

## How Predictable Ratings are: The Role of Personality Traits

Rajab Esfandiari<sup>1\*</sup>

<sup>1\*</sup> Assistant Professor of Applied Linguistics, Department of English Language, Imam Khomeini International University, Qazvin, Iran, [esfandiari@hum.ikiu.ac.ir](mailto:esfandiari@hum.ikiu.ac.ir)

### Abstract

Previous research has shown that raters' personality characteristics exert an influence on their ratings, skewing the results and invalidating the decisions made about students' future life. Although the exact mechanism of these factors, their precise effect on ratings, and the interaction between the traits and ratings are yet to be empirically demonstrated, anecdotal evidence coupled with research findings suggests raters' ratings do not necessarily reflect students' abilities and may be affected by other construct-irrelevant variances, including personality traits. The purpose of the present study, therefore, was to examine the extent to which some selected personality traits would predict the ratings awarded by the raters to students' written performance. To that end, teacher raters rated students' essays on 13 assessment criteria using a 5-point analytic rating scale. Big five inventory and student essays were used to measure teacher raters' personality traits and to collect data. Data were computed and analysed using SPSS (version 25). Results from linear regression showed that extroversion, agreeableness, openness, neuroticism, and conscientiousness did not significantly contribute to the ratings. The findings suggest that such personality factors may not account for the ratings, and rater variability should be explained in terms of other personality variables. The implications of the study are discussed.

**Keywords:** Linear Regression, Personality Traits, Ratings, Rating Scale

---

Received 08 June 2019

Accepted 08 August 2019

Available 05 September 2019

DOI: 10.30479/jmrels.2019.11338.1414

---

## 1. Introduction

Individual differences (IDs) play a key role in second, or foreign, language learning. Findings from previous research have pointed to the significance of these factors as correlates of L2 learning success (Dörnyei, 2005). Similarly, Ellis (2008) observed that the study of individual learner differences would contribute to theory development in second language acquisition (SLA) better accounting for the diverse set of learner factors affecting the ultimate success in L2. Dörnyei and Ryan (2015) noted that ID research had moved in diverse directions and that the “seemingly straightforward conceptualization” of IDs as “enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree” (p. 3) would not be necessarily tenable. The intervening years since the publication of this landmark book have witnessed a number of studies unpacking the complexities of IDs and their relation to language learning, a testimony to the importance of such variables in language learning.

One of the most important IDs in SLA research is personality traits. As Dörnyei (2005) remarked, “personality is the most individual characteristic of a human being” (p. 10), which “constitutes a major factor contributing to success or failure in language learning” (Ellis, 2008, p. 672). Personality traits may refer to those characteristics that may differ from person to person but may remain stable on different occasions. In Robinson, Lopez, and Ramos’ words (2014), personality traits are “consistencies in how a person typically behaves across different contexts of life” (p. 180). Such traits, as Hockenbury and Hockenbury (2011) neatly summarised, are viewed as “relatively stable, enduring predisposition[s] to consistently behave in a certain way” (p. 439).

Over the years, researchers have tried to present different taxonomies of personality traits. One of the most promising taxonomies ever proposed is that of Big Five Model (BFM), which was originally coined by Goldeberg (1981), but in recent years it has been identified with the work of McCrae and Costa (Dörnyei & Ryan, 2015). Such a model has gained traction and become ubiquitous in personality studies and has made consistent inroads in SLA in recent years. The BFM, as the name suggests, consists of five broad dimensions: extroversion-introversion, openness, agreeableness, neuroticism, and conscientiousness (John, Naumann, & Soto, 2008).

Previous research has shown that personality traits may be related to academic achievement, communicative language use, and learning measures. However, as Ellis (2008) concluded, “the research overall has been somewhat disappointing” (p. 676). Dörnyei and Ryan (2015) arrived at similar conclusions, commenting that the results were somewhat varied and inconclusive and the situation did not change from 2005 to 2015.

Much of the research in SLA on personality traits has focused on the relationship between these factors and success in L2 learning. How such dispositions may affect the ratings raters award to students' written performances is yet to be empirically examined. There are good reasons to investigate personality traits and ratings. The first reason has to do with the quality of ratings. In Schaefer's (2008) words, "it is desirable that raters rate consistently and objectively" (p. 465). Therefore, the accuracy of ratings is particularly important in assessment, as a result of which fair decisions can be made about the success or failure of students. Moreover, focusing on personality traits, raters can be aware of their own personality types, so they will become cautious about the sources of bias.

The present study aimed to investigate how raters' personality traits would predict their ratings. As such, the present study sought to answer the following research question: To what extent do raters' personality traits (extroversion-introversion, openness, agreeableness, neuroticism, and conscientiousness) predict their ratings?

## **2. Literature Review**

In the following two subsections, a selective review of some concepts relevant to the present study is presented. First, the notions of personality, traits, and personality traits are briefly explained. Next, BFM and the five dimensions related to this personality theory are outlined. Finally, some sources of rater variability are presented, and the results of some studies done on the link between personality variables and rater variability are summarised.

### **2.1. Personality Traits in SLA**

Trait theory in psychology plays a pivotal role in personality psychology. To better understand what personality traits refer to, we need to first explicate the notions of personality and traits. Pervin and John (2001) defined personality as those characteristics that "account for consistent patterns of feeling, thinking, and behaving" (p. 4). Such a standard, though broad definition, primarily lays emphasis on the behavior that people manifest consistently in different situations, although a person's consistent behavior may differ from that of another person. Traits are also "largely stable over time, and it is difficult to detect patterns of systematic change across populations of individuals. The usual overall pattern is that traits become more stable with age" (Matthews, Deary, & Whiteman, 2009, p. 83). Building on the features of personality and trait, we can now define personality traits as those attributes of people that seem stable across situations and over time.

Researchers have proposed so many theories to account for the consistent characteristics of people. BFM, as outlined in the Introduction, is one of the competing trait theories, which have proved influential in accounting for traits. BFM focuses on the highest hierarchical level of trait description and tries to elaborate on the total character of the person at a very broad level (John & Srivastava, 1999; McCrae & John, 1992). As John (1990) argued, through analysing self-report and ratings, natural languages, children or adults' personality, and men and women's samples, five factors were regarded as the essential aspects of personality. BFM encapsulates different personality traits into five major domains, namely: extroversion-introversion, openness, agreeableness, conscientiousness, and neuroticism.

BFM is composed mainly of two elements: basic tendencies and characteristic adaptations (McCrae, 2011). As McCrae and Lockenhoff (2010) argued, basic tendencies are heritable predispositions derived from biological bases. Unlike basic tendencies, characteristic adaptations are not innate (McCrae, 2011). The interaction between basic tendencies and external influences creates characteristic adaptations which result in individuals' behaviors and emotions (McCrae, 2011; McCrae & Lockenhoff, 2010). In the following, paragraphs, more explanation is given about those five traits BFM consists of. The discussion begins with extroversion-introversion and ends with neuroticism.

As Patrick (2011) neatly put it, "a combination of interpersonal interaction skills, positive affect and energy level make up the domain of extraversion" (p. 242). For the first time, Jung (1971) introduced extroversion-introversion as features demonstrating different attitudes toward the world (Watson & Clark, 1997). More specifically, traits such as venturesome, affiliation, positive affectivity, energy, ascendance, and ambition were regarded as the six components of extroversion (Watson & Clark, 1997).

Various facets specify extroversion. Tendency toward gregariousness, warmth, assertiveness, and excitement seeking behaviors displays extroversion (Costa & McCrae, 1992). The positive effect of being an extrovert on environmental behaviors and environmental engagement has been revealed in many studies (e.g., Milfond & Sibley, 2012). Ross, Orr, Susic, Arseneaul, Simmering, and Orr (2009) found that people high on extroversion joined online groups more frequently. Predictably, extrovert people have more online friends; moreover, extroversion can predict the number of Facebook friends (Chen, 2014). As Mondak, Hibbing, Seligson, and Anderson (2010) found, extrovert people tend to participate in political activity. Therefore, extroverts join political systems and stay loyal to them (Bakker, Klemmenson, Norgaard, Schuamacher, 2015).

The performance of extroverts and introverts are different in terms of tasks and contextual variables (Matthews, Deary & Whiteman, 2009). In a study, Eysenk (1997) explained that extroverts were more inclined to respond sooner in comparison with introverts who were slower to answer. While working on challenging and memory tasks, extroverts tended to produce a more superb performance than introverts did (Furnham & Strbac, 2002). Unlike introverts, extroverts are good at verbal communication and are not easily distracted (LePine & Van Dyne, 2001). On the other hand, introverts are good at long-term memory and reflective tasks such as problem solving (Matthews, Deary, & Whiteman, 2009).

Agreeableness as the second component of BFM is “a tendency to be compassionate and cooperative rather than suspicious and antagonist toward others; reflects individual differences in general concern for social harmony” (Stangor, 2010, p. 602). According to this definition, agreeable people have the propensity to be friendly, helpful, and considerate. Interestingly, agreeableness is particularly an appealing personality attribute among people since it is related to pro-social behaviors (Courbalay, Deroche, Prigent, Chalabaev & Amorim, 2015).

Agreeableness offers some unique features. Agreeable people strive for being trusting, altruistic, and tender-minded (Swickert, 2009). Moreover, agreeable people value social or interpersonal relationships and positive interpersonal correspondence (Gleason, Jensen-Campbell, & Richardson, 2004; Yao & Moskowitz, 2015). In the same vein, Gleason, Jensen-Campbell, and Richardson (2004) posited that agreeableness could be the predictor of individuals’ aggression. In other words, agreeable people tend not to show aggressive behaviors. As a result, people high on agreeableness are soft-hearted whereas people who score low on agreeableness are somehow ruthless (Feist & Feist, 2008).

Openness (also known as openness to experience) is widely regarded as one of the basic aspects of personality, emphasising art, emotion, adventure, imagination, and curiosity (McCrae, 1994; Stangor, 2010). Following the literature in grasping the concepts of openness to experience, McCrae and Costa (1997) explained the concept of openness to experience in terms of three aspects: traditional, structural, and motivational. Substantially, culture comprised of education and family influences was the first feature of traditional conception of openness to experience. (McCrae & Costa, 1997). According to McCrae and Costa (1997), people high on openness are capable of obtaining intellectual information very well.

Conceptually, McCrae and Costa (1997) explained that openness as psychic means “a matter of inner experience, a mental phenomenon related to

the scope of awareness or the depth and intensity of consciousness” (p. 835). Regarding motivational concepts, openness is regarded as a need for experience (McCrae & Costa, 1997). Open people are able to explore new and vast experiences (McCrae & Costa, 1997); therefore, openness includes motivation, art, and cognition (Maddi & Berne, 1964; Osberg, 1987). Costa and McCrae (1992) reiterated that openness is a broad personality trait. Accordingly, they provided six sub-dimensions for openness to experience: fantasy, aesthetics, feelings, actions, ideas, and values.

Openness to experience differentiates people who tend to be spontaneous, imaginative, and adventurous and people who prefer to be cautious and conservative. High scorers mostly seek divergent thinking and new experiences; they are also curious about traditional values and attempt to challenge them (Feist & Feist, 2008; Silvia, Nusbaum, Berg, Martin, & O’Connor, 2009). In sharp contrast, low scorers of openness to experience prefer familiar things rather than new and are conventional (Feist & Feist, 2008). Moreover, people low on openness to experience tend to have fixed or routine lives (Feist & Feist, 2008). Openness to experience is the predictor of creative accomplishments in art (Kaufman, Quilty, Grazioplene, Hirsh, Gray, Peterson, & DeYoung, 2015). When scrutinised in detail, Mussel, Winter, Gelléri, and Schuler (2011) found that openness to experience and its sub-dimensions corresponded to job performance and work-related criteria.

Among different components of BFM, conscientiousness has been seen as a category encompassing conformity and socially impulse control. More specifically, adjectives such as practical, organised, thorough, and responsible conceptualise conscientiousness (Kruglanski, Orehek, Higgins, Pierro, & Shalev, 2010). DeYoung and Gray (2009) explained that “conscientiousness appears to reflect the tendency to maintain motivational stability within the individual, to make plans, and carry them out in an organized and industrious manner” (p. 335). Over time, different facets of conscientiousness have been developed, but Patrick (2011) identified competence, order, dutifulness, achievement striving, self-discipline, and deliberation as the core components of conscientiousness.

People high on conscientiousness display underlying tendencies. As Stangor (2010) posited, “individuals who are conscientious have a preference for planned rather than spontaneous behavior” (p. 602). As to the relationship between personality and job performance, Salgado, Moscoso, and Berges (2013) found that conscientiousness could be a predictor of three job performance criteria including overall job performance, task performance, and orderliness. Highly conscientious individuals care about providing their personal resources which lead them to have different reactions to work pressure (Grawitch, Barber, & Justice, 2010). Moreover, Lin, Ma, Wang, and Wang (2015) concluded that people who score in the direction of

conscientiousness tended to keep their job performance standards. However, most crucially, conscientious people in their study suffered from job stress brought about in tense situations (Lin, Ma, Wang, & Wang, 2015).

Neuroticism is conceptualised as a negative trait (Feist & Feist, 2008). Moreover, Neuroticism is a broad personality trait which is related to super-ordinate dissociative behaviors (McCrae & Costa, 1984; Triandis, 1997). More specifically, DeYoung and Gray (2009) noted that “neuroticism appears to reflect sensitivity to threat and the whole range of negative emotions and cognitions that accompany experiences of threat and punishment, including anxiety, depression, anger, irritation, self-consciousness and vulnerability” (p. 332). Researchers regard neuroticism as a personality trait positively related to emotional lability, impulsiveness, and self-consciousness (Swichert, 2009). Different facets of neuroticism are anxiety, hostility, depression, self-consciousness, vulnerability, and impulsiveness (Wiggins & Trapnell, 1997).

People standing on neuroticism are characterised with some descriptors. Neurotic people are unstable, moody, tense, anxious, nervous, and self-punishing (Kruglanski, et al., 2010). Being pre-occupied with negative thoughts, neurotic individuals avoid taking a risk of setting goals with positive outcomes (Elliot & Thrash, 2002). Instead, they reflect on their own tendencies toward having avoidance goals which are identified as less challenging and not clearly defined (Elliot & Thrash, 2002; McCrae & Lockenhoff, 2010). As Szymura and Nęcka (2005) explained, those with high scores on neuroticism respond to emotional stimuli quickly and nervously. Moreover, seeing the negative side of different events causes neurotic people to mostly put the blame on others rather than themselves (Bollmer, Harris, & Milich, 2006).

## **2.2. Rater Variability in Rater-Mediated Assessments**

Since rater variability has been related to unwanted score variability, it has raised the further issue of focusing on extraneous factors which affect the assessment of writing (Huang, 2008; Yan, 2014). As Barkaoui (2011), Eckes (2012), Weigle (1999), and Weigle (2002) reported, factors such as rater background, rater mother tongue, training, rater cognition, rating scales, and rater experience may affect raters' rating. In the subsequent section, some of these factors are explained.

Rating scales seem to be at play among different factors, which affect raters' ratings. As Myford and Wolfe (2003) explained, “a rating scale is a measurement instrument used to record the results of the rater's observations” (p. 388). Sometimes, instructions regarding what raters are supposed to do may not be clear enough or the definitions of scale categories

may be ambiguous. Therefore, such problems may adversely affect raters' consistency of ratings. As Schoonen (2005) emphasised, general impression scores elicited from a holistic scoring may exhibit some positive features, which may seem more important and outstanding compared to weaker features. As a result, weaker features of an examinee might not be paid enough attention to (Schoonen, 2005). On the other hand, as Goulden (1994) believed, an analytic scoring manages the amount of importance or scores raters give to each category.

Focusing on the effect of using analytic rating scale and holistic rating scale, Barkaoui (2007) found that raters gave higher scores to students' writing, using an analytic rating scale. However, Barkaoui explained that impressionistic criteria provided in the analytic rating scale might be the reason behind the higher level of scores the raters gave. Barkaoui (2011) extended his research on marking methods, raters' experience, and raters' severity and found that experienced raters were less severe, using an analytic rating scale. He further explained that in analytic rating scales, unlike holistic rating scales, examinees' weaknesses are not overemphasised to affect the final assessment. However, using an analytic rating scale, novice raters were more lenient giving varying degrees of importance to various scale categories.

Rater backgrounds may affect raters' decision-making behaviors in rating and contribute to raters' construct irrelevant variance as well (Wiseman, 2012). Focusing on rater experience as a facet of rater background and drawing on marking methods, namely holistic and analytic scales, Barkaoui (2011) emphasized that novice raters were more lenient than experienced raters. Moreover, novice raters found argumentation as the most important assessment criterion whereas experienced raters gave higher scores to linguistic accuracy (Barkaoui, 2010, 2011). Building on this line of research, Wiseman (2012) conducted a study to examine raters' judgment on rating narrative and persuasive essays, focusing on rater background. Comparatively, raters having a creative writing background showed a high degree of severity when rating the narrative essays. In the same vein, Sweedler-Brown (1985) explained that experienced raters rated the ESL essays severely compared to inexperienced ones.

Focusing on assertiveness and non-assertiveness as two sub-categories of extroversion, Ockay (2009) attempted to analyse how far assertive and non-assertive test takers' scores of an oral exam were affected by being assertive or non-assertive of their group members. The findings of this study showed that test takers' personality traits in a particular group could affect their scores in the same group. In other words, when grouped with non-assertive test takers, the assertive test takers received higher scores;

however, when assertive test takers were assigned to the group having assertive members, they received lower score.

Bernardin, Tyler, and Villanova (2009) carried out a study to work on rater level and rating accuracy regarding raters' agreeableness and conscientiousness in a peer rating. One hundred and twenty-six students participated in the peer rating sessions to rate each other's written performance. The findings suggested that there was a relationship between raters' personality traits and rating accuracy. Low accuracy and more leniency were seen among agreeable raters. However, high accuracy and low leniency were found among conscientious raters. Moreover, the worst ratings were assigned by more agreeable and less conscientious raters in the study. Some factors used in this study affected the findings. As to management issues, the researchers concluded that less accurate and more lenient ratings might mask differences at operational systems.

In another study, Grahek (2007) explored the relationship between agreeableness and conscientiousness and raters' leniency. To that end, Grahek drew on 226 undergraduate participants. Through using an instructor evaluation form and personality scale, Grahek gathered the data. Moreover, raters' leniency was measured through mean ratings. There was a positive relationship between agreeableness and leniency. The more lenient the raters were, the higher ratings they assigned. On the other hand, a negative relationship was found between conscientiousness and leniency. Conscientious raters were not particularly keen to rate leniently. This may be associated with assertion that the conscientious people to be scrupulous, determined, and purposeful.

### **3. Method**

#### **3.1. Participants**

The raters, in the current study, consisted of 85 teachers. Using convenience sampling, we selected teachers from different language institutes in Qazvin. The teachers ranged in age from 21 to 50. Two of the raters (2.4%) were AA (Associate of Arts) holders, 32 of raters (37.6%) BA holders, 49 (57.6%) MA holders, and two of them (2.4%) PhD holders. Forty-one (48.2%) had degrees in English teaching, 19 (22.4%) had translation degree, six (7.1%) had English literature degree, and 19 (22.4%) studied other fields except English. Participants' teaching experience ranged from one to 20 three years. Moreover, their writing experience ranged from one to 15. Fifty-two of them (61.2%) were female and 33 (38 %) were male. Seventy-nine teachers (92.9%) were native-Persian speakers, five (5.9) were

native-Turkish speakers, and one (1.2 %) was native-Kurdish speaker. Only three of them (3.5%) had experience living in an English speaking country.

Twenty-four university male and female students were chosen as the second group of participants to write one-paragraph essays. We randomly selected the writings of ten students. Students were undergraduate university students majoring in English Translation enrolled in an Advanced Writing class at Imam Khomeini International University in Qazvin. The students ranged in age from 19 to 22. Six students (60%) were male and four (40%) were female. All of them were native-Persian speakers. All of them were second-year students. Only one of them had experience living in an English-speaking country. The number of years they had studied English ranged from two to 17 years. Seventy percent of students had studied the English language in language institutes before entering the university.

### **3.2. Materials and Instruments**

The big five inventory (BFI) by John, Naumann, and Soto (2008) was used to identify raters' personality traits. BFI has 44 items with five categories (extroversion, agreeableness, openness to experience, conscientiousness, and neuroticism) to measure the personality trait of each rater. John, Naumann and Soto (2008) noted that the reliability of BFI ranges from .75 to .90, and the mean convergent validity of BFI is .92. Therefore, high validity and reliability of BFI were the reasons this inventory was used in this study.

One-paragraph essays were another type of instrument used in this study. One-paragraph essays were preferred to five-paragraph ones because one-paragraph essays were easier for the raters to rate, shorter in length, and less time-consuming to rate. One-paragraph essays, as the name suggests, include only a single paragraph and develop a single main idea. Structurally, they begin with a topic sentence, continue with supporting ideas, and end with a reworded topic sentence. Like other pieces of well-structured writing, they must be coherent and unified: They are grammatically and semantically well-formed, and only a central idea is developed through the paragraph. First, students brainstormed some topics which were common and popular among them. Second, we rated the topics based on their frequency and chose the most frequent topics presented by students. Finally, each student wrote one-paragraph essay based on the chosen topic.

A researcher-made analytic rating scale was used to rate the essays. The following steps were taken in order to develop the scale. Choosing the appropriate criteria is the most important goal when facing the task of developing a rating scale. Therefore, to begin with, we met some students and explained the procedures and the purpose of our study. Then we used the students' interpretation of primary criteria necessary for assessing writing.

The students came up with their ideas for expressing the major criteria and we gathered those criteria. Moreover, in order to provide considerable support for the previous step, we asked some teachers to rate an essay and enumerate necessary criteria for writing assessment. Finally, the first draft of the scale was constructed using the above steps and some of the criteria for previously published scales. Therefore, based on students' major criteria, teachers' own interpretation of assessing writing criteria, and the researcher's own list, the analytic rating scale should include topic sentence, rewarded topic sentence, unity, coherence, grammar, vocabulary, content, format, expression, punctuation, mechanics of writing, spelling, and sentence variety.

In addition, the scale was given to ten raters to use the scale and rate ten one-paragraph essays. After piloting the analytic rating scale, the ultimate criteria were chosen: topic sentence, rewarded topic sentence, unity, coherence, grammar, vocabulary, content, format, expression, punctuation, mechanics of writing, spelling, and sentence variety. As some raters had problems with understanding the meaning of the criterion expression, the wording of expression after piloting was changed. Also our scale categories were very poor (1), poor (2), fair (3), good (4), and excellent (5). Moreover, the reliability coefficient of this scale is .94, suggesting very good internal consistency reliability. According to Pallant (2011), values of reliability which are above .7 are acceptable and values above .8 are preferable.

### **3.3. Procedure**

The students of an Advanced Writing class at Imam Khomeini International University were asked to write a one-paragraph essay. Before asking students to write a paragraph, it was important to come up with a specific topic. To that end, each student was asked to write ten interesting topics. Then, topics were categorized and the following eighteen categories were identified: nutrition, computer games, smoking, the Internet, college education, urban or rural areas, using cell phones, wearing school uniforms, co-ed education, pollution, language learning, movies, transportation, types of books, abortion, deforestation, technology, and men and women. After that, students were asked to choose the most interesting topic. Finally, language learning was the most interesting category. Following the structures of writing one-paragraph essays, 24 students wrote a one-paragraph essay about the following topic: What are three reasons people learn a foreign language? Please, support your reasons as specifically as possible? After the essays were collected, ten of the essays were randomly selected and given the raters to rate, using an analytic rating scale. Accordingly, we formulated the following topic for essay writing.

A one-hour training session was held to familiarise teachers with the analytic scale, categories, and accurate rating. First of all, the researcher clarified the purpose of the study to the raters and assured them of complete confidentiality. The researcher handed raters an analytic rating scale which included the scale criteria ranging from one to five and a rubric explaining each criterion in detail. The researcher asked raters to go through the scale and the rubric, read them to see if they fully understood the criteria. Then the researcher gave raters two one-paragraph essays which had already been rated using the analytic rating scale. One sample was a well-written essay and the other was a poorly written essay. These samples were given to the raters to let them understand accurate rating. The raters read the essays first and the ratings and then compared the ratings of the strong and weak essays.

After making sure that the raters found the criteria and rating procedures easy to understand, the researcher gave each rater a sample to rate in the training session. Moreover, the researcher asked raters to be aware of their possible errors while rating and asked them to rate as accurately as possible, paying attention to all categories. Finally, ten essays were given to the raters to rate at home using the analytic rating scale. Within two weeks, raters were supposed to hand in the essays and ratings. Raters were also given the BFI to fill in.

### **3.4. Data Analysis**

SPSS (version 25) was used to analyse the data for the present study. Teachers' ratings on the rating scale and the BFI were computed using the Compute option of SPSS. Standard multiple regression was used to answer the research question. All the assumptions of this statistical test were checked before running the test.

## **4. Results and Discussion**

### **4.1. Results**

Standard multiple regression is a test with many assumptions. Before using the test, its assumptions should be checked. Otherwise, it may not be a suitable technique for the data. In what follows, the assumptions of this test are analysed, and then the research question of the study is examined.

According to Pallant (2011) and the formula he provided, the number of participants should be as follows:  $N > 50 + 8m$  ( $m$  is the number of independent variables). Five independent variables were used in this study. Moreover, 85 raters participated in the study; therefore, the number of participants was enough and the first assumption was met.

As Pallant (2011) noted, tolerance values which are very low (less than .1) may cause problems. As can be seen in table 1, tolerance values for

the variables is greater than .1, and the VIF value is less than 10. This shows that multicollinearity did not exist.

Table 1

*Collinearity Statistics of Personality Traits*

Traits	Tolerance	VIF
Extroversion	.964	1.038
Agreeableness	.830	1.204
Conscientiousness	.828	1.207
Neuroticism	.885	1.130
Openness	.863	1.159

It is important to pay attention to outliers since standard multiple regression is sensitive to them (Pallant, 2011). According to Tabachnick and Fidell (2014), outliers are identified by the scores with standardized residual values greater 3.3 or less than -3.3. However, in this study, the scatterplot (Figure 1) does not show the presence of any outliers.

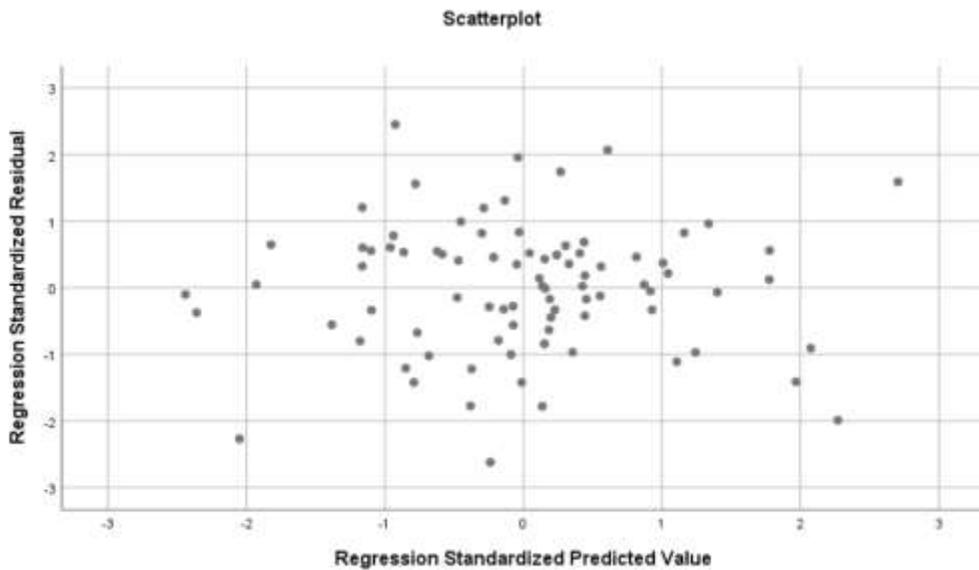
Normality, linearity, homoscedasticity, and independence of residuals through inspecting the Normal Probability Plot (P-P) of the Regression Standardized Residual were also checked. Most crucially, Figure 2 shows that the residuals were distributed “roughly rectangularly” (Pallant, 2011, p. 158), and no deviations were found from the centralized rectangle. As can be seen in figure 2, the points are laid from bottom left to top right in an almost reasonably straight diagonal line and it suggests “no major deviations from normality” (Pallant, 2011, p. 159). Therefore, all of the assumptions were met.

After making sure that all the assumptions were met, the standard multiple regression was run to answer the research question. The results in table 2 show that R-square value is .026, suggesting only 2.6 percent of personality traits might predict raters’ ratings.

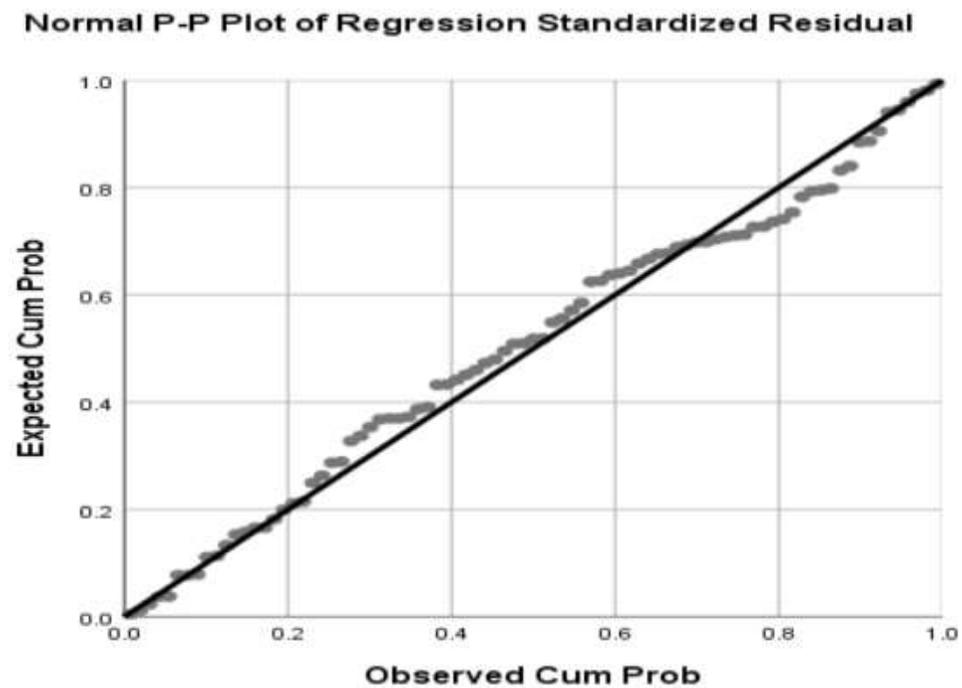
Table 2

*Model Summary of Predictors*

R	R square	Adjusted R square	Standard error of estimate
.160	.026	-.036	42.53795



*Figure 1.* Scatterplot Showing the Absence of Outliers in the Data



*Figure 2.* Normal Probability Plot Showing Normality in the Data

According to table 3, the results of the ANOVA test show that the predictive power of the model is not statistically significant ( $F_{(5, 79)} = .415, p > .837$ ).

Table 3

*ANOVA Test for Predictors*

Model	Sum of squares	Df	Mean square	F	Sig.
Regression	3756.500	5	751.300	.415	.837
Residual	142948.700	79	1809.477		
Total	146705.201	84			

In Table 4, the largest beta value belongs to neuroticism. However, this value is negative, so it does not contribute to raters' ratings. The researcher also checked the Sig. value for each independent variable. If it was less than .05, the variable was regarded to make a significant unique contribution to the prediction of the dependent variable. As can be seen in table 4, none of the personality traits made statistically significant contribution to raters' ratings.

Table 4

*Coefficients for Predictors*

Model	Unstandardized coefficient		Standardized coefficient		
	B	Std. Error	Beta	t	Sig.
1					
(Constant)	394.378	60.247		6.546	.000
Extroversion	-.343	.842	-.046	-.407	.685
Agreeableness	-.436	1.150	-.046	-.379	.705
Conscientiousness	-.661	.943	-.086	-.701	.485
Neuroticism	-.922	.826	-.132	-1.116	.268
Openness	.647	.861	.090	.752	.454

## 4.2. Discussion

The present study set out to examine whether personality traits of the teacher raters would predict the ratings they awarded to students' essays. The results of the standard multiple regression showed that extroversion, openness, agreeableness, neuroticism, and conscientiousness did not predict the ratings. The findings support those of Yun, Donahue, Dudley, and McFarland (2005), who focused on rater personality, rating format, and two types of feedback and found that agreeableness could not predict raters' leniency when it came to ratees' poor performance. In other words, in their study, agreeableness did not reliably predict raters' rating. The results of the present study, however, are incompatible with those of Randall and Sharples' (2012), who found that agreeableness could predict raters' leniency. In other words,

agreeable raters assigned higher ratings to the ratees' performance. Moreover, the findings of the present study ran counter to those of Bernardin, Tyler, and Villanova (2009), who found personality traits such as agreeableness and conscientiousness could predict raters' rating as more agreeable raters were more lenient and less conscientious raters were less accurate.

According to the results of the study, personality traits cannot predict raters' ratings. Although we do not know the actual reasons for the absence of prediction from the findings of the present study, it seems that personality traits may be combined, or interact, with other factors for more satisfactory results. Alternatively, other individual differences may mediate the role of such personality factors in predicting the ratings. As Dörnyei and Ryan (2015) argued, to better account for the effect of personality traits, they should be supported by good, existing theories.

Whatever the reasons, the following main arguments can be presented to acknowledge why raters' personality traits failed to predict the ratings they awarded to students' essays. Most of past studies just worked on two personality traits including agreeableness and conscientiousness (e.g., Bernardin, Thomason, Buckley, & Kane, 2015; Bernardin, Tyler, & Villanova, 2009; Randall & Sharples, 2012). Most crucially, focusing on five personality traits in our study may lead to such results. Obviously, five personality traits imposed more limitations on the study while the previous studies used two personality traits which reflected predictable results. Aside from personality traits, previous published research investigated the role of individual personality trait, such as just agreeableness or conscientiousness, in the prediction of raters' rating. Such a reason may also potentially lead to the current result in our study. It is noteworthy that researchers in previous studies used different types of personality questionnaires. Given the increasing significance of personality traits and their sub-categories, Chapman (2007) found that focusing on subcomponents of personality traits and utilizing particular personality questionnaires could enhance the prediction of personality traits. Therefore, the type of personality questionnaires may influence the results as well.

The setting might affect the results as well. When it came to rating the essays, the raters rated the essays at home and brought them back after one week. It would have been ideal if we had asked all raters to rate the essays in one sitting. However, as it was evident from the limitations, making raters rate the essays in one session was not practical. Therefore, such a limitation might adversely have affected the results.

## 5. Conclusion and Implications

The findings of the present study showed that personality traits did not contribute to the ratings. Curiously, personality traits may not be among the factors that may lead to clarifying raters' ratings. Of the studies available, different questionnaires were drawn on to elicit the participants' personality traits. Using different questionnaires may lead to various results in different studies including the present study. Although rating an essay might be related and influenced by psychological factors, personality traits do not seem to be parts of those factors. In other words, traits could be the causes for many behaviors among teachers, but they could not predict their ratings. Moreover, focusing too much on teachers' awareness of their personality traits may let them believe that such a factor could be influential; therefore, teachers try to mitigate the possible effects.

The present study had some limitations that should be recognised. One shortcoming was related to the number of raters. Due to time constraints, unavailability of language teachers, and reluctance of teachers to cooperate with the researcher, the present study did not use a large number of teacher raters. In the future, researchers should use a larger number of raters to come up with findings that are more robust. Another limitation of the present study had to do with the varying degrees of experience of teacher raters study. Studies of this type should deal with the teachers who possess similar levels of rating experience.

## References

- Bakker, B. N., Klemmensen, R., Nørgaard, A. S., & Schumacher, G. (2015). Stay loyal or exit the party? How openness to experience and extroversion explain vote switching. *Political psychology*, 37(3), 419-429.
- Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12(2), 86-107.
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54-74.
- Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice*, 18(3), 279-93.
- Bernardin, H. J., Thomason, S., Buckley, M. R., Kane, J. S. (2015). Rater rating-level bias and accuracy in performance appraisals: The impact of rater personality, performance management competence, and rater accountability. *Human Resource Management*, 55(2), 321-340.

- Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment*, 17(3), 300-310.
- Bollmer, J. M., Harris, M. J., & Milich, R. (2006). Reactions to bullying and peer victimization: Narratives, physiological arousal, and personality. *Journal of Research in Personality*, 40(5), 803–28.
- Chapman, B. P. (2007). Bandwidth and fidelity on the NEO Five Factor Inventory: Replicability and reliability of Saucier's (1998) item cluster subcomponents. *Journal of Personality Assessment*, 8 (1), 220–234.
- Chen, G. M. (2014). Revisiting the social enhancement hypothesis: Extroversion indirectly predicts number of Facebook friends operating through Facebook usage. *Computers in Human Behavior*, 39(1), 263-269.
- Costa, P. T., Jr., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: the NEO personality inventory. *Psychological Assessment*, 4(1), 5–13.
- Courbalay, A., Deroche, T., Prigent, E., Chalabaev, A., & Amorim, M. A. (2015). Big five personality traits contribute to prosocial responses to others' pain. *Personality and Individual Differences*, 78 (1), 94-99.
- DeYoung, C. G., & Gray, J. R. (2009). Personality neuroscience: explaining individual differences in affect, behavior and cognition. In Corr, P. J., & Matthews, G. (Eds.), *The Cambridge handbook of personality psychology*, (pp. 323-346). New York: Cambridge University Press.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of language learner revisited*. New York and London: Routledge.
- Eckes, T. (2012). Operational Rater Types in Writing Assessment: Linking rater cognition to rater behavior. *Language Assessment Quarterly*, 9(3), 270–292.
- Elliot, A. J., & Thrash, T. M. (2002). Approach avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, 82(5), 804–818.
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- Eysenck, H. J. (1997). Personality and experimental psychology: the unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73(6), 1224–37.
- Feist, J., & Feist, G. J. (2008). *Theories of personality (7<sup>th</sup> ed.)*. The United States of America: McGraw–Hill Primis.
- Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive

- test performance of introverts and extraverts. *Ergonomics*, 45(3) 203–17.
- Gleason, K. A., Jensen-Campbell, L. A., & Richardson, D. S. (2004). Agreeableness as a predictor of aggression in adolescence. *Aggressive Behavior*, 30(1), 43-61.
- Goldberg, L.R. (1981). Language and individual differences: The search for universals in personality lexicon. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills: Sage.
- Goulden, N. R. (1994). Relationship of analytic and holistic methods to raters' scores for speeches. *The Journal of Research and Development in Education*, 27(2), 73–82.
- Grahek, M. (2007). Personality and rater leniency: Comparison of broad and narrow measures of conscientiousness and agreeableness (Master's thesis). Retrieved from [http://digital.library.unt.edu/ark:/67531/metadc3668/m2/1/high\\_res\\_d/thesis.pdf](http://digital.library.unt.edu/ark:/67531/metadc3668/m2/1/high_res_d/thesis.pdf)
- Grawitch, M. J., Barber, L. K., & Justice, L. (2010). Rethinking the work–life interface: It's not about balance, it's about resource allocation. *Applied Psychology: Health and Well-Being*, 2(2), 127–159.
- Hockenbury, D. H., & Hockenbury, S. E. (2011). *Discovering psychology* (5<sup>th</sup> ed.). United States of America: Worth Publishers.
- Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing*, 13(3), 201-218.
- John, O. P. (1990). The Big Five factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. Pervin (Ed.), *Handbook of personality theory and research* (pp. 66-100). New York: Guilford.
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: 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. John, Robins, R. & L. Pervin. (Eds.), *Handbook of personality: Theory and research* (3<sup>rd</sup> ed. pp., 114-158). New York: Guilford Press.
- Jung, C. J. (1971). *Psychological types*. In *collected works* (vol. 6). Princeton, NJ: Princeton University Press.
- Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, G. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2015). Openness to experience and

- intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, 84(2), 248-258.
- Kruglanski, A. W., Orehek, E., Higgins, E. T., Pierro, A., & Shalev, I. (2010). Modes of self-regulation: assessment and locomotion as independent determinants in goal pursuit. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp.375-402). London: Blackwell.
- LePine, J. A. & Van Dyne, L. (2001). Voice and cooperative behaviour as contrasting forms of contextual performance: evidence of differential relationships with big five personality characteristics and cognitive ability. *Journal of Applied Psychology*, 86(2), 326–336.
- Lin, W., Ma, J., Wang, L., & Wang, M. (2015). A double-edged sword: The moderating role of conscientiousness in the relationships between work stressors, psychological strain, and job performance. *Journal of Organizational Behavior*, 36(1), 94–111.
- Maddi, S. R., & Berne, N. (1964). Novelty of productions and desire for novelty as active and passive forms of the need for variety. *Journal of Personality*, 22(2), 270-277.
- Matthews, J., Deary, L., & Whiteman, M. (2009). *Personality Traits (3<sup>rd</sup> ed)*. New York: Cambridge University Press.
- McCrae, R. R. (1994). Openness to experience: Expanding the boundaries of Factor V. *European Journal of Personality*, 8(4), 251-272.
- McCrae, R. R. (2011). Personality theories for the 21st Century. *Teaching of Psychology*, 38(3), 209-214.
- McCrae, R. R. (2011). Personality theories for the 21st Century. *Teaching of Psychology*, 38(3), 209-214.
- McCrae, R. R., & Costa, Jr. P.T. (1997). Conceptions and correlations of openness to experience. In Hogan, R., Johnson, J., & Briggs, S. (Eds.), *Handbook of personality psychology* (pp.825-847). The United States of America: Academic Press.
- McCrae, R. R., & John, O. P. (1992). An Introduction to the five-factor Model and its Applications. *Journal of Personality*, 60(2), 175-215.
- McCrae, R. R., & Lockenhoff, C. E. (2010). Self-regulation and the five-factor model of personality traits. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp. 145-168). Singapore: Blackwell Publishing.
- Milfont, T. L., & Sibley, C. G. (2012). The big five personality traits and environmental engagement: Associations at the individual and societal level. *Journal of Environmental Psychology*, 32(2), 187–195.
- Mondak, J. J., Hibbing, M. V., Seligson, M. A., & Anderson, M. R. (2010). Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *American Political Science Review*, 104(1), 1–26.

- Mussel, P., Winter, C., Gelléri, P., Schuler, H. (2011). Explicating the openness to experience construct and its subdimensions and facets in a work setting. *International Journal of Selection and Assessment*, 19(2), 145-156.
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Ockay, G. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing*, 26(2), 161-186.
- Osberg, T. M. (1987). The convergent and discriminant validity of the need for cognition scale. *Journal of Personality Assessment*, 51(3), 441-450.
- Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS (4<sup>th</sup> ed.)*. Crows Nest, NSW, Australia: Allen & Unwin.
- Patrick, C. L. (2011). Student evaluations of teaching: Effects of the big five personality traits, grades and the validity hypothesis. *Assessment and Evaluation in Higher Education*, 36(2), 239-249.
- Pervin, L. A., & John, O. P. (2001). *Personality, theory and research (8th ed.)*. New York: John Wiley & Sons, Inc.
- Randall, R., & Sharples. D. (2012). The impact of rater agreeableness and rating context on the evaluation of poor performance. *Journal of Occupational and Organizational Psychology*, 85(1) 42–59.
- Robinson, O., Lopez, F., & Ramos, K. (2014). Parental antipathy and neglect: Relations with big five personality traits, cross-context trait variability and authenticity. *Personality and Individual Differences*, 56(1), 18-185.
- Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, A. G., & Orr, R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), 578–586.
- Salgado, J. F., Moscoso, S., Berges, A. (2013). Conscientiousness, its facets, and the prediction of job performance: Evidence against the narrow measures. *International of Journal Selection and Assessment*, 21(1), 74-84.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465- 493.
- Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing*, 22(1) 1–30.
- Silvia, P. J., Nusbaum, E. C., Berg, C., Martin, C., & O'Connor, A. (2009). Openness to experience, plasticity, and creativity: Exploring lower-order, higher-order, and interactive effects. *Journal of Research in Personality*, 43(6), 1087–1090.

- Stangor, C. (2010). *Introduction to psychology*. The United States: Flat World Knowledge.
- Sweedler-Brown, C.O. (1985). The influence of training and experience on holistic essay evaluation. *English Journal*, 74(5), 49–55.
- Swickert, R. (2009). Personality and social support processes. In Corr, P. J., & Matthews, G. (Eds.), *The Cambridge handbook of personality psychology*, (pp. 524-540). New York: Cambridge University Press.
- Szymura, B., & Nećka, E. (2005). Three super factors of personality and three aspects of attention. In Elias, A., Hampson, S. E., DeRaad, B. (Eds.). *Advances in Personality psychology*, (pp. 65-79). The United States of America: Psychology Press.
- Tabachnick, B. G. and Fidell, L.S. (2014). *Using multivariate statistics (6th ed.)*. Boston: Pearson Inc.
- Triandis, H. C. (1997). Cross-cultural perspectives on personality. In Hogan, R., Johnson, J., & Briggs, S. (Eds.), *Handbook of personality psychology*, (pp. 439-464). The United States of America: Academic Press.
- Watson, D., & Clark, L. A. (1997). Extroversion and its positive emotional core. In Hogan, R. Johnson, J. & Briggs, S. (Eds.), *Handbook of personality psychology*, (pp. 767-793). The United States of America: Academic Press.
- Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, 6(2), 145–178.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.
- Wiggins, J. S., & Trapnell, P. D. (1997). Personality structure: The return of the big five. In Hogan, R. Johnson, J. & Briggs, S. (Eds.), *Handbook of personality psychology*, (pp. 737- 765). The United States of America: Academic Press.
- Wiseman, C. S. (2012). Rater effects: Ego engagement in rater decision-making. *Assessing Writing*, 17(3), 150-173.
- Yan, X. (2014). An examination of rater performance on a local oral English proficiency test: A mixed-methods approach. *Language Testing*, 31(4), 501-527.
- Yao, Q., & Moskowitz, D. S. (2015). Trait Agreeableness and social status moderate behavioral responsiveness to communal behavior. *Journal of Personality*, 83(2), 191- 201.
- Yun, G. J., Donahue, L.M., Dudley, N. M., & McFarland, L. A. (2005). Rater personality, rating format, and social context: implications for performance appraisal ratings. *International Journal of Selection and Assessment*, 13(2), 97-107.

***Bibliographic information of this paper for citing:***

Esfandiari, R. (2019). How predictable ratings are: The role of personality traits. *Journal of Modern Research in English Language Studies*, 6(3), 33-55.

Copyright© 2019, Esfandiari, R.