ***high***.

Conscientiousness

Conscientiousness concerns the way in which we control, regulate, and direct our impulses. Impulses are not inherently bad; occasionally time constraints require a snap decision, and acting on our first impulse can be an effective response. Also, in times of play rather than work, acting spontaneously and impulsively can be fun. Impulsive individuals can be seen by others as colorful, fun-to-be-with, and zany.

Nonetheless, acting on impulse can lead to trouble in a number of ways. Some impulses are antisocial. Uncontrolled antisocial acts not only harm other members of society, but also can result in retribution toward the perpetrator of such impulsive acts. Another problem with impulsive acts is that they often produce immediate rewards but undesirable, long-term consequences. Examples include excessive socializing that leads to being fired from one's job, hurling an insult that causes the breakup of an important relationship, or using pleasure-inducing drugs that eventually destroy one's health.

Impulsive behavior, even when not seriously destructive, diminishes a person's effectiveness in significant ways. Acting impulsively disallows contemplating alternative courses of action, some of which would have been wiser than the impulsive choice. Impulsivity also sidetracks people during projects that require organized sequences of steps or stages. Accomplishments of an impulsive person are therefore small, scattered, and inconsistent.

A hallmark of intelligence, what potentially separates human beings from earlier life forms, is the ability to think about future consequences before acting on an impulse. Intelligent activity involves contemplation of long-range goals, organizing and planning routes to these goals, and persisting toward one's goals in the face of short-lived impulses to the contrary. The idea that intelligence involves impulse control is nicely captured by the term prudence, an alternative label for the Conscientiousness domain. Prudent means both wise and cautious. Persons who score high on the Conscientiousness scale are, in fact, perceived by others as intelligent.

The benefits of high conscientiousness are obvious. Conscientious individuals avoid trouble and achieve high levels of success through purposeful planning and persistence. They are also positively regarded by others as intelligent and reliable. On the negative side, they can be compulsive perfectionists and workaholics. Furthermore, extremely conscientious individuals might be regarded as stuffy and boring. Unconscientious people may be criticized for their unreliability, lack of ambition, and failure to stay within the lines, but they will experience many short-lived pleasures and they will never be called stuffy.

*Your score is: **average***

- *If your score on Conscientiousness is low, indicating you like to live for the moment and do what feels good now. Your work tends to be careless and disorganized.*
- *If your score on Conscientiousness is average. This means you are reasonably reliable, organized, and self-controlled.*
- *If your score on Conscientiousness is high. This means you set clear goals and pursue them with determination. People regard you as reliable and hard-working.*

Conscientiousness Facets

- Self-Efficacy. *Self-Efficacy describes confidence in one's ability to accomplish things. High scorers believe they have the intelligence (common sense), drive, and self-control necessary for achieving success. Low scorers do not feel effective, and may have a sense that they are not in control of their lives. Your level of self-efficacy is **low**.*
- Orderliness. *Persons with high scores on orderliness are well-organized. They like to live according to routines and schedules. They keep lists and make plans. Low scorers tend to be disorganized and scattered. Your level of orderliness is **high**.*
- Dutifulness. *This scale reflects the strength of a person's sense of duty and obligation. Those who score high on this scale have a strong sense of moral obligation. Low scorers find contracts, rules, and regulations overly confining. They are likely to be seen as unreliable or even irresponsible. Your level of dutifulness is **high**.*
- Achievement-Striving. *Individuals who score high on this scale strive hard to achieve excellence. Their drive to be recognized as successful keeps them on track toward their lofty goals. They often have a strong sense of direction in life, but extremely high scores may be too single-minded and obsessed with their work. Low scorers are content to get by with a minimal amount of work, and might be seen by others as lazy. Your level of achievement striving is **high**.*
- Self-Discipline. *Self-discipline-what many people call will-power-refers to the ability to persist at difficult or unpleasant tasks until they are completed. People who possess high self-discipline are able to overcome reluctance to begin tasks and stay on track despite distractions. Those with low self-discipline procrastinate and show poor follow-through, often failing to complete tasks-even tasks they want very much to complete. Your level of self-discipline is **high**.*
- Cautiousness. *Cautiousness describes the disposition to think through possibilities before acting. High scorers on the Cautiousness scale take their time when making decisions. Low scorers often say or do first thing that comes to mind without deliberating alternatives and the probable consequences of those alternatives. Your level of cautiousness is **high**.*

Extraversion

Extraversion is marked by pronounced engagement with the external world. Extraverts enjoy being with people, are full of energy, and often experience positive emotions. They tend to be enthusiastic, action-oriented, individuals who are likely to say "Yes!" or "Let's go!" to opportunities for excitement. In groups they like to talk, assert themselves, and draw attention to themselves.

Introverts lack the exuberance, energy, and activity levels of extraverts. They tend to be quiet, low-key, deliberate, and disengaged from the social world. Their lack of social involvement should not be interpreted as shyness or depression; the introvert simply needs less stimulation than an extravert and prefers to be alone. The independence and reserve of the introvert is sometimes mistaken as unfriendliness or arrogance. In reality, an introvert who scores high on the agreeableness dimension will not seek others out but will be quite pleasant when approached.

Your score is: ***low***

- *If your score on Extraversion is low, indicating you are introverted, reserved, and quiet. You enjoy solitude and solitary activities. Your socializing tends to be restricted to a few close friends.*
- *If your score on Extraversion is average, indicating you are neither a subdued loner nor a jovial chatterbox. You enjoy time with others but also time alone.*
- *If your score on Extraversion is high, indicating you are sociable, outgoing, energetic, and lively. You prefer to be around people much of the time.*

Extraversion Facets

- Friendliness. *Friendly people genuinely like other people and openly demonstrate positive feelings toward others. They make friends quickly and it is easy for them to form close, intimate relationships. Low scorers on Friendliness are not necessarily cold and hostile, but they do not reach out to others and are perceived as distant and reserved. Your level of friendliness is* ***high****.*
- Gregariousness. *Gregarious people find the company of others pleasantly stimulating and rewarding. They enjoy the excitement of crowds. Low scorers tend to feel overwhelmed by, and therefore actively avoid, large crowds. They do not necessarily dislike being with people sometimes, but their need for privacy and time to themselves is much greater than for individuals who score high on this scale. Your level of gregariousness is* ***high****.*
- Assertiveness. *High scorers Assertiveness like to speak out, take charge, and direct the activities of others. They tend to be leaders in groups. Low scorers tend not to talk much and let others control the activities of groups. Your level of assertiveness is* ***low****.*
- Activity Level. *Active individuals lead fast-paced, busy lives. They move about quickly, energetically, and vigorously, and they are involved in many activities. People who score low on this scale follow a slower and more leisurely, relaxed pace. Your activity level is* ***low****.*
- Excitement-Seeking. *High scorers on this scale are easily bored without high levels of stimulation. They love bright lights and hustle and bustle. They are likely to take risks and seek thrills. Low scorers are overwhelmed by noise and commotion and are adverse to thrill-seeking. Your level of excitement-seeking is* ***low****.*
- Cheerfulness. *This scale measures positive mood and feelings, not negative emotions (which are a part of the Neuroticism domain). Persons who score high on this scale typically experience a range of positive feelings, including happiness, enthusiasm, optimism, and joy. Low scorers are not as prone to such energetic, high spirits. Your level of positive emotions is* ***low****.*

Agreeableness

Agreeableness reflects individual differences in concern with cooperation and social harmony. Agreeable individuals' value getting along with others. They are therefore considerate, friendly, generous, helpful, and willing to compromise their interests with others'. Agreeable people also have an optimistic view of human nature. They believe people are basically honest, decent, and trustworthy.

Disagreeable individuals place self-interest above getting along with others. They are generally unconcerned with others' well-being, and therefore are unlikely to extend themselves for other people. Sometimes their skepticism about others' motives causes them to be suspicious, unfriendly, and uncooperative.

Agreeableness is obviously advantageous for attaining and maintaining popularity. Agreeable people are better liked than disagreeable people. On the other hand, agreeableness is not useful in situations that require tough or absolute objective decisions. Disagreeable people can make excellent scientists, critics, or soldiers.

Your score is: ***high***

- *If your score on Agreeableness is low, indicating less concern with others' needs than with your own. People see you as tough, critical, and uncompromising.*
- *If your level of Agreeableness is average, indicating some concern with others' needs, but, generally, unwillingness to sacrifice yourself for others.*
- *If your high level of Agreeableness indicates a strong interest in others' needs and well-being. You are pleasant, sympathetic, and cooperative.*

Agreeableness Facets

- Trust. *A person with high trust assumes that most people are fair, honest, and have good intentions. Persons low in trust see others as selfish, devious, and potentially dangerous. Your level of trust is* ***low***.
- Morality. *High scorers on this scale see no need for pretense or manipulation when dealing with others and are therefore candid, frank, and sincere. Low scorers believe that a certain amount of deception in social relationships is necessary. People find it relatively easy to relate to the straightforward high-scorers on this scale. They generally find it more difficult to relate to the unstraightforward low-scorers on this scale. It should be made clear that low scorers are* <u>*not*</u> *unprincipled or immoral; they are simply more guarded and less willing to openly reveal the whole truth. Your level of morality is* ***high***.
- Altruism. *Altruistic people find helping other people genuinely rewarding. Consequently, they are generally willing to assist those who are in need. Altruistic people find that doing things for others is a form of self-fulfillment rather than self-sacrifice. Low scorers on this scale do not particularly like helping those in need. Requests for help feel like an imposition rather than an opportunity for self-fulfillment. Your level of altruism is* ***high***.
- Cooperation. *Individuals who score high on this scale dislike confrontations. They are perfectly willing to compromise or to deny their own needs in order to get along with others. Those who score low on this scale are more likely to intimidate others to get their way. Your level of compliance is* ***high***.
- Modesty. *High scorers on this scale do not like to claim that they are better than other people. In some cases this attitude may derive from low self-confidence or self-esteem. Nonetheless, some people with high self-esteem find immodesty unseemly. Those who* <u>*are*</u> *willing to describe themselves as superior tend to be seen as disagreeably arrogant by other people. Your level of modesty is* ***high***.
- Sympathy. *People who score high on this scale are tenderhearted and compassionate. They feel the pain of others vicariously and are easily moved to pity. Low scorers are not affected strongly by human suffering. They pride themselves on making objective judgments based on reason. They are more concerned with truth and impartial justice than with mercy. Your level of tender-mindedness is* ***high***.

Neuroticism

Freud originally used the term neurosis to describe a condition marked by mental distress, emotional suffering, and an inability to cope effectively with the normal demands of life. He suggested that everyone shows some signs of neurosis, but that we differ in our degree of suffering and our specific symptoms of distress. Today neuroticism refers to the tendency to experience negative feelings. Those who score high on Neuroticism may experience primarily one specific negative feeling such as anxiety, anger, or depression, but are likely to experience several of these emotions. People high in neuroticism are emotionally reactive. They respond emotionally to events that would not affect most people, and their reactions tend to be more intense than normal. They are more likely to interpret ordinary situations as threatening, and minor frustrations as hopelessly difficult. Their negative emotional reactions tend to persist for unusually long periods of time, which means they are often in a bad mood. These problems in emotional regulation can diminish a neurotic's ability to think clearly, make decisions, and cope effectively with stress.

At the other end of the scale, individuals who score low in neuroticism are less easily upset and are less emotionally reactive. They tend to be calm, emotionally stable, and free from persistent negative feelings. Freedom from negative feelings does not mean that low scorers experience a lot of positive feelings; frequency of positive emotions is a component of the Extraversion domain.

Your score is: ***average***

- *If your score on Neuroticism is low, indicating that you are exceptionally calm, composed and unflappable. You do not react with intense emotions, even to situations that most people would describe as stressful.*
- *If your score on Neuroticism is average, indicating that your level of emotional reactivity is typical of the general population. Stressful and frustrating situations are somewhat upsetting to you, but you are generally able to get over these feelings and cope with these situations.*
- *If your score on Neuroticism is high, indicating that you are easily upset, even by what most people consider the normal demands of living. People consider you to be sensitive and emotional.*

Neuroticism Facets

- Anxiety. *The "fight-or-flight" system of the brain of anxious individuals is too easily and too often engaged. Therefore, people who are high in anxiety often feel like something dangerous is about to happen. They may be afraid of specific situations or be just generally fearful. They feel tense, jittery, and nervous. Persons low in Anxiety are generally calm and fearless. Your level of anxiety* ***is low.***
- Anger. *Persons who score high in Anger feel enraged when things do not go their way. They are sensitive about being treated fairly and feel resentful and bitter when they feel they are being cheated. This scale measures the tendency to feel angry; whether or not the person expresses annoyance and hostility depends on the individual's level on Agreeableness. Low scorers do not get angry often or easily. Your level of anger is* ***high.***
- Depression. *This scale measures the tendency to feel sad, dejected, and discouraged. High scorers lack energy and have difficult initiating activities. Low scorers tend to be free from these depressive feelings. Your level of depression is* ***low.***
- Self-Consciousness. *Self-conscious individuals are sensitive about what others think of them. Their concern about rejection and ridicule cause them to feel shy and uncomfortable abound others. They are easily embarrassed and often feel ashamed. Their fears that others will criticize or make fun of them are exaggerated and unrealistic, but their awkwardness and discomfort may make these fears a self-fulfilling prophecy. Low scorers, in contrast, do not suffer from the mistaken impression that everyone is watching and judging them. They do not feel nervous in social situations. Your level or self-consciousness is* ***low.***
- Immoderation. *Immoderate individuals feel strong cravings and urges that they have difficulty resisting. They tend to be oriented toward short-term pleasures and rewards rather than long- term consequences. Low scorers do not experience strong, irresistible cravings and consequently do not find themselves tempted to overindulge. Your level of immoderation is* ***high.***
- Vulnerability. *High scorers on Vulnerability experience panic, confusion, and helplessness when under pressure or stress. Low scorers feel more poised, confident, and clear-thinking when stressed. Your level of vulnerability is* ***high.***

Report for the BSRI

Dear LieYen, you just completely filled in the BEM Sex Role Inventory. This is a detailed report of your results based on your given answers.

The Bem Sex Role Inventory was developed in 1971 by Dr. Sandra Lipsitz Bem. It characterizes your personality as masculine, feminine, androgynous, or undifferentiated. The BSRI is based on gender stereotypes, so what it's actually measuring is how well you fit into your traditional sex role. It measures masculinity and femininity as well as the individual's propensity to view the world using gender as a lens.

Each respondent receives both a masculinity score and a femininity score which is in the range of [0, 7]. The higher the score the more your fit within that gender type.

- *Femininity: Measures personality characteristics determined to be significantly more socially desirable for a woman than for a man.*
- *Masculinity: Measures personality characteristics determined to be significantly more socially desirable for a man than for a woman.*
- *The neutral items were used to constitute a measure of Social Desirability in response, which is an item do not differ significantly between males and females.*
- *The androgyny score looks at the degree of difference between a respondent's masculine and feminine score. It is a reflection of sex-typed, self-perception. This score is the highest if there is less bias in either directions, the score varies between [-6, 6].*

Additional:

- *Those who score above the median on the sex-congruent scale and below the median on the sex-incongruent scale are defines as sex typed.*
- *Those who show the opposite pattern are defines as cross-sex-typed.*
- *Those who score above the median on both scales are defined as androgynous.*
- *Those who score below the median on both scales are defined as undifferentiated.*

These are your scores for the BSRI:

Type	Score	Total score
Masculinity	5+6+1+4+2+2+2+1+2+1+5+2+5+6+5+5+2+2+6+1	3.25
Femininity	5+6+6+7+6+6+6+4+5+5+3+5+3+7+5+2+2+6+6	5.05
Neutral	6+4+6+4+6+2+6+2+4+3+5+2+3+1+6+2+2+6+6+7	4.15
Androgyny	-	1.8

According to your given answers in the BSRI questionnaire, your personality is characterized as follow:

- *Your score for masculinity is low, and your score for femininity is low,*
- *Your masculinity score is 3,25 which is lower than the median score of this group.*
- *Your femininity score is 5,05 which is higher than the median score of this group.*
- *This means that your personality can be categories as: Femininity.*

As defined by Silberstein et al. (1985):

"Sex-typed individuals are highly attuned to cultural expectations and prescriptions. Moreover this group of people are motivated to keep their behavior consistent with them, a goal they accomplish both by selecting behaviors and attributes that are consistent with their gender and by avoiding behaviors and attributes that are inconsistent with it.
In contrast, androgynous individuals are less attuned to these cultural definitions of masculinity and femininity and are less likely to regulate their behavior in accordance with them. "

According to Bem (1987), a sex-typed individual is someone whose self-concept incorporates prevailing cultural definitions of masculinity and femininity.

References

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. Psychological review, 88(4), 354.
Bem, S. L. (1987). Gender schema theory and the romantic tradition. In P. Shaver & C. Hendrick (Eds.), Review of personality and social psychology (Vol. 7, pp. 251-271). Newbury Park, CA: Sage.
Rodin, J., Silberstein, L., Striegel-Moore, R., Sonderegger, T. B., & Anastasi, A. (1985). Psychology and gender. Psychology and gender.

Appendix C

Additional tables and figures

Test of normality and homogeneity for the Big Five traits and the gender roles

Trait	Shapiro Wilk			Levene's Test for Equality of Variances between male and female in the IBM population	
	Statistic	df	Sig.[a]	F	Sig.[b]
Openness	,975	177	,003	3,528	,062
Conscientiousness	,956	177	,000	4,561	,034
Extroversion	,941	177	,000	6,235	,013
Agreeableness	,945	177	,000	10,956	,001
Neuroticism	,989	177	,210	,019	,891
Masculine	,983	177	,228	1,620	,205
Feminine	,990	177	,218	2,805	,096

[a]. Variable is normal distributed if sig. value is > 0.05.
[b]. Variable has equal variance between males and females if sig. value is > 0.05.

Table C.1: Test of normality and homogeneity IBM population

Test of normality and homogeneity for the Big Five traits

Trait	Kolmogorov-Smirnova			Levene's Test for Equality of Variances between females and males	
	Statistic	df	Sig. [a]	F	Sig[b]
Openess	,025	20992	,000	16,003	,000
Conscientiousness	,022	20992	,000	4,456	,035
Extroversion	,026	20992	,000	6,582	,010
Agreeableness	,039	20992	,000	32,498	,000
Neuroticism	,022	20992	,000	,064	,801

[a]. Variable is normal distributed if sig. value is > 0.05.
[b]. Variable has equal variance between males and females if sig. value is > 0.05.

Table C.2: Test of normality and homogeneity Penstate population

Personality scores Big 5 Facets - Penstate

Trait	Facet	TotalMean N=20993	TotalSD	MaleMean N=7743	MaleSD	FemaleMean N=13249	FemaleSD
Agreeableness	Trust	3,28	0,10	3,23	0,07	3,31	0,13
	Morality	4,09	0,23	3,89	0,23	4,20	0,24
	Altruism	4,04	0,07	3,84	0,08	4,16	0,07
	Cooperation	3,50	0,35	3,36	0,45	3,59	0,31
	Modesty	3,14	0,43	2,92	0,46	3,27	0,42
	Sympathy	3,73	0,20	3,46	0,19	3,88	0,21
Conscientiousness	Self-Efficacy	3,89	0,10	3,91	0,12	3,88	0,10
	Orderliness	2,96	0,21	2,92	0,17	2,99	0,23
	Dutifulness	3,96	0,29	3,88	0,38	4,01	0,24
	Achievement-Striving	3,80	0,19	3,67	0,22	3,88	0,18
	Self-Discipline	3,25	0,21	3,20	0,24	3,28	0,19
	Cautiousness	3,18	0,16	3,24	0,14	3,14	0,17
Extroversion	Friendliness	3,45	0,25	3,35	0,23	3,51	0,26
	Gregariousness	2,96	0,08	2,88	0,08	3,01	0,08
	Assertiveness	3,51	0,10	3,57	0,07	3,48	0,12
	Activity Level	3,09	0,50	2,97	0,49	3,16	0,51
	Excitement-Seeking	3,33	0,58	3,38	0,57	3,30	0,58
	Cheerfulness	3,66	0,27	3,54	0,34	3,72	0,23
Neuroticism	Anxiety	3,13	0,44	2,86	0,45	3,28	0,44
	Anger	3,08	0,19	2,93	0,21	3,16	0,18
	Depression	2,67	0,32	2,59	0,33	2,72	0,32
	Self-Consciousness	3,05	0,13	3,04	0,16	3,06	0,13
	Immoderation	3,06	0,12	2,96	0,15	3,12	0,11
	Vulnerability	2,57	0,19	2,31	0,18	2,73	0,21
Openness	Imagination	3,99	0,15	4,00	0,14	3,98	0,16
	Artistic Interests	3,85	0,17	3,66	0,22	3,96	0,15
	Emotionality	3,88	0,06	3,61	0,08	4,03	0,05
	Adventurousness	3,19	0,34	3,22	0,35	3,16	0,34
	Intellect	3,81	0,17	3,93	0,25	3,75	0,13
	Liberalism	2,96	0,47	2,90	0,45	3,00	0,49

Table C.3: The avarages scores for the facets by Penstate population.

Personality analysis IBM Population

	Cronbach's Alpha	Female (n=42)		Male (n=135)		Total (n=177)		F	p-value
		Mean	SD	Mean	SD	Mean	SD		
Openness	**,746**	3,3991	,68141	3,3470	,58710	3,3594	,60917	,234	,630
IPIPO1_Fantasy	,766	3,7024	,71619	3,4000	,78059	3,4718	,77463	4,992	**,027***
IPIPO2_Aesthetics	,765	3,9345	,64406	3,5407	,82363	3,6342	,80084	8,056	**,005***
IPIPO3_Feelings	,701	3,8988	,57398	3,5963	,63553	3,6681	,63318	7,586	**,007***
IPIPO4_Actions	,712	3,5833	,66412	3,6674	,70804	3,6475	,69693	,465	,496
IPIPO5_Ideas	,753	4,0476	,53601	3,9648	,67205	3,9845	,64191	,532	,467
IPIPO6_Values	,696	3,5774	,51052	3,3981	,59404	3,4407	,57900	3,106	,080
Conscientiousness	**,887**	3,5208	,79958	3,6499	,66591	3,6193	,69970	1,090	,298
IPIPC1_Competence	,632	4,0655	,50624	4,1167	,50998	4,1045	,50812	,324	,570
IPIPC2_Order	,832	3,6012	,86437	3,5315	,91087	3,5480	,89813	,192	,662
IPIPC3_Dutifulness	,691	4,1071	,49431	4,2148	,50201	4,1893	,50091	1,484	,225
IPIPC4_Achievement_Striving	,792	4,2202	,49449	4,1907	,70526	4,1977	,66017	,064	,801
IPIPC5_Self_Discipline	,744	3,6190	,80440	3,6481	,69533	3,6412	,72042	,052	,820
IPIPC6_Deliberation	,867	3,8214	,92276	3,7574	,76169	3,7726	,80051	,204	,652
Extraversion	**,882**	3,1313	,80898	3,2627	,65388	3,2315	,69363	1,150	,285
IPIPE1_Warmth	,812	3,8036	,67520	3,8333	,67884	3,8263	,67618	,062	,804
IPIPE2_Gregariousness	,791	2,9286	,76963	3,0056	,84777	2,9873	,82842	,276	,600
IPIPE3_Assertiveness	,876	3,5000	,80395	3,7593	,65819	3,6977	,70188	4,457	**,036***
IPIPE4_Activity	,756	3,4881	,70051	3,3611	,54265	3,3912	,58433	1,517	,220
IPIPE5_Excitement_Seeking	,773	2,9286	,63266	2,9778	,70345	2,9661	,68588	,164	,686
IPIPE6_Cheerfulness	,800	4,0833	,67324	4,0056	,59965	4,0240	,61681	,508	,477
Agreeableness	**,872**	3,5833	,71671	3,5104	,54838	3,5277	,59126	,486	,487
IPIPA1_Trust	,856	3,7560	,65678	3,8481	,63341	3,8263	,63836	,667	,415
IPIPA2_Straightforwardness	,763	4,2381	,69833	4,0963	,67951	4,1299	,68470	1,377	,242
IPIPA3_Altruism	,746	4,4107	,39748	4,1111	,53660	4,1822	,52189	11,167	**,001***
IPIPA4_Compliance	,787	4,2560	,51646	4,0185	,54089	4,0749	,54327	6,303	**,013***
IPIPA5_Modesty	,761	3,4048	,86946	3,0037	,74312	3,0989	,79109	8,588	**,004***
IPIPA6_Sympathy	,746	3,9048	,62701	3,5630	,58088	3,6441	,60807	10,678	**,001***
Neuroticism	**,873**	2,4468	,53015	2,1544	,54090	2,2238	,55117	9,444	**,002***
IPIPN1_Anxiety	,789	2,9762	1,07325	2,3889	,82049	2,5282	,91851	14,069	**,000***
IPIPN2_Angry_Hostility	,851	2,3452	,86418	2,1130	,83906	2,1681	,84842	2,420	,122
IPIPN3_Depression	,841	2,2857	,88959	1,9574	,74909	2,0353	,79448	5,614	**,019***
IPIPN4_Self_Consciousness	,756	2,8750	,89927	2,5815	,76529	2,6511	,80621	4,326	**,039***
IPIPN5_Impulsiveness	,722	2,9226	,69293	2,5593	,69661	2,6455	,71088	8,737	**,004***
IPIPN6_Vulnerability	,775	2,3929	,77133	1,9667	,63024	2,0678	,68853	13,119	**,000***

significant p-values are bold and marked with a *

Table C.4: The differences in the average scores of females and males on the facet level of the Big Five personality traits.

Appendix D

Game descriptions

The game elements

Game behavior	Game element	Description
Engagement	Total time	To total amount of time spent on the IBM game.
	Average time session	The average time spend per session on the IBM game.
	Number of sessions	The total number of sessions playing the IBM game.
	Total IBM questions	The total number of IBM questions answered.
	Total social questions	The total number of social questions answered.
	Total world wide technology questions	The total number of answered world wide technology questions.
	Max. score	The maximum score in the IBM game.
Social behavior	News shared on LinkedIn	The total number of news shared on LinkedIn.
	Invites sent	The total number of invites sent.
	Invites sent pm	The total number of invites sent via personal message.
Popularity	LinkedIn connections	The total number of LinkedIn connections.
	Facebook Connections	The total number of Facebook connections.
	In game connections	The total number of in game connections.
Expertise	Correct IBM questions	The total number of correct answered IBM questions.
	Correct social questions	The total number of correct answered social questions.
	Correct wwt questions	The total number of correct answered social questions.
Curiosity	Peers impression	The total number of times looking at your peers.
	News impression	The total number of times looking at your news.
	Invites accepted	The total number of invites accepted.
	Invites accepted pm	The total number of accepted via personal message.
Controlled behavior	Leaderboard impressions	The total number of times looking at your leaderboard.
	Achievement impressions	The total number of times looking at your achievements.
	Badges earned	The total number of badges earned.

Table D.1: A description of the game behaviors

Description of treatment groups in the IBM game

Group	Description
0	received the default game mechanics: an IBMer score and feedback for answering quiz questions.
1	received, in addition to Group 0, leaderboards.
2	received, in addition to Group 0, badges.
3	received, in addition to Group 0, both leaderboards and badges.

Table D.2: The treatment groups

The game statistics

Game Element	Group 1 (N=100)					Group 2 (N=112)				
	Mean	Variance	Skewness	Min.	Max.	Mean	Variance	Skewness	Min.	Max.
total_time	21,152	2073,436	4,468	1,000	306,000	19,413	3205,411	7,935	0,000	547,000
avg_session	9,065	87,655	2,567	0,330	55,000	7,051	31,796	1,051	0,000	23,000
num_sessions	2,495	26,028	4,602	1,000	30,000	2,578	49,617	8,811	0,000	71,000
total_ibm_questions	9,818	415,293	3,982	0,000	106,000	10,165	376,898	3,829	0,000	106,000
total_wwt_questions	8,960	521,896	4,138	0,000	122,000	7,633	419,401	4,890	0,000	122,000
total_social_questions	20,222	4109,481	6,185	0,000	510,000	18,156	6538,633	8,794	0,000	798,000
max_score	40,899	15149,255	6,067	10,000	959,000	38,239	20129,609	8,280	10,000	1365,000
correct_ibm_questions	4,939	113,976	3,852	0,000	59,000	5,514	134,011	4,017	0,000	71,000
correct_wwt_questions	5,000	171,490	4,042	0,000	74,000	4,349	154,285	5,031	0,000	83,000
correct_social_questions	17,566	3087,024	6,177	0,000	442,000	16,147	5926,441	9,113	0,000	771,000
invites_sent	2,040	41,713	3,809	0,000	36,000	2,037	93,128	8,814	0,000	96,000
invites_sent_pm	1,091	15,553	5,266	0,000	30,000	1,615	89,054	9,243	0,000	95,000
news_shared_linkedin	1,354	42,476	7,779	0,000	59,000	1,018	13,426	5,577	0,000	26,000
invites_accepted	0,384	1,872	4,203	0,000	8,000	0,257	1,378	7,567	0,000	11,000
invites_accepted_pm	0,212	0,801	4,965	0,000	6,000	0,183	1,244	8,808	0,000	11,000
linkedin_connections	148,848	16713,252	1,713	2,000	773,000	187,890	30810,358	2,465	1,000	1126,000
facebook_connections	2,626	69,910	3,579	0,000	50,000	6,312	376,568	4,386	0,000	142,000
ingame_connections	15,111	189,222	1,267	0,000	71,000	14,330	176,186	1,474	0,000	64,000
total_impressions	27,899	4111,133	4,484	2,000	422,000	23,193	5128,787	8,746	0,000	717,000
play_impressions	7,323	522,405	5,328	0,000	149,000	6,037	381,850	6,673	0,000	166,000
peers_impressions	4,545	138,863	4,040	0,000	71,000	3,651	131,266	7,580	0,000	109,000
news_impressions	1,434	11,881	7,443	0,000	32,000	1,211	10,612	8,101	0,000	32,000
badges_earned	1,949	6,375	1,797	0,000	13,000	2,303	9,417	2,103	0,000	19,000
leaderboards_impressions	3,010	154,337	5,817	0,000	92,000	1,211	28,631	9,127	0,000	54,000
achievements_impressions	0,879	6,006	6,932	0,000	22,000	1,376	30,866	8,669	0,000	55,000

Group 1: gamers who did not completed the survey
Group 2: gamers who completed the survey

Table D.3: Descriptive statistics for the two samples.

Game element	Median Group 1[a]	Median Group 2[b]	Mann-Whitney U	p value[c]
total_time	7,000	7,000	5189,000	,634
avg_session	6,000	5,000	4892,000	,246
num_sessions	1,000	1,000	5041,000	,344
total_ibm_questions	3,000	4,000	5079,000	,464
total_social_questions	2,000	2,000	5022,000	,386
total_wwt_questions	6,000	4,000	5221,000	,682
max_score	12,000	13,000	5346,500	,910
correct_ibm_questions	1,000	2,000	5020,500	,379
correct_social_questions	1,000	0,000	4938,500	,288
correct_wwt_questions	5,000	4,000	5341,500	,895
invites_sent	0,000	0,000	5290,500	,738
invites_sent_pm	0,000	0,000	5389,000	,979
news_shared_linkedin_total	0,000	0,000	5382,500	,965
linkedin_connections	0,000	0,000	4687,000	,102
facebook_connections	0,000	0,000	5041,500	,203
ingame_connections	127,000	149,000	5253,500	,744
peers_impressions	10,000	10,000	5157,500	,569
news_impressions	2,000	2,000	4940,500	,263
invites_accepted	1,000	1,000	5325,000	,692
invites_accepted_pm	1,000	1,000	5352,000	,816
badges_earned	0,000	0,000	5120,000	,490
leaderboards_impressions	0,000	0,000	5148,500	,495
achievements_impressions	0,000	0,000	5334,000	,870

[a]. Group 1: gamers who did not completed the personality questionnaire
[b]. Group 2: gamers who completed the personality questionnaire
[c]. Significance value is set at 0.05, if $p > 0.5$ then there is no statistical difference between the groups.

Table D.4: Mann-Whitney test statistics for comparing game statistics between two groups of gamers.