ORIGINAL RESEARCH

How Personality Traits Affect Peer Assessment in Distance Learning

Tomás Alves[1,2] · Francisco Sousa[1] · Sandra Gama[1,2] · Joaquim Jorge[1,2] · Daniel Gonçalves[1,2]

Abstract

Recent research has leveraged peer assessment as a grading system tool where learners are involved in learning and evaluation. However, there is limited knowledge regarding individual differences, such as personality, in peer assessment tasks. We analyze how personality factors affect the peer assessment dynamics of a semester-long remote learning course. Specifically, we investigate how psychological constructs shape how people perceive user-generated content, interact with it, and assess their peers. Our results show that personality traits can predict how effective the peer assessment process will be and the scores and feedback that students provide to their peers. In conclusion, we contribute design guidelines based on personality constructs as valuable factors to include in the design pipeline of peer assessment systems.

Keywords Peer assessment · Personality psychology · Remote teaching · Moodle · Design guidelines

1 Introduction

Research shows that learning modes leveraging remote settings impact teaching and learning positively (Gilbert and Flores-Zambada, 2011; Morris, 2014). For instance, blended learning (Garrison and Kanuka, 2004; Graham, 2013) enriches traditional class lectures with virtual learning environments (Dillenbourg et al., 2002) to provide increased convenience, flexibility, exchange of materials, individualized learning, and feedback (Chou and Liu, 2005). All these prospects led online education to be an established feature of higher education (Puška et al., 2021), mainly driven by the recent coronavirus pandemic (Dhawan, 2020). However, this practice usually leads to an untenable grading burden, as more

✉ Tomás Alves
tomas.alves@tecnico.ulisboa.pt

[1] Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

[2] INESC-ID, R. Alves Redol 9, 1000-029 Lisboa, Portugal

students require extra effort to assess all their assignments (Piech et al., 2013; Gernsbacher, 2015).

Generally, the course faculty is responsible for evaluating student performance in their classes since individualized feedback is an integral part of education. Given the considerably large number of students that need to receive feedback in open and distance learning modes such as Massive Open Online 30 Courses (MOOCs), research has proposed methods to provide suitable feedback analogous to those of feedback activities in a traditional classroom (see Suen (2014) for further discussion). Among these methodologies, we highlight automated formative assessment (Gikandi et al., 2011) and peer feedback (Reinholz, 2016). Recent studies leverage automated assessment frameworks showing that they can foster student performance, perception, and engagement with valuable learning experiences (Núñez-Peña et al., 2015; Sangwin and Köcher 2016; Moreno and Pineda, 2020). However, several disadvantages of online assessment, such as software costs to develop educational content and supporting infrastructure, may hinder its use (Tuah and Naing, 2021). Researchers widely use peer assessment techniques to save time and resources during the formative assessment by offloading part of the grading work to students (Mahanan et al., 2021).

Peer assessment consists of a set of activities through which individuals judge the work of others (Reinholz, 2016). By playing the roles of both assessor and assessed to grade and comment on each other's work (Na and Liu, 2019), students can also reflect and discover new understandings by finding the difference between others and themselves (Chang et al., 2020). Recent meta-analyses confirm that peer assessment is essential in teaching and learning nowadays (Zheng et al., 2020; Li et al., 2020; Yan et al., 2022). In particular, formative assessment literature shows that peer assessment has great potential to improve students' academic performance (Black and Wiliam, 2009; Yan et al., 2022) by applying pedagogical activities that facilitate learning (Adachi et al., 2018; Double et al., 2020) and promote social-affective development (Li et al., 2020). Researchers demonstrated that this methodology provides positive results at all educational levels (Li et al., 2020; Yan et al., 2022) and can be a useful pedagogic tool for blended learning and MOOCs across a broad range of disciplines (Liu and Sadler, 2003; Price et al., 2013). Peer assessment fosters student autonomy and self-regulation capabilities (Li et al., 2020; Reinholz, 2016) by allowing students to develop a habit of reflection and a constructive critical spirit as evaluators (Panadero et al., 2013). Furthermore, it encourages students to have a more robust engagement in the overall learning process since the awareness of an audience encourages more thoughtful authoring (Wheeler et al., 2008). Nevertheless, peer assessment is susceptible to several pitfalls and hindrances that may introduce noisy data in the evaluation results (Adachi et al., 2018).

State-of-the-art research tackled limitations, including reliability, perceived expertise, and power relations (Liu and Carless, 2006; Li et al., 2018; Panadero et al., 2019). However, recent studies highlight the need to understand how other idiosyncratic factors may explain the variance observed in peer assessment environments (Schmidt et al., 2021). For instance, individual differences may lead students to follow distinct scoring tactics (Chan and King, 2017) and perceive user-generated content differently (Shao, 2009), yielding biased and unreliable grades. Educational psychology recognizes that individual differences play a significant role in learning and achievement (An and Carr, 2017). However, there is an empirical gap regarding the impact of psychological constructs such as personality on peer assessments (Chang et al., 2021; Rivers, 2021). Recent studies focus on whether personality influences peer assessment evaluation perceptions (Cachero et al., 2022; Rod et al., 2020) or how a target's personality traits predicted how peers rate them (Martin and

Locke, 2022). Nevertheless, to the best of our knowledge, there needs to be more research studying how the assessor's personality traits affect their peer assessment despite researchers discussing the effect of personality traits as a possible factor in the quality of peer assessment techniques (Murray et al., 2017; Na and Liu, 2019).

Our goal is to address this research gap by *investigating the impact of personality in peer assessment in distance learning*. If personality traits offer substantial evidence, that would suggest that peer assessments may hinge more on the assessor's personality than on the work they grade, warranting further inquiry. Our work presents an analysis through the scope of the personality on the peer assessment in a real dataset collected from a university course. This research might be helpful for practitioners as the first steps to include personality factors in the development of peer assessment environments by (a) Helping them to predict how personality introduces variance in the grading process and (b) Helping them to choose the set of student evaluators most appropriate to minimize the noise introduced by personality.

The remainder of the paper is organized as follows: Sect. 2 explains the conceptual model of our study. Section 3 reviews the literature on peer assessment and how past research used personality to understand the impact of individual differences in the peer assessment process. Then, Sect. 4 elaborates on the experimental design and details its execution. Next, Sect. 5 cover the analysis of the results, and Sect. 6 draws the implications of our work to the education community. Finally, Sect. 7 outlines the main conclusions and proposes future research directions.

2 Conceptual Model

We draw on an existing theoretical model of personality to empirically assess how an individual's personality traits affect their grading process in a peer assessment environment. We also discuss the metrics to evaluate peer assessment environments that interest our study.

2.1 Personality

The American Psychological Association defines *personality* as *the enduring configuration of characteristics and behavior that comprises an individual's unique adjustment to life, including major traits, interests, drives, values, self-concept, abilities, and emotional patterns.* There is a broad range of theories and models, each with differing perspectives on particular topics when defining personality constructs (Corr & Matthews, 2009). Among them, several models have been developed based on different personality theories such as the Five-Factor Model (FFM) (McCrae and John, 1992; Costa and McCrae, 2008), the HEXACO (Lee and Ashton, 2004), and the Eysenck's Model (Eysenck, 1963; Carducci, 2015).

Among the first trait-based personality models, Eysenck (1963) created the PEN model, where personality divides into three dimensions: Psychoticism, Extraversion, and Neuroticism. However, many researchers in the 1980 s began to agree that terms of five broad, roughly independent dimensions were better fit to summarize personality variation. These five dimensions led to the creation of the FFM (McCrae and John, 1992). This model is a hierarchical organization of personality traits in five fundamental dimensions: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and

Conscientiousness. Each of the dimensions is divided then into six subdimensions (facets). It is usually referred to as the OCEAN model – the acronym of the five presented dimensions –, or simply by Big Five. Regarding the HEXACO model (Lee and Ashton, 2004), it shares four traits with the FFM and added two new personality traits. It is composed of Honesty-Humility (H), Emotionality (E), eXtraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O).

Several personality researchers agree that the FFM personality traits represent cross-cultural differences in normal behavior, and studies have replicated this taxonomy in a diversity of samples (Digman, 2003; Chamorro-Premuzic and Furnham, 2014). These findings led the FFM to be considered one of the theories that best represents the personality construct (Avia et al., 1995; Feldt et al., 2010) and dominate the personality research landscape (Terzis et al., 2012; Cruz et al., 2015). More importantly, researchers applied the FFM in education settings (Bergold and Steinmayr, 2018; Rivers, 2021), and the community considers this helpful model to predict both achievement and behaviors. Costa and McCrae (2008) derived the FFM from the lexical hypothesis, which claims that the model can express important personality characteristics succinctly, and the more relevant a trait is, the pertinent words for that trait will appear in the dictionary (Kortum and Oswald, 2018). The FFM consists of five general dimensions to describe personality as follows:

Neuroticism   is often referred to as emotional instability, addressing the tendency to experience mood swings and negative emotions such as anxiety, worry, fear, anger, frustration, envy, jealousy, guilt, depressed mood, and loneliness (Thompson, 2008). Highly neurotic people are more likely to experience stress and nervousness and are at risk for the development and onset of common mental disorders (Jeronimus et al., 2016). In contrast, people with lower neuroticism scores tend to be calmer and more self-confident, but, at the extreme, they may be emotionally reserved.

Extraversion   measures a person's imagination, curiosity, seeking of new experiences, and interest in culture, ideas, and aesthetics. People high on Openness to Experience tend to have a greater art appreciation, devise novel ideas, hold unconventional values, and willingly question authority (Costa and McCrae, 2008). On the contrary, those with low openness to experience tend to be more conventional, less creative, and more authoritarian.

Openness to Experience   measures a person's tendency to seek stimulation in the external world, the company of others, and to express positive emotions. Extroverts tend to be more outgoing, friendly, and socially active, which may explain why they are more prone to boredom when alone. Introverts are more comfortable in their own company and appreciate solitary activities such as reading, writing, using computers, hiking, and fishing.

Agreeableness   addresses how much a person focuses on maintaining positive social relations. On the one hand, people with high agreeableness scores tend to be friendly and compassionate but may find it difficult to tell the hard truth. On the other hand, disagreeable people show signs of negative behavior, such as manipulation and competing with others rather than cooperating.

Conscientiousness assesses the preference for an organized approach to life compared to a spontaneous one. People high on conscientiousness are more likely to be well-organized, reliable, and consistent. Individuals with low conscientiousness are generally more easy-going, spontaneous, and creative.

2.2 Peer Assessment Factors

There are several scopes through which we can observe the quality of the peer assessment and, thus, how personality may affect a selection of metrics that reflect such quality. For instance, in a peer assessment process, students are assigned an arbitrary amount of peer reviews to complete. We can measure two typical performance metrics in the peer assessment process. One approach to measuring how effective students are in their peer assessment is to calculate the percentage of the completed peer reviews per student from all assigned to them (*peer assessment efficacy*). Another promising performance variable to consider is how long students complete the peer assessment (*peer assessment efficiency*).

Previous research has extensively used the differences between grades provided by students and faculty members as metrics to assess the quality of the peer assessment environment (AlFallay, 2004; Kulkarni et al., 2013; Yan et al., 2022). We also consider it relevant to tackle how accurate and fair the grades given by students are. We decided to consider three variables for this assessment: (i) The *grade* that the student provides to a specific post, (ii) The difference between the grade that the student provided to a specific post and the grade resulting from weighing all the peer grades for that post (*grade difference from final*), and (iii) The difference between the grade that the student provided to a specific post and the grade the faculty provided for that post (*grade difference from faculty*). These three measures allow us to validate if there is a perceiver effect solely based on the grader or if it generalizes compared to the remaining grades for a post and the faculty assessment.

Finally, the additional constructive feedback students provide to their peers improves the quality of the peer assessment process (Dochy et al., 1999; Basheti et al., 2010). Not all students may provide feedback to their peers or lack the creativity to elaborate their feedback further than the primary, formal requirements of the submission. In order to analyze the feedback that students provide in their peer assessments, we decided to count the number of characters present in the feedback (*feedback length*). Although it does not allow us to assess the quality of the feedback itself, this approach provides a broad analysis of whether students were engaged in the process.

3 Related Work

We present a literature review related to our work.

3.1 Peer Assessment Validity and Reliability

Nowadays, practitioners use peer assessment as both an assessment and a learning tool. However, there are several drawbacks to consider that may affect the validity and accuracy of the assessment. For instance, experimental results depict that peer grades usually correlate with faculty-assigned grades (Liu et al., 2004), but the former may be slightly higher (Kulkarni et al., 2013). Furthermore, students may give lower grades than the faculty to the best-performing students (Sadler and Good, 2006), and peer assessment completed by undergraduate students may not be as reliable (Gielen et al., 2011). These effects usually stem from students having less grading experience than faculty. Educators have often relied on rubrics to counter this lack of experience, thus assisting assessors in judging the quality of student performance (Panadero and Jonsson, 2020). However, even when the

faculty lists specific criteria to support the assessment, students' different backgrounds and knowledge levels question the eligibility and grading accuracy (Kulkarni et al., 2013; Na and Liu, 2019), leading peer assessment itself, as a valid assessment form, to be sometimes challenged by course participants (Glance et al., 2013; Suen, 2014).

Student perceptions may also hinder the peer assessment process (Adachi et al., 2018; To and Panadero, 2019). In particular, student attitudes and perceptions directly affect the reliability of peer assessment since it is highly dependent on objective and honest evaluation. Papinczak et al. (2007) explored the social pressure that makes students hesitate to criticize their peers and score them honestly. Kaufman and Schunn (2011) reported that students often regard peer assessment as unfair and question their peers' ability to evaluate students' works. Hoang et al. (2016) also found that students may occasionally not fully invest in peer assessment activities by providing poor knowledge sharing in their feedback.

Other factors affecting the peer assessment validity and reliability are the assessment task complexity and the reviewer load. Tong et al. (2023) found that higher complexity assessment tasks had lower validity. The authors also showed that increasing reviewer load can decline or improve single-rater reliability depending on task complexity. Researchers can counteract this subjectiveness bias with several strategies such as anonymity, multiple reviews per peer (where the final peer grade is an average of each student grade), and training (Bostock, 2000; Glance et al., 2013; Hoang et al., 2022). Nevertheless, other individual factors, such as personality, are more pervasive and still under-researched (Chang et al., 2021; Rivers, 2021). We consider that personality may provide an accurate and deep understanding relevant to the design of peer assessment environments as it may affect the fairness, accuracy, and reliability of peer assessment (Vickerman, 2009).

3.2 Personality Factors in Peer Assessment Environments

Few studies focused on understanding the role of personality in the peer assessment environments (Chang et al., 2021; Rivers, 2021). Recently, Martin and Locke (2022) focused on the association between peer ratings, personality traits, and self-ratings. The authors leveraged the HEXACO personality model to test if those members' personality traits could predict the peer ratings of team members. Results showed that conscientiousness predicted higher peer ratings, suggesting practitioners may want to assign one highly conscientious person to every team in this setting.

In another example, using machine learning techniques, Cachero et al. (2022) studied the influence of personality and modality on peer assessment evaluation perceptions. The authors measured personality through the FFM. They found that agreeableness was the best predictor of peer assessment usefulness and ease of use, extraversion of compatibility, and neuroticism of intention of use. Furthermore, individuals with low consciousness scores may be more resistant to introducing peer assessment processes in the classroom. However, the value of peer assessment improves the positive feelings of those scoring high on neuroticism. These studies demonstrate that personality may bias how the assessor perceives the work of their peers (Martin and Locke, 2022) but also how one perceives the whole peer assessment process.

Besides these two factors, we hypothesize that personality may also affect how one conducts the peer grading process. For instance, students more prone to emotional instability may frequently be more anxious about being peer assessed and how to assess other students. Despite their influence on how individuals perceive user-generated content differently (Shao, 2009), little to no research focused on the dynamics of personality factors in

students' behavior in a peer assessment environment. Only AlFallay (2004) conducted a study to evaluate the effect of self-esteem, classroom anxiety, and motivational intensity personality traits on the accuracy of self- and peer-assessment in oral presentation tasks. The author found some exciting results, with learners possessing the positive side of a trait being more accurate than those with its negative side, except for students with high classroom anxiety. Nevertheless, the traits that AlFallay (2004) focused on are more relevant for oral, physical presentation settings rather than remote learning environments. Considering our domain, we believe that leveraging personality traits from well-researched models such as the FFM may provide a more substantial and broader understanding to help us integrate personality-based design guidelines in developing peer assessment mechanisms for remote learning.

4 Research Method

As we mentioned in the past sections, we need to carefully design a peer assessment environment to not only act as an effective formative assessment tool (Wanner and Palmer, 2018) but allow us to understand whether personality traits can bias the peer assessment process. This section presents the context of the remote learning course, our research question and hypotheses, research procedure, and data analysis.

4.1 Research Context

To accomplish the primary goal of this study, which is understanding whether personality factors affect the peer assessment environment in a remote learning course, we need to collect data from a course that supports both remote learning and peer assessment. In our study, we leverage a course named Multimedia Content Production (MCP) designed for MSc students in Information Systems and Computer Engineering. In particular, there is already an extensive body of research based on this course (e.g., Barata et al., 2016; Barata et al., 2017; Nabizadeh et al., 2021). As such, we believe that the MCP course provides a stable, well-designed course system where we can implement a peer assessment environment to study our research question.

In this edition of MCP, 69 students (86.25%) out of the 80 students that enrolled in MCP completed the course. Although MCP is traditionally a blended learning course, the COVID-19 pandemic led the faculty to run the course in a remote setting. In this case, the course setting has students attending theoretical lectures and practical laboratories through a videoconference platform. While in the theoretical lectures, students learn about different media formats (e.g., audio, video, and image) from an engineering standpoint (e.g., compression algorithms and encoding formats), the laboratory classes focus on creating high-quality media using the Processing programming language. In addition to these classes, students also join the discussions and complete online assignments via Moodle.[1]

---

[1] Moodle is a web-based Learning Management System (LMS) that allows faculty to provide and share, among other education materials, documents, graded assignments, and quizzes, with students (Al-Ajlan and Zedan, 2008) Research has shown how this technology is crucial for creating high-quality online courses, especially in higher education (Wharekura-Tini and Aotearoa, 2004; Cole and Foster, 2007). This system is free to download, use, modify, and distribute under the terms of GNU (Brandl, 2005). It can be found at www.moodle.org.

Fig. 1 Visual depiction of the Skill Tree used in MCP

Among the various educational components in MCP, students can earn grade points through a *Skill Tree*. The Skill Tree is a selection of learning activities students can complete autonomously during the semester. It is a precedence tree where each node represents a skill (see Fig. 1). Students start the semester with a set of unlocked skills that they can complete right away. Subsequent nodes can be unlocked if students complete a subset of anterior skills, thus acting as requirements. In order to complete a skill, students submit their work on Moodle. Then, a member of the faculty grades their work and provides feedback. The student completes the skill and earns some grade points if the grade is above a fixed threshold. If students receive a poor grade, they may try up to three times to resubmit their work and complete the skill.

This Skill Tree submission system is public, and every student in the course can see it. Moreover, it has been a constant feature in MCP where students invest significant effort in completing the course (Barata et al., 2016). As such, this component contains the appropriate characteristics to deploy a peer assessment environment.

4.2 Research Question and Hypotheses

Our work focuses on whether personality affects the peer assessment dynamics of a remote learning course. Thus, we formulate our research question as follows:

RQ: How does personality affect students' behavior in the peer assessment process of a semester-long remote learning course?

Since this metric is intrinsically related to each student's level of personal organization, we believe that the conscientiousness personality trait may have a mediator role regarding peer assessment efficacy. As we mentioned, conscientiousness suggests the self-use of socially prescribed restraints that facilitate goal completion, following norms and rules,

and prioritizing tasks. Personality Psychology research has highlighted how this trait is decisive in behavioral control. The general rule is that individuals with high levels of conscientiousness have a more robust task performance than those who are low on conscientiousness (Barrick and Mount, 1991; Witt et al., 2002), and these performance qualities transfer to other domains such as health (Booth-Kewley and Vickers, 1994; Courneya and Hellsten, 1998) and gaming (Liao et al., 2021). Therefore, we project that students with higher levels of conscientiousness will complete a more significant percentage of the peer assessments assigned to them than their counterparts. Thus, we formulate our first hypothesis as follows:

H1: Conscientiousness positively affects peer assessment efficacy.

Besides peer assessment efficacy, the conscientiousness trait may also influence students' time to complete a peer assessment. In particular, there is a similar effect compared to efficacy, i.e., individuals with higher levels of conscientiousness will complete their peer-assigned reviews faster than those low on conscientiousness. In other words, people with higher scores in conscientiousness will be more efficient, as they take less time to complete the process. As such, we formulate our second hypothesis as follows:

H2: Conscientiousness positively affects peer assessment efficiency.

The agreeableness trait may provide some an accurate and deep understanding regarding the grades students provide their peers since it distinguishes cooperation from competition. Recent research (Yee et al., 2011) has shown how this trait modulates behavior in game theory, with agreeable people preferring non-combat gameplays such as exploration and crafting and disagreeable individuals focused more on the competitive and antagonistic aspects of gameplay. Therefore, we believe that the personality trait of agreeableness will create a bias in the peer assessment process, acting as a moderator of interpersonal conflict (Jensen-Campbell and Graziano, 2001). In particular, we expect that agreeable students will provide higher scores to their peers, while more competitive individuals will assess submissions with lower grades. We formulate our third hypothesis as follows:

H3: Agreeableness positively affects peer assessment grades.

Finally, it is essential to leverage personality traits that explain how students interact with the outer world, i.e., their peers, in the overall peer assessment environment. As such, we focus on the traits of extraversion and openness to experience, as these constructs positively affect creativity (Sung and Choi, 2009; Filippi et al., 2017). We formulate our fourth hypothesis as follows:

H4: Extraversion and openness to experience positively affect the amount of feedback in peer assessment.

4.3 Data Collection Tools

We defined the variables leveraged in this study in Sect. 2. The personality variables of the FFM were collected with the (IPIP-BFM-50) (Goldberg, 1999; Goldberg et al., 2006; Oliveira, 2019) The International Personality Item Pool (IPIP) is a large-scale collaborative repository of public domain personality items[2] to measure personality constructs. In particular, the IPIP-BFM-50 provides several measures of the FFM personality traits, including the ones we target in this study. It has 50 items, with ten items for each trait. For each personality trait, we calculate responses by the sum of the Likert scales and direction of

---

[2] https://ipip.ori.org/index.htm (Retrieved in 2023/10/11 01:56:46).

scoring based on assertions semantically connected to behaviors and five possible alternatives of agreement between *very inaccurate* and *very accurate*.

Regarding the metrics to measure the peer assessment process, we created a Moodle plug-in to simulate the peer assessment environment. The Moodle plug-in is responsible for assigning the reviewers to a post and for storing the data. We can aggregate all peer reviews that students performed and compute for each student their (i) Peer assessment efficacy, (ii) Peer assessment efficiency, (iii) Average peer assessment grade, (iv) Average grade difference from final, (v) Average grade difference from faculty, and (vi) Average feedback length.

4.4 Research Procedure

The Skill Tree was open during the whole duration of the course. Whenever a student posts a submission for a skill in Moodle, the plug-in pseudo-randomly assigns five other students to peer review it. We follow a pseudo-random approach to even the number of assigned peer reviews per student. We removed students that gave up or showed no activity in the course at half the semester length from the pool of students that could peer grade.

Each reviewer has two days to grade an assigned submission. The reviewer must provide a grade from 0 to 5, where a higher grade represents better work and as much written feedback as they want. The peer assessment is single-anonymized, i.e., the reviewer knows which student they are grading, but the assessed student does not know who does it. We decided to follow this approach to allow reviewers to critique submissions without any influence exerted by the authors. Although the double-anonymized method offers more advantages (Tomkins et al., 2017), it was impossible to apply since students submit their work publicly, and it is available to the whole course before being graded. In that case, reviewers could check Moodle anytime and find out the author, which would render the double-anonymized method strictly as good as the single-anonymized approach. Furthermore, the faculty could grade the submission at any time. However, the grading and feedback would only be available either (i) after a minimum of three student reviewers completed the peer assessment or (ii) after the two days limit has passed since the original submission.

After either of these conditions is satisfied, the student who submitted some work could see the feedback and grades from both the faculty and the peers that reviewed their work (see Fig. 2). In particular, each submission presents the grade that the faculty attributed as well as a *weighted peer assessment grade*, which is the average grade from all the peer assessments of that post. In the example of Fig. 2, the weighted peer assessment grade is 3.6 based on the pool of peer assessments (3, 3, 4, 4, 4).

In addition, MCP contains a training feature where students can check and grade test examples for each skill to gain an initial level of knowledge necessary to assess their peers' submissions. This feature aims to minimize the subjectiveness bias related to knowledge levels (Kulkarni et al., 2013) and provide a scaffold for students to base their evaluations. A student must first complete the training of that skill to be eligible to grade a post. This enforcing strategy aims at minimizing rogue views.[3], which increase with the anonymous nature of the review as well as with the decreased feeling of community affiliation (Hamer et al., 2005; Lu

---

[3] Rogue reviews are insufficient reviews due to laziness, collusion, dishonesty, retaliation, competition, or malevolence (Reily et al., 2009; Glance et al., 2013).
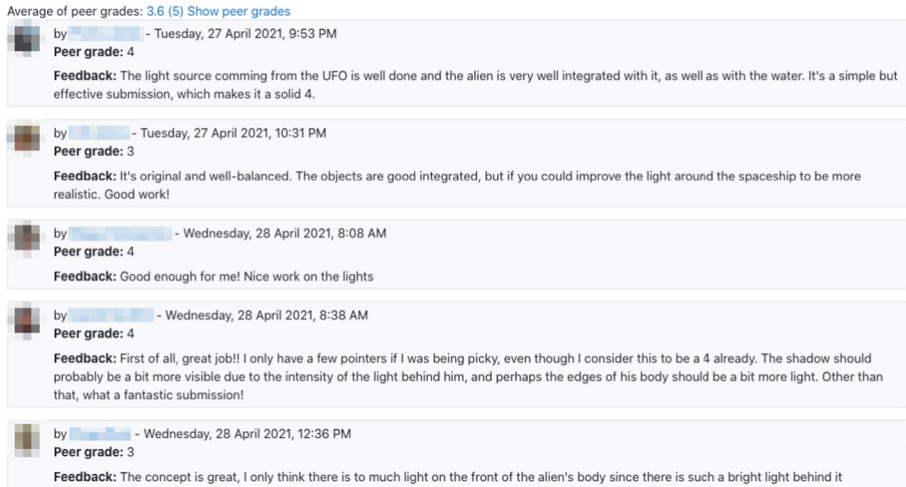
Fig. 2 Example of a container with the peer assessments for a given post

and Bol, 2007). By the end of the semester, we asked all native Portuguese speaker students to complete the Portuguese version of the IPIP-BFM-50. Finally, students received extra credit for participating in this experiment.

4.5 Data Analysis

First, we removed from the dataset all peer assessments that contained no feedback related to the post. We continued by aggregating all peer reviews that students performed that filled in the personality questionnaire and computed the peer assessment environment metrics. Then, we merged these data with the personality data to produce the final dataset. Our final dataset includes information from 806 posts and 2688 peer grades, resulting in an average of 3.33 peer assessments per post. Regarding our participants, the dataset contains personality information from 45 students.

Table 1 presents the descriptive statistics of our study variables, including a 95% confidence interval and a Shapiro-Wilk normality test. In addition, Fig. 3 illustrates the distribution of personality data used in our study. Emotional stability and extraversion present a more extensive interquartile range than the other traits. Moreover, we can observe two outliers in the openness to experience distribution. Although outliers may distort the statistical analysis, such a short number in our sample (4%) can be neglected.

We ran correlation methods to study the relationships between our quantitative values. We decided which method to run based on the Shapiro-Wilk test and preliminary visual inspection. We ran a Pearson's product-moment correlation if both quantitative values presented normal distributions and no significant outliers. In case data fails the assumption, we run Spearman's rank-order correlation.

Table 1 Descriptive statistics of the study variables

	Variable	Mean	STD	95% CI	p-value
Personality	Emotional Stability	27.980	1.200	[25.560; 30.400]	.044
	Extraversion	29.290	1.003	[27.270; 31.310]	.049
	Openness to Experience	36.890	0.677	[35.530; 38.250]	.042
	Agreeableness	39.440	0.702	[38.030; 40.860]	.225
	Conscientiousness	35.240	0.782	[33.670; 36.820]	.324
	Efficacy (%)	0.703	0.225	[0.636; 0.771]	$< .001$
	Efficiency (min)	33.305	16.872	[28.236; 38.373]	.004
Peer Assessment	Grade	3.604	0.314	[3.509; 3.698]	.291
	Grade Difference from Final	−0.023	0.223	[−0.090; 0.044]	.337
	Grade Difference from Faculty	0.457	0.326	[0.358; 0.555]	.385
	Feedback Length	175.328	83.928	[150.113; 200.543]	.148

"p-value" refers to the level of significance of a Shapiro-Wilk normality test. Variables with a non-normal distribution have gray rows
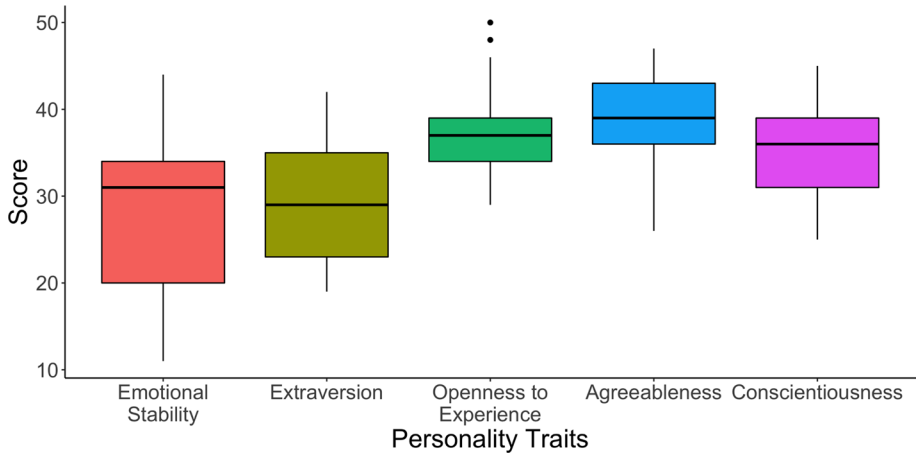
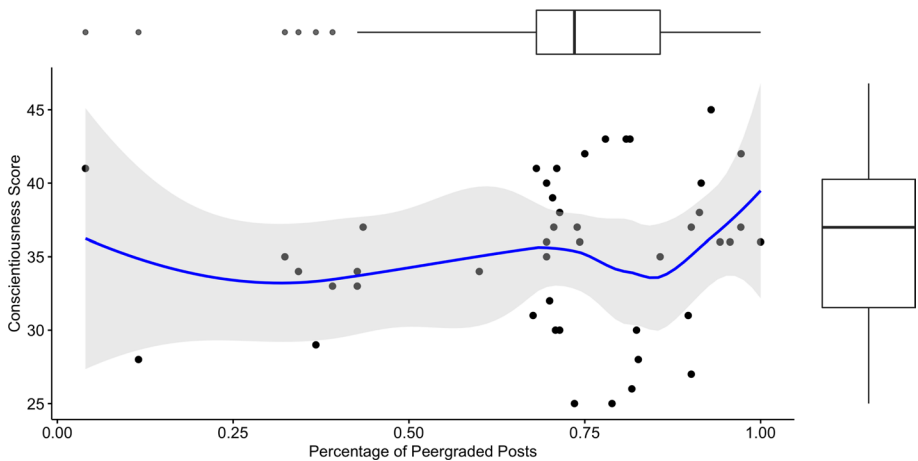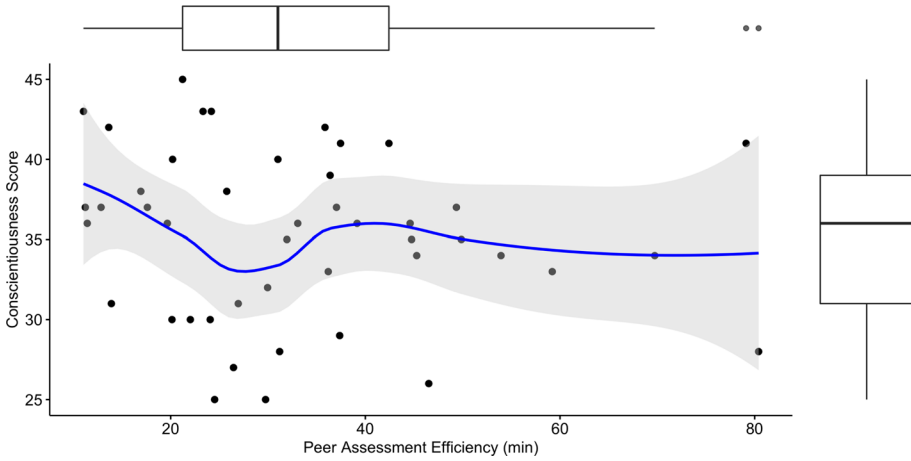**Fig. 3** Distributions of personality traits of the FFM from the sample

Fig. 4 Scatterplot showing the relationship between peer assessment efficacy (x axis) and conscientiousness scores (y axis)

5 Results

This section describes the results of the experiment. It starts by analyzing the effect of personality on peer assessment efficacy and efficiency and continues with the grades and the feedback that students provide their peers. We also answer the research question and state the limitations of our study.
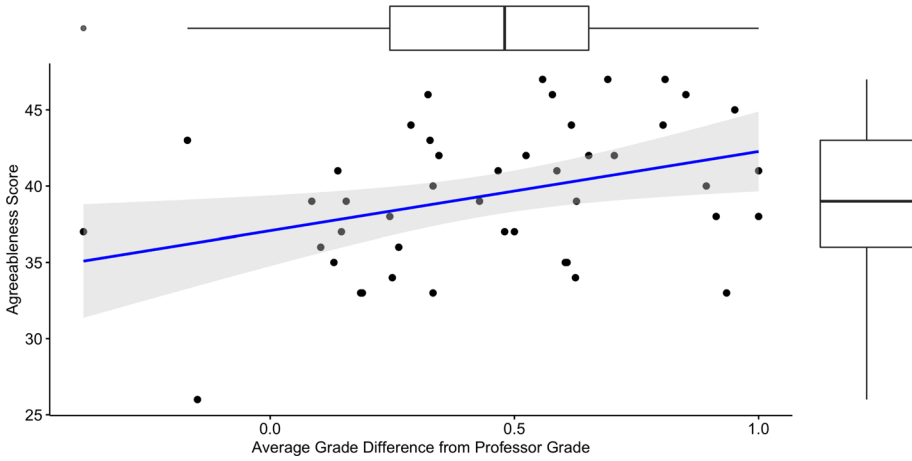
Fig. 5 Scatterplot showing the relationship between peer assessment efficiency (x axis) and conscientiousness scores (y axis)
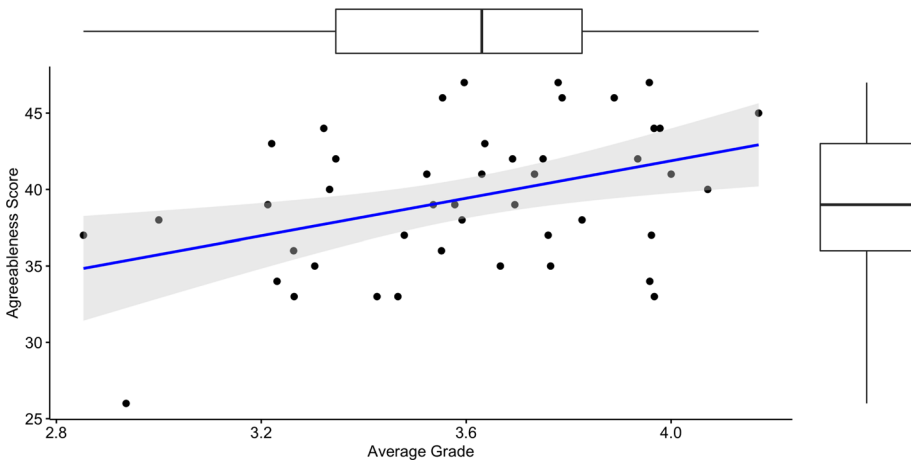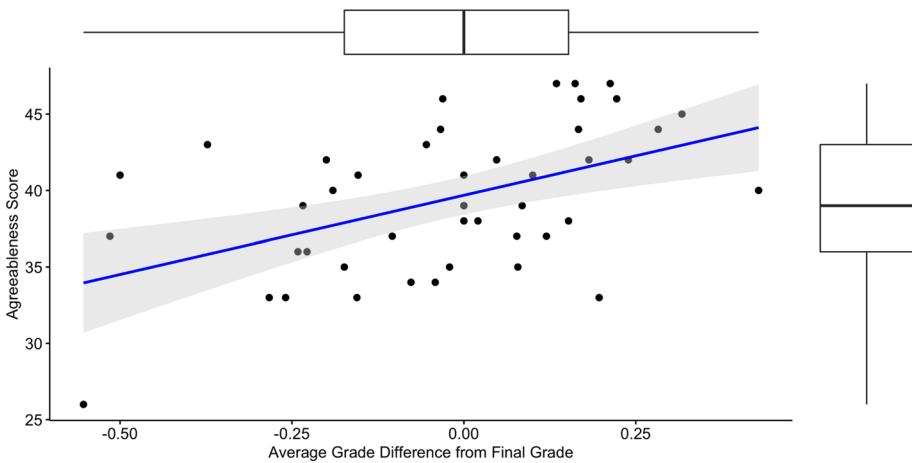

Fig. 6 Scatterplot showing the relationship between peer assessment grades (x axis) and agreeableness scores (y axis)

5.1 Peer Assessment Efficacy

As we mentioned, peer assessment efficacy refers to the percentage of assigned peer reviews completed by each student. By checking Table 1, we can see that peer assessment efficacy shows a non-normal distribution. Additionally, a preliminary analysis showed that the relationship is not linear, as assessed by visual inspection of a scatterplot (Fig. 4). We can observe that the right-hand half of locally estimated scatterplot smoothing (LOESS) line (Jacoby, 2000) presents a distorted sine wave without a general trend in the data. As such, we ran a Spearman's rank-order correlation that showed a statistically non-significant, very weak positive correlation between peer assessment

Fig. 7 Scatterplot showing the relationship between the average grade difference from the faculty grade (x axis) and agreeableness scores (y axis)
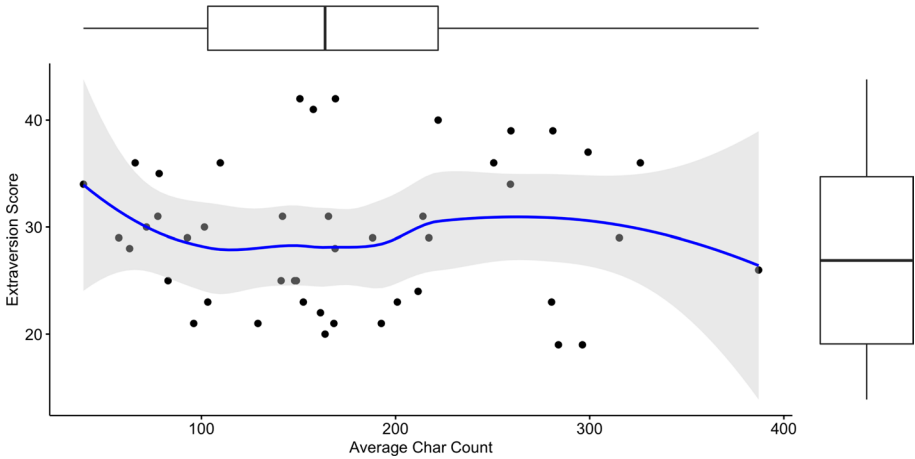

Fig. 8 Scatterplot showing the relationship between the average grade difference from the final grade (x axis) and agreeableness scores (y axis)
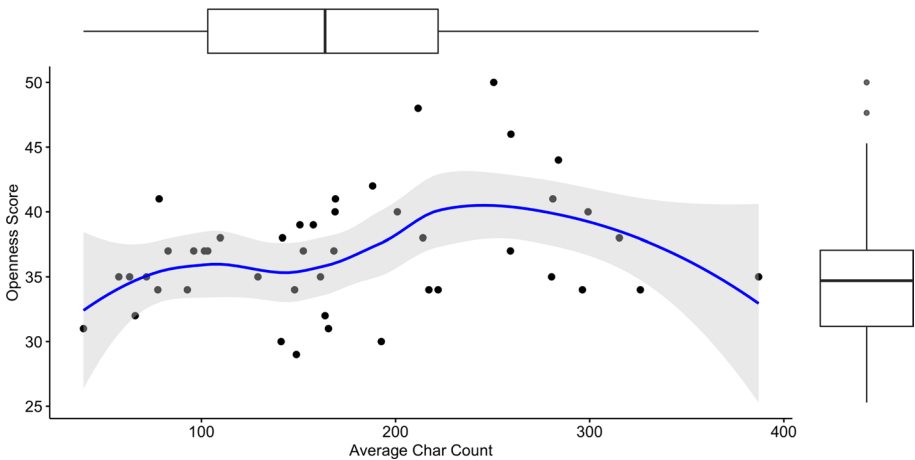
efficacy and conscientiousness scores, $r_s(45) = .196, p = .198$. In this light, we cannot accept H1, since conscientiousness does not affect peer assessment efficacy.

5.2 Peer Assessment Efficiency

Next, we want to verify whether the same personality trait influences the peer assessment efficiency of each student, i.e., the average time students take to complete a peer assessment. In particular, students are more efficient as their average minutes to complete a peer assessment decreases. Besides peer assessment efficiency not showing a normal distribution, a preliminary visual inspection of the scatterplot (Fig. 5) showed a non-linear relationship. In particular, the LOESS line is quite a symmetric instance of the peer assessment
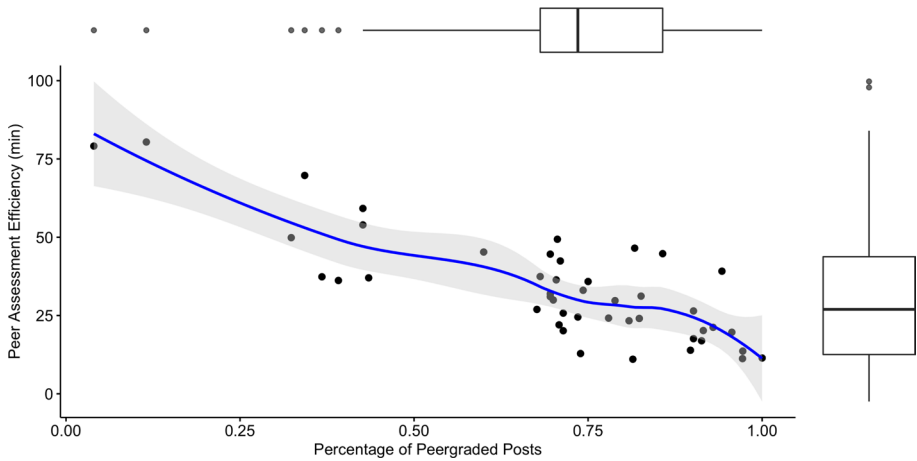
Fig. 9 Scatterplot showing the relationship between average characters in feedback (x axis) and extraversion scores (y axis)

Fig. 10 Scatterplot showing the relationship between average characters in feedback (x axis) and openness to experience scores (y axis)

efficacy's line, which may hint at a correlation between peer assessment efficiency and efficacy. Indeed, Spearman's rank-order correlation showed a statistically non-significant, weak negative correlation between peer assessment efficiency and conscientiousness scores, $r_s(45) = -.216, p = .154$. We cannot accept H2 as well since conscientiousness does not affect peer assessment efficiency.

Fig. 11 Scatterplot showing the relationship between peer assessment efficacy (x axis) and peer assessment efficiency (y axis)

5.3 Peer Assessment Grades

Regarding the grades that students provide to their peers, we started by exploring how the level of agreeableness models these grades without considering any other factor. All variables of interest show a normal distribution. Moreover, a preliminary analysis through visual inspection showed the relationship to be linear and without any outliers (Fig. 6). Therefore, we opted for a Pearson's product-moment correlation, which showed a statistically significant, weak positive correlation between peer assessment grades and agreeableness scores, $r_s(45) = .358, p = .016$. Therefore, we accept H3, since we found that agreeableness is positively associated with peer assessment grades.

We decided to follow up our analysis by considering two different baselines regarding the grades that students provide to their peers. Given that the distributions of average differences from the final and the faculty's grades also presented normal distributions (Table 1), we decided to continue applying Pearson's product-moment correlation method. First, we checked for each student whether agreeableness also regulated the average difference between the grade that faculty attributes to a post and those that students give to that post (Fig. 7). Again, we found a statistically significant, moderate positive correlation between these variables, $r_s(45) = .360, p = .015$. In addition, we found a statistically significant, weak positive correlation between the average difference for the final grade of the post and the student's grades, and agreeableness scores, $r_s(45) = .490, p = .001$. See Fig. 8 for a visual inspection of this relationship. Both these results support the acceptance of H3, since agreeableness consistently led students to rate their assigned peer reviews differently independently of the baseline.

5.4 Peer Assessment Feedback

Finally, we want to check whether personality affects the feedback students provide in their reviews. Since extraversion and openness to experience did not present a normal distribution (see Table 1), we opted for a Spearman correlation method

to tackle these relationships. In addition, we found in a preliminary analysis through visual inspection that both relationships presented a sine-based correlation (Figs. 9 and 10). We started by finding a statistically non-significant, very weak negative correlation between the amount of feedback that students provide and extraversion scores, $r_s(45) = .017, p = .913$. In contrast, there was a statistically significant, weak positive correlation between the amount of feedback that students provide and openness to experience scores, $r_s(45) = .299, p = .046$. As such, we cannot accept H4, taking into account that, although openness to experience is positively associated with peer assessment feedback, extraversion is not.

5.5 Additional Findings

As mentioned in Sects. 5.1 and 5.2, results hinted at a relationship between peer assessment efficacy and efficiency. As both variables have a non-normal distribution (see Table 1), we used a Spearman correlation to study their relationship. A statistically significant, strong negative correlation existed between peer assessment efficacy and efficiency, $r_s(45) = -.717, p < .001$. This result shows that people that complete their assigned peer reviews do it faster than their counterparts (Fig. 11).

6 Discussion

In this section, we answer our research question by summarizing the significant results, deriving design implications, and discussing the limitations of this work.

6.1 Answering the Research Question

Results have shown that personality traits can bias the peer assessment process. The most noteworthy effect was how the agreeableness trait modulated the grades of the peer assessments. As mentioned, this psychological construct measures the disposition to maintain positive social relations (Halko and Kientz, 2010). As expected, people leaning more towards helping others assessed their peers' submissions, on average, with higher grades than people more prone to compete. This raises a severe drawback of the overall reliability of the peer assessment process. If we consider the faculty's grade as the true score, students' grades are less likely to depend on the value of the submission and more on the grader. Although most grade differences are not at a grade point value, the pool of reviewers containing students with similar agreeableness scores may exacerbate the grading disparity. For instance, the grading system becomes less fair, considering that if the majority of the reviewers have high scores on agreeableness, the peer assessment grade may be higher than it was supposed to be. We also conducted another correlation test to compare the peer assessment grades with a different baseline. The association of agreeableness was still present when we compared the average difference between a post's weighted peer assessment grade and the one a student provides. As we mention, the accuracy of peer grades is often questioned (AlFallay, 2004; Gielen et al., 2011; Kulkarni et al., 2013; Yan et al., 2022) and some students believe it to be unfair because of these grade discrepancies (Kaufman and Schunn, 2011). Our results confirm previous work regarding the peer

grades being slightly higher than those given by the professors (Kulkarni et al., 2013). Our findings show that agreeableness is associated and may be responsible for biasing the peer assessment.

Openness to experience also showed a significant association on the feedback length students provide to their peers. In particular, a sin-based shape effect reflects an increase of feedback characters in higher values of openness. This result is in line with our expectations since people high on openness to experience tend to have a greater appreciation for art (Dollinger, 1993; Rawlings and Ciancarelli, 1997) as well as for new or unusual ideas (Halko and Kientz, 2010). Moreover, it shows that our feature to allow the students to train their peer assessment and which type of feedback they should provide did not affect the effort that students applied in the overall process. These results may shed more light on the findings from previous work since several studies reported a lack of engagement from the students in the peer assessment. Our findings suggest that openness to experience may play a significant role in this relationship.

Nevertheless, some personality traits we investigated did not affect the peer assessment. Regarding extraversion, research has shown that extroverts desire social attention and a tendency to display positivity (Bowden-Green et al., 2020). In addition, extroverts are also more likely to use social media, spend more time using one or more social media platforms, and regularly create content (Bowden-Green et al., 2020). In our case, we found contrasting results, with extraversion having no significant effect on the feedback length students provided each other. Taking into account the remote setting of the course and the features that Moodle has in common with a social media platform, we expected that extroverts would be more forthcoming in their feedback. We hypothesize that the lack of a significant relationship may arise for the other pole of extraversion, based on the remote setting, and the ability to participate in the course without much exposure may have led introverts to be as motivated as extroverts to provide feedback to their peers.

Moreover, in contrast to other researchers (Huels and Parboteeah, 2019; Joyner et al., 2018), conscientiousness, the degree to which one prefers an organized life or a spontaneous one (Halko and Kientz, 2010; Morizot, 2014), did not have a significant effect on the student's behaviors. In particular, we found that this trait had no significant effect on the efficacy or the efficiency of the peer assessment, although results hinted towards weak negative correlations. There are several explanations for these results. For instance, the faculty periodically reminded the students in the theoretical classes and Moodle to complete their peer assessments, which provided exterior stimuli to the students' behavior patterns and led them to complete the peer assessments independently of the conscientiousness score. Another factor may be the reward that students received for completing peer assessments. In either case, these factors were a normalizing strategy that tuned down the conscientiousness effect. Nevertheless, our findings show that agreeableness and openness to experience play a role in peer assessment. As such, practitioners must devise strategies to reduce the polarizing effect of personality traits to promote a stable and reliable peer assessment environment.

6.2 Design Implications

Based on our results, we devised a set of implications that can be useful for the pipeline of peer assessment systems, particularly for learning and training. *Reminders to complete peer assessment assignments are positive.* Results suggest a non-significant relationship between conscientiousness scores and efficiency and efficacy in the peer

assessment process. Although people with high scores in conscientiousness are more likely to complete their assignments on time, we cannot expect the same from people with lower scores. Tooltips in the form of reminders that appear after a certain period of the assignment may help the latter individuals, accompanied by a reminder in class and grade incentives to complete the peer assignments. These tooltips may contain a set of targeted reminder texts since state-of-the-art research has shown how different personality types have distinct preferences for persuasive messages (Halko and Kientz, 2010; Anagnostopoulou et al., 2018).

Personality may jeopardize the fairness and reliability of the peer assessment. Although the previous strategy of persuasive periodic reminders to prompt an impartial grading process may also work for the agreeableness effect, there are other strategies to keep in mind. For instance, the system can also provide examples next to the peer assessment input screen to prompt the students to look at them and refresh their reviewing skills with the baseline set by the faculty. Another example is that when the system picks the students that should grade a post, this selection can leverage agreeableness to balance the number of agreeable and disagreeable students in the subset. This balancing strategy may reduce the effect of agreeableness by weighting the same grades that tend to undervalue and overvalue a submission.

The peer assessment system should empower closed-to-experience individuals. Since these individuals tend to be more conventional and less creative, peer assessment systems should provide features that allow people with lower scores in openness to experience to provide feedback on their peers' submissions. For instance, the system can contain predefined fields covering a range of criteria that the student can quickly fill in. The student can provide complete feedback by leveraging auto-generated sentences based on their input options.

6.3 Limitations

Although we found exciting trends in our results, some essential factors may explain the lack of significance observed in some of our results. For instance, MOOC students are strangers to their peers. In our case, the MCP course is part of a Master's degree, which inadvertently increases the chance of the students being already familiar with each other. As such, they may have created social bonds, inevitably influencing peer grading with some bias. As previously discussed, a double-anonymized assessment could suppress this effect. However, the number of students in the course needs to be more significant to guarantee that the sample of reviewers and the assessed student does not talk with one another while the review process is active. This approach would produce a flawed double-anonymized process. Moreover, the number of participants in this experience could have been more significant, which would provide conclusions with a more substantial impact and allowed us to investigate whether the relationship between the participants or any combination of values from personality factors of both participants affected the interactions. Another bias to consider is that our sample mainly comprises Portuguese individuals, which may render a cultural bias in our results.

7 Conclusions and Future Work

Our work aimed to investigate whether personality traits from the FFM influence how students perform their peer assessment in a semester-long online course. The results show that the personality traits of agreeableness and openness to experience are associated with how students evaluate or provide feedback to their peers. Therefore, these results indicate that in the context of distance learning, peer assessment in terms of reliability and fairness may be compromised by student personality traits.

Future work includes investigating whether some personality constructs with finer granularity, such as the facets from the FFM, play a role in this context. Indeed, facets such as trust and cooperation from the agreeableness trait, or assertiveness and friendliness from the extraversion trait, could play a decisive role in peer-grading. Other psychological constructs may also be interesting to consider in this type of setting, such as creativity (Amabile, 2018) and the Locus of Control (LoC) (Rotter, 1954, 1966). In addition, we would like to conduct another study with a larger sample. Another factor to consider is the remote setting. Therefore, we would like to study whether our results can be applied to in-person courses or blended learning environments. Finally, our findings can be applied in the design pipeline of peer assessment systems as well as in systems that simulate training and education of people, such as GIMME (Gomes et al., 2019). Furthermore, we should investigate how to account for inexperienced graders in a fair manner. Indeed, experience suggests that grading patterns evolve with experience and knowledge of subject matter. Leveraging how personality constructs affect the dynamics of these environments can empower researchers to increase the expressiveness of their models while taking a more human approach to integrate people into learning environments.

Data Availability statement Data is available as supplemental material.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability Not applicable.

Authors contribution statement TA: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing—Original Draft, Visualization, Funding acquisition. FS: Methodology, Software, Validation, Investigation, Writing—Review & Editing. SG: Methodology, Writing—Review & Editing, Supervision, Funding acquisition. JJ: Methodology, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. DG: Methodology, Writing—Review & Editing, Supervision, Project administration, Funding acquisition.

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adachi, C., Tai, J.H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education, 43*(2), 294–306.

Al-Ajlan, A., & Zedan, H. (2008). Why moodle. In: *2008 12th IEEE International Workshop on Future Trends of Distributed Computing Systems*, IEEE, pp. 58–64.

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.

Amabile, T. M. (2018). *Creativity in context: Update to the social psychology of creativity*. Routledge.

An, D., & Carr, M. (2017). Learning styles theory fails to explain learning and achievement: Recommendations for alternative approaches. *Personality and Individual Differences, 116*, 410–416.

Anagnostopoulou, E., Urbančič, J., Bothos, E., Magoutas, B., Bradesko, L., Schrammel, J., & Mentzas, G. (2018). From mobility patterns to behavioural change: Leveraging travel behaviour and personality profiles to nudge for sustainable transportation. *Journal of Intelligent Information Systems, 54*, 157–178.

Avia, M., Sanz, J., Sánchez-Bernardos, M., Martínez-Arias, M., Silva, F., & Graña, J. (1995). The five-factor model-ii. relations of the neo-pi with other personality variables. *Personality and Individual Differences, 19*(1), 81–97.

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2016). Early prediction of student profiles based on performance and gaming preferences. *IEEE Transactions on Learning Technologies*, *9*(3), 272–284. https://doi.org/10.1109/TLT.2016.2541664

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2017). Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior*, *71*, 550–585. https://doi.org/10.1016/j.chb.2016.08.049

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1–26.

Basheti, I. A., Ryan, J., Woulfe, J., & Bartimote-Aufflick, K. (2010). Anonymous peer assessment of medication management reviews. *American Journal of Pharmaceutical Education, 74*(5), PMC42.

Bergold, S., & Steinmayr, R. (2018). Personality and intelligence interact in the prediction of academic achievement. *Journal of Intelligence, 6*(2), 27.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education), 21*, 5–31.

Booth-Kewley, S., & Vickers, R. R., Jr. (1994). Associations between major domains of personality and health behavior. *Journal of Personality, 62*(3), 281–298.

Bostock, S. (2000). Student peer assessment. *Learning Technology, 5*(1), 245–249.

Bowden-Green, T., Hinds, J., & Joinson, A. (2020). How is extraversion related to social media use? a literature review. *Personality and Individual Differences, 164*, 110040.

Brandl, K. (2005). Review of are you ready to "moodle"? *Language Learning & Technology, 9*(2), 16–23.

Cachero, C., Rico-Juan, J. R., & Macià, H. (2022). Influence of personality and modality on peer assessment evaluation perceptions using machine learning techniques. *Expert Systems with Applications, 29*, 119150.

Carducci, B. J. (2015). *The psychology of personality: Viewpoints, research, and applications* (3rd ed.). Wiley.

Chamorro-Premuzic, T., & Furnham, A. (2014). *Personality and intellectual competence*. Psychology Press.

Chan, H. P. & King, I. (2017). Leveraging social connections to improve peer assessment in moocs. In: *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 341–349.

Chang, C.-Y., Lee, D.-C., Tang, K.-Y., & Hwang, G.-J. (2021). Effect sizes and research directions of peer assessments: From an integrated perspective of meta-analysis and co-citation network. *Computers & Education, 164*, 104123.

Chang, S.-C., Hsu, T.-C., & Jong, M.S.-Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers & Education, 146*, 103758.

Chou, S.-W., & Liu, C.-H. (2005). Learning effectiveness in a web-based virtual learning environment: A learner control perspective. *Journal of Computer Assisted Learning, 21*(1), 65–76.

Cole, J., & Foster, H. (2007). *Using Moodle: Teaching with the popular open source course management system*. O'Reilly Media Inc.

Corr, P. J., & Matthews, G. (2009). *The Cambridge handbook of personality psychology*. Cambridge University Press Cambridge.

Costa, P., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE Handbook of Personality Theory and Assessment, 2*, 179–198.

Courneya, K. S., & Hellsten, L.-A.M. (1998). Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. *Personality and Individual differences, 24*(5), 625–633.

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

Dhawan, S. (2020). Online learning: A panacea in the time of covid-19 crisis. *Journal of Educational Technology Systems, 49*(1), 5–22.

Digman, J. M. (2003). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Dillenbourg, P., Schneider, D., Synteta, P., et al. (2002). Virtual learning environments. In: *Proceedings of the 3rd Hellenic Conference Information & Communication Technologies in Education*, pp. 3–18.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education, 24*(3), 331–350.

Dollinger, S. J. (1993). Research note: Personality and music preference: Extraversion and excitement seeking or openness to experience? *Psychology of Music - PSYCHOL MUSIC, 21*, 73–77.

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*, 481–509.

Eysenck, H. J. (1963). The biological basis of personality. *Nature, 199*, 1031–4.

Feldt, R., Angelis, L., Torkar, R., & Samuelsson, M. (2010). Links between the personalities, views and attitudes of software engineers. *Information and Software Technology, 52*(6), 611–624.

Filippi, S., Barattin, D., et al. (2017). Evaluating the influences of heterogeneous combinations of internal/external factors on product design. In *DS 87-8 Proceedings of the 21st International Conference on Engineering Design* (ICED 17) Vol 8: Human Behaviour in Design, Vancouver, Canada, 21-25.08.2017, pp. 001–010.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education, 7*(2), 95–105.

Gernsbacher, M. A. (2015). Why internet-based education? *Frontiers in Psychology, 5*, 1530.

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137–155.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351.

Gilbert, J. A., & Flores-Zambada, R. (2011). Development and implementation of a "blended" teaching course environment. *MERLOT Journal of Online Learning and Teaching, 7*(2), 244–260.

Glance, D. G., Forsey, M., & Riley, M. (2013). The pedagogical foundations of massive open online courses. First monday.

Goldberg, L. R., et al. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*(1), 7–28.

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

Gomes, S., Dias, J., and Martinho, C. (2019). Gimme: Group interactions manager for multiplayer serious games. In *2019 IEEE Conference on Games (CoG)*, IEEE, pp. 1–8.

Graham, C. R. (2013). Emerging practice and research in blended learning. *Handbook of Distance Education, 3*, 333–350.

Halko, S. & Kientz, J. A. (2010). Personality and persuasive technology: An exploratory study on health-promoting mobile applications. In *International Conference on Persuasive Technology*, Springer, pp. 150–161.

Hamer, J., Ma, K. T., & Kwong, H. H. (2005). A method of automatic grade calibration in peer assessment. In *Proceedings of the 7th Australasian Conference on Computing*, education-Volume 42, pp. 67–72.

Hoang, L. P., Arch-Int, S., Arch-Int, N., et al. (2016). Multidimensional assessment of open-ended questions for enhancing the quality of peer assessment in e-learning environments. In *Handbook of Research on Applied e-Learning in Engineering and Architecture Education*, IGI Global, pp. 263–288.

Hoang, L. P., Le, H. T., Van Tran, H., Phan, T. C., Vo, D. M., Le, P. A., Nguyen, D. T., and Pong-Inwong, C. (2022). Does evaluating peer assessment accuracy and taking it into account in calculating assessor's final score enhance online peer assessment quality? Education and Information Technologies, pp. 1–29.

Huels, B., & Parboteeah, K. P. (2019). Neuroticism, agreeableness, and conscientiousness and the relationship with individual taxpayer compliance behavior. *Journal of Accounting and Finance, 19*(4), 453. https://doi.org/10.33423/jaf.v19i4.2181

Jacoby, W. G. (2000). Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies, 19*(4), 577–613.

Jensen-Campbell, L. A., & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality, 69*(2), 323–362.

Jeronimus, B., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders: A meta-analysis on 59 longitudinal/prospective studies with 443,313 participants. *Psychological Medicine, 46*, 2883–2906.

Joyner, C., Rhodes, R. E., & Loprinzi, P. D. (2018). The prospective association between the five factor personality model with health behaviors and health behavior clusters. *Europe's Journal of Psychology, 14*(4), 880.

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science, 39*, 387–406.

Kortum, P., & Oswald, F. L. (2018). The impact of personality on the subjective assessment of usability. *International Journal of Human-Computer Interaction, 34*(2), 177–186.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI), 20*(6), 1–31.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the hexaco personality inventory. *Multivariate Behavioral Research, 39*(2), 329–358. PMID: 26804579.

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? a meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193–211.

Li, K., Xu, B., Gao, K., Yang, D., & Chen, M. (2018). Self-paced learning with identification refinement for spoc student grading. In *Proceedings of ACM Turing Celebration Conference-China*, pp. 79–84.

Liao, G.-Y., Cheng, T., Shiau, W.-L., & Teng, C.-I. (2021). Impact of online gamers' conscientiousness on team function engagement and loyalty. *Decision Support Systems, 142*, 113468.

Liu, E.Z.-F., Yi-Chin, Z., & Yuan, S.-M. (2004). Assessing higher-order thinking using a networked portfolio system with peer assessment. *International Journal of Instructional Media, 31*(2), 139.

Liu, J., & Sadler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes on l2 writing. *Journal of English for Academic Purposes, 2*(3), 193–227.

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning, 6*(2), 100–115.

Mahanan, M. S., Talib, C. A., & Ibrahim, N. H. (2021). Online formative assessment in higher stem education; A systematic literature review. *Asian Journal of Assessment in Teaching and Learning, 11*(1), 47–62.

Martin, C. C. & Locke, K. D. (2022). What do peer evaluations represent? A study of rater consensus and target personality. In Frontiers in Education, Frontiers, volume 7, pp. 746457.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

Moreno, J., & Pineda, A. F. (2020). A framework for automated formative assessment in mathematics courses. *IEEE Access, 8*, 30152–30159.

Morizot, J. (2014). Construct validity of adolescents' self-reported big five personality traits: Importance of conceptual breadth and initial validation of a short measure. *Assessment, 21*(5), 580–606.

Morris, N. P. (2014). *How Digital Technologies*. ERIC: Blended Learning and MOOCs Will Impact the Future of Higher Education.

Murray, D. E., McGill, T. J., Toohey, D., & Thompson, N. (2017). Can learners become teachers? evaluating the merits of student generated content and peer assessment. *Issues in Informing Science and Information Technology, 14*, 21–33.

Na, J. and Liu, Y. (2019). A quantitative revision method to improve usability of self-and peer assessment in moocs. In *Proceedings of the ACM Turing Celebration Conference*, China, pp. 1–6.

Nabizadeh, A. H., Jorge, J., Gama, S., & Gonçalves, D. (2021). How do students behave in a gamified course?—A ten-year study. *IEEE Access*, *9*, 81008–81031. https://doi.org/10.1109/ACCESS.2021.3083238

Núñez-Peña, M. I., Bono, R., & Suárez-Pellicioni, M. (2015). Feedback on students' performance: A possible way of reducing the negative effect of math anxiety in higher education. *International Journal of Educational Research, 70*, 80–87.

Oliveira, J. P. (2019). Psychometric properties of the portuguese version of the mini-ipip five-factor model personality scale. *Current Psychology, 38*(2), 432–439.

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: Examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education, 44*(3), 379–397.

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review, 30*, 100329.

Panadero, E., Romero, M., & Strijbos, J.-W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation, 39*(4), 195–203.

Papinczak, T., Young, L., & Groves, M. (2007). Peer assessment in problem-based learning: A qualitative study. *Advances in Health Sciences Education, 12*, 169–186.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in moocs. arXiv preprint arXiv:1307.2579.

Price, E., Goldberg, F., Patterson, S., & Heft, P. (2013). Supporting scientific writing and evaluation in a conceptual physics course with calibrated peer review. In *AIP Conference Proceedings*, American Institute of Physics, volume 1513, pp. 318–321.

Puška, E., Ejubović, A., Dalić, N., & Puška, A. (2021). Examination of influence of e-learning on academic success on the example of Bosnia and herzegovina. *Education and Information Technologies, 26*, 1977–1994.

Rawlings, D., & Ciancarelli, V. (1997). Music preference and the five-factor model of the neo personality inventory. *Psychology of Music, 25*(2), 120–132.

Reily, K., Finnerty, P. L., and Terveen, L. (2009). Two peers are better than one: aggregating peer reviews for computing assignments is surprisingly accurate. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pp. 115–124.

Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education, 41*(2), 301–315.

Rivers, D. J. (2021). The role of personality traits and online academic self-efficacy in acceptance, actual use and achievement in moodle. *Education and Information Technologies, 26*(4), 4353–4378.

Rod, R., Joshua, W., Melissa, P., Dandan, C., & Adam, J. (2020). Modeling student evaluations of writing and authors as a function of writing errors. *Journal of Language and Education, 6*(22), 147–164.

Rotter, J. B. (1954). *Social learning and clinical psychology*. Prentice-Hall, Inc.

Rotter, J. B. (1966). General expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*, 1–28.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.

Sangwin, C. J., & Köcher, N. (2016). Automation of mathematics examinations. *Computers & Education, 94*, 215–227.

Schmidt, J. A., O'Neill, T. A., & Dunlop, P. D. (2021). The effects of team context on peer ratings of task and citizenship performance. *Journal of Business and Psychology, 36*, 573–588.

Shao, G. (2009). Understanding the appeal of user-generated media: A uses and gratification perspective. Internet research.

Suen, H. K. (2014). Peer assessment for massive open online courses (moocs). *International Review of Research in Open and Distributed Learning, 15*(3), 312–327.

Sung, S. Y., & Choi, J. N. (2009). Do big five personality factors affect individual creativity? the moderating role of extrinsic motivation. *Social Behavior and Personality: An International Journal, 37*(7), 941–956.

Terzis, V., Moridis, C. N., & Economides, A. A. (2012). How student's personality traits affect computer based assessment acceptance: Integrating bfi with cbaam. *Computers in Human Behavior, 28*(5), 1985–1996.

Thompson, E. R. (2008). Development and validation of an international english big-five mini-markers. *Personality and Individual Differences, 45*(6), 542–548.

To, J., & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year under-graduates. *Assessment & Evaluation in Higher Education, 44*(6), 920–932.

Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences, 114*(48), 12708–12713.

Tong, Y., Schunn, C. D., & Wang, H. (2023). Why increasing the number of raters only helps sometimes: Reliability and validity of peer assessment across tasks of different complexity. *Studies in Educational Evaluation, 76*, 101233.

Tuah, N. A. A., & Naing, L. (2021). Is online assessment in higher education institutions during covid-19 pandemic reliable? *Siriraj Medical Journal, 73*(1), 61–68.

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education, 34*(2), 221–230.

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education, 43*(7), 1032–1047.

Wharekura-Tini, H. & Aotearoa, K. (2004). Technical evaluation of selected learning management systems. Master's thesis, Catalyst IT Limited, Open Polytechnic of New Zealand.

Wheeler, S., Yeomans, P., & Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology, 39*(6), 987–995.

Witt, L., Burke, L. A., Barrick, M. R., & Mount, M. K. (2002). The interactive effects of conscientiousness and agreeableness on job performance. *Journal of Applied Psychology, 87*(1), 164.

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A pairwise and network meta-analysis. *Educational Research Review, 2022*, 100484.

Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011). Introverted elves & conscientious gnomes: the expression of personality in world of warcraft. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 753–762.

Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(3), 372–386.