Investigate relationship between personality and programming characteristics through favourite programming language

A dissertation submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) in Computing

by E. Jankauskas

Department of Computing & Information Systems

Cardiff School of Management

Cardiff Metropolitan University

April 2017

Declaration

I hereby declare that this dissertation entitled **Investigate relationship between personality and programming characteristics through favourite programming language** is entirely my own work, and it has never been submitted nor is it currently being submitted for any other degree.

Candidate: Egidijus Jankauskas

Signature: Egidijus Jankauskas

Date: 20-04-2017

Supervisor: Mohamed Mostafa

Signature:

Date:

Abstract

The variance in programming languages is growing rapidly. Therefore, it is getting harder to choose the perfect language, which could be described as preferred language to work with. This research was done to explore if those choices can be facilitated. The relationship between programming languages and personality traits has been measured. Big Five questionnaire (Ipip.ori.org, 2017) has been used to measure the personality traits, and the top six programming languages have been chosen for this research based on their popularity as per survey done in 2016 by one of most recognisable programmer's forums (StackOverflow.com, 2017). Big Five questionnaire was used due to its usability and recognisability in other researches. The top six programming languages chosen were: JavaScript, Java, Python, PHP, VisualBasic and C#. The results were analysed adopting ordinal logistic regression using statistical software (SPSS) and mean and standard deviation analysis with Microsoft Excel. Both methods are using different variations of data, therefore it was explored whether the outcomes would be similar. After data analysis was completed some correlations were noticed. However, these correlations were noticed by using SPSS software only. Correlations between three programming languages and two personality traits were discovered.

Acknowledgements

Thank you to my supervisor, Mohamed Mostafa, for guidance and feedback throughout this research project. Thanks to all the participants who participated in the survey.

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

Each year numerous programming languages are competing against each other to appear at the top of the popularity list and it is getting harder and harder to choose the most suitable one (Stackoverflow.com, 2017). Raises few questions of how programmers are choosing the programming language. Whether certain aspects are considered such as readability, usability and the different features language may have. Furthermore, another question is whether programmers choosing the language based on the preference, meaning it will be enjoyable experience to learn it, or the choice is made due to workplace requirements or the popularity of the language in general.

Therefore, it has been decided to do this research with the interest of how programmers choosing the preferred languages and if this preference is impacted by one's personality. The main question of this research is to explore whether there is a correlation between programming languages and programmer's personality. To challenge this research question, literature will be reviewed, to find out whether the answer to the question already exists.

1.1 Literature review

Literature review could be divided into three categories. Firstly, considering the personality nature. Gathering information about personality will help to understand what aspects separates each person. Therefore, trying to find suitable personality test measurement in order to use it for this study. Secondly, choosing programming languages and focusing on their characteristics. Separating each individually and looking into specifics, what makes each language unique as well as attractive or unappealing. Third and last part in literature review will combine previous two factors together. Therefore, searching for prior researches that have answered the same or similar hypothesis.

1.2 Methodologies

Approach to correlation will be determined by the factors of the data. Hence, initially methodologies section will explore the broader view of methods used. Going forward to specifics such as understanding what data is required for the study, how it will be conducted and evaluating various data gathering methods. Understanding information is critical therefore this cannot be overlooked and data needs to be examined to have a good indication for the type of analysis that could be used. Finally, biggest emphases will be related to data exploration. To understand the concept and use the appropriate techniques to analyse the correlation between two variables.

1.3 Findings and Discussions

Analysed data will be interpreted and discussed in this section. The results will be evaluated and looked at in more depth, focusing on the main aspect whether the is any correlation between participant's personality and chosen programming languages.

1.4 Research Aim

This research aim is to try and understand whether there is any link between programmer's personality and whether he/she likes or dislikes the programming language. To explore and analyse whether the personality has any impact on the preferred programming language that has been chosen.

1.5 Objectives

First objective is to find suitable personality test that can be used for this research to measure the personality.
Second objective of this study is to explore programming languages that are preferred by programmers.
Third objective is to find appropriate analysis technique for the data investigation.

Fourth objective is to find whether there is a correlation between two variables – programming language and personality.

2 LITERATURE REVIEW

2.1 Personality and Big Five History

Personality is a list of characteristics that creates an individuals' character and defines an individual's behaviour. This definition of personality answers the main question of why one persons' behaviour differs from other persons' behaviour. Gordon Allport (1937) considers that human personality is determined at birth - biologically and then carved by person's environmental involvement, which changes human's factors such as intelligence, habits, skills, temperament, attitude and traits. Many researches have been done already to define the personality more specifically. Icek Ajzen (2005) has noted that a majority of personality features have already been determined, such as conscientiousness, emotional stability, sociability, helpfulness, independence, ambitiousness and the new traits are being added to the list continuously. Each of us has different personalities and each of our behaviour would be different in the same situation. For example, some individuals are very social, open-minded, and talkative and some are complete opposite – shy, distant, quiet. To find the perfect answer to these differences, researchers needed a descriptive model or taxonomy of traits (Oliver and Srivastava, 1999).
The history of taxonomy of traits started when Cattell (1943) used the Allport and Odbert (1936) list as a starting point for his multidimensional model of personality structure (Lawrence and Oliver, 1999). That structure was immoderate, so Cattell introduced 35 variable trait terms instead of 4,500. The research continued and Cattell had identified 12 personality factors, which eventually became part of his 16 Personality Factors (16PF) questionnaire (Cattell and Eber and Tatsuoka, 1970). Later, the other researchers and investigators were involved in the discovery of the Big Five dimensions. The Cattell's 22 variables were constructed to have more simplified descriptions. Those descriptions were first simplified by Donald Fiske in 1949. However, in 1961, Ernest Tupes and Raymond Christal re-examined the connection of eight factors and discovered five relatively strong factors (Cattell and Eber and Tatsuoka, 1970). This five-factor structure has been emulated by Warren Norman in 1963 where the personality traits were defined as Extraversion or Surgency (communicative, confident, active); Agreeableness (friendly, helpful,

gullible); Conscientiousness (disciplined, accountable, faithful); Emotional Stability (peaceful, well-balanced, easy-going); and Culture (well educated, polite, self-reliant) (John, 2008). The title of Big Five was chosen as per introduction of five main factors, which nowadays a better known as extraversion, agreeableness, conscientiousness, neuroticism, and openness. However, all of these five dimensions imply the personality notionally and each dimension generalise a large number of more specific characteristics. The Big Five model is prevailing approach, which shows an individual's personality trait conformation (Roccas, 2002).

Sometimes when the time is limited the easiest way to learn about the individual's personality is to ask direct question of how extraverted the person is rather than ask multiple questions whether he is talkative, outgoing, and enthusiastic (Gosling and Rentfrow and Swann, 2003). On the other hand, not everybody can understand the meaning of each personality trait. It can be difficult to say whether you fall into category of extraversion, agreeableness, openness, conscientiousness, or neuroticism. Individuals, without testing themselves cannot be sure which personality trait category he/she should consider as theirs. For this reason, the Big Five personality dimensions were developed to 5 and 10-item inventories. These inventories are normally used when the personality is not the primary topic of interest and the brief measures can be accepted. It might not be the most accurate way of testing, but it would give the brief description of the personality (Gosling and Rentfrow and Swann, 2003).

On the other hand, Myers-Briggs Type Indicator or MBTI (Briggs-Myers and Briggs, 1985) is a high-profile personality measure as well. It is used widely in job orientated and psychology consultation. MBTI is determined as a self-report questionnaire drafted to evaluate personality types. The personality type is described as a four-letter code, and a synoptic explanation is provided on the back of the report form. These four-letters may refer individuals either as extraverted (E) or introverted (I), sensing (S) or intuitive (N), thinking (T) or feeling (F), and judging (J) or perceiving (P). Once each of the four preferences is conflated, the personality type can be identified (Boyle, 1995). The Myers-Briggs Type Indicator tests can be found online for personal assessment, same as Big Five Personality tests.

Nevertheless, comparing The Big Five model and The Myers-Briggs Type Indicator it was noticed that The MBTI's four factors measure is reinforced by its connection with

the Big Five. Also, it was mentioned by Rowan Bayne (1994) that both modules could add value to each other by studying them closely.
However, it is noticeable that Big Five tests are commonly used for testing individuals for academic purposes, occupational choice, political choice, programming choice and many other arrears. Also, McCrae and John (1992) believe that the model should be useful for both psychologists who are measuring different individual's personalities and individual assessment. However, there is always a question whether Five-Factor model is reliable and accurate. Although, this cannot be answered precisely, Big Five is arguably the most identified personality test model that can be offered today by personality psychology (McAdams and Pals, 2006).

2.2 Programming language's History and Characteristics

Almost all programming was done in machine or assembly language before 1954 as stated by John Backus (1978). Programmers fairly observed their work as complicated, human creativeness was required to produce an efficient program. John Backus talks about economics which played huge role in 1954, that a cost of computers was the same of the cost of good programmers. Mainly because there were not a lot of programmers that could understand machine code. Cost being the prime incentive of FORTRAN creation. FORTRAN was the very first programming language created by John Backus and IBM in 1954 (Backus, 1978).
At the end of the decade programmers could write programs which would include mathematical formulas and in some rare events even anticipated for the same program to run on different machines. 1960s has been a year where programming languages were more than just mere tools but have evolved into articles of knowledge in their own right which led into the development of software industry and the practice of computer science. And by 1972 there were over 200 higher level languages created (Nofre and Priestley and Alberts, 2014).
Nowadays there are so many programming languages to choose from with varies implementations, frameworks, and also different purposes dependent of what is required from the end product. This also gives an enormous burden to programmers to differentiate which programming language is suitable for the job or even for his/her

own needs (Lovrenčić and Konecki and Orehovački, 2009).

Therefore, to obtain the most recent list of popular programming languages Stackoverflow[1] has been chosen.

Six chosen programming languages includes: JavaScript, Java, PHP, PYTHON, C#, and VisualBasic. Some explanation about commonly used expressions are given, followed by each language characteristics.

Scripting languages are deemed to be easier to learn, require less time and less programming effort to produce the program (Eckel, 2006). Often running with interpreter that does not require for the code to be compiled into the machine code therefore process of the program is faster (Stavrianos and Syropoulos, 2014). The source code is directly imbedded into the HTML page therefore while HTML page is being displayed plug-in that interprets the scripting language is automatically activated (Eckel, 2006). Scripting languages are often referred to as the tools for fast developing, as it is easier to glue components together.

Weakly typed programming languages do not require variable type to be defined. They are more portable but more prone for typos being missed and slipped into product code (Pierce, 2002; Ortin, 2011).

Strongly typed programming languages require variable types such as int, string, char, etc. to be defined before compiling the code (Pierce, 2002; Ortin, 2011).

Statically typed – Variables have to be pre-defined before they could be used in the program. Also shows the errors for the programmer at compile time giving an option for developer to fix them immediately rather than finding it later at run time. Therefore, it offers performance and robustness (Pierce, 2002; Ortin, 2011).

Dynamically typed - allows compiling or running the code without giving errors on variables that type has not been assigned to, because type is decided at run time. However, typos made by programmer are very costly. Nevertheless, these languages allows programmer to have better customisation, faster developing and flexibility when writing code (Pierce, 2002; Ortin, 2011).

JavaScript is an easy-implemented, weakly typed, dynamically typed, scripting language with object-oriented capabilities largely used for web development on the client-side, which was developed by Brendan Eich from Netscape in 1995. Although it sounds very similar to Java, should not to be confused by the name, because in

[1] StackOverflow is a popular forum for enthusiasts and professionals to ask questions and provide answers. Could be accessed at http://stackoverflow.com

general the two languages are entirely unrelated (Flanagan and Ferguson, 2002). Most of general-purpose core of the language (interpreter) has been integrated into the web browser (Internet Explorer, Chrome, Firefox) and so it does not require for the program to be compiled to the machine language (binary code) giving more flexibility and faster implementation (Flanagan and Ferguson, 2002).
However other side of JavaScript is that it is a full-featured programming language, in some instances might be even more complicated than others. Programmers who pursuit to use JavaScript for nonstandard tasks more often than less finds the process confronting if they do not have enough experience and solid understanding of the language (Flanagan and Ferguson, 2002).
Security being one of the big threats in JavaScript has some limitations within the language: restricts programmers from creating programs that could perform damaging acts, such as complete network operations or read local files to name a few (Flanagan and Ferguson, 2002).
To conclude, it can be said that JavaScript is still number one choice for most of client side scripting and web page dynamics. It is also really appealing for beginners, as well as powerful and agile for experienced programmers (Flanagan and Ferguson, 2002).
Python is fully scalable, dynamically typed, extensible, expressive, robust, fully object-oriented and agile programming language, which was created by Guido van Rossum and released in early 1991 for public distribution (Chun, 2006; Lutz, 2008). Python programming language does not require extensive time to learn and it is relatively easy. As defined by Chun (2006), Python has plain structure, relatively few keywords, clearly determined syntax that is easy to read. Readability is key as it does not give much flexibility to write unclear code, therefore makes it easier for the other programmers to understand ones' code faster. Chun (2006) claims that Python code is reasonably understandable even for those who never seen any Python code before. For example, language does not require end-of-line finish (more often semicolon) (Chun, 2006).
The syntax and semantics of this language are basic, but library is very large and it could be said that a lot of common tasks would already have sufficient code behind it. Also, syntax of Python enforces the programmer to write clean code. Python is a dynamic programming language and most common platforms support Python (Lutz, 2008).

Furthermore, robustness of Python language is a big advantage as it gives a full trace of error that has occurred during run and leads developer straight to the source where it came from (Chun, 2006).

Java is object-oriented, statically and strongly typed, multiplatform programming language created by James Gosling and his co-workers at Sun Microsystems in 1995 (Lindholm and Yellin, 1999; Evans and Flanagan, 2014). It is also statically and strongly typed programming language (Eckel, 2006; Evans and Flanagan, 2014). Because of its robustness and performance Java is hugely popular today, especially in areas of programming such as: assurance societies, banks that would require reliable, secure and portable product (Evans and Flanagan, 2014). It allows programmers to create exclusive programs called applets which are downloaded and runs in web browser. Security of such applet is outstanding as it does not allow writing to user's addresses in memory or disk, preventing a virus or system crash (Lindholm and Yellin, (1999). Portability of the code is one of Java's strongest sides all of the code is compiled to machine code, which runs on JVM (Java Virtual Machine) that lives on any platform, which is why it is referred to as platform independent (Lindholm and Yellin, 1999; Gosling and Joy and Steele and Bracha, 2000).

Java is a complete and very powerful language with numerous features, frameworks, design-patterns, plug-ins, libraries and concepts. It can run in a browser or by itself. However, all these technologies around Java programming language makes it quite complex and difficult for beginners or just programming enthusiast (Evans and Flanagan, 2014).

Nerveless programming in Java language is increasing with new libraries, new features and design-patterns that are added by separate communities almost every day (Lindholm and Yellin, (1999).

C# - is quite easy to pick up for the programmers who already are familiar with object-oriented programming language which has been developed by Anders Hejlsberg as part of .NET package. Programming language relates to Microsoft .NET framework which gives a quick access to simple applications almost instantly, although for more robust applications one would require a more in depth knowledge about this strongly typed language including Microsoft Framework (Nash, 2007). The knowledge of so called 'Toolbox' in .NET is crucial as it makes it easier for the programmer to navigate and exclude some of the coding necessary to implement

such things as: textboxes, buttons, labels. C# code compilation works differently from other languages, it only compiles the methods and functions that are actually required at the run time. For example, created program has a print function to print some text onto the screen but if the program never hits that stage the compiler never compiles the code in the first place. This is to do with the fact that C# source code is compiled into managed code which is intermediate level between machine code and high level language. Another good feature is high level “Garbage collector “helps you worry less about memory allocation and management which according to Nash is one of the common errors of coding in software development. C# applications are known to be generally slower (due to its compiler) than an application that uses native languages which either compiles program to machine code or keeps it at a high programming level (Nash, 2007).

PHP – simple yet powerful open source, dynamically typed, scripting language, primarily designed for developing HTML content. Rasmus Lerdorf created it back in 1994. This programming language is mainly used for server-side scripting, but could also be used in graphical applications as a standalone application (Lerdorf and Tatroe and MacIntyre, 2006).

There are millions of websites powered by PHP, though showing its popularity and ease of use. This programming language requires PHP parser and a web server to be able to send code documents. Flexibility of PHP programming language spans to all major platforms such as: Linux, FreeBSD, Windows, and Mac OS so it could be called a multiplatform language and also could be used with all major web servers. Moreover, a wide variety of databases that it supports such as MySQL, Oracle, MS-SQL, DB2 with its provided libraries for an easy interaction (Lerdorf and Tatroe and MacIntyre, 2006).

However, PHP that we know today is significantly different from when it was first introduced. Further versions of PHP have been dramatically enhanced with object orientation, XML and support of SQLite database connection (Lerdorf and Tatroe and MacIntyre, 2006).

The motto in PHP development by Razmus’ was simple ‘Be Lazy’. His intentions for PHP were to be as simple as possible for programming language. Using its object-oriented functionality, the language could also be used as a powerful tool for much more complex tasks together with many embedded functions for many different purposes (Lerdorf and Tatroe and MacIntyre, 2006).

Visual Basic has been created by Microsoft Inc. in 1991. It is an Event-Driven programming which communicates with the system through the system of events. Visual Basic has a large variety of built in procedures that have complete control of operating system's triggers and objects. As explain by Willis and Newsome (2010) the program acknowledges your steps and takes the actions to complete your wanted tasks. Because it is tied with Microsoft it cannot be said that Visual Basic is platform independent language. Although it is used throughout Microsoft platform for example Microsoft Office has incorporated Visual Basic for programming macros which gives a scripting version of the programming language. Visual Basic was the first rapid application development (RAD) setting for the operating system that was event-based. Moreover, speed is a big pitfall for Visual Basic because of its interpretation which makes programs into object code and then interprets during run. Therefore, it takes a lot of time to run complex program. Another bad feature is lack of backwards compatibility which increases work for the programmer to rewrite the code to be compatible with its previous versions. Nevertheless, explicit data typing and its capability to connect to database make it more attractive for wider variety of programmers. Also, Visual Basic allows the programmer to essentially draw the User Interface to however he wants and then add the code behind that will react to events. This makes it inviting for the programmers who are beginners or just enjoys creating something simple and quick (Stephens, 2008; Willis and Newsome 2010).

2.3 Personality correlation with programming

Correlation between personality and programming might be a wide area for a discussion. It is because more and more computer applications and programs are being used in most of the organisations. Therefore, more and more programmers are being needed.

A few researches have been done to measure how the personality traits can be connected with programming in general. One of the researches was done by Catherine Bishop-Clark (1995). She has analysed the correlation between cognitive style, personality and programming. It was mentioned that personality and computer programming is less evolved and more dispersed rather than cognitive styles and computer programming. Nevertheless, based on the other few studies (Chin and

Zecker, 1985; Kagan and Douthat, 1985; Kagan and Esquerra, 1984; Newsted, 1975) author discussed few findings in concatenation between personality and computer programming. The biggest side of discussion was done in connection with introversion and extroversion association with achievements in programming. It was discovered that there was no connectivity between students who can be described as extroverts and achievements on their programming exams. Moreover, the follow up studies indicated that the achievements in other programming exams were unrelated with introversion either. However, those studies had limited responses from volunteer students and may not be very accurate. It would be beneficial to research the correlation between personality and programming more intensely to be able to make this conclusion. However, the author mentioned that a high percentage of computer programmers are judging, introverted, sensing, and thinking. Yet, it is worth to mention that most of these studies used Myers-Briggs Type Indicator to measure personality traits (Bishop-Clark, 1995).

Timo Gnambs (2015) has also done a meta-analysis by linking personality traits and programming capability. A meta-analysis can be described as a technique used for methodically combined data from several different researches to draw interference from it with a greater statistical power. The purpose of this research was to measure which personality trait is associated with programmers who are doing less errors, are more creative while developing new program, and are more focused on certain tasks. This meta-analysis concluded that programming aptitude was associated with three personality traits, conscientiousness, openness and introversion.

Further within the study The Big Five personality test has been used to measure work-place behaviours of the programmers. Firstly, the conscientiousness of programmers was measured and this describes the person as careful, scrupulous, accurate, and detail-oriented. These features are very important for programmers, as coding requires a huge focus on minor details. Although, some features can be more useful for one programming language, were the coding should very specific and accurate whilst the other features can be more beneficial for other languages (Gnambs, 2015).

The other designated personality trait concluded in the meta-analysis was openness. The long known stereotype is that a programmer must always be a loner, who do not have social skills, and might also have psychological deficiencies. However, there are many programmers who are open, friendly and socialise a lot in their free time.

Openness helps programmer to be more open-minded and look at the bigger picture to create the perfect code (Gnambs, 2015).

In comparison with Timo Gnambs (2015) and Catherine Bishop-Clark (1995) meta-analysis it can be noted both of the outcomes were different. Those differences might be due to the fact that both researches had significantly different amount of respondents. Also, the different personality tests have been used. Myers-Briggs Type Indicator was used in Catherine Bishop-Clark's research and Big Five personality test was used to measure personality traits in research done by Timo Gnambs.

In conjunction with personality and programming the similar research has been done to measure the influence of personality on programming styles (Karimi, 2016). There are four basic programming styles – depth-first, breadth-first, top-down style, and bottom-up style. The programmers who are using depth-first style are conformist, and to solve the problem they are applying first alternative. Whereas, the breadth-first programming style users are methodical and analysing all problem solutions before pursuing. In the top-down style, the programmers examine the model space to determine the problem, whereas in the bottom-up style, programmers examine code space to find the solution of the problem (Cox and Fisher, 2009).

Although, the programming styles and programming languages are different, the concept of the research done by Zahra Karimi (2016) is very similar. The self-assessed survey was used to examine 65 programmers correlation between personality measures and their chosen or preferable programming styles. The academic achievement, programming experience and attitude towards programming was taken into consideration as well. The findings were that programmers with higher openness have broader views and saw more alternatives used breadth-first style. Moreover, is was concluded that the programmers with conscientiousness used depth-first style, which means that individuals are goal oriented and are good at finding fast solution. This research actual showed how personality can be affiliated with programming styles. Individuals who can describe their personality trait as openness, which means they lack of restrictions can actually use breadth-first style, as this style requires programmer to be more open-minded and search for different solutions and proceed with the best one that can be found. As well as programmers who can describe their personality trait as conscientiousness - individuals, who are

very faithful, honest and dedicated, used depth-first style, where the solution needs to be found fast, however it must be the right solution (Karimi, 2016).

The other research done by Shirley Cruz, Fabio da Silva and Luiz Fernando Capretz (2015) was to measure the correlation between personality and software engineering. The Big Five, Myers-Briggs Type Indicator and Kersey Temperament Sorter were used. The authors discussed that the final outcome could not be reached due to the fact that different tests were applied. Those tests had various outcomes, and it would have been difficult to reach a conclusion as the tests were slightly different from each other. Although all models are used to measure personality, it is hard to establish which test can describe the personality of individual at its best. However, the most used test in this research was Myers-Briggs Type Indicator, but The Big Five test is described as most recognisable test for personality. Yet, it does not mean that researchers should not take into account the other personality test seriously, but the conclusion can be reached as there is still a wide area for testing personality correlation with programming and software engineering (Cruz and Da Silva and Capretz, 2015).

Also, the research done by Venkatesan and Sankar (2014) measured the groups of pair programming with analogous personality and mixed personality pairs. The results were that conscientiousness and openness showed a positive connection with student's accomplishment. However, the groups with mixed personality traits achieved the higher scores in their tasks rather than students with the same personality traits. The Big Five or Five Factor model has been used in this research to measure the personality of the students (Venkatesan and Sankar, 2014).

Although on various occasions Myers-Briggs Type Indicator were used to determine the personality traits. As per above literature review it can be concluded that personality traits in the recent years and researches are measured by Big Five personality tests.

The aim of this research is to see whether there is a correlation between personality and programming languages using Big Five module. Although in can be noted that the gap in the studies exist, in understanding whether programmer's personality defines chosen favourable programing language.

In addition, from the standpoint of employability, to become successful software engineer, individual should not only show his/her technical knowledge and academic qualifications, but have certain soft skills desired by employers. One of the most

important personality traits for software engineer to have is openness. Software engineers should have good communication skills to be able to communicate with others in meetings, over the phone and with work colleagues. For example, Microsoft has a special competency toolkit, which helps to specify the soft skills needed for specific positions. Also, Microsoft has separated most desirable soft skills required for Software Development Engineers – communication skills, self-development, composure- stress management, change management, problem solving skills, interpersonal skills, and drive for results (Orsted, 2000).

The other research has been done to investigate the personality traits that are described in job advertisements for software architectures. By analysing 124 job adverts it has been discovered nine most frequently used soft skills were analytical and problem solving skills, communicational skills, organisational skills, team player, interpersonal skills, fast learner, innovative and open, ability to work on own initiative, and adaptable to changes (Ahmed and Campbell and Beg and Capretz, 2011). It can be noted that most of the soft skills are similar to the ones mentioned above by the Software Development Engineer of Microsoft.

However, technical skills and academic knowledge is most important criteria, which is being looked at by the employers. Software engineer candidate would be more successful if he/she could show the knowledge of more than one programming language (Java, C++, C, Visual Basic, C#), also would have knowledge of more than one database (Oracle, SQL Server, DB2, Sybase, Access) and even would have knowledge on using different platforms (Windows, Unix, Mainframe/midrange, Cross-platform, Macintosh) (Surakka, 2005).

Therefore, this research could add value to the computer science job industry by looking into programmer's personality and analysing whether there are any traits and connections with the programming language that he/she choose and likes or dislikes.

3 METHODOLOGIES

Main question has been asked whether personality factors correlate with programming characteristics through programming languages. To be able to answer the question firstly, secondary data has been gathered. Mainly to see whether the same question has been answered before or any previous similar researches conducted on the topic. However, there were no previous researches that could be found exploring the correlation between personality and programming languages. As a result, this research has been conducted. To analyse the correlation theory, data has been gathered measuring participant personality and ones most or least favourable programming language. Adopting positivism as a research philosophy the hypothesis could be analysed and explained, in whole or part, if necessary further development could be done by additional research. It is emphasised that data collected is quantifiable and can be used in statistical analysis. Deductive approach used has few important characteristics, search of the relationship between two variables and the use of quantitative data (Saunders and Lewis and Thornhill, 2007). Both of which will be discussed further in '3.1 Structure' section. Experiment using data gathered by survey has been accomplished where relationship between two variables has been examined (Saunders and Lewis and Thornhill, 2007). More precisely five independent variables and one dependent variable. In depth discussion about data collection and analysis will be in section '3.3 Data categorisation and analyses'.

3.1 Structure

At first, the comparison was used between few different personality tests researched previously to be able to find the most suitable test for this study. According to literature review Big Five model is used in more recent researches than Myers-Briggs or any other personality tests. Also, as stated by Cruz, Da Silva, and Capretz (2015) Big Five is most recognisable personality test. In addition, books, journals and other researches have been reviewed to strengthen the decision to use the Five Factor model. Therefore, it was adopted for this research to be used in questionnaire. Secondly, to incorporate programming languages into the same questionnaire as a second part, it required the examination of the most used

programming languages. Although this questionnaire will only cover a very little fraction of the top programming languages that are out in the public for everyone to pick and start learning or even develop with. So, to pick the top tier languages Stackoverflow has been chosen. Stackoverflow (2017) is being visited by 40 million people every month and has a large population of programmers. As shows, the survey[2] on stackoverflow (2016) website has been completed by 56033 coders in 173 countries in 2016. The forum reports most popular programming languages every year so programmers can follow the newest trends in programming if they choose to. However, for this research 2016 top programming language list has been selected. It has been obtained and reported by Stackoverflow on assumption that the most popular programming languages are the ones that receives highest number of questions and answers on their website (Stackoverflow.com, 2017). Once all of the details of the relevant data have been decided upon, questionnaire had to be created.

Creation of the questionnaire has involved some structural decisions in terms of ease of use for the participants. It consisted of three parts: years of programming experience, Big Five personality test and list of six programming languages.

To start off, the research was done to find the most appropriate personality test. Few different available online tests were reviewed and examined, which concluded with selection of International Personality Item Pool's (Ipip.ori.org, 2017) Big Five personality test. The International Personality Item Pool (IPIP) is a public-domain compilation of information for use in personality tests. The IPIP project was firstly introduced at the 8th European Conference on Personality in 1996. The concept is now administered by Oregon Research Institute (Ipip.ori.org, 2017). Also, it was mentioned that to use the IPIP tests, there are no need to ask for a permission to do so. Moreover, the research was done on IPIP as a project itself and it has been noticed that the project is recognisable by the famous personality measure experts such as Goldberg (Goldberg et al. 2006). The personality Big Five test used in questionnaire had 50 statements with the numerical answer options: 1 being ‘Disagree’, 2 – ‘Slightly disagree’, 3 – ‘Neutral’, 4 – ‘Slightly agree’, 5 – ‘Agree’ where only one option per statement can be used. This later on was restricted to have only one option, available to choose by bullet points, per statement in online survey.

[2] Survey link http://stackoverflow.com/research/developer-survey-2016

Using Stackoverflow as a main source for the most popular programming languages that have been chosen to be used in questionnaire were: JavaScript, Java, PHP, Python, C#, and VisualBasic. For each of the programming language there were six options to choose from to be more precise six numbers from 0 through to 5. Where 0 option is ‘Never used it’, 1 – ‘Do not like it’, 2 – ‘Slightly dislike it’, 3 – ‘Like it’, 4 – ‘Like it a lot’, 5 – ‘My favourite’. Zero value has been introduced so that participant could have an option to choose if he/she has never used the language before. Furthermore, as with Big Five, answers were restricted in online survey to have the availability of only one bullet point option for each programming language.
And lastly years of experience although was not used in the analysis of the research have gave a good indication of the variety of participants with a different levels of programming skills. On the other hand, the years of experience cannot be divided into Beginners, Intermediates, and Advanced levels due to the fact that it takes one to learn and master the programming longer than others or vice versa. Also, it is not known whether the experience gain was in the industry or university or as an enthusiast which greatly reduces the assumption of different levels (Feigenspan et al. 2012).
Crucially there were no open ended or qualitative questions asked, therefore quantitative method has been chosen throughout this research for the data collection. It would have been more difficult to evaluate the most favourable programming language and least favourable programming language if they would have been asked to answer it as an open-ended question. This is because different respondents can describe the language and the choice of the language from their perspective, whereas the surveyor can interpret that answer from completely different angle. This might result in incorrect evaluation of the study. Additionally, most of the respondents prefer numeric based answers rather than write an essay about each programming language. It is due to the fact that respondents who agreed to participate in the survey have a limited amount of time for these questionnaires. Also, it is more likely to receive the completed questionnaires, where the overall completion does not take more than 10 minutes of each respondent’s time.
Moreover, it is worth to mention that as stated in questionnaire it was mandatory for participants to have at least one years’ experience in programming (programming study years also counted) in order to participate in the survey.

After questionnaire has been developed it was sent for a review to Cardiff Metropolitan Ethics Committee. The Ethics approval has been successfully granted and reference number issued (2016D0277). It is essential to have such an approval, especially when this questionnaire was mostly based on personality. Although the survey was anonymous and no identity could be recognised, it is always worth to be sure that questionnaires could be used in this study without any disruption. Also, the information sheet and consent form have been included in this questionnaire, so the respondents were fully aware of the concept of this survey.

There were few options considered, how the respondents could complete the questionnaire in the most suitable and convenient manner. The paper based interviews where time would need to be allocated with each respondent. Paper or hard copy based questionnaire, where the physical hand-out needed to be conducted to reach the participants. Online based survey, where it could be accessed by a wider variety of respondents. Firstly, interviews would have been more time consuming than questionnaires, each respondent should have been met face to face to take the interview and would not be considered as anonymous. Secondly, hard copies would have been circulated in various locations such as university or work place. However, to receive the questionnaires back from the respondents might have been a challenge and also it would have been a time-consuming option. Thirdly, online survey option is easy accessible for both surveyor and respondents. The questionnaire could be completed from various locations where internet connection can be found, also by using laptop, desktop, tablet or smartphone. Therefore, considering all the options, online survey method has been chosen to gather the data.

3.2 Distribution

For the distribution of the questionnaires Surveymonkey[3] online website has been used. Surveymonkey (2017) is the provider of web-based survey solutions. It is widely known and trustful portal used by millions of individuals, organisations and companies (Surveymonkey.com, 2017). From the respondent point of view the easy access of the survey is one of the most important criteria. Easy access has been provided by Surveymonkey application on the smartphone, once the questionnaire is

[3] Can be accessed at www.surveymonkey.com

submitted by respondent the notification comes instantly notifying about completed questionnaire. This option is very useful, when the time is limited and the surveyor does not need to login to Survey Monkey portal each time to check the number of questionnaires completed. Distribution of the survey has been made easy by this online option. Emails were sent to specific respondents, including work colleagues and friends; social media was used to target the programming forums and groups. This really helped to target the relevant respondents giving the assumption of at least one years' experience to participate. However, the down side of online questionnaire it is easy to exit the survey if there is no interest as showed by the number of uncompleted samples in the Surveymonkey. Overall the data from 49 participants have been collected which was used in the analysis.

3.3 Data categorisation and analyses

All data gathered in the online survey have been transferred into Microsoft Excel spread sheet for preparation for the analysis. Firstly, Big Five answers from the questionnaire had to be calculated by the specific formula created by IPIP (2017) as shown in Figure 3.1.

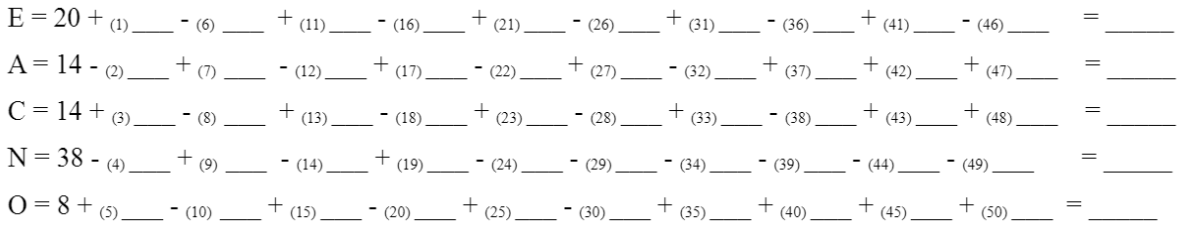


Figure 3.1: E is Extroversion, A - Agreeableness, C - Conscientiousness, N - Neuroticism, O - Openness

In the equation above small number in parentheses represents the number from 50 statements. Also, the scores calculated should be between zero and forty. Each respondent's questionnaire has been added to separate spreadsheet tab. All the Big Five formulas were included in each tab and used to calculate the respondents' score of each personality trait. To be able to analyse the scores they had to be transformed into percentages. Therefore, each score of personality trait for each respondent has been converted as follows 'score/40x100'.
However, with the programming languages the percentage conversion would diminish the situation of the calculation as it is not a continuous value. Programming language has essentially two states, options 1 and 2 person does not like the

language and options 3 to 5 inclusive - person likes the language it only falls down to the strength of each side. Figure 3.2 shows data sample of 20 participants.

Participants	Years of Experience	Big Five Factors %					Programming languages (0-4)					
		Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness	JavaScript	Java	PHP	Python	C#	Visual Basic
1	6	68	83	50	53	90	4	4	4	3	3	3
2	3	43	65	45	48	60	2	4	3	3	2	4
3	4	33	83	50	65	68	0	3	4	3	3	3
4	3	58	78	73	68	80	4	4	3	4	0	3
5	6	75	55	53	80	65	4	1	4	0	4	4
6	1	20	43	45	40	43	0	0	3	4	3	2
7	4	45	70	63	53	60	2	4	3	4	4	3
8	3	65	65	98	78	70	3	4	3	3	4	3
9	3	45	65	73	50	55	3	4	0	2	3	4
10	3	85	95	53	48	73	4	4	3	3	0	4
11	4	53	98	48	43	75	4	3	2	3	2	1
12	3	63	48	33	65	60	2	3	0	3	3	4
13	6	23	40	75	88	88	1	4	1	4	3	4
14	3	90	73	30	28	80	3	4	0	3	2	1
15	6	45	60	60	50	55	3	4	1	2	4	1
16	6	53	93	48	60	60	0	2	0	3	3	4
17	3	55	53	58	45	58	0	1	0	0	0	4
18	2	43	75	60	30	63	3	3	0	3	0	4
19	4	75	85	55	55	78	2	4	2	3	0	1
20	3	63	88	73	58	73	3	2	0	3	1	4

Figure 3.2: Excel data sample

Crucial point is that zero values had to be removed from the options once gathered from the survey. It cannot be justified if one likes the programming language or not, because he or she has never used it. Therefore, different count of the participants for each of the language will be shown in the findings and discussions section and will be used to analyse data.

Whilst gathering data, various different analyses have been tested and looked at how the data could be analysed. Since the data is collected in a numeric format, statistical methods were used to investigate and calculate it. Regression analysis has been chosen to analyse the data conducted by the online survey using IBM SPSS tool. Statistical Package for the Social Sciences or SPSS is a dynamic, easily operated software package for operation and statistical analysis of data. This package is convenient for researchers and students in sociology, psychology and other behavioural sciences, as those contain univariate and multivariate procedures (Landau and Everitt, 2004).

Regression analysis is usually explained as a relationship which can predict the dependent variable from the independent variable. Dependent variable sometimes also referred to as target, outcome or criterion. This variable is a variable that depends on the independent variable. In this research case, it means that

programming language option (dislike - like) might depend on whether the person scores high or low in any of the particular Big Five Factor. An independent variable sometimes called as explanatory or predictor is a variable that is used to observe or predict the outcome variable (Field, 2009; Statistics.laerd.com, 2017). In this research, independent variables are five personality traits. Worth noting that this research is non-experimental meaning that the independent variables will not be manipulated in any way and therefore will be used as submitted by the participants for the accurate outcome. Furthermore, considering various different variable types two have met the criteria for this analysis to fit the data. Independent variables classed as the type of continuous and dependent variable in this case is classed as the type of ordinal. Continuous variables are known as quantitative and can be organised as ratio or interval (Field, 2009; Statistics.laerd.com, 2017). For this study personality, Big Five percentages are measured from 0 through to 100. Ordinal variables are also referred to as categorical variables which would have graded options or categories such as disagree or agree (Field, 2009; Statistics.laerd.com, 2017). For this research programming language options are classed as categorical variable. It cannot be continuous due to the fact that there may not be a level in between, that being said, for example it can only be 1 but not 1.6.

There are many different variations of regression analysis, however for this research due to the nature of the variables ordinal logistic regression analysis has been chosen. This type of regression analysis will allow for verification which of the independent variables (if any) have a statistically significant impact on the dependent variable. Furthermore, it determines how sufficiently model predicts the outcome variable. Ordinal regression analysis deals with categorical dependent variables (Agresti, 2010). In this case, answers from the questionnaire whether one likes programming language or does not like it. However, because regression analysis does not accept categorical variables as in words ‘dislike it’ or ‘like it’ they must stay as numbers.

Firstly, in order to use ordinal logistic regression for the analysis of data four assumptions have to be met. First assumption is looking at the dependent variable which has to be ordinal, categorical type. As previously mentioned this assumption is met due to the categorical variable of programming languages: 1 – ‘Do not like it’, 2 – ‘Slightly dislike it’, 3 – ‘Like it’, 4 – ‘Like it a lot’, 5 – ‘My favourite’. Second assumption states that one or more independent variables have to be continuous or

nominal. Big Five personality traits measured as percentage values making second assumption tenable. Third assumption tests for multicollinearity. Multicollinearity means that one or more explanatory variables highly correlate with each other. This introduces issues when trying to understand which independent variable contributes to the justification of outcome variable and further develops problems in calculating and ordinal logistic regression (Agresti, 2010). Test for this assumption is made by running linear regression analysis and viewing the 'Tolerance' levels which will be explained further in findings and discussions section (Statistics.laerd.com, 2017). Fourth assumption of proportional odds deemed to be the fundamental in ordinal logistic regression. The essence of such an assumption is that the consequences of the independent variables are proportional across different cumulative odds. Cumulative odds are different thresholds of the dependent variable, separated out and made into dichotomous (binary: 1 or 0, 'Yes' or 'No') categories. It is explained in Figure 3.3, Figure 3.4 and Figure 3.5 by dividing outcome variable into 3 dichotomous categories. It was created using Microsoft Visio tool for graphical presentations, example taken from Laerd Statistics website (2017) which could be found in the appendix C.

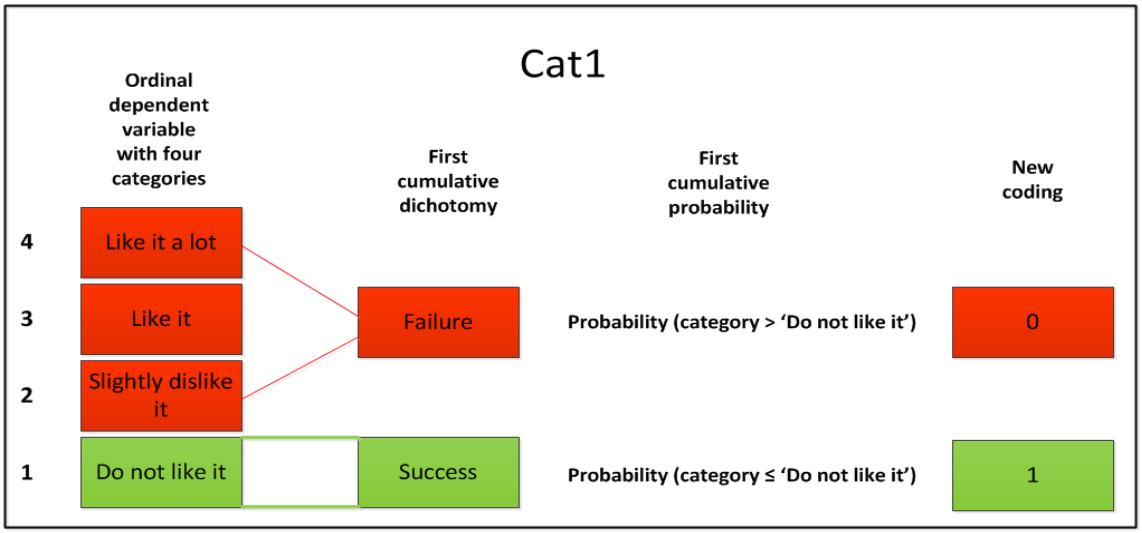


Figure 3.3 First Cumulative category

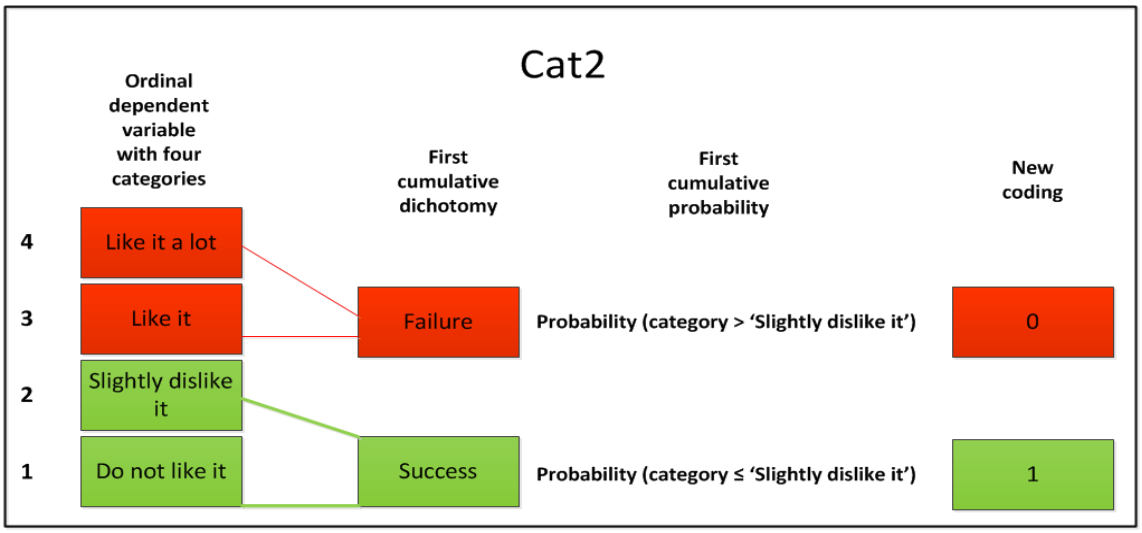


Figure 3.4 Second Cumulative category

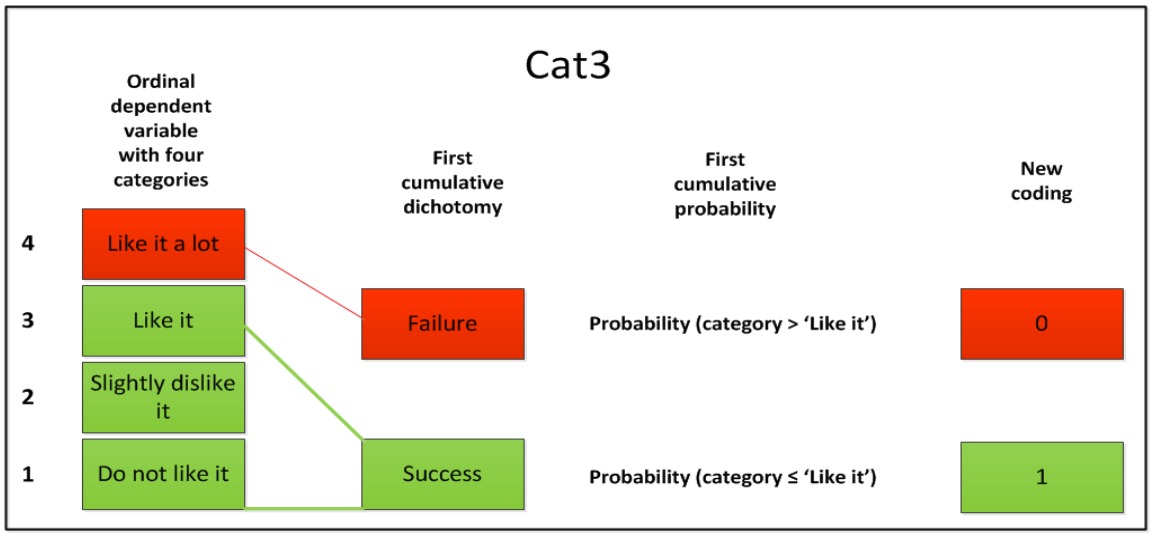


Figure 3.5 Third Cumulative category

Another test for proportional odds could be made in SPSS by examining parallel lines, which could be easily executed as part of the ordinal regression analysis. However, it is not as accurate as running separate binomial regressions as it mostly results in rejection of the assumption (O'Connell, 2006).

Binomial logistic regression has been executed on each of the dichotomous category using all the independent variables. Binomial regression is also referred to as binary regression which is used for binary variables ('Yes' or 'No'). Once binary regression is completed for each different category the proportions must be similar to meet the final assumption. Therefore, it could be stated that each individual independent variable has the same effect on each individual dichotomous category. Moreover, it

could be said that the result of a variable is to have higher or lower category of the dependent variable, despite of the value of the dependent variable.
However, due to the ordinal logistic regression assumption failure with the limitation of the participant count, decision has been made to merge programming language options ‘Like it a lot’ and ‘My Favourite’ into ‘Like it a lot’ in order to proceed with analysis. Results of fourth and final assumption will be reviewed in findings and discussion section.
To further explore the correlation, it has been decided to use additional analysis to see whether there is a correlation between personality Big Five factors and chosen programming language in the calculation of mean and standard deviation.
Firstly, data has been split into two groups for each of the programming language. Answers from the questionnaire of programming language part, stated with numeric values 1 and 2 have been classed as do not like the language. And values 3 and 4 have been classed as participant who likes the chosen programming language. Each of the Big Five personality traits has been associated with relevant category.
Microsoft Excel program was used to calculate mean, also called average and standard deviation for each personality trait of each programming language group. Analysis then observed in detail and explained in findings and discussions section.

4 FINDINGS AND DISCUSSIONS

In this section, the analysis will be reviewed and discussed. Starting with, looking at participant's years of experience, number of participant's preference for each programming language, assumptions for ordinal regression. Also, the findings of two analyses will be evaluated and explained including ordinal regression and standard deviation and mean.

4.1 Years of Programming Experience

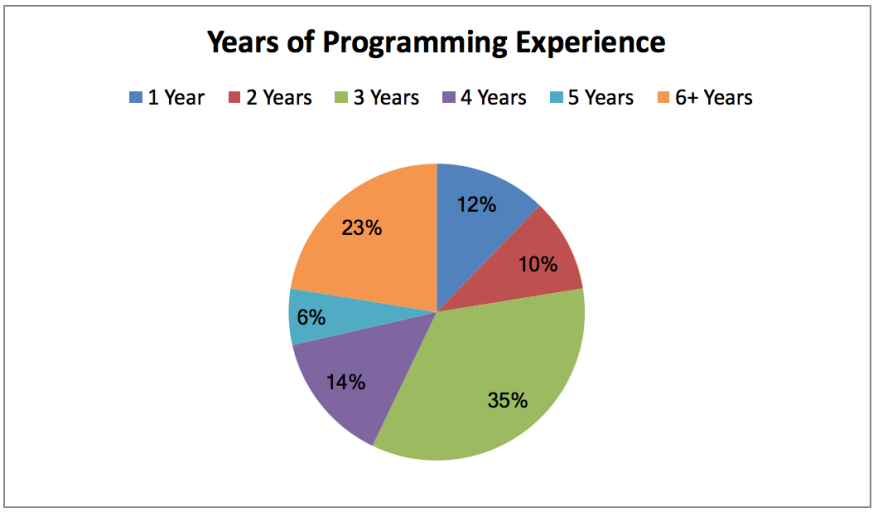


Figure 4.1: Years of Programming Experience

The chart for 'Years of Programming Experience' (Figure 4.1) shows that 35% of participants have 3 years of experience in programming, and 23% - have 6 years of experience. Those are the two largest proportions for the respondents' programming practice who took part in this research. The smallest share was 5 years with the percentage of 6% overall. As stated in the questionnaire, study years have been counted as an experience. Moreover, this chart illustrates that various respondents with various backgrounds in programming were willing to participate in this study.

4.2 Exclusions of Zero values

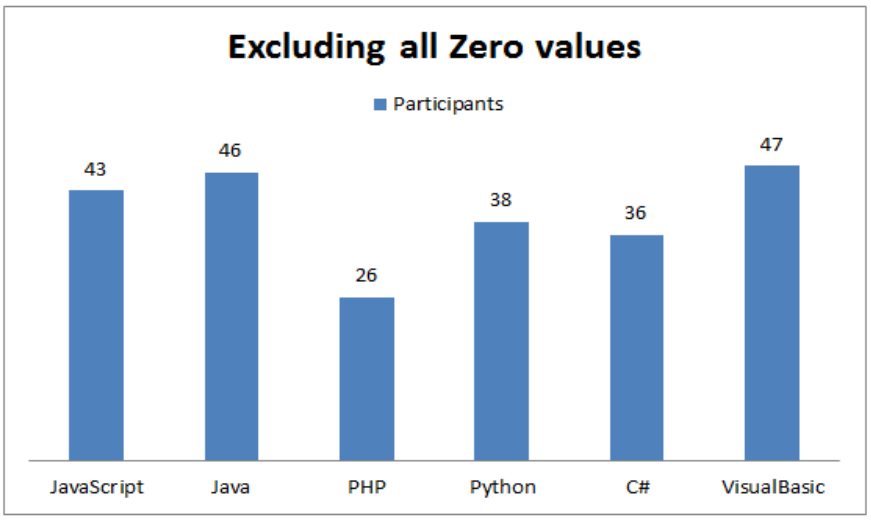


Figure 4.2: Numeric zero values excluded for each programming language

'Excluding all Zero Values' (Figure 4.2) demonstrates that out of 49 participants zero values were excluded from each of the programming language choices were zero means 'never used the language'. Outcome is 43 participants are using JavaScript, 46 – Java, 26 – PHP, 38 – Python, 36 – C#, and 47 – VisualBasic.

4.3 Assumptions

First, before data can be analysed and the ordinal logistic regression analysis conducted, certain assumptions had to be met to qualify for the model and interpretation. Two out of four assumptions have been met for all of the languages as discussed in methodologies section. The other two assumptions will be looked at in this section.

As mentioned in methodologies, to be able to qualify for the ordinal logistic regression analysis giving the limitations of the participant count, decision has been made to merge two options from the programming language section in questionnaire. That is, answer 'Like it a lot' has been merged with answer 'My Favourite' and made into one category called 'Like it a lot'. Given this change all the programming languages passed all four assumptions.

4.3.1 3rd assumption – 'Multicollinearity'

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.819	1.220
Agreeableness	0.922	1.085
Conscientiousness	0.732	1.366
Neuroticism	0.847	1.181
Openness	0.914	1.094

a. Dependent Variable: C#

Table 4.1: Multicollinearity test for C# language

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.903	1.108
Agreeableness	0.906	1.103
Conscientiousness	0.779	1.283
Neuroticism	0.819	1.221
Openness	0.904	1.107

a. Dependent Variable: Java

Table 4.2: Multicollinearity test for Java language

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.794	1.260
Agreeableness	0.880	1.136
Conscientiousness	0.717	1.395
Neuroticism	0.757	1.322
Openness	0.920	1.087

a. Dependent Variable: JavaScript

Table 4.3: Multicollinearity test for JavaScript language

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.748	1.336
Agreeableness	0.689	1.451
Conscientiousness	0.714	1.400
Neuroticism	0.668	1.497
Openness	0.815	1.228

a. Dependent Variable: PHP

Table 4.4: Multicollinearity test for PHP language

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.836	1.196
Agreeableness	0.829	1.206
Conscientiousness	0.799	1.252
Neuroticism	0.875	1.143
Openness	0.863	1.159

a. Dependent Variable: Python

Table 4.5: Multicollinearity test for Python language

Coefficients[a]

Model	Collinearity Statistics	
	Tolerance	VIF
Extroversion	0.849	1.178
Agreeableness	0.891	1.122
Conscientiousness	0.800	1.250
Neuroticism	0.862	1.160
Openness	0.872	1.147

a. Dependent Variable: VisualBasic

Table 4.6: Multicollinearity test for VisualBasic language

Firstly, tolerance level for each programming language must be above 0.1 (the lowest tolerance level is 0.668) and VIF (Variance Inflation Factor) levels must be

below 10 (with the highest being 1.497), which is shown for PHP programming language (Table 4.4).

This indicates that each individual independent variable is not related with each other independent variable. Therefore, it could be said that each individual independent variable will not jeopardise the understanding of how each independent variable affects dependent variable and also no technical issues associated with calculation of ordinal logistic regression (Field, 2009). All the tables above have met the assumption of no multicollinearity.

4.3.2 4th and fundamental assumption – Proportional odds

C#	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	1.061	1.105	1.017
Agreeableness	1.031	1.029	1.011
Conscientiousness	1.077	1.033	0.960
Neuroticism	0.888	0.934	1.006
Openness	1.028	0.977	0.834

Table 4.7: Proportional Odds Ratio for C# language

Java	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	1.055	1.011	0.997
Agreeableness	0.952	0.978	0.978
Conscientiousness	1.024	1.011	0.962
Neuroticism	1.015	1.026	1.015
Openness	0.844	0.943	0.962

Table 4.8: Proportional Odds Ratio for Java language

JavaScript	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	0.927	0.965	0.991
Agreeableness	0.956	1.005	0.906
Conscientiousness	1.011	0.942	1.039
Neuroticism	1.130	1.057	0.937
Openness	0.833	0.972	0.936

Table 4.9: Proportional Odds Ratio for JavaScript language

PHP	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	0.927	0.974	0.997
Agreeableness	0.975	0.988	0.952
Conscientiousness	1.013	0.977	1.129
Neuroticism	1.106	0.997	0.915
Openness	0.864	1.007	0.857

Table 4.10: Proportional Odds Ratio for PHP language

Python	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	0.974	0.994	1.001
Agreeableness	0.879	0.938	1.023
Conscientiousness	1.101	1.082	1.008
Neuroticism	0.890	0.914	0.993
Openness	1.029	0.982	0.975

Table 4.11: Proportional Odds Ratio for Python language

VisualBasic	Exp(B) (Odds Ratio, OR)		
	Cat1	Cat2	Cat3
Extroversion	1.026	1.020	0.986
Agreeableness	1.039	0.986	1.006
Conscientiousness	0.957	0.952	0.988
Neuroticism	0.912	0.881	0.974
Openness	1.015	1.021	1.007

Table 4.12: Proportional Odds Ratio for VisualBasic language

The odds ratio tables (Table 4.7 to 4.12) are testing the differences for each cumulative odds categories of each personality trait. Biggest variation is in C#

programming language (Table 4.7). Where Openness measured against each of the three categories showed the sizable difference between ‘Cat1’ and ‘Cat3’ being 0.194.

Even though the difference was 0.194, the assumption of proportional odds is still tenable. Overall, no major differences in values have been found. Therefore, it could be stated that estimated parameters are equally valid. Explaining each binomial logistic regression run on each dichotomised cumulative category were; Cat1 is ‘Do not like it’ versus ‘Slightly dislike it’, ‘Like it’, ‘Like it a lot’. Cat2 is ‘Do not like it’, ‘Slightly dislike it’ versus ‘Like it’, ‘Like it a lot’ and Cat3 is ‘Do not like it’, ‘Slightly dislike it’, ‘Like it’ versus ‘Like it a lot’. Tables above have met the final assumption of proportional odds. Therefore, further analyses of ordinal logistic regression could be investigated.

4.4 Exploring and Interpreting Outcome

Number of explanations about limitations and table meanings are given, followed by interpretation.

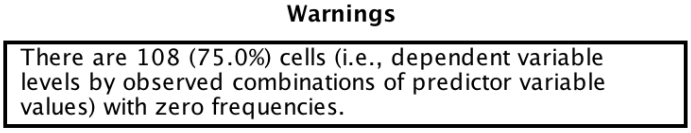


Figure 4.3: SPSS warning of cell frequency

At first, the limitations had to be discussed before the outcome could be analysed. When each of the run on each of the programming language has been completed warning message has been issued by SPSS tool. Figure 4.3 shows that there are not enough cell frequencies. Cell frequency is how frequent is each case. In this research because independent variables are continuous and there are five traits it is very implausible that participant with each exact Big Five trait values will choose the same dependent category.

Therefore, the Goodness-of-Fit data will not be interpreted as it cannot be relied upon because of the cell frequencies (Statistics.laerd.com, 2017).

Model Fitting Information, Pseudo R-Square and Parameter Estimates (Coefficients) are the main tables from ordinal logistic regression that will be interpreted.

Model fitting information table is checking the comparison between the model and the baseline. In other words, it shows whether the model has improved when all five independent variables were added comparing to just intercept with no predictors. The '-2 Log Likelihood' (deviance) is used to measure the difference between 'Intercept only' and 'Final' models giving the value of Chi-Square. This means the outcome prediction is improved when the Chi-Square is higher. With significance (Sig.) level being below 0.05 indicating that independent variables improve the prediction of the dependent variables (Field, 2009; Statistics.laerd.com, 2017).

Pseudo R-Square model includes Cox and Snell's and Nagelkerke's statistics, which are known to be similar for the interpretation as they both provide indication of the substantive significance of the model, or so called 'variance explained' measures (Field, 2009). In other words, these measures show the percentage that is explained by independent variables. However due to their similarities, only Nagelkerke will be referenced in this research.

Parameter Estimates table (other known as 'Coefficients') explains individual influence of each independent variable on the outcome variable within the context of the model. The 'B' column for continuous independent variables is not as intuitive as it would be if the variables were dichotomous and polytomous (more than two options). Therefore, 'Exp(B)' column value of the odds ratio is used, which is an indicator of the change in odds that occurs from a unit change in the predictor (Field, 2009; Statistics.laerd.com, 2017).

All tables below will only include the information necessary for the analyses and interpretation. However, all the original table screenshots for each of the language could be found in Appendix B.

4.4.1 C# programming language

Model Fitting Information - C#

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	85.809		
Final	80.015	5.794	0.327

Table 4.13: Model Fitting Information C# language

'Model Fitting Information' table (Table 4.13) illustrates that final model is not statistically significant in predicting the dependent variable against the intercept only model, with Chi-Square being 5.794 and the significance is above 0.05 (0.327 > 0.05).

Adding all Big Five variables as shown by the 'Model Fitting Information' it can be summarised that given all variables together there is no improvement against the model that has no independent variables.

Pseudo R-Square - C#

Nagelkerke	0.164

Table 4.14: Pseudo R-Square C# language

This table (Table 4.14) displays the Nagelkerke outcome of 0.164 or 16.4% (0.164x 100). According to Nagelkerkle 16.4% of the variance in the model could be explained by explanatory variables. That is 83.6% of the variance is defined by other factors.

Parameter Estimates - C#

Parameter		Hypothesis Test		95% Wald Confidence Interval for Exp(B)	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	-0.028	0.213	0.972	0.931	1.016
Agreeableness	-0.010	0.624	0.990	0.950	1.032
Conscientiousness	0.021	0.359	1.021	0.977	1.067
Neuroticism	0.012	0.582	1.012	0.969	1.058
Openness	-0.012	0.687	0.988	0.933	1.047
(Scale)	1				

Table 4.15: Parameter Estimates - C# language

Parameter Estimates (Table 4.15) presents that none of the independent variables are statistically significant with the closes being Extraversion with Sig. of 0.213 (0.213 > 0.05).
Running the analysis for the C# programming language data suggests that the value of dependent variable cannot be predicted by any of the independent variables given the current figures.

4.4.2 Java programming language

Model Fitting Information - Java

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	113.254		
Final	104.361	8.893	0.113

Table 4.16: Model Fitting Information Java language

Testing 'Model Fitting Information' for Java programming language it can be seen that the model we are trying to fit is not statistically significant with the value of Sig. 0.113 > 0.05 and the Chi-Square value of 8.893 (Table 4.16).
It could be recapped that independent variables together do not significantly enhancing the intercept only model.

Pseudo R-Square - Java

Nagelkerke	0.192

Table 4.17: Pseudo R-Square Java language

Nagerkerke measures that 19.2% variance is explained by independent variables (Table 4.17).
As per Nagelkerkle 19.2% of the variance in the model could be clarified by predictors. The other factors are responsible for the remaining 80.8%.

Parameter Estimates - Java

Parameter		Hypothesis Test		95% Wald Confidence	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	-0.006	0.739	0.994	0.960	1.029
Agreeableness	0.025	0.191	1.025	0.988	1.065
Conscientiousness	0.014	0.520	1.014	0.972	1.057
Neuroticism	-0.011	0.577	0.989	0.951	1.028
Openness	0.055	0.043	1.057	1.002	1.114
(Scale)	1				

Table 4.18: Parameter Estimates - Java language

Four out of five independent variables did not have statistically significant prediction of dependent variable. Significance values are 0.739; 0.191; 0.520; 0.577 respectively. Openness as an independent variable is statistically significant to the model with 0.043 < 0.05 (Table 4.18).

Model in general with all independent variables together has no significant value to the dependent variable. However, openness measured as an individual variable in this model has significance over the dependent variable as illustrated in 'Parameter Estimates' table.

This could be further analysed by looking at the value of Exp(B) (odds ratio). This means that with every unit increase in openness, odds of being in a higher dependent category increases by 1.057 times with the 95% confidence value being between lower 1.002 and upper 1.114.

In other words, it could be explained that programmers who have more intellectual interest, are creative and more interested in how the process works (openness) (John and Srivastava, 1999), prefers Java programming language. Java language to be known as reasonably hard for beginners requires exploration as it has numerous features and covers every angle of the programming world with lots of plug-ins, libraries, design patterns and concepts (Evans and Flanagan, 2014).

4.4.3 JavaScript programming language

Model Fitting Information - JavaScript

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	109.132		
Final	100.015	9.116	0.105

Table 4.19: Model Fitting Information JavaScript language

Analysing 'Model Fitting Information' table indicates that final model for JavaScript language is not statistically significant in predicting the dependent variable over the intercept only model with the value of 0.105 > 0.05 and the Chi-Square value of 9.116 (Table 4.19).

Five independent variables that have been fitted into this model showed no significant prediction of the outcome. Although significance value of 0.105 is fairly close to 0.05 of being significant.

Pseudo R-Square - JavaScript

Nagelkerke	0.207

Table 4.20: Pseudo R-Square JavaScript language

For JavaScript language Nagelkerke shows that 20.7% variance is explained by independent variables (Table 4.20).

The Nagelkerke percentage could be explained by explanatory variables and the remaining 79.3% depends on other factors.

Parameter Estimates - JavaScript

Parameter		Hypothesis Test		95% Wald Confidence Interval for Exp(B)	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	0.019	0.290	1.019	0.984	1.055
Agreeableness	0.030	0.134	1.031	0.991	1.072
Conscientiousness	0.024	0.276	1.025	0.981	1.070
Neuroticism	-0.021	0.308	0.979	0.941	1.019
Openness	0.035	0.172	1.035	0.985	1.088
(Scale)	1				

Table 4.21: Parameter Estimates - JavaScript language

Parameter Estimates table – Coefficients shows that none of the independent variables are statistically significant with the closes significance being 0.134 > 0.05 (Table 4.21).
It could be interpreted that the value of dependent variable cannot be predicted by the independent variables given the current figures for programming language of JavaScript.

4.4.4 PHP programming language

Model Fitting Information - PHP

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	67.981		
Final	64.122	3.859	0.570

Table 4.22: Model Fitting Information PHP language

By examining the final model it can be noticed that it did not statistically significantly predict the outcome variable over the baseline model. Significance in this model is 0.570> 0.05, and Chi-Square being 3.859 (Table 4.22).
Out of all six languages, the outcome for PHP programming language shows least significance in predicting the dependent variable, given all five independent variables.

Pseudo R-Square - PHP

Nagelkerke	0.149

Table 4.23: Pseudo R-Square PHP language

Nagelkerke in this configuration shows that 14.9% variance is justified by independent variables (Table 4.23).
Other factors are accountable for the remaining 85.1% variance according to Nagelkerke Pseudo R-Square measure.

Parameter Estimates - PHP

Parameter		Hypothesis Test		95% Wald Confidence Interval for Exp(B)	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	0.022	0.378	1.022	0.974	1.072
Agreeableness	0.017	0.556	1.018	0.960	1.078
Conscientiousness	-0.006	0.831	0.994	0.943	1.048
Neuroticism	0.000	0.993	1.000	0.949	1.054
Openness	0.033	0.284	1.034	0.973	1.099
(Scale)	1				

Table 4.24: Parameter Estimates - PHP language

For PHP language Parameter Estimates table demonstrates that none of the independent variables are statistically significant with the closes being Sig. 0.284 > 0.05 (Table 4.24).

It could be defined that the value of dependent variable cannot be forecasted by the independent variables given the current figures for programming language of PHP.

4.4.5 Python programming language

Model Fitting Information - Python

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	95.185		
Final	88.088	7.097	0.214

Table 4.25: Model Fitting Information Python language

Chi-Square measure in this model is 7.097 and significance above 0.05 (0.214>0.05). Therefore, the final model did not statistically significantly predict the dependent variable (Table 4.25).

This means that intercept model is not complemented by all of the independent variables together.

Pseudo R-Square - Python

Nagelkerke	0.186

Table 4.26: Pseudo R-Square Python language

Pseudo R-Square measure of Nagelkerke shows 18.6% variance (Table 4.26).

These are defined by independent variables. Other variances of 83.4% are dependent on other factors.

Parameter Estimates - Python

Parameter		Hypothesis Test		95% Wald Confidence Interval for Exp(B)	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	0.013	0.438	1.013	0.980	1.048
Agreeableness	0.012	0.561	1.012	0.972	1.053
Conscientiousness	-0.031	0.197	0.970	0.925	1.016
Neuroticism	0.046	0.032	1.047	1.004	1.091
Openness	0.026	0.310	1.027	0.976	1.080
(Scale)	1				

Table 4.27: Parameter Estimates - Python language

The Parameter Estimates table displays that neuroticism independent variable is statistically significant in predicting the outcome variable with the Sig. value being 0.032<0.05. The other four explanatory variables are not significant towards the prediction of the dependent variable with the second closest significance value 0.197 of conscientiousness (Table 4.27).

Although all independent variables together have no significant importance to the outcome variable, neuroticism as a single explanatory variable has statistically significant prediction.

The further analysis of Exp(B) value for neuroticism can be explained as a prediction of odds ratio. This implies that with every unit increase in neuroticism it increases the odds of choosing higher dependent category by 1.047 times with the 95% confidence value being between lower 1.004 and upper 1.091.

The increase in neuroticism means that programmers who are more anxious, emotional, nervous, worries a lot, and feels tense (John and Srivastava, 1999) prefers Python programming language. Python programming language does not require a lot of time to learn, syntax is known to be one of the easiest to read as it defines its readability as key feature, with relatively few keywords and plain structure (Chun, 2006).

4.4.6 VisualBasic programming language

Model Fitting Information - VisualBasic

Model	-2 Log Likelihood	Chi-Square	Sig.
Intercept Only	108.884		
Final	100.225	8.619	0.125

Table 4.28: Model Fitting Information VisualBasic language

The Model Fitting Information table for VisualBasic programming language illustrates that final model did not statistically significantly predict the dependent variable against the baseline model. The significance value being 0.125 > 0.05 and the Chi-Square value – 8.619 (Table 4.28).

Pseudo R-Square - VisualBasic

Nagelkerke	0.186

Table 4.29: Pseudo R-Square VisualBasic language

This Pseudo R-Square table displays Nagelkerke outcome of 18.6% variance which is explained by independent variables (Table 4.29).
The variance of 81.4% is defined by other elements.

Parameter Estimates - VisualBasic

Parameter		Hypothesis Test		95% Wald Confidence Interval for Exp(B)	
	B	Sig.	Exp(B)	Lower	Upper
Extroversion	0.005	0.776	1.005	0.972	1.039
Agreeableness	0.000	0.992	1.000	0.962	1.039
Conscientiousness	0.022	0.315	1.023	0.979	1.068
Neuroticism	0.045	0.030	1.046	1.004	1.088
Openness	-0.024	0.352	0.976	0.928	1.027
(Scale)	1				

Table 4.30: Parameter Estimates - VisualBasic language

There was no statistically significant indication of dependent variable prediction of the four out of five independent variables. On the other hand, neuroticism is

statistically significant towards the prediction of target variable with the Sig. being 0.030<0.05 (Table 4.30).
Neuroticism is the only variable as shown in this model that has influence over the dependent variable.
In addition, investigation of Exp(B) value for neuroticism can be summarised as forecast of odds ratio. This suggests that every unit increase in neuroticism, the odds increases of choosing higher dependent category by 1.046 times with the 95% confidence value being between lower 1.004 and upper 1.088.
This could be explained that programmers who have higher score in neuroticism are more nervous, emotionally unstable, worries a lot, and feels tense (John and Srivastava, 1999) prefers VisualBasic programming language. VisualBasic – event driven programming language, allows user to navigate easier in user interface creation by dragging and dropping certain required features without any need to code (Stephens, 2008; Willis and Newsome 2010).

4.5 Correlation analysis using Mean and Standard Deviation calculations

For the second analysis of data, standard deviation and mean calculations were used. Firstly, each programming language answers from the survey were divided into two groups. First group of 'Do not like it' refers to the values of 1 and 2, and second group of 'Like it' refers to the values of 3 and 4. Tables below shows three values where N is the number of instances or count, mean is an average of each personality trait and standard deviation is how much the value deviates from the mean. High deviation value would not show correlation and low value of deviance would show correlation.

4.5.1 JavaScript programming language

JavaScript "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	16	16	16	16	16
Mean	50.16	68.59	56.88	56.41	66.41
Standard Deviation	15.95	13.57	13.18	16.46	13.13

Table 4.31: JavaScript 'Do not like it' category

JavaScript "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	27	27	27	27	27
Mean	53.61	72.13	62.87	48.98	71.67
Standard Deviation	20.00	16.67	16.48	15.78	11.20

Table 4.32: JavaScript 'Like it' category

JavaScript ‘Do not like it’ category table (Table 4.31) shows high dispersion for each of the personality trait which could not be interpreted either way because of the mean values. Mean values are too close to the 50% of personality trait which would indicate that the one is neither high or low in particular trait. Even when taking the highest mean value, in this case Agreeableness with 68.59, standard deviation being 13.57. It would explain that the lowest value of the mean would be 55.02 which drop too close to the middle range where the level of Agreeableness could not be explained.

JavaScript ‘Like it’ table (Table 4.32) is slightly different, even with the dispersion being too high for each of the trait. Examining openness trait, mean value is 71.67 which would indicate that participants with higher openness are more likely to like the JavaScript programming language than do not like it. However, it is not significant given the value of the standard deviation of 11.20. Although the Agreeableness mean value is slightly higher than Openness it will not be looked at as the standard deviation value is also higher.

4.5.2 Java programming language

Java "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	12	12	12	12	12
Mean	53.33	65.42	60.21	58.75	64.58
Standard Deviation	15.39	19.18	16.08	13.16	10.27

Table 4.33: Java 'Do not like it' category

Java "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	34	34	34	34	34
Mean	51.47	73.46	58.82	51.25	71.47
Standard Deviation	18.24	14.01	14.79	16.57	12.03

Table 4.34: Java 'Like it' category

Java ‘Do not like it’ table (Table 4.33) illustrates that there is no correlation between each of the personality traits and respondents that do not like the language. All the mean values are close to the middle range with a high standard deviation. With the openness mean value 64.58 and standard deviation of 10.27 the deviance range is too wide for the correlation.

Java ‘Like it’ category table (Table 4.34) demonstrates high standard deviation value in each personality trait. This means that deviance range is too ample. On the other hand, agreeableness and openness mean values are 73.46 and 71.47 respectively. This explains that a participant with higher agreeableness and openness likes Java language, although standard deviations are too high with 14.01 and 12.03 accordingly, which is not significant to say that there is a correlation.

4.5.3 PHP programming language

PHP "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	10	10	10	10	10
Mean	47.00	67.50	54.50	53.25	68.25
Standard Deviation	18.70	16.24	13.63	19.76	14.44

Table 4.35: PHP 'Do not like it' category

PHP "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	16	16	16	16	16
Mean	54.53	71.56	58.59	56.72	69.38
Standard Deviation	16.91	14.72	17.65	15.29	13.24

Table 4.36: PHP 'Like it' category

PHP ‘Do not like it’ category table (Table 4.35) explains that the mean values are too close to the half-way point for each of the personality trait. Also, standard deviation values are considerably sizable to be able to have any correlation.
PHP ‘Like it’ table (Table 4.36) shows that agreeableness has high value of mean – 71.56 which explains that programmers with high agreeableness are more expected to like PHP programming language. However, the standard deviation of 14.72 overrules this model.

4.5.4 Python programming language

Python "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	10	10	10	10	10
Mean	40.75	61.75	64.75	43.25	67.25
Standard Deviation	19.33	18.56	14.83	15.14	9.46

Table 4.37: Python 'Do not like it' category

Python "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	28	28	28	28	28
Mean	53.48	72.59	56.79	55.00	70.98
Standard Deviation	18.96	15.08	14.35	16.07	14.03

Table 4.38: Python 'Like it' category

Python ‘Do not like it’ table (Table 4.37) illustrates that each of the personality trait mean values are not significant enough to interpret with the high standard deviation. However, the ‘Like it’ category table (Table 4.38) demonstrates that Agreeableness and Openness mean values are in a higher range being 72.59 and 70.98 respectively. Explaining that participants who likes Python language would be more friendly, warm and likable (agreeableness) and curious, imaginative and open minded (openness) (John and Srivastava, 1999). Yet it cannot be concluded because of the high standard deviation.

4.5.5 C# programming language

C# "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	8	8	8	8	8
Mean	59.06	74.06	55.94	49.69	69.06
Standard Deviation	16.14	15.58	20.26	18.59	11.87

Table 4.39: C# 'Do not like it' category

C# "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	28	28	28	28	28
Mean	45.45	66.96	59.73	57.05	66.79
Standard Deviation	15.40	16.75	16.32	15.06	11.68

Table 4.40: C# 'Like it' category

Table 'Do not like it' (Table 4.39) for C# programming language shows no correlation between each personality trait and respondents who do not like the language. Nevertheless, agreeableness shows high mean value of 74.06 defining that programmers who score high in agreeableness would not like the language. However, there are two limitations to this argument. First the standard deviation is huge with the value of 15.58 and secondly, participant count is eight.
C# 'Like it' category table (Table 4.40) displays no significantly low or high mean values. Also, standard deviation values for each personality trait are high. And it can be summarised that latter table is not showing any correlation.

4.5.6 VisualBasic programming language

VisualBasic "Do not like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	10	10	10	10	10
Mean	53.00	71.75	51.50	37.00	72.50
Standard Deviation	24.83	15.28	10.88	11.83	16.54

Table 4.41: VisualBasic 'Do not like it' category

VisualBasic "Like it" category					
Personality Traits	**Extroversion**	**Agreeableness**	**Conscientiousness**	**Neuroticism**	**Openness**
N	37	37	37	37	37
Mean	50.81	69.39	61.42	56.89	66.96
Standard Deviation	16.77	16.11	15.55	14.45	10.58

Table 4.42: VisualBasic 'Like it' category

VisualBasic ‘Do not like it’ category table (Table 4.41) presents three aspects. Firstly, high agreeableness mean value – 71.75, low neuroticism value of 37.00 and another high value in openness of 72.50. Interpreting this data it could be explained that programmers with high score in agreeableness and openness and low score in neuroticism do not like VisualBasic programming language. However, looking at standard deviation values for each of the three personality traits mentioned above, no significance was found.

VisualBasic ‘Like it’ table (Table 4.42) shows no notable mean or standard deviation values. Therefore, to summarise there is no correlation between each of the personality traits and participant preference towards VisualBasic programming language.

5 CONCLUSION AND FUTURE WORK

5.1 Conclusion

Primary data (questionnaires) have been collected for the purpose of finding the answer to the main question of this study. Secondary data (literature review) was used for theoretical knowledge improvement and for comparison of previous researches done on similar projects. Lastly, results from both analyses were examined to provide the answer about the correlation between programming languages and personality.

Firstly, ordinal logistic regression analysis was used to analyse the results from online questionnaire. It has been noticed that using this statistical analysis three out of six languages have not showed any correlation with personality traits. The outcome variable was not predicted for JavaScript, PHP and C# programming languages. However, the other three programming languages showed correlation with some personality traits where explanatory variables being statistically significant for the prediction of outcome. Java programming language had a correlation with openness personality trait. Java programming language can be considered as difficult to learn for beginners with many features and libraries which require in depth understanding and attention to detail is a necessity (Eckel, 2006; Evans and Flanagan, 2014). Programmers who had high score in openness can be described as intelligent, open-minded, and eager to understand the process (John and Srivastava, 1999). Data suggests that respondents with openness could benefit if they would decide to choose Java programming language. Likewise, there was a correlation between Python programming language and neuroticism. Python code is easy to read and understand with excellent error traceability (Chun, 2006). Respondents who had high score in neuroticism can be described as anxious, inpatient and nervous (John and Srivastava, 1999). Programmers who have high neuroticism would enjoy working with easier programming languages which do not require in depth analyses, has simple structure and is easy to learn. The last correlation observed was VisualBasic with neuroticism. VisualBasic gives and option to draw user interface which fastens the process and removes the need for coding of visual appearance of the program (Willis and Newsome, 2010). As mentioned

previously, programmers who had high score in neuroticism can be described as anxious, angry and frustrated (John and Srivastava, 1999). Data suggests that the higher percentage of neuroticism, influences programmer's positive preference for VisualBasic programming language.

Secondly, mean and standard deviation calculations were used to analyse results. Conversely to ordinal logistic regression analysis, there was no significance found in the sense of correlation with any of the language and personality trait. Mean values suggested that certain aspects can be interpreted as a sign of correlation with the values being objectively high or low. However, the high values of standard deviation showed that the correlations were not statistically significant. Nevertheless, it can be said that mean and standard deviation calculations were based on limited (merged) results whereas ordinal logistic regression analysis measured each score of the likeliness separately. This resulted in different outcome from both analyses.

Similar researches completed, could be linked to findings of this study. Research studied on correlation between programmers' work-place behaviours and personality has been reviewed (Gnambs, 2015). The results showed that mostly mentioned personality traits for programmers were conscientiousness, openness and introversion. However, this study was based on programming languages rather than programing as overall concept, this previous research explains that programming and personality can be related. Moreover, the other research by Zahra Karimi (2016) has been completed to explore the correlation between programming styles and personality traits. Karimi (2016) research concluded that two out of five personality traits can be connected with two different programming styles. Openness personality trait was associated with breadth-first style which requires programmers to be more open-minded.

To summarise the findings, it can be noted that using ordinal logistic regression analysis, correlation found between three languages and two personality traits can be a start for further research to explore this question in more depth. Also, the previous research findings, correlation between openness and breadth-first programming style can be supplemented by the findings of this research with correlation between openness and Java programming language.

5.2 Limitations

Several limitations were detected during this study. Firstly, the participant count limits validity of this research. The number of respondents also limited the accuracy of the calculations when using mean and standard deviation analysis. Secondly, only top six programming languages were used in this research. Some of the participants may use different programming languages which were not included in this research.

5.3 Future work

This study can be described as an introduction in further future research which can be performed in finding the correlation between programming languages and five personality traits. As a future work, ordinal logistic regression could be looked at in more depth. For more ample analysis only four independent variables could be used by leaving one out to see how it would change the prediction of the outcome variable. Also, the independent variables count, for example giving only two or three variables, can be reformed to examine the difference in the outcome variable. Furthermore, each personality trait that showed significant correlation could be separated and applied to see how the model as a whole is improved. Moreover, a wider range of programming languages could be added to the model to improve the variety of choice.

In addition, binomial regression investigation can be used for examination of the model. Binomial regression analysis in a sense is similar to mean and standard deviation analysis, but have some extra features when it comes to data analysis, such as optimising the outcome variables with proportions and different calculations. Therefore, minimising the options for respondents to choose, to only two where participant either likes the language or does not like it.

5.4 Summary

There is a wide range of different approaches that could have been used for this study to analyse the question in more depth. However, the aim of this research has been achieved. The sign of correlation has been found for three languages and no significant correlation has been found for other three languages. This research can

be used as a sample model of how personality traits can be connected with certain programming languages. It can also be beneficial for programming industry, in understanding whether personality can affect the choice of programming language.

6 REFERENCES

Agresti, A. (2010). *Analysis of ordinal categorical data.* 2nd ed. Hoboken: John Wiley & Sons.

Ahmed, F., Campbell, P., Beg, A., Capretz, L. F. (2011). Soft Skills Requirements in Software Architecture's Job: An Exploratory Study. In: *International Conference on Computer Communication and Management (ICCCM)*, Sydney, Australia, 2011, 565-569.

Ajzen, I. (2005). *Attitudes, Personality and Behaviour*. 2nd ed. Berkshire: Open University Press.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: H. Holt and Company.

Allport, G.W., Odbert, H.S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1), p.i-171.

Backus, J. (1978). The history of FORTRAN I, II, and III. *ACM SIGPLAN Notices*, 13(8), pp.165-180.

Bayne, R. (1994). The "Big Five" versus the Myers-Briggs. *The Psychologist*, 7 (1), pp.14-16.

Bishop-Clark, C. (1995). Cognitive Style, Personality, and Computer Programming. *Computers in Human Behaviour*, 11(2), pp.241-260.

Boyle, G. J. (1995). *Myers-Briggs Type Indicator (MBTI):* Some psychometric limitations. *Australian Psychologist*, 30(1), pp.71-74.

Briggs-Myers, I., & Briggs, K.C. (1985). *Myers-Briggs Type Indicator (MBTI)*. Palo Alto, Calif: Consulting Psychologists Press.

Cattell, R.B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 38, pp.476-506.

Chun, W. J. (2006). *Core Python Programming*. 2nd ed. London: Prentice Hall.

Cox, A., Fisher, M. (2009). Programming style: influences, factors and elements. *Proceedings of the Second International Conferences on Advances in Computer-Human Interactions*. IEEE, pp.82-89.

Cruz, S., Da Silva, F.Q.B., Capretz, L.F. (2015). Forty years of research on personality in software engineering: A mapping study. *Computers in Human Behaviour*, 46, pp.94-113.

Eckel, B. (2006). *Thinking in Java,* 4th ed. Upper Saddle River, Prentice Hall.

Evans, J.B., Flanagan, D. (2014). *JAVA in a Nutshell,* 6th ed. Sebastopol: O'Reilly Media.

Feigenspan, J., Kastner, C., Liebig, J., Apel, S., Hanenberg, S. (2012). Measuring programming experience, (Program Comprehension, Germany, 11-13 June 2012) *IEEE*. Publication Year: 2012, pp.73-82

Field, A. (2009). *Discovering statistics using SPSS*. 3rd ed. London: SAGE Publications Ltd.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *Journal of Abnormal and Social Psychology*, 44, pp.329-344.

Flanagan, D., Ferguson, P. (2002). *JavaScript: The Definitive Guide*, 4th ed. Sebastopol: O'Reilly Media.

Gnambs, T. (2015). What makes a computer wiz? Linking personality traits and programming aptitude. *Journal of Research in Personality*, 58, pp.31-34.

Goldberg, L., Johnson, J., Eber, H., Hogan, R., Ashton, M., Cloninger, C. and Gough, H. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), pp.84-96.

Gosling, J., Joy, B., Steele, G., Bracha, G. (2000). *Java(TM) Language Specification*, 2nd ed. Reading: Addison-Wesley.

Gosling, S.D., Rentfrow P.J., Swann Jr., W.B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality,* 37, pp.504-528.

Ipip.ori.org. (2017). *IPIP Home*. [online] Available at: http://ipip.ori.org/ [Accessed 11 Jan. 2017].

John, O. P., Lawrence, A. (1999). *Handbook of Personality*, 2nd ed. New York/ London: The Guilford Press.

John, O.P., Naumann, L.P., Soto. C.J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O.P. John, W. Robins, L.A. Pervin (Eds.), *Handbook of personality: Theory and research*, New York: Guilford Press, pp.114-158

Karimi, Z., Baraani-Dastjerdi, A., Ghasem-Aghaee, N., Wagner, S. (2016). Links between the personalities, styles and performance in computer programming. *The Journal of Systems and Software*, 111, pp.228-241.

Landau, S., Everitt, B. (2004). *A handbook of statistical analyses using SPSS*. 1st ed. Boca Raton, FL: CRC Press.

Lerdorf, R., Tatroe, K., MacIntyre, P. (2006). *Programming PHP*, 2nd ed. Sebastopol: O'Relly Media.

Lerdorf, R., Tatroe, K., MacIntyre, P. (2013). *Programming PHP*, 3rd ed. Sebastopol: O'Relly Media.

Lindholm, T., Yellin, F. (1999). *The Java Virtual Machine Specification*, 2nd ed. Palo Alto: Prentice-Hall Int.

Lovrenčić, A., Konecki, M., Orehovački, T. (2009). 1957 - 2007: 50 Years of Higher Order Programming Languages. *Journal of Information and Organizational Sciences*, 33(1) [Peer Reviewed Journal]

Lutz, M. (2007). *Learning Python*, 3rd ed. Sebastopol: O'Reilly Media.

McAdams, D. P., Pals, J. L. (2006). A New Big Five, Fundamental Principles for an Integrative Science of Personality. *American Psychologist*, 61(3), pp.204-217.

McCrae, R.R., John, O.P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60(2), pp.175-215.

Nash, T. (2007). *Accelerated C#,* New York: Apress.

Nofre, D., Priestley, M., Alberts, G. (2014). When Technology Became Language: The Origins of the Linguistic Conception of Computer Programming, 1950-1960. *Technology and Culture*, 55(1), pp.40-75 [Peer Reviewed Journal].

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 66(6), pp.574-583.

O'Connell, A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks: Sage publications.

Orsted, M. (2000). Software Development Engineer in Microsoft – A subjective view of soft skills required, In: *Proceedings of the 22nd international conference on software engineering*. Dublin, Ireland, 01 June 2000. 539-540.

Ortin, F. (2011). Type Inference to Optimize a Hybrid Statically and Dynamically Typed Language. *The Computer Journal*, 54(11), pp.1901-1924 [Peer Reviewed Journal].

Pierce, B.C. (2002). *Types and Programming Languages*. Cambridge: The MIT Press.

Roccas, S., Sagiv, L., Schwartz, S., Knafo, A. (2002). The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin*, 28(6), pp.789-801.[Peer reviewed journal]

Saunders, M., Lewis, P. and Thornhill, A. (2007). *Research methods for business students.* 1st ed. Harlow: Pearson.

Stackoverflow.com. (2017). *Stack Overflow*. [online] Available at: http://stackoverflow.com [Accessed 10 Jan. 2017].

Statistics.laerd.com. (2015). *Ordinal logistic regression using SPSS Statistics. Statistical tutorials and software guides*. [online] Available at: https://statistics.laerd.com [Accessed 20 Feb. 2017].

Stavrianos, A., Syropoulos, A. (2014). *Using Scripting Languages to Teach Programming*. Cornell University.

Stephens, R. (2008). *Visual Basic 2008 Programmer's Reference (Programmer to Programmer)*. Indianapolis: Wiley Publishing.

Surakka, S. (2005). Analysis of Technical Skills in Job Advertisements Targeted at Software Developers. *Informatics in Education*, 4(1), pp.101-122.

Surveymonkey.com. (2017). *SurveyMonkey: Free online survey software & questionnaire tool*. [online] Available at: http://www.surveymonkey.com [Accessed 5 Feb. 2017].

Tupes, E. C., Christal, R. C. (1961). *Recurrent personality factors based on trait ratings*. Technical Report, U.S. Air Force, Lackland Air Force Bace, TX.

Venkatesan, V., Sankar, A. (2014). Investigation of Student's personality on pair programming to enhance the learning activity in the academia. *Journal of Computer Sience*, 10(10), pp. 2020-2028.

Willis, T., Newsome, B. (2010). *Beginning Visual Basic 2010*. 1st ed. Hoboken: John Wiley & Sons.

7 APPENDICES

7.1 Appendix A

Ethics form and questionnaire

Ethics approval number: 2016D0277

When undertaking a research or enterprise project, Cardiff Met staff and students are obliged to complete this form in order that the ethics implications of that project may be considered.

If the project requires ethics approval from an external agency (e,g., NHS), you will not need to seek additional ethics approval from Cardiff Met. You should however complete Part One of this form and attach a copy of your ethics letter(s) of approval in order that your School has a record of the project.

The document ***Ethics application guidance notes*** will help you complete this form. It is available from the Cardiff Met website. The School or Unit in which you are based may also have produced some guidance documents, please consult your supervisor or School Ethics Coordinator.

Once you have completed the form, sign the declaration and forward to the appropriate person(s) in your School or Unit.

PLEASE NOTE:
Participant recruitment or data collection MUST NOT commence until ethics approval has been obtained.

PART ONE

Name of applicant:	Egidijus Jankauskas
Supervisor (if student project):	Mohamed Mostafa
School / Unit:	School of Management
Student number (if applicable):	St20092233
Programme enrolled on (if applicable):	BSc (Hons) Computing
Project Title:	Investigate relationship between personality and programming characteristics through favourite programming language
Expected start date of data collection:	15/01/2017
Approximate duration of data collection:	6 Weeks
Funding Body (if applicable):	N/A
Other researcher(s) working on the project:	None

Will the study involve NHS patients or staff?	No
Will the study involve human samples and/or human cell lines?	No

Does your project fall entirely within one of the following categories:	
Paper based, involving only documents in the public domain	No
Laboratory based, not involving human participants or human samples	No

Practice based not involving human participants (eg curatorial, practice audit)	No
Compulsory projects in professional practice (eg Initial Teacher Education)	No
A project for which external approval has been obtained (e.g., NHS)	No
If you have answered YES to any of these questions, expand on your answer in the non-technical summary. No further information regarding your project is required. If you have answered NO to all of these questions, you must complete Part 2 of this form	

In no more than 150 words, give a non-technical summary of the project
Our world cannot be imagined without computers and especially without certain computer programs. Those programs are created by computer programmers, but each computer programmer uses different programming languages and can approach same project from different angles. This research will be based on the multiple surveys of programing students and programmers to answer the question if the different personality traits can be affiliated with certain programming languages. Tests will be taken based on Big Five Personality questionnaires to examine each personality trait and compare the results with different programing languages the respondents are using mostly. Furthermore, this research will involve the literature review of academic journals and books to compare the findings. The results will then be analysed and findings discussed whether the research will be positive or negative in answering the question.

DECLARATION:
I confirm that this project conforms with the Cardiff Met Research Governance Framework

I confirm that I will abide by the Cardiff Met requirements regarding confidentiality and anonymity when conducting this project.

STUDENTS: I confirm that I will not disclose any information about this project without the prior approval of my supervisor.

Signature of the applicant: Egidijus Jankauskas	Date: 05/12/2016

FOR STUDENT PROJECTS ONLY	
Name of supervisor: Mohammed Mostafa	Date: 05/12/2016
Signature of supervisor: 	

Research Ethics Committee use only	
Decision reached:	Project approved ☐ Project approved in principle ☐ Decision deferred ☐ Project not approved ☐ Project rejected ☐
Project reference number: Click here to enter text.	
Name: Click here to enter text.	Date: Click here to enter a date.
Signature:	
Details of any conditions upon which approval is dependant: Click here to enter text.	

PART TWO

A RESEARCH DESIGN	
A1 Will you be using an approved protocol in your project?	No
A2 If yes, please state the name and code of the approved protocol to be used[4]	
N/A	
A3 Describe the research design to be used in your project	
The overall project will have 4 main phases consisting of secondary data gathering as in literature review, primary data collection, analysis and conclusion. Primary data gathering utilising qualitative method (surveys) for the 2 main outcomes: personality using Big Five questionnaire and relevant programming language questionnaire. A positivistic research philosophy will be adopted involving a deductive research strategy for gathering qualitative data and will be obtained utilising convenience sampling technique. Following by analysing data that has been found in literature review and primary data gathering and comparing it. Finally, the conclusion of the overall research and findings as well as the answer to the main research question. Respondents that will undertake the survey are working with programming languages for	

[4] An Approved Protocol is one which has been approved by Cardiff Met to be used under supervision of designated members of staff; a list of approved protocols can be found on the Cardiff Met website here

preferably at least a year or more. It is necessary to use exactly them as they have the knowledge with most experience in the field. Data will be taken from approximately 30-40 programmers who will be accessed online or approached directly. All data gathered will be confidential and kept in a password protected computer system or if it is hand written then it will be kept in a locked cabinet. All participants will remain anonymous and any data provided will not be traceable back to specific people. Analysing the result – using systematic content analysis.

A4 Will the project involve deceptive or covert research?	No
A5 If yes, give a rationale for the use of deceptive or covert research	
N/A	
A6 Will the project have security sensitive implications?	No
A7 If yes, please explain what they are and the measures that are proposed to address them	
N/A	

B PREVIOUS EXPERIENCE
B1 What previous experience of research involving human participants relevant to this project do you have?
Previously conducted project in a work environment going through steps where meetings and group discussions were held to obtain relevant information necessary in order to gain some knowledge and data about a particular project for software artefact creation.
B2 **Student project only** What previous experience of research involving human participants relevant to this project does your supervisor have?
Mohamed has 4 years of previous human participation experiences as part of his PhD research and previously supervised undergraduate's projects to success.
C POTENTIAL RISKS
C1 What potential risks do you foresee?
• Approach with the questionnaire may be thought as intrusive where the person might think he needs to make himself available for the project thus speeding up the questionnaire completion resulting in the data not being clear. • The most desirable information may not be obtained therefore only possible data could be available. • Enough respondents to participate • Clarity of data
C2 How will you deal with the potential risks?
• Obtaining information about respondent available time and also willingness to participate. • Utilising possible data to make outcome possible. • Utilising other survey distributing systems. • Making Survey questions as easy to understand as possible.

When submitting your application you **MUST** attach a copy of the following:

- All information sheets
- Consent/assent form(s)

An exemplar information sheet and participant consent form are available from the Research section of the Cardiff Met website.

Online questionnaire has been built below are the screenshots

Welcome to Investigation of relationship between personality and programming characteristic survey

Survey Information Sheet

Participant Information Sheet

Investigate relationship between personality and programming characteristics through favourite programming language

My name is Egidijus Jankauskas and I am a final year student conducting this research at Cardiff Metropolitan University, Cardiff, United Kingdom.

What is the study about?

The purpose of this study is to try and understand is there any correlation between individual personality and chosen programming language.

Why have I been approached?

You have been approached because the study requires information from people who are working in this field or are familiar with the programming and used it for varies reasons.

Do I have to take part?

No. It's completely up to you to decide whether or not you take part, but it would be very helpful for me to progress with my research and to maintain as much information as possible for this project to be successful.

What will I be asked to do if I take part?

If you decide you would like to take part, you would be asked to complete the survey.

Will my data be identifiable?

The information you provide is confidential. The data collected for this study will be stored securely and only the researcher conducting this study will have access to this data:
The files on the computer will be encrypted (that is no-one other than the researcher will be able to access them) and the computer itself password protected. It will be destroyed and/or deleted once the project has been submitted for publication/examined.
The written version of your questionnaire will be made anonymous by removing any identifying information including your name.
All your personal data will be confidential and will be kept separately from your questionnaire responses.

What will happen to the results?

The results will be summarised and reported and may be submitted for publication in an academic or professional journal, also if necessary the results may be submitted for further research.

Are there any risks?

There are no risks anticipated with participating in this study. However, if you experience any distress following participation you are encouraged to inform the researcher and contact the resources provided at the end of this sheet.

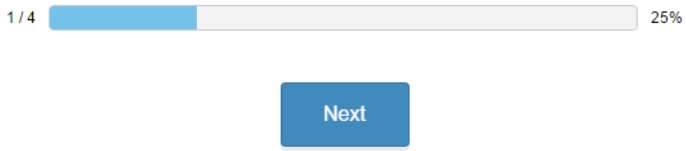
Are there any benefits to taking part?

Although you may find participating interesting, there are no direct benefits in taking part.

Where can I obtain further information about the study if I need it?

If you have any questions about the study, please contact the main researcher:
[Egidijus Jankauskas, Cardiff Metropolitan University student. Email:st20092233@outlook.cardiffmet.ac.uk]
[Mohamed Mostafa, Cardiff Metropolitan University staff. Email: momostafa@cardiffmet.ac.uk]

Thank you for taking the time to read this information sheet.



Welcome to Investigation of relationship between personality and programming characteristic survey

Survey Information Sheet

The purpose of this research project is establish if there is correlation between programmers personality and programming language that he/she has chosen. This is a research project being conducted by Egidijus Jankauskas, final year student at Cardiff Metropolitan University. You are invited to participate in this research project because you are a programmer or have been involved with programming for at least a year. The survey questions will be about personality Big Five and programming languages.

Your participation in this research study is voluntary. You may choose not to participate. If you decide to participate in this research survey, you may withdraw at any time. If you decide not to participate in this study or if you withdraw from participating at any time, you will not be penalized.

Your responses will be confidential and we do not collect identifying information such as your name, email address or IP address. We will do our best to keep your information confidential. All data is stored in a password protected electronic format. To help protect your confidentiality, the surveys will not contain information that will personally identify you.

The results of this study will be used for scholarly purposes only and may be shared with Cardiff Metropolitan University representatives.

If you have any questions about the research study, please contact: [Egidijus Jankauskas, Cardiff Metropolitan University student. Email:st20092233@outlook.cardiffmet.ac.uk] or [Mohamed Mostafa, Cardiff Metropolitan University staff. Email: momostafa@cardiffmet.ac.uk]

Clicking on the "NEXT" button below indicates that:

- you have read the above information
- you voluntarily agree to participate
- You have at least one year's experience in programming (study years count as experience).

If you do not wish to participate in this research study, please decline participation by exiting the survey.

* 1. Please tick the box if you agree with terms listed above

◯ Agree

2 / 4 50%

Prev Next

Welcome to Investigation of relationship between personality and programming characteristic survey

Personality Big Five and programming language questions

* 2. Years of experience in programming (study years count as experience)

- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6+

* 3. In the table below, for each statement 1-50 select how much you agree with on the scale 1-5, where 1=disagree, 2=slightly disagree, 3=neutral, 4=slightly agree and 5=agree, in the box to the left of it. I...

	1 = disagree	2 = slightly disagree	3 = neutral	4 = slightly agree	5 = agree
Am the life of the party.	◯	◯	◯	◯	◯
Feel little concern for others.	◯	◯	◯	◯	◯
Am always prepared.	◯	◯	◯	◯	◯
Get stressed out easily.	◯	◯	◯	◯	◯
Have a rich vocabulary.	◯	◯	◯	◯	◯
Don't talk a lot.	◯	◯	◯	◯	◯
Am interested in people.	◯	◯	◯	◯	◯
Leave my belongings around.	◯	◯	◯	◯	◯
Am relaxed most of the time.	◯	◯	◯	◯	◯
Have difficulty understanding abstract ideas.	◯	◯	◯	◯	◯
Feel comfortable around people.	◯	◯	◯	◯	◯
Insult people.	◯	◯	◯	◯	◯
Pay attention to details.	◯	◯	◯	◯	◯
Worry about things.	◯	◯	◯	◯	◯

Have a vivid imagination.	○	○	○	○	○
Keep in the background.	○	○	○	○	○
Sympathize with others' feelings.	○	○	○	○	○
Make a mess of things.	○	○	○	○	○
Seldom feel blue.	○	○	○	○	○
Am not interested in abstract ideas.	○	○	○	○	○
Start conversations.	○	○	○	○	○
Am not interested in other people's problems.	○	○	○	○	○
Get chores done right away.	○	○	○	○	○
Am easily disturbed.	○	○	○	○	○
Have excellent ideas.	○	○	○	○	○
Have little to say.	○	○	○	○	○
Have a soft heart.	○	○	○	○	○
Often forget to put things back in their proper place.	○	○	○	○	○
Get upset easily.	○	○	○	○	○
Do not have a good imagination.	○	○	○	○	○
Talk to a lot of different people at parties.	○	○	○	○	○
Am not really interested in others.	○	○	○	○	○
Like order.	○	○	○	○	○
Change my mood a lot.	○	○	○	○	○
Am quick to understand things.	○	○	○	○	○
Don't like to draw attention to myself.	○	○	○	○	○
Take time out for others.	○	○	○	○	○

Shirk my duties.	○	○	○	○	○
Have frequent mood swings.	○	○	○	○	○
Use difficult words.	○	○	○	○	○
Don't mind being the center of attention.	○	○	○	○	○
Feel others' emotions.	○	○	○	○	○
Follow a schedule.	○	○	○	○	○
Get irritated easily.	○	○	○	○	○
Spend time reflecting on things.	○	○	○	○	○
Am quiet around strangers.	○	○	○	○	○
Make people feel at ease.	○	○	○	○	○
Am exacting in my work.	○	○	○	○	○
Often feel blue.	○	○	○	○	○
Am full of ideas.	○	○	○	○	○

* 4. In the table below, for each name 1-6 select **how much do you like certain programming language** on the scale 0-5, where 0= never used it, 1= do not like it, 2=slightly dislike it, 3=like it, 4=like it a lot and 5=my favorite

	0 = never used it	1 = do not like it	2 = slightly dislike it	3 = like it	4 = like it a lot	5 = my favourite
JavaScript	○	○	○	○	○	○
Java	○	○	○	○	○	○
PHP	○	○	○	○	○	○
Python	○	○	○	○	○	○
C#	○	○	○	○	○	○
Visual Basic	○	○	○	○	○	○

3 / 4 75%

Prev Next

Welcome to Investigation of relationship between personality and programming characteristic survey

Big Thank You

Thank you for participating in my survey. Your feedback is important.
If you have any questions about the research study, please contact:

[Egidijus Jankauskas, Cardiff Metropolitan University student. Email:st20092233@outlook.cardiffmet.ac.uk] or

[Mohamed Mostafa, Cardiff Metropolitan University staff. Email: momostafa@cardiffmet.ac.uk]

4 / 4 100%

Prev Done

7.2 Appendix B

Programming language C# original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.819	1.220
	Agreeableness	.922	1.085
	Conscientiousness	.732	1.366
	Neuroticism	.847	1.181
	Openness to Experience	.914	1.094

a. Dependent Variable: C#

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.059	.066	.796	1	.372	1.061
	Agreeableness	.031	.073	.178	1	.673	1.031
	Conscientiousness	.074	.089	.697	1	.404	1.077
	Neuroticism	-.119	.100	1.418	1	.234	.888
	Openness	.028	.097	.081	1	.777	1.028
	Constant	-8.913	9.240	.930	1	.335	.000

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.100	.053	3.517	1	.061	1.105
	Agreeableness	.029	.035	.674	1	.412	1.029
	Conscientiousness	.032	.035	.865	1	.352	1.033
	Neuroticism	-.068	.048	2.004	1	.157	.934
	Openness	-.024	.048	.241	1	.623	.977
	Constant	-5.210	4.464	1.362	1	.243	.005

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.017	.026	.445	1	.505	1.017
	Agreeableness	.011	.024	.214	1	.643	1.011
	Conscientiousness	-.041	.026	2.392	1	.122	.960
	Neuroticism	.006	.026	.057	1	.812	1.006
	Openness	.015	.033	.202	1	.653	1.015
	Constant	-.181	3.068	.003	1	.953	.834

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	85.809			
Final	80.015	5.794	5	.327

Link function: Logit.

Pseudo R-Square

Cox and Snell	.149
Nagelkerke	.164
McFadden	.068

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[C#=1]	-4.092	2.9022	-9.780	1.597	1.987	1	.159	.017	5.658E-5	4.937
	[C#=2]	-2.357	2.8150	-7.875	3.160	.701	1	.402	.095	.000	23.568
	[C#=3]	-.560	2.7866	-6.022	4.901	.040	1	.841	.571	.002	134.456
Extroversion		-.028	.0224	-.072	.016	1.554	1	.213	.972	.931	1.016
Agreeableness		-.010	.0211	-.052	.031	.240	1	.624	.990	.950	1.032
Conscientiousness		.021	.0226	-.024	.065	.843	1	.359	1.021	.977	1.067
Neuroticism		.012	.0225	-.032	.056	.302	1	.582	1.012	.969	1.058
Openness		-.012	.0294	-.069	.046	.163	1	.687	.988	.933	1.047
(Scale)		1[a]									

Dependent Variable: C#
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

Programming language Java original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.903	1.108
	Agreeableness	.906	1.103
	Conscientiousness	.779	1.283
	Neuroticism	.819	1.221
	Openness to Experience	.904	1.107

a. Dependent Variable: Java

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.054	.056	.926	1	.336	1.055
	Agreeableness	-.049	.041	1.441	1	.230	.952
	Conscientiousness	.024	.039	.363	1	.547	1.024
	Neuroticism	.015	.057	.068	1	.794	1.015
	Openness	-.170	.091	3.525	1	.060	.844
	Constant	6.732	6.496	1.074	1	.300	839.194

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.011	.024	.208	1	.648	1.011
	Agreeableness	-.022	.023	.893	1	.345	.978
	Conscientiousness	.011	.026	.183	1	.669	1.011
	Neuroticism	.025	.027	.889	1	.346	1.026
	Openness	-.059	.037	2.575	1	.109	.943
	Constant	1.861	3.464	.288	1	.591	6.427

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.003	.019	.024	1	.877	.997
	Agreeableness	-.023	.022	1.105	1	.293	.978
	Conscientiousness	-.039	.026	2.174	1	.140	.962
	Neuroticism	.015	.022	.454	1	.500	1.015
	Openness	-.039	.029	1.818	1	.178	.962
	Constant	6.092	3.144	3.755	1	.053	442.201

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	113.254			
Final	104.361	8.893	5	.113

Link function: Logit.

Pseudo R-Square

Cox and Snell	.176
Nagelkerke	.192
McFadden	.079

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[Java=1.00]	3.205	2.7201	-2.127	8.536	1.388	1	.239	24.644	.119	5094.503
	[Java=2.00]	4.392	2.7369	-.973	9.756	2.575	1	.109	80.770	.378	17253.708
	[Java=3.00]	5.666	2.7903	.197	11.135	4.124	1	.042	288.920	1.218	68524.732
Extroversion		-.006	.0176	-.040	.029	.111	1	.739	.994	.960	1.029
Agreeableness		.025	.0192	-.013	.063	1.709	1	.191	1.025	.988	1.065
Conscientiousness		.014	.0214	-.028	.056	.413	1	.520	1.014	.972	1.057
Neuroticism		-.011	.0199	-.050	.028	.311	1	.577	.989	.951	1.028
Openness		.055	.0272	.002	.108	4.079	1	.043	1.057	1.002	1.114
(Scale)		1[a]									

Dependent Variable: Java
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

Programming language JavaScript original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.794	1.260
	Agreeableness	.880	1.136
	Conscientiousness	.717	1.395
	Neuroticism	.757	1.322
	Openness to Experience	.920	1.087

a. Dependent Variable: JavaScript

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.076	.059	1.616	1	.204	.927
	Agreeableness	-.045	.042	1.174	1	.279	.956
	Conscientiousness	.011	.046	.054	1	.817	1.011
	Neuroticism	.122	.077	2.508	1	.113	1.130
	Openness	-.182	.124	2.176	1	.140	.833
	Constant	8.189	7.797	1.103	1	.294	3600.656

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.036	.024	2.309	1	.129	.965
	Agreeableness	.005	.026	.034	1	.853	1.005
	Conscientiousness	-.060	.031	3.720	1	.054	.942
	Neuroticism	.056	.028	4.123	1	.042	1.057
	Openness	-.029	.032	.792	1	.374	.972
	Constant	3.610	3.333	1.173	1	.279	36.960

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.009	.028	.107	1	.744	.991
	Agreeableness	-.098	.041	5.686	1	.017	.906
	Conscientiousness	.038	.040	.920	1	.338	1.039
	Neuroticism	-.065	.039	2.693	1	.101	.937
	Openness	-.066	.044	2.199	1	.138	.936
	Constant	15.158	5.562	7.427	1	.006	3828727.06

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	109.132			
Final	100.015	9.116	5	.105

Link function: Logit.

Pseudo R-Square

Cox and Snell	.191
Nagelkerke	.207
McFadden	.084

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[JavaScript=1]	3.378	2.6744	-1.863	8.620	1.596	1	.207	29.324	.155	5541.878
	[JavaScript=2]	5.335	2.7083	.027	10.643	3.881	1	.049	207.505	1.027	41913.037
	[JavaScript=3]	7.489	2.8472	1.909	13.070	6.919	1	.009	1788.799	6.745	474394.971
Extroversion		.019	.0178	-.016	.054	1.120	1	.290	1.019	.984	1.055
Agreeableness		.030	.0202	-.009	.070	2.244	1	.134	1.031	.991	1.072
Conscientiousness		.024	.0222	-.019	.068	1.188	1	.276	1.025	.981	1.070
Neuroticism		-.021	.0205	-.061	.019	1.039	1	.308	.979	.941	1.019
Openness		.035	.0255	-.015	.085	1.862	1	.172	1.035	.985	1.088
(Scale)		1[a]									

Dependent Variable: JavaScript
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

Programming language PHP original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.748	1.336
	Agreeableness	.689	1.451
	Conscientiousness	.714	1.400
	Neuroticism	.668	1.497
	Openness to Experience	.815	1.228

a. Dependent Variable: PHP

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.075	.047	2.588	1	.108	.927
	Agreeableness	-.025	.048	.270	1	.604	.975
	Conscientiousness	.013	.042	.096	1	.757	1.013
	Neuroticism	.101	.069	2.107	1	.147	1.106
	Openness	-.146	.106	1.896	1	.169	.864
	Constant	6.649	5.275	1.589	1	.208	772.058

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.027	.029	.857	1	.355	.974
	Agreeableness	-.012	.034	.132	1	.717	.988
	Conscientiousness	-.023	.033	.496	1	.481	.977
	Neuroticism	-.003	.030	.009	1	.923	.997
	Openness	.007	.034	.041	1	.840	1.007
	Constant	2.730	3.318	.677	1	.411	15.332

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.003	.038	.005	1	.946	.997
	Agreeableness	-.049	.052	.871	1	.351	.952
	Conscientiousness	.121	.088	1.890	1	.169	1.129
	Neuroticism	-.089	.070	1.591	1	.207	.915
	Openness	-.155	.097	2.533	1	.111	.857
	Constant	14.906	10.078	2.187	1	.139	2976182.29

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	67.981			
Final	64.122	3.859	5	.570

Link function: Logit.

Pseudo R-Square

Cox and Snell	.138
Nagelkerke	.149
McFadden	.057

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[PHP=1]	3.069	2.9310	-2.676	8.813	1.096	1	.295	21.515	.069	6723.647
	[PHP=2]	3.867	2.9645	-1.943	9.678	1.702	1	.192	47.805	.143	15954.562
	[PHP=3]	5.898	3.0852	-.149	11.945	3.655	1	.056	364.421	.862	154072.003
Extroversion		.022	.0245	-.026	.070	.776	1	.378	1.022	.974	1.072
Agreeableness		.017	.0295	-.040	.075	.347	1	.556	1.018	.960	1.078
Conscientiousness		-.006	.0271	-.059	.047	.045	1	.831	.994	.943	1.048
Neuroticism		.000	.0268	-.052	.053	.000	1	.993	1.000	.949	1.054
Openness		.033	.0312	-.028	.095	1.147	1	.284	1.034	.973	1.099
(Scale)		1[a]									

Dependent Variable: PHP
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

Programming language Python original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.836	1.196
	Agreeableness	.829	1.206
	Conscientiousness	.799	1.252
	Neuroticism	.875	1.143
	Openness to Experience	.863	1.159

a. Dependent Variable: Python

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.026	.040	.436	1	.509	.974
	Agreeableness	-.129	.096	1.786	1	.181	.879
	Conscientiousness	.096	.078	1.532	1	.216	1.101
	Neuroticism	-.117	.069	2.845	1	.092	.890
	Openness	.029	.050	.333	1	.564	1.029
	Constant	4.944	5.516	.804	1	.370	140.378

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.006	.026	.060	1	.806	.994
	Agreeableness	-.064	.044	2.115	1	.146	.938
	Conscientiousness	.079	.044	3.194	1	.074	1.082
	Neuroticism	-.090	.038	5.572	1	.018	.914
	Openness	-.018	.035	.271	1	.602	.982
	Constant	4.569	3.720	1.508	1	.219	96.482

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.001	.020	.002	1	.965	1.001
	Agreeableness	.023	.024	.897	1	.343	1.023
	Conscientiousness	.008	.028	.089	1	.765	1.008
	Neuroticism	-.007	.023	.095	1	.758	.993
	Openness	-.025	.030	.689	1	.407	.975
	Constant	.953	3.037	.098	1	.754	2.592

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	95.185			
Final	88.088	7.097	5	.214

Link function: Logit.

Pseudo R-Square

Cox and Snell	.170
Nagelkerke	.186
McFadden	.075

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[Python=1.00]	1.761	2.5444	-3.226	6.748	.479	1	.489	5.816	.040	852.018
	[Python=2.00]	2.869	2.5716	-2.171	7.909	1.245	1	.265	17.623	.114	2722.678
	[Python=3.00]	5.061	2.6684	-.169	10.291	3.598	1	.058	157.792	.845	29475.671
Extroversion		.013	.0173	-.020	.047	.602	1	.438	1.013	.980	1.048
Agreeableness		.012	.0205	-.028	.052	.338	1	.561	1.012	.972	1.053
Conscientiousness		-.031	.0239	-.078	.016	1.663	1	.197	.970	.925	1.016
Neuroticism		.046	.0212	.004	.087	4.620	1	.032	1.047	1.004	1.091
Openness		.026	.0258	-.024	.077	1.031	1	.310	1.027	.976	1.080
(Scale)		1[a]									

Dependent Variable: Python
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

Programming language VisualBasic original data screenshots

Multicollinearity

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Extroversion	.849	1.178
	Agreeableness	.891	1.122
	Conscientiousness	.800	1.250
	Neuroticism	.862	1.160
	Openness to Experience	.872	1.147

a. Dependent Variable: VisualBasic

Proportional odds through binomial logistic regression

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.025	.032	.615	1	.433	1.026
	Agreeableness	.039	.044	.778	1	.378	1.039
	Conscientiousness	-.044	.044	1.006	1	.316	.957
	Neuroticism	-.092	.052	3.044	1	.081	.912
	Openness	.015	.051	.085	1	.771	1.015
	Constant	-.335	4.741	.005	1	.944	.716

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	.020	.032	.401	1	.526	1.020
	Agreeableness	-.014	.036	.155	1	.694	.986
	Conscientiousness	-.049	.040	1.539	1	.215	.952
	Neuroticism	-.127	.053	5.650	1	.017	.881
	Openness	.021	.049	.177	1	.674	1.021
	Constant	5.961	4.388	1.846	1	.174	388.163

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Extroversion	-.014	.018	.557	1	.456	.986
	Agreeableness	.006	.020	.079	1	.779	1.006
	Conscientiousness	-.012	.023	.259	1	.611	.988
	Neuroticism	-.027	.021	1.600	1	.206	.974
	Openness	.007	.027	.061	1	.805	1.007
	Constant	1.889	2.557	.545	1	.460	6.610

a. Variable(s) entered on step 1: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness.

Ordinal logistic regression output tables

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	108.844			
Final	100.225	8.619	5	.125

Link function: Logit.

Pseudo R-Square

Cox and Snell	.168
Nagelkerke	.186
McFadden	.079

Link function: Logit.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Threshold	[VisualBasic=1]	.279	2.4189	-4.462	5.020	.013	1	.908	1.322	.012	151.469
	[VisualBasic=2]	.833	2.4159	-3.902	5.568	.119	1	.730	2.300	.020	261.949
	[VisualBasic=3]	2.318	2.4337	-2.452	7.088	.907	1	.341	10.159	.086	1197.764
Extroversion		.005	.0171	-.029	.038	.081	1	.776	1.005	.972	1.039
Agreeableness		.000	.0196	-.039	.038	.000	1	.992	1.000	.962	1.039
Conscientiousness		.022	.0223	-.021	.066	1.009	1	.315	1.023	.979	1.068
Neuroticism		.045	.0205	.004	.085	4.715	1	.030	1.046	1.004	1.088
Openness		-.024	.0260	-.075	.027	.868	1	.352	.976	.928	1.027
(Scale)		1[a]									

Dependent Variable: VisualBasic
Model: (Threshold), Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness

a. Fixed at the displayed value.

7.3 Appendix C

Division of categorical dependent variable into dichotomous categories