

The Predictive Power of Big Five Traits on EFL Teacher Ratings of University Students' Academic Essays

Research Paper

Received: 2021-01-05

Abstract

¹Rajab Esfandiari*

IJEAP- 2101-1679

Accepted: 2020-04-06

DOR: [20.1001.1.24763187.2021.10.1.5.9](https://doi.org/10.1.24763187.2021.10.1.5.9)

Published: 2020-04-06

Results from some previous L2 studies have pointed to the positive correlation between psychological variables and L2 proficiency, fluency, and academic success. Such research findings have contributed considerably to our knowledge of the field. However, the extent of such psychological factors as conscientiousness, and their relation with teacher ratings is still unknown. The ultimate goal of the present study was to examine the predictive power of five psychological variables, known as big five traits, on the ratings Iranian EFL teachers awarded to student essays. To that end, 150 teacher raters were asked to rate ten randomly selected five-paragraph essays (once analytically and once holistically) students of English had written in their essay writing classes using analytic and holistic rating scales. Two separate standard multiple regression procedures, as implemented in SPSS (version 25), were used to analyse the data for the present study. Results from regression analyses showed big five traits, including extroversion, conscientiousness, openness to experience, neuroticism, and agreeability, did not predict analytic and holistic ratings. The findings suggest that such psychological variables do not statistically contribute to the ratings awarded to expository text types.

Keywords: Analytic Scales, Big Five, Holistic Scales, Ratings, Traits

1. Introduction

Although personality appears to play a key role in second language acquisition (SLA), it has not always received the attention it merits. Ellis (2004), for example, noted that “[i]ntuitively, personality is a key factor for explaining individual differences in L2 learning” (p. 541). Researchers have presented some cogent reasons why such a useful construct has most often been neglected in SLA, including methodological rigour to examine it (Piechurska-Kuciel, 2020) and the absence of psychometrically sound instruments to measure it (Dörnyei & Ryan, 2015). Such inadequacies notwithstanding, researchers have analysed personality and tried to establish its relationship with other SLA constructs such as motivation since a long time ago (Krashen, 1981). Cook summarised (1991) the reasons for researchers’ interest in personality as follows: “First, to gain scientific understanding, second, to access people and next, to change people” (p. 3).

Personality traits, reflecting the personal features of individual people, account for many learning outcomes. Using trait theories, researchers have conceptually defined personality simply as a set of dispositions encompassing a wide range of psychological constructs (see Hiriyappa, 2012). Although research results have been somewhat mixed regarding the relationship between personality and learning, such inconsistencies seem to stem from several other intervening factors, not the least among them the interaction of context-specific variables (see Piechurska-Kuciel, 2020, for a long list of the reasons). As such, the study of personality traits merits further scrutiny.

Big five traits, conceived as instances of personality traits, have made consistent inroads into SLA. From psychological point of view, such traits are considered individual differences, which appear to be unique, relatively stable characteristics of people across different occasions (Dörnyei & Ryan, 2015). In their words, “enduring personal characteristics ... are assumed to apply to everybody and on which people differ by degree” (p. 3). Similarly, Ellis (2008) noted that the study of individual differences helps researchers to appreciate the infinite number of variables which may account for language learners’ inconsistent performance in SLA. Such features reflect the personality of

¹ Associate Professor of TEFL (Corresponding Author), esfandiari@hum.ikiu.ac.ir; Imam Khomeini International University, Qazvin, Iran.

individuals and are researched in personality psychology (Dörnyei, 2005). The big five traits, as the name suggests, include extroversion, openness, neuroticism, conscientiousness, and agreeableness, which the present researcher has focused on, as research studies on the predictability of such traits on teacher ratings can be very rarely found in the literature.

Although the relationship between the big five traits and learning has proved promising (e.g., Harari, Rudolph, & Laginess, 2015; AlFallay, 2004), we constantly encounter situations in SLA in which L2 raters may be asked to assess the written, or spoken, performance of language learners. L2 research studies have shown L2 raters may be affected by a number of factors including rater experience, fatigue, and rating scales. However, the extent to which such big fives may correlate with and predict raters' analytic and holistic ratings is a topic, which seems to have been somewhat neglected in SLA studies. Esfandiari (2019) is an exception. Esfandiari examined the extent to which big five traits were capable of predicting the ratings Iranian teacher raters awarded to one-paragraph essays 24 Iranian BA students wrote in their Advanced Writing classes. Using a 5-point, researcher-made analytic rating scale to rate student essays, 85 Iranian EFL teachers rated ten essays on some assessment criteria. Results from a standard multiple regression procedure showed big five traits—extroversion, agreeableness, openness, neuroticism, and conscientiousness—did not contribute statistically significantly to the analytic ratings, suggesting that these traits were not predictors of those traits. Although Esfandiari outlined some reasons to account for the lack of prediction, he concluded that "personality traits may not be among the factors that may lead to clarifying raters' ratings" (p. 49). This study was one of the first to investigate the predictive power of big five factors, but it included some shortcomings. The raters were relatively limited in number, and essays were rated using only an analytic rating scale. Therefore, the need for more studies using more EFL raters and both analytic and holistic ratings scales is strongly felt to come up with firm results. This study was done to fill this gap, focusing on the following two research question. Therefore, the present study is aimed at answering the following research questions.

Research Question One: To what extent do big five traits predict L2 raters' holistic ratings?

Research Question Two: To what extent do big five traits predict L2 raters' analytic ratings?

2. Literature Review

2.1. Personality Traits

Cervone and Pervin (2009) defined personality as "psychological qualities that contribute to an individual's enduring and distinctive patterns of thinking, feeling and behaving" (p. 8). Personality is also defined as an individual's consistent patterns of feeling, thinking, and behaving (Caprara & Cervone, 2000). In other words, personality is one way to understand human behavior and experience through individual differences in relatively consistent thoughts, feelings and actions across situations (John & Srivastava, 1999; McCrae & Costa, 1999).

Personality traits are "dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings and actions" (McCrae & Costa, 1999, p. 142). Roberts (2009), who is a leading figure in the field of personality psychology, defined personality traits as "the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances" (p. 140). Thus, traits are relatively enduring characteristics that influence our behavior across many situations. Personality traits such as extroversion, introversion, conscientiousness, agreeableness, modesty, friendliness, honesty, and helpfulness are important because they explain certain consistencies in behavior. Personality traits tend to describe an individual in terms of general predispositions which are broader than specific behaviors, moods and experiences, but more specific than any universal characteristics (Roberts, Kuncel, Shiner, Caspi, & Golberg, 2007).

The trait approach to personality was pioneered by early psychologists and scholars (Cattell, 1990; Chaplin, John, & Goldberg, 1988; Eysenck, 1990). Each of these psychologists regarded traits as the stable units of personality, and they tried to provide a taxonomy of the most important trait dimensions. They typically provided people with a self-report measure and used statistical analyses

to look for the underlying factors or clusters of traits. They did so by scrutinizing the frequency and the co-occurrence of traits in the respondents.

The fundamental work on trait dimensions conducted by Cattell, Eysenck, and many others has led to contemporary trait models, the most important and well-validated of which is the Big Five Model of personality (McCrae & Costa, 2003). According to this model, there are five fundamental underlying trait dimensions that are stable across time, cross-culturally shared, and explain a substantial proportion of behavior (Roberts, et al., 2007), which include agreeableness, conscientiousness, extroversion, neuroticism, and openness to experience. In the following section, these traits are explained in great detail.

2.2. The Big Five Inventory (BFI)

Goldberg (1981) coined the term Big Five to indicate that extroversion, agreeableness, conscientiousness, neuroticism and openness constituted the core of personality. The model provides a framework for describing personality at its broadest level. Big Five personality model is absolutely parsimonious in encapsulating personality in just five factors (Dornyei, 2005). As McCrae and Costa (1999) noted, these five factors are the broadest categorization of personality, which are comprised of some underlying facets. As Judge et al. (2013) noted, the typology Costa and McCrae offered is popular, well-researched, and still present in contemporary taxonomies. Therefore, using this typology, Costa and McCrae argued that extroversion includes warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotion. Agreeableness consists of trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness. Conscientiousness comprises competence, order, dutifulness, achievement striving, self-discipline, and deliberation. Neuroticism is composed of anxiety, angry hostility, depression, self-consciousness, impulsiveness, and vulnerability. Openness contains fantasy, aesthetics, feelings, actions, ideas, and values. BFI is a 44-item questionnaire designed to measure the Big Five personality traits. It has reasonable reliability and convergent validity (Simms & Clark, 2005). Interested readers may decide to refer to Piechurska-Kuciel (2020), who has provided a detailed discussion of this inventory in second language acquisition, the most single important and recent treatment of the topic. In what follows, components of Big Five are fully described.

According to John and Srivastava (1999), “extroversion implies an energetic approach toward the social and material world and includes traits such as sociability, activity, assertiveness, and positive emotionality” (p. 121). Adjective markers of extroversion include sympathetic, warm, kind, cooperative, cold, cooperative, cold, unsympathetic, rude, and harsh (Saucier, 1994, p. 516). According to Dornyei (2005), high-scoring people on extroversion are sociable, gregarious, assertive, active, talkative, and passionate, whereas low-scoring people are quiet, reserved, sober, aloof, withdrawn, passive and restrained. L2 research has shown that extroversion is generally positively related to L2 success, although some inconsistencies have been observed in previous studies (Liang & Kelsen, 2018).

Agreeableness is in contrast with a prosocial and communal orientation toward others with antagonism. It includes traits such as altruism, tender-mindedness, trust, and modest (John & Srivastava, 1999). As Dörnyei and Ryan (2015) concluded, high scores on agreeableness indicate being good-natured, friendly, likeable, kind, forgiving, generous, modest, trusting, and cooperative. Conversely, low scores are representative of being cold, rude, cynical, antagonistic, unpleasant, critical, vengeful, irritable, and uncooperative. As Dewaele (2012) commented, agreeableness seems to be the least researched trait in L2 as an independent variable, but the very few research studies which can be found in the literature point to the positive relationship between agreeableness and willingness to communicate, greater L2 fluency, and positive attitudes to language learning (Vural, 2019).

Conscientiousness displays socially prescribed impulse control that facilitates task-, and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks (John, Naumann, & Soto, 2008). Piechurska-Kuciel (2020) reminded us that high scorers on conscientiousness are systematic, meticulous, efficient,

organized, reliable, responsible, hard-working, persevering, and self-disciplined. Low scorers, on the other hand, are unreliable, aimless, careless, disorganized, late, lazy, negligent, and weak-willed. Like agreeableness, L2 research on conscientiousness is also very scarce, but research from very few previous research studies in L2 shows that conscientiousness increases L2 proficiency and positively affects L2 reading, writing, grammar, and spelling (Novikova et al., 2020).

Neuroticism contrasts emotional stability and even-temperedness with negative emotionality, such as feeling anxious, nervous, sad, and tense (John, Naumann, & Soto, 2008). This trait, as previous research has shown, is related to negative consequences of almost all aspects of peoples' lives (Oznska-Ponikwia, 2018). According to Dornyei (2005), high scorers on neuroticism, typically, tend to be anxious, worrying, depressed, insecure, self-conscious, emotional, unstable, and moody, whereas low scorers have been found to be calm, stable, relaxed, unemotional, comfortable, hardy, content, even-tempered, and self-satisfied. Piechurska-Kuciel (2020) neatly summarized the research findings on neuroticism in L2 research. In her words, "to sum up, the existing research on the role of neuroticism in SLA appears to confirm the negative effects of the trait on foreign language knowledge and use processes" (p. 117).

John, Naumann, and Soto (2008) pointed out that "openness to experience (vs. close-mindedness) describes the breadth, depth, originality, and complexity of an individual's mental and experiential life" (p. 138). In personality psychology research, the most single most important feature identified in openness to experience is independence (Sutin, 2017). According to Dornyei (2005), high scores on openness to experience imply being head-in-clouds, curious, flexible, move by art, creative, original, untraditional, and novelty seeking, while low scores suggest being down-to-earth, conservative, conventional, practical, and unartistic. Like extroversion, openness to experience has been extensively examined in SLA, and research findings from previous L2 studies have indicated that openness to experience is strongly related to L2 proficiency, positively correlates with L2 willingness to communicate, facilitates the learning process, and positively shapes L2 learners' social and cultural lives (Lindblom-Ylännne et al., 2019; Obralic & Mulalic, 2017).

Alibakhshi, Qaracholloo and Mohammadi (2017) used BFI to analyse the predictive power of personality factors on the choice of language learning strategies among 100 Iranian BA and MA students of English. Results from correlational analyses showed that neuroticism negatively predicted memory and metacognitive strategies. The authors attributed lack of such a prediction to the inherent properties of the trait itself.

Ghapanchi, Khajavy, and Asadpour (2011) were interested in examining how big five traits contributed to L2 proficiency and motivational L2 self-system of Iranian EFL students, using Goldberg's (1992) personality inventory. Results from this study showed that extroversion strongly correlated with L2 proficiency and predicted ideal L2 self and L2 learning experience. The findings of the study also pointed to the assertion that when extroversion was coupled with openness, it explained 13% of variance of L2 proficiency.

Some researchers have demonstrated the relationship between openness to experience and speaking ability in L2 English. For example, Vural (2019) found that less anxious participants while speaking in English were those who were more open to experience. Similarly, Khany and Ghoreyshi (2013) reported that openness to experience strongly predicted foreign language speaking confidence in the classroom.

Some other studies have examined the correlation between agreeableness and willingness to communicate (WTC). Oz (2014) demonstrated the relationship between agreeableness and L2 WTC, reporting that agreeableness was one of the strongest predictors of L2 WTC. Šafranj and Katić (2019) also reported a relatively small correlation between agreeableness and L2 WTC. The findings of these two studies show more agreeable language learners are more likely to be willing to communicate in L2 situations. Finally, the findings from Ghyasi, Yazdani, and Farsani's (2013) study showed to interesting findings. The first finding pointed to the relationship between conscientiousness and language learning strategies, and the second result was that contentious language learners' study habit

and elements of the contentiousness promoted learning. Alibakhshi, Qaracholloo, and Mohammadi's (2017) study also showed the predictive power of contentiousness on compensation strategies.

Although researchers have examined the relationship between big five traits and other key L2 variables in SLA (Dewaele, 2012; Piechurska-Kuciel, 2020; Sutin, 2017), as the review in this section shows, research studies analyzing the correlation between, or the predictive power of, these traits and raters' ratings are very rare.

3. Methodology

3.1. Participants

Participants in the present study included two groups: students and raters. Forty-five male and female BA students studying English Translation and English Language teaching in two intact Essay Writing classes at Imam Khomeini International University in Qazvin participated in the study. These students had already passed Advanced Writing. These students provided the researcher with rating data.

In addition to students, one hundred and fifty English as a Foreign Language (EFL) Iranian non-native English speaking teachers purposefully selected for the purposes of the present study (Dornyei, 2007) at language institutes participated in this study. The teacher raters were both female and male (one hundred and four female assessors and forty-six male assessors). They ranged in age from 24 to 60. They came from five language backgrounds: one hundred and thirty-seven teacher assessors were native-Farsi speakers (91.33%), ten teacher assessors were native-Turkish speakers (6.67%), and three teacher assessors were native-Armenian speakers (2%). Twenty-four of them (16%) had experience living in an English-speaking country. They had taught writing courses from 1 to 30 years. They were 39 BA (26%), 98 MA (65.3%), and 13 PhD (8.7%) holders in English Language Teaching, English Literature, Translation Studies, and other fields.

3.2. Instrumentation

The researcher used three instruments to collect the needed data: BFI, students' essays, and rating scales. BFI developed by John, Donahue, and Kentle (1991) was used in the current study. It consisted of 44 items and five major factors: Neuroticism, extroversion, openness to experience, agreeableness, and conscientiousness (see Appendix). Each factor has six facets. Extroversion includes warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotion. Agreeableness consists of trust, straightforwardness, altruism, compliance, modesty, and tender-Mindedness. Conscientiousness comprises competence, order, dutifulness, achievement striving, self-discipline, and deliberation. Neuroticism is composed of anxiety, angry hostility, depression, self-consciousness, impulsiveness, and vulnerability. Openness contains fantasy, aesthetics, feelings, actions, ideas, and values.

The scale categories included strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5). Eight items were used to measure extroversion. Nine items were used to measure agreeableness. Another nine items were used to measure conscientiousness. Eight items were used to measure neuroticism. Ten items were used to measure openness to experience.

The inventory is a reliable instrument. In American and Canadian samples, the reliability coefficients of the BFI scales range from .75 to .90 (John & Srivastava, 1999); three-month test-retest reliabilities range from .80 to .90, with a mean of .85. (John & Srivastava, 1999; Rammstedt, Lechner, & Danner, 2018). Validity evidence includes substantial convergent and divergent relations with other Big Five instruments as well as with peer ratings (Pervin & John, 2001). The reliability of the BFI in present study is as follows: Cronbach's Alpha: 0.60, Cronbach's Alpha based on standardized items is 0.65.

In addition to BFI, Forty-five five-paragraph essays collected from undergraduate (BA) students were also used in the study. The students enrolled in two Essay Writing courses at Imam Khomeini International University in Qazvin, Iran. Students had 90 minutes to write an expository comparative-contrast, five-paragraph essay ranging in length from 500 to 700 words on the following topic: An e-mail and a letter are both used to transfer information. There are, however, some

differences between these two communication systems. Discuss three differences between them. Ten essays (every five essay) were randomly selected for teacher raters to rate.

Finally, two rating scales were used to rate the essays. An analytic rating scale was developed based on the ESL Composition Profile (Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey, 1981), the Composition Grading Scale (Bailey & Brown, 1984) as cited in Farhady, Jafarpour, and Birjandi (1994), and the principles of the comparison and contrast five-paragraph essay, to rate expository essays in the current study. The main criteria in the analytic rating scale were: (1) organization, (2) content, (3) mechanics, (4) grammar, (5) vocabulary, and (6) coherence and transitions. Each criterion consisted of five descriptors. The scale categories ranged from 1 (very poor), 2 (poor), 3 (good), 4 (very good), to 5 (excellent).

A holistic rating scale was used to rate expository essays in this study. In the analytic scale, the six assessment criteria included five categories each. In designing a holistic rating scale, five categories were also used to score the essays. The scale categories ranged from 1 (very poor), 2 (poor), 3 (good), 4 (very good), 5 (excellent). Each category had its own distinctive descriptor. The raters were free to select one of the above five categories, depending on how effectively the essay was written.

3.3. Procedure

The researcher managed to conduct training sessions for raters before the rating process. The rater trainer was the researcher himself. He had a thirty-minute, one-to-one training session for each and every of the teacher raters and provided explanations on how to rate the essays analytically and holistically. Teacher raters rated expository essays, using analytic and holistic rating scales. The trainer also presented some previously rated expository essays rated both holistically and analytically to raters. In order to enhance the effectiveness of the training program, the trainer then asked raters to rate some essays individually holistically and analytically. In cases where raters assigned completely different ratings, they were asked to explain their highly unexpected ratings.

Expository essays written by the students were handed in to the teacher raters to rate both holistically and analytically based on what they acquired during the training session. The raters were asked to rate the first five essays based on the analytic rating scale and the second five essays based on the holistic rating scales. They were also asked to leave comments when necessary about various elements and features of the scripts and correct the students' errors if necessary. They were supposed to hand in the rated essays within two weeks.

3.4. Statistical Analyses

To analyse the data for the present study, the researcher used SPSS (version 25, a general-purpose, versatile computer programme commonly used in social sciences for data analysis. Two separate standard multiple regression procedures were used to answer the research questions of the present study. The main reason for the choice of these statistical procedures was due to the presence of five independent variables to predict holistic ratings and analytic ratings as dependent variables.

4. Results

4.1. Investigation of the First Research Question

The first research question was concerned with whether big five traits would predict teacher raters' analytic ratings. A standard multiple regression was used to answer this question. Based on the results shown in Table 1, the R square value is 0.057. This shows that these traits explain only 5.7% of variance in teacher raters' analytic ratings.

Table 1: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.238	.057	.024	13.782

The results of the ANOVA test (Table 2) show that the predictive power of the model is not statistically significant ($F_{1,144} = 1.726, p > .132$).

Table 2: ANOVA test for predictors

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1639.654	5	327.931	1.726	.132
Residual	27352.906	144	189.951		
Total	28992560	149			

Dependent variable: Total Analytic Score Predictors: Extroversion, Agreeableness, Conscientiousness, Neuroticism, Openness to Experience.

Because results from ANOVA test are not statistically significant, it is not possible to meaningfully interpret information in Table 3 concerning the predictability of these traits.

Table 3: Coefficients for Predictors

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	110.042	15.083		7.296	.000
Extroversion Raw	-.215	.219	-.083	-.984	.327
Agreeableness Raw	.134	.271	.045	.493	.623
Conscientiousness Raw	-.040	.240	-.016	-.166	.868
Neuroticism Raw	-.535	.208	-.230	-2.566	.011
Openness Raw	.014	.189	.006	.072	.943

4.1. Investigation of the Second Research Question

The second research question related to the extent to which big five traits were able to predict raters' holistic ratings. Another multiple regression analysis was run to answer this question. Based on the results shown in 4, the R square value is 0.046. This shows that the big five traits explain only 4.6% of variance in teacher raters' total ratings.

Table 4: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.215 ^a	.046	.013	15.110

Dependent Variable: Total Score Predictors: Openness, Agreeableness, Extroversion, Neuroticism, Conscientiousness

The results of the ANOVA test (Table 5) show that the predictive power of the model is not statistically significant ($F_{5,144} = 1.379, p > .229$).

Table 5: ANOVA test for predictors

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1594.754	5	318.951	1.397	.229
Residual	32878.606	144	228.324		
Total	34473.360	149			

Dependent Variable: Total Score Predictors: (Constant), Openness, Agreeableness, Extraversion, Neuroticism, Conscientiousness

Based on the information in Table 6, only neuroticism can predict the total score in a statistically significant way, as did for analytic ratings, but because results from ANAOVA test are not statistically significant, the meaningful interpretation of such a prediction is ambiguous and misleading.

Table 6: Coefficients for predictors

Model	Unstandardized Coefficients			t	Sig.
	B	Std. Error	Beta		
(Constant)	126.697	16.537		7.661	.000
Extraversion	-.187	.240	-.066	-.778	.438
Agreeableness	.133	.297	.041	.446	.656
Conscientiousness	-.029	.263	-.010	-.111	.912
Neuroticism	-.525	.229	-.207	-2.297	.023
Openness	-.043	.208	-.018	-.206	.837

Dependent Variable: Total Score

5. Discussion

The ultimate goal of the present study was to determine the extent to which big five traits could predict rater's analytic and holistic ratings. The findings from two separate standard multiple regression procedures showed no statistically significant contributions of big five traits to ratings. Therefore, these traits were not able to predict the ratings raters awarded to students' essays. In the following paragraphs, the findings are compared with those of previous similar studies, discussed in light of relevant theories, and furnished with possible explanations.

The findings of the present study partially lend support to those of Esfandiari (2019), who found similar results when Iranian EFL raters rated ten BA students' essays on an analytic rating scales. However, Bernardin, Cooke, Villanova (2000) found different results in a different context. They used the five-factor model of personality to predict raters' lenient ratings. Raters high on agreeableness provided more elevated ratings, while raters high on conscientiousness awarded less elevated ratings. Bernardin et al. used the conditional approach to dispositional constructs proposed by Kane et al. (1995) in order to interpret their results. This approach indicates that raters' traits interact with the context to influence the rating behavior. As Kane et al. noted, raters' agreeableness might predict leniency when raters anticipate future interaction with the examinees/test takers after the ratings, or when the raters are solely responsible for the ratings, or when the ratings have administrative significance.

Similarly, Yun et al. (2005) manipulated the social context. They examined the interactive effect of three factors, namely, the rating social context, rater personality, and rating format on rating elevation. These researchers predicted that the context which requires raters to provide feedback to students/test takers would affect the relationship between rater personality and rating elevation. The results of their study indicated that both the social context in which ratings took place and the rating scale might affect the relationship between personality and rating elevation. The results of the

manipulation checks indicated that the participants in the face-to-face feedback condition were significantly less motivated to provide accurate ratings than those in the no face-to-face feedback condition. Yun et al. found a two-way interaction between agreeableness and social context on rating elevation where raters high on agreeableness provided more elevated ratings than raters low on agreeableness when they anticipated providing a provide face-to-face feedback. This interaction was found when raters assessed students' performance of moderate and high performance.

In the present study, the teacher raters were teachers of English as a foreign language, and they were not required to provide face-to-face feedback to the students who wrote the essays. Face-to-face feedback requires interaction between the raters and students, and it may elicit motives to avoid negative reactions on the part of the students. In the context of face-to-face feedback, raters may be motivated to give uniformly positive feedback (enhance their ratings) to prevent any negative reactions from students/test takers (Latham, 1986).

The reasons for the absence of any prediction of both analytic and holistic ratings are unknown in the present study, but, speculatively, some possible reasons can be given why big five traits did not predict ratings. First, in the very few past studies, not all big five traits were simultaneously examined to measure their predictive power on raters' ratings. Previous studies (e.g., Bernardin, et al., 2009) limited themselves to only a single, or at best two personality factors, to see its results on ratings, so they did not give us a complete picture of all the traits. Possibly, when the concurrent examination of traits is in order, the results will be different, as were the case in this study. Second, instrumentation may also account for the findings of the present study. Researchers examining traits do not necessarily use the same trait instruments and trait theories. Because the underlying underpinnings of various trait models may differ considerably from each other due to paradigmatic differences, results may not necessarily be similar across research studies. Even in studies in which big five models traits been used, as was the case in the present study, researchers have employed different versions of BFI which are conceptually different and have followed different validation procedures. Last, but not least, native versus not-native status of raters also affects the ratings, with raters from different native background focusing on different textual elements, as has been documented in previous studies (see Wind, 2020, for an update). Although raters in the present study were carefully trained before rating the essays, they were non-native English speaking raters who might have developed their own interpretation of the scales as rating unfolded, thereby exerting an influence on ratings.

5. Conclusions and implications

The researcher of the present study set out to examine the degree of predictability of big five traits of EFL rater ratings in a EFL setting. Although the predictive power of big five traits was not confirmed in the present study, it seems that these traits by themselves do not adequately contribute to the ratings, and contextual, social, and affective factors may mediate the role of big five traits in best accounting for any ratings awarded to students' performances. Factors such as assessment setting, raters' own interpretation of the rating scales, and other individual differences may interact with raters' traits to affect the role these traits play in predicting the ratings. The conclusion which is warranted in the present context points to the inadequacy of big five traits, when considered alone, to help researchers to establish their predictive power in rater-mediated assessments in EFL settings, implying researchers should not primarily focus on these personality factors in justifying the unexpected ratings EFL raters assign.

The present study has some limitations which need to be acknowledged to help us better appreciate the findings and suggest some areas for further research. Although a larger number of teacher raters participated in the rating process, researchers should consider carrying out a comparative study between native and non-native speaking raters to observe the results on ratings. Such a cross-comparison of ratings might yield more informative results. Second, in future studies, more varied, innovative research designs, including the incorporation of stimulated recalls, need to be used to complement the findings and help researchers to more meaningfully explain the results. Finally, the predictability of big five traits seems to be mediated by a number of intervening variables,

so future research studies should aim to examine the extent to which such contextual factors may affect the role of traits in predicting ratings.

References

- Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, 85(2), 232–234.
- Caprara, G. V., & Cervone, D. (2000). *Personality: Determinants, dynamics, and potentials*. Cambridge: Cambridge University Press.
- Cattell, R. B. (1990). Advances in Cattellian personality theory. In L. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 101-110). New York: Guilford.
- Cervone, D., & Pervin, L. A. (2009). *Personality: Theory and research*. Hoboken, NJ: John Wiley and Sons.
- Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: dimensional attributes with ideals as prototypes. *Journal of personality and social psychology*, 54(4), 541.
- Cook, V. (1991). *Second language learning and language teaching*. London: Edward Arnold.
- Dewaele, J.-M. (2012). Personality in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–8). Oxford: Wiley-Blackwell Publishing Ltd.
- 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.
- Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). Malden, MA: Blackwell.
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- Eysenck, H. J. (1990). Biological dimension of personality: In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 244-276). New York, NY: The Guilford Press.
- Frahady, H., Jafarpour, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. Tehran: SAMT Publications.
- Hiriyappa, B. (2012). *Development of personality and its theories*. Bloomington, IN: Booktango.
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- 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., Donahue, E. M., & Kentle, R. (1991). *The Big Five Inventory—Versions 4a and 54* (Technical report). Berkeley: Institute of Personality and Social Research, University of California.
- John, O. P., Nauman, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114-158). New York: Guilford.
- Judge, et al. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98(6), 875–925.

- Kane, et. al. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal*, 38(4), 1036-1051.
- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Latham, G. P. (1986). Job performance and appraisal. In C. L. Cooper & I. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 1, pp. 117–155). London: John Wiley.
- Liang, H-Y., & Kelsen, B. (2018). Influence of personality and motivation on oral presentation performance. *Journal of Psycholinguistic Research*, 47(4), 755–776.
- Lindblom-Ylännne, S., et al. (2019). What constitutes the surface approach to learning in the light of new empirical evidence? *Studies in Higher Education*, 44(12), 2183–2195.
- McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed). Guilford Press.
- McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality theory and research* (Vol. 2, pp. 139 –153). New York: Guilford Press.
- Novikova, I. A., Berisha, N. S., Novikov, A. L., & Shlyakhta, D. A. (2020). Creativity and personality traits as foreign language acquisition predictors in university linguistics students. *Behavioral Sciences*, 10(1). <https://doi.org/10.3390/bs10010035>.
- Obralic, N., & Mulalic, A. (2017). Correlation between personality traits and language learning strategies among IUS students. *Journal of Applied Linguistics and Language Research*, 4(5), 76–84.
- Ozanska-Ponikwia, K. (2018). *Personality and emotional intelligence in second language learning*. Newcastle upon Tyne: Cambridge Scholars.
- Pervin, L. A., & John, O. P. (2001). *Personality: Theory and research*. New York: Wiley.
- Piechurska-Kuciel, E. (2020). *Big Five in SLA*. Cham: Springer.
- Rammstedt, B., Lechner, C. M., & Danner, D. (2018). Relationships between personality and cognitive ability: A facet-level analysis. *Journal of Intelligence*, 6(2), 1–13.
- Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, 43(2), 137-145.
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345.
- Saucier, G. (1994). Mini-Markers: A brief version of Goldberg's unipolar big-five markers. *Journal of personality assessment*, 63(3), 506-516.
- Simms, L. J., & Clark, L. A. (2005). Validation of a computerized adaptive version of the Schedule for Nonadaptive and Adaptive Personality (SNAP). *Psychological Assessment*, 17(1), 28-43.
- Sutin, A. R. (2017). Openness. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model* (pp. 83–104). Oxford: Oxford University Press.
- Vural, H. (2019). The relationship of personality traits with English speaking anxiety. *Research in Educational Policy and Management*, 1(1), 55–74.
- Wind, S. A. (2020). Do raters use rating scale categories consistently across analytic rubric domains in writing assessment? Assessing Writing, 43. Doi: 10.1016/j.asw.2019.100416.

- Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, 43, 1-13.
- Yun, et al. (2005). Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, 13(2), 97-107.
- Harari, Rudolph, & Laginess, (2015). Does rater personality matter? A meta-analysis of rater big five–performance rating relationships. *Journal of Occupational and Organizational Psychology*, 88(2), 387-414.
- AlFallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self-and peer-assessment. *System*, 32(3), 407-425.
- Alibakhshi, G., Qaracholloo, M., & Mohammadi, M. J. (2017). A cultural inquiry into personality factors and language learning strategies in an Iranian EFL context. *The International Journal of Humanities*, 24(4), 1–16.
- Ghapanchi, Z., Khajavy, G. H., & Asadpour, S. F. (2011). L2 motivation and personality as predictors of the second language proficiency: Role of the Big Five traits and L2 motivational self-system. *Canadian Social Science*, 7(6), 148–155.
- Khany, R., & Ghoreyshi, M. (2013). The nexus between Iranian EFL students' Big Five personality traits and foreign language speaking confidence. *European Online Journal of Natural and Social Sciences: Proceedings*, 2(2), 601-611.
- Oz, H. (2014). Big Five personality traits and willingness to communicate among foreign language learners in Turkey. *Social Behavior & Personality: An International Journal*, 42(9), 1473–1482.
- Šafranj, J., & Katić, M. (2019). Five personality traits and willingness to communicate. *Andragogija*, 65(1), 69–81.
- Ghyasi, M., Yazdani, M., & Farsani, M. A. (2013). The relationship between personality types and self-regulated learning strategies of language learners. *International Journal of Applied Linguistics and English Literature*, 2(4), 74–82.

Appendix. Big Five Personality Inventory

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 Disagree Strongly	2 Disagree a little	3 Neither agree nor disagree	4 Agree a little	5 Agree strongly
----------------------------------	----------------------------------	---	-------------------------------	-------------------------------

I see myself as someone who ...

- 1. is talkative
- 2. tends to find fault with others
- 3. does a thorough job
- 4. is depressed, blue
- 5. is original, comes up with new ideas
- 6. is reserved
- 7. is helpful and unselfish with others
- 8. can be somewhat careless
- 9. is relaxed, handles stress well
- 10. is curious about many different things
- 11. is full of energy
- 12. starts quarrels with others
- 13. is a reliable worker
- 14. can be tense
- 15. is ingenious, a deep thinker
- 16. generates a lot of enthusiasm
- 17. has a forgiving nature
- 18. tends to be disorganized
- 19. worries a lot
- 20. has an active imagination
- 21. tends to be quiet
- 22. is generally trusting
- 23. tends to be lazy
- 24. is emotionally stable, not easily upset
- 25. is inventive
- 26. has an assertive personality
- 27. can be cold and aloof
- 28. perseveres until the task is finished
- 29. can be moody
- 30. values artistic, aesthetic experiences
- 31. is sometimes shy, inhibited
- 32. is considerate and kind to almost everyone
- 33. does things efficiently
- 34. remains calm in tense situations
- 35. prefers work that is routine
- 36. is outgoing, sociable
- 37. is sometimes rude to others
- 38. makes plans and follows through with them
- 39. gets nervous easily
- 40. likes to reflect, play with ideas
- 41. has few artistic interests
- 42. likes to cooperate with others
- 43. is easily distracted
- 44. is sophisticated in art, music, or literature