# The Role of Personality in Preference for Different Types of Academic Assessment

Juraj Takács

s4731077

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 2324_2a_08

Supervisor: prof. dr. Rob Meijer

Second evaluator: dr. Stacey M. Donofrio

In collaboration with: Silke Kok, Hage Niggendijker, Matsen Piepers, Annika Selesnew, Zhang Ziying

June 30, 2024

**Acknowledgements**

*A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned*.

I acknowledge the use of QuillBot (https://quillbot.com/) to assist with improving my writing for comprehension and grammar, without changing my original structure of the work. The AI was not used to generate novel ideas.

No content generated by AI technologies has been presented as my own work.

No prompts were used for the assisted writing and the utilization of AI was based upon thorough personal consideration of what constitutes a concise academic writing without resorting to plagiarism.

**Abstract**

Learning is a fundamental aspect of life, necessitating thorough exploration to optimize efficiency. Assessment is a critical element of the learning process. Contemporary education utilizes a variety of assessment methods, necessitating the identification of the most effective ones to align with changing academic standards. The present study examines potential influencing factors and student preferences for various assessment types (performance tasks, open-ended exams, and multiple-choice). We posited that students would prefer multiple-choice exams. Furthermore, we investigated the rationale for this preference. In light of the existing evidence that suggests a correlational relationship, we also examined the relationship between the Big Five personality traits and assessment preferences. We administered a survey to 128 university students at the University of Groningen in the Netherlands to assess their personality traits and preferred assessment formats. The results confirmed our hypothesis: students preferred multiple-choice exams over other exam types because of their higher objectivity, positive valence, and lower perceived difficulty. The assessment preferences were minimally influenced by personality traits, with only the trait of open-mindedness exhibiting a significant correlation. Our results suggest a consistent preference for multiple-choice exams. The significance of incorporating personality factors in academic assessments is called into question by the apparent negligible influence of personality traits on assessment preferences. Theoretical and practical implications, as well as limitations and suggestions for future research, are addressed.

*Keywords:* assessment, multiple-choice exams, open-question exams, performance tasks, the Big Five Inventory, personality, familiarity

**The Role of Personality in Preference for Different Types of Academic Assessment**

Despite the fact that assessment is widely despised, its importance in promoting learning is undeniable (Bjork, 1975). In the contemporary era, the ability to master a variety of life domains is contingent upon the ability to learn effectively. As a result, there is an increasing emphasis on the optimization of learning processes. Recent trends have sought to reconceptualize the role of tests, making them essential to the learning process rather than solely evaluating learning (Dochy et al., 1997). This paradigm shift has incited inquiries regarding the optimal allocation of testing resources and the variables that influence its efficacy.

At present, the majority of academic institutions prioritize the use of multiple-choice (MC) testing formats, which necessitate selecting the appropriate response from a list of alternatives, and essay testing formats, also referred to as open-ended formats (OQ exams), which enable test-takers to compose their own responses (Gronlund, 1976). The goal of this study is to investigate students' views on these assessment methods. It is crucial to acknowledge that the efficacy of any assessment format can fluctuate substantially due to the design of the questions (Murphy et al., 2023; Schuwirth et al., 1996; Zeidner, 1987).

Nevo (1985) stresses the importance of emphasizing students' perspectives, as their perspectives are instrumental in establishing face validity and offer valuable insights into the various factors that may affect the effectiveness of assessments. These factors are critical for ensuring high-quality education, which may include motivation for exam preparation, classroom interactions, and preparation methods. Nevertheless, the current body of research on the influence of students' preferences is still limited, frequently indicating a widespread preference for MC examinations (Zeidner, 1987). Furthermore, students frequently regard OQ examinations as more indicative of their actual knowledge (Zeidner, 1987).

The present study aims to investigate whether these preferences persist and identify potential predictors of assessment preferences. Personality has been a persistent focus of study in academic environments due to its large impact on variations in performance (Piedmont et al.,

1999; You et al., 2021). Research has demonstrated that personality traits such as openness to experience and conscientiousness are linked to academic achievement and performance (Furnham & Chamorro-Premuzic, 2004; You et al., 2021). It is logical to infer that this correlation also applies to academic preferences. Distinctive individual characteristics can suggest an inclination towards specific academic methodologies, as they are linked to educational aspirations and interests (Gasser et al., 2004). Ackerman (1999) and Chamorro-Premuzic et al. (2005) have produced evidence that personality may influence preferences for particular categories of exams. In their 2005 study, Chamorro-Premuzic et al. discovered that students who score high on openness to experience tend to prefer OQ examinations, while those who score low on neuroticism exhibit a preference for these exams over MC formats. These findings suggest that personality traits may be one of the factors that improve the alignment between students as test-takers and educators. Highlighting this aspect can facilitate the adjustment of testing procedures to align with students' preferences, consequently boosting motivation and cultivating a favorable learning atmosphere. Evidence substantiates this concept, showing enhanced performance when assessment categories align with student characteristics (Whittle et al., 2018). While personality is not the sole determinant, its significance in academic environments is unquestionable. Currently, there is only a limited amount of research regarding the correlation between personality traits and student preferences. In addition, the evaluation methods analyzed have been restricted to OQ and MC exams. Therefore, the present study is centered on broadening the range of methods for assessment and addressing this deficiency in current understanding by elucidating the relationship between personality and assessment preference. This will help inform the construction of tests and educational practices.

We adopt Murphy et al.'s (2023) definition of assessment as "a tool for learning encompassing both formal and informal activities that prompt students to answer questions related to course content". The present study investigates OQ exams, MC exams, and performance tasks (PT). The latter, defined as tasks requiring demonstration of knowledge,

understanding, and proficiency through projects, essays, or presentations, is particularly relevant to our sample of university students (see Method section), as it is occasionally used in their courses. We employed Birenbaum's (1994) Assessment Preferences Inventory (API) in conjunction with a translated inventory developed by Lindner et al. (2018). We utilized the condensed version of the Big Five Inventory (Soto et al., 2017) to evaluate personality. The inventory evaluates individual distinctions in neuroticism, extroversion, open-mindedness, agreeableness, and conscientiousness, all of which have been validated and are employed effectively in academic environments (McCrae et al., 1987; Mammadov, 2021).

Our research investigates several questions derived from the studies of Chamorro-Premuzic et al. (2005) and Zeidner (1987). We know that in general, students prefer MC over OQ exams; however, it remains unclear what the relationship is with personality traits. Continuously, we included the less common form of assessment, PT, to investigate the added effect on preference. Additionally, we hypothesize that prior experience with assessment methods may affect the relationship between personality and assessment preferences. Despite the largely unexplored impact of prior experience on academic preferences, Liao et al. (2011) found that familiarity was associated with increased liking for certain items. Furthermore, potential underlying mechanisms that underpin familiarity could evoke a sense of truth, leading to the perception of familiar information as credible and preferred (Pan & Hu, 2024). In accordance with this idea, we hypothesize that familiarity may evoke a preference for familiar assessment formats, with the most prevalent type of assessment demonstrating the highest influence of familiarity. In contrast, Teigen (1987) discovered a preference for novelty in a familiar situation, providing some evidence for the contrary effect. In general, the existing research on familiarity in academic preference is constrained to broad conclusions. Therefore, the additional purpose of this study is to determine if familiarity can act as a mediator in the connection between personality and academic preference. Following this research, we speculate that this relationship can also be found in academic settings. The anticipated findings are as follows:

**Hypothesis 1**: There will be a higher general preference for the multiple-choice test format than for the open-question exams and performance tasks.

**Hypothesis 2**: Students higher on the trait of openness will show a decreased preference for the multiple-choice test format.

**Hypothesis 3**: Students lower on the trait of neuroticism will show an increased preference for open-question exams.

**Hypothesis 4**: Students higher on the trait of extroversion will show an increased preference for performance tasks.

**Hypothesis 5**: Prior experience will influence the strength of the relationship between personality traits and preference for an assessment method.

## Method

### Sample

The present study was conducted with students from the faculty of Behavioral and Social sciences at the University of Groningen. The original sample consisted of 143 students, but 12 students failed to complete the survey. Throughout the survey, five attention checks were incorporated. Three participants failed to pass two or more attention checks. If someone failed to complete the survey or the attention checks, that participant was removed from the final data set. This resulted in a sample size of 128 students, which was used for the data analysis.

There were 37.5% Psychology (EN), 61.7% Psychology (NL) and 0.8% Sociology students. For practical reasons, we limited our sample to students from this faculty. Of the sample, 72.7% were female, 23.4% were male, 3.1% were non-binary/third gender, and 0.8% preferred not to say. Moreover, 35.2% of the students were in their second year of studying or higher, while there were 64.8% first-year students. A higher number of first-year students was expected since the sampling was mostly done through Sona Systems (https://www.sona-systems.com/).

**Incentive**

As an incentive to participate in the study, first-year Psychology students received 0.6 course credits through Sona Systems. These credits are used to pass the first-year course 'A Practical Introduction to Research Methods'. Students in higher years did not receive an incentive for their participation.

**Procedure**

Participants were recruited through convenience sampling. The survey made available through Sona Systems. Furthermore, a link to the study was shared in several group chats on WhatsApp to obtain more participants who were further along in their studies. Researchers were not physically present when participants filled out the survey, and participants were asked to complete the survey independently. The survey was available from April 16th, 2024, to April 24th, 2024.

Before filling out the questionnaire, participants were informed about the content and aim of the research, their choice to participate or refrain from participating, the incentive for completing the questionnaire, and the confidentiality of handling the data. The researchers' contact information was provided in case of any questions about the research. Subsequently, participants were required to fill out the informed consent form about participating in the study and processing their data.

The questionnaire started with demographic questions, followed by questions measuring several constructs. Due to the collaborative nature of the study, some of these constructs are not relevant to the current research and will only be mentioned briefly. For further information on these sections, please see the theses written by collaborators mentioned in the headline.

After the demographic questions, students were presented with questions assessing their personality type and levels of trait test anxiety. Participants were then asked to indicate their preferences for different exam formats: OQ exams, MC exams, and PT (both individual and group). Subsequently, prior experience and learning strategies for the different formats were

assessed. Next, respondents answered questions regarding their levels of state evaluation anxiety for each examination method. The order in which these different sections were shown was randomized for each participant. After completing the questionnaire, participants were directed to the Sona Systems website to receive credits, if applicable.

**Ethics**

Before recruiting participants, we obtained ethics approval from the Ethical Committee (EC-BSS). Based on a checklist developed by the EC-BSS at the University of Groningen, the study was exempt from full ethical review.

**Materials**

First, participants were required to provide demographic information, namely gender, year of study, and subject of study.

**Overall preference**

To assess the preferences for different examination methods, an Examination Preference Inventory (EPI) was developed (see Appendix A), which was presented separately for each type of examination.

The inventory consisted of four scales, each consisting of two to four items, where participants indicated their level of agreement with the displayed statements on a five-point Likert scale. Three scales were based on an existing inventory by Lindner et al. (2018): 1) potential to show performance, 2) objectivity, and 3) valence. The scale measuring potential to show performance was directly adopted from Lindner's research, where it consisted of four items. An example item of this scale is "Open question exams/MC exams/Performance tasks allow me to express my knowledge precisely." Perceived objectivity was assessed through two items, one adopted directly from Lindner et al. (2018) and another added to enhance the scale's reliability. Three items were included to measure participants' valence (liking) of the different exam formats. One item's phrasing was adjusted to prevent misinterpretations, and one was added. An example item is "Open question exams/MC exams/Performance tasks should be the

main method of examination." The fourth scale measured difficulty and included two items adapted from Zeidner (1987) and one more item added for construct validity.

The means of the different subscales were combined to form the general preference score. Reliability was estimated using Cronbach's alpha and proved adequate (larger than .7) for all scales. Moreover, the reliability of all scales combined as a measure of preference was also adequate (see Table 1).

**Table 1**

*Cronbach's Alpha Scores for the EPI scale and subscales*

| Measure | Open Question Exams | Multiple-Choice Exams | Performance Tasks |
|---|---|---|---|
| | α | α | α |
| Difficulty | .78 | .74 | .81 |
| Potential to show performance | .80 | .71 | .71 |
| Objectivity | .91 | .81 | .91 |
| Valence | .76 | .74 | .69 |
| Total | .72 | .75 | .71 |

*Note.* The variables difficulty, potential to show performance, objectivity, and valence were measured on a bipolar five-point Likert scale.

**Personality**

To assess personality the Big Five Inventory-2 Short Form (BFI-2-S) was included in the survey. The domain scales in our study were proven to be reliable for extraversion (α = .79), agreeableness (α = .71), conscientiousness (α = .78), negative emotionality (α = .84), and open-

mindedness (α = .74). All domain scales had six items. We measured the traits across 30 questions, asking participants to indicate how much the displayed characteristics apply to them on a five-point Likert scale, ranging from disagree strongly to agree strongly.

### Familiarity

To assess prior experience, a modified scale of familiarity by Leidner et al. (2018) was used. The participants were asked to indicate how familiar they were with specific assessment methods (see Appendix A). Familiarity with each assessment method was assessed through three items. The reliability of the scales for MC exams (α = .63) and OQ exams (α = .63) were found to be questionable; however, because the item-rest correlations were larger than .2 and because content-wise it was difficult, there were no reasons to remove an item. We proceeded with the analyses, but the results should be interpreted with care. Reliability for the PT scale was found to be reliable (α = .75).

## Results

### General Preference

As was hypothesized, a higher general preference EPI score was found for multiple-choice exams ($M$ = 3.45, $SD$ = .54) in comparison to OQ exams ($M$ = 3.01, $SD$ = .46) and PT ($M$ = 3.00, $SD$ = 0.48). Supporting the first hypothesis, a paired sample t-test between MC exams and PT showed that preference for MC exams was significantly higher than for PT $t(127) = 6.16$, $p < .001$, $d = 0.54$, 99.5% CI [0.25, 0.65][1,2]. Also, students prefer MC exams over OQ exams, $t(127) = 6.21$, $p < .001$, $d = 0.55$, [0.24, 0.63] (see Footnotes 1 and 2). The effect sizes showed moderate effects ($d > 0.37$). The difference between the means of PT ($M$ = 3.00, $SD$ = 0.48) and OQ exams ($M$ = 3.01, $SD$ = .46) was negligible, and significance testing was thus not conducted.

---

[1] The assumptions of independence between subjects, same-subject paired measurements, and normal distribution of differences between pairwise comparisons were met. Some outliers were found; therefore, a Kruskal-Wallis test would be an alternative for the current research. However, this statistical technique is not part of the curriculum. Therefore, paired-sample t-tests were still run. The results must be interpreted with caution.

[2] Due to the increased risk of a type I error when conducting multiple statistical tests, we corrected the original alpha value of 0.05 according to the Bonferroni correction. This means that the critical p-value was corrected to 0.006 when taking into account the t-tests for the subscales later in this section.

**Subscales of General Preference**

The means of the four subscales for each examination format can be found in Table 2. It can be seen that MC exams were rated more positively on objectivity and valence than the other two formats. Moreover, MC exams were rated as the least difficult. OQ exams were rated as the most difficult, but they also had the most potential to show one's performance. As for PT, this format was placed between the other two formats on all scales except objectivity, where it scored the lowest.

Based on the mean differences, we decided to conduct paired-sample t-tests for objectivity, difficulty, and potential to show performance between selected exam formats (see Footnotes 1 and 2).

The objectivity scores for MC exams were significantly higher compared to PT, $t(127) = 16.562$, $p < .001$, $d = 1.46$, 99.5% CI [1.51, 2.13], and compared to OQ exams. $t(127) = 13.870$, $p < .001$, $d = 1.23$, [1.11, 1.68]. Similarly, the differences in difficulty between MC exams and PT and MC exams and OQ exams were significant, $t(127) = 4.204$, $p < .001$, $d = .37$, [0.11, 0.53] and $t(127) = 13.559$, $p < .001$, $d = 1.20$, [0.90, 1.36], respectively. PT and OQ exams were rated as more difficult. The last paired sample t-tests tested the differences in potential to show performance for MC exams and OQ exams, and MC exams and PT. OQ exams and PT were rated significantly higher on that subscale compared to MC exams, $t(127) = 7.37$, $p < .001$, $d = .65$, [0.52, 1,16] and $t(127) = 5.427$, $p < .001$, $d = 0.48$, [0.29, 0.91], respectively. The effect showed medium to large effects.

**Table 2**

*Means and Standard Deviations of the Different Scales of the EPI*

| Measure | Multiple Choice | | Open Question | | Performance Tasks | |
|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Difficulty | 2.95 | 0.84 | 4.08 | 0.61 | 3.27 | 0.29 |
| Potential to show performance | 3.14 | 0.83 | 3.98 | 0.75 | 3.75 | 0.72 |
| Objectivity | 4.49 | 0.73 | 3.09 | 0.97 | 2.67 | 0.99 |
| Valence | 3.55 | 0.87 | 2.75 | 0.86 | 3.00 | 0.80 |

**Personality and Familiarity**

The data suggest that people from our sample scored highest on the domains of agreeableness ($M = 3.81$, $SD = 0.62$), and open-mindedness ($M = 3.66$, $SD = 0.73$), and there is a very high level of familiarity with MC exams ($M = 3.45$, $SD = 0.54$) in comparison with other assessment methods.

The correlations for the personality traits were significant only in the case of openness-mindedness and overall MC exam preference ($r = -0.19$, $p < .05$), although only indicating a weak relationship (Table 3). Other correlations were not significant enough to warrant further exploration.

**Table 3**

*Means, Standard Deviations, and Correlations of the Overall Preference, BFI-2-S and Familiarity*

| | *M* | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. | 11. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Overall MC preference | 3.45 | 0.54 | - | -.25** | -.32** | -.03 | .08 | -.05 | -.10 | -.19* | .21* | .05 | .01 |
| 2. Overall OQ preference | 3.01 | 0.46 | -.25** | - | .30** | .01 | .03 | .12 | -.09 | .15 | .07 | .07 | -.05 |
| 3. Overall PT preference | 3.00 | 0.48 | -.32** | .30** | - | .10 | .00 | .10 | -.04 | .05 | -.02 | .02 | .26** |
| 4. Extraversion | 3.13 | 0.76 | -.03 | .01 | .10 | - | .05 | .15 | -.26** | .22* | -.08 | -.03 | .17 |
| 5. Agreeableness | 3.81 | 0.62 | .08 | .03 | .00 | .05 | - | .23** | -.12 | .02 | .22* | .10 | .26** |
| 6. Conscientiousness | 3.18 | 0.79 | -.05 | .12 | .10 | .15 | .23** | - | -.16 | -.05 | .00 | .10 | .04 |
| 7. Negative emotionality | 3.23 | 0.86 | -.10 | -.09 | -.04 | -.26** | -.12 | -.16 | - | .10 | .19* | .05 | -.07 |
| 8. Open-mindedness | 3.66 | 0.73 | -.19* | .15 | .05 | .22* | .02 | -.05 | .10 | - | -.02 | -.05 | .08 |
| 9. Familiarity MC | 4.52 | 0.59 | .21* | .07 | -.02 | -.08 | .22* | .00 | .19* | -.02 | - | -.14 | .04 |
| 10. Familiarity OQ | 3.18 | 0.94 | .05 | .07 | .02 | -.03 | .10 | .10 | .05 | -.05 | -.14 | - | .07 |
| 11. Familiarity PT | 3.13 | 0.84 | .01 | -.05 | .26** | .17 | .26** | .04 | -.07 | .08 | .04 | .07 | - |

*Note.* All Pearson's correlations are based on the $N = 128$.

*$p < .05$, **$p < .01$.

A series of three multiple regressions were performed[3], to replicate the findings of Chamorro-Premuzic et al. (2005) and to further investigate for each exam type whether the preference is predicted by the Big Five domain scales (second, third, and fourth hypothesis). For a complete model of coefficients for multiple regression analysis, see Table 4. The results of multiple regression analysis for the preference for MC exams show that personality domains explained 5.7% of the variance ($R^2$ = .057, $F$(5,122) = 1.47, $p$ = 0.2). Only the trait of open-mindedness was found to be significant under $p < 0.05$ (β = -.14, $p$ = .045, 99% CI [-0.31, 0.04])[4]. The results of multiple regression analysis for the preference for OQ exams suggest that personality domains explained 1.3% of the variance ($R^2$ = .013, $F$(5,122) = 1.33, $p$ = .26). Open-mindedness was found to be the only significant trait under $p < 0.05$ (β = .12, $p$ = .047, [-0.04, 0.27]). The multiple regression analysis for the preference for PT yielded results of 1.8% explained variance ($R^2$ = .018, $F$(5,122) = 0.457, $p$ = .807). No trait was found to be significant at $p < 0.05$. It is important to mention that the results are not significant under the Bonferroni correction of $p < 0.01$.

**Moderator Analysis**

In accordance with hypothesis five, we investigated whether familiarity moderated the relationship between personality traits and preference for an assessment method. We followed the suggestions discussed in Murphy (2021). The correlations between the moderator and preference for MC exams (r = .21), OQ exams (r = .26), and PT (r = .07) were found to be reasonably small enough. However, the correlations between the dependent variable (preference for an assessment type) and independent variables (the Big Five scales) were found to be smaller than the dependent variables and moderator (familiarity) for all three examination types (see Table 3). Due to this violation, there was no reason to suspect a moderator effect of familiarity, and we did not further test this relationship.

---

[3] The assumptions of linearity, homoscedasticity, independence, and normality of residuals were met.

[4] Bonferonni-corrected 99% CI was used due to the increased type I error.

**Table 4**

*Hierarchical Multiple Regression: Big Five Personality traits as predictors of preference for assessment methods*

| | Multiple Choice | | | | | | Open Question | | | | | | Performance Task | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE* | β | 99% CI | | *p* | *B* | *SE* | β | 99% CI | | *p* | *B* | *SE* | β | 99% CI | | *p* |
| | | | | *LL* | *UL* | | | | | *LL* | *UL* | | | | | *LL* | *UL* | |
| Constant | 4.02 | 0.49 | | 2.75 | 5.29 | <.001 | 2.74 | 0.42 | | 1.64 | 3.84 | <.001 | 2.68 | 0.44 | | 1.53 | 3.83 | <.001 |
| E | -0.00 | 0.07 | -0.00 | -0.18 | 0.18 | .98 | -0.05 | 0.06 | -0.08 | -0.20 | 0.10 | .39 | 0.05 | 0.06 | 0.08 | -0.11 | 0.21 | .44 |
| A | 0.08 | 0.08 | 0.10 | -0.12 | 0.29 | .29 | -0.01 | 0.07 | -0.01 | -0.19 | 0.17 | .92 | -0.02 | 0.07 | -0.03 | -0.21 | 0.17 | .78 |
| C | -0.07 | 0.06 | -0.10 | -0.23 | 0.10 | .30 | 0.07 | 0.05 | 0.12 | -0.07 | 0.21 | .19 | 0.06 | 0.06 | 0.09 | -0.09 | 0.20 | .32 |
| N | -0.06 | 0.06 | -0.09 | -0.21 | 0.10 | .34 | -0.06 | 0.05 | -0.11 | -0.19 | 0.07 | .23 | -0.01 | 0.05 | -0.01 | -0.14 | 0.13 | .93 |
| O | -0.14 | 0.07 | -0.19 | -0.31 | 0.04 | .05 | 0.12 | 0.06 | 0.18 | -0.04 | 0.27 | .05 | 0.02 | 0.06 | 0.04 | -0.14 | 0.18 | .71 |

*Note.* Total $N = 128$. CI = confidence interval; LL = lower limit, UL = upper limit. E = Extraversion, A = agreeableness, C = conscientiousness, N = negative emotionality, O = open-mindendess.

## Discussion

Following the current research on the importance of assessment preferences in a student environment, we revisited the question of students' examination preferences. We base this study on the idea that understanding students' perspectives can enhance the academic environment and learning efficacy (Whittle et al., 2018). Despite the widespread support for the multiple-choice format of examination, there is a significant lack of investigation into additional interactive factors (Zeidner, 1987). We conducted an analysis of the role of personality in assessment preferences to address this gap in contemporary knowledge, which was supported by prior research (Ackerman, 1999; Chamorro-Premuzic et al., 2005).

This study examined the general preference for assessment formats among university students and their relationship with the Big Five personality traits. The assessment formats included the three most common methods at the university level. Consistent with Zeidner (1987), our findings showed a general preference for multiple-choice exams, supporting our first hypothesis. We mainly attribute this preference to higher objectivity, positive valence, and lower perceived difficulty. Furthermore, students gave higher ratings to performance tasks and open-question exams for their ability to demonstrate performance. This implies that students generally favor assessments that are more objective and have a lower level of difficulty than those that enable them to demonstrate more control over the material they have learned.

The variations illustrated in the preference subscales suggest a subjective diversity in the required knowledge for each type of test. While past research demonstrated differences in types of knowledge that are applicable to academic settings (De Jong & Ferguson-Hessler, 1996), this connection with the field of assessments remains largely unexplored. Research indicates that there is a distinction between procedural and conceptual knowledge in the terminology used in academic fields (Star & Stylianides, 2013). Therefore, it is recommended to recognize these differences in various assessment formats in order to enhance the learning process. Consequently, we suggest a potential differentiation among assessment formats based on their

capacity to evaluate different categories of knowledge. While not grounded in research, our results suggest a multiple-choice style may be more suitable for courses that demand a substantial amount of factual information. On the other hand, courses that measure the students' understanding, especially in topics where the display of originality is assessed, might gain more from performance activities or open-ended question forms.

Our findings provided limited support for a link between personality and academic preference. The only significant correlation indicated that individuals scoring higher on openness-mindedness showed a lower preference for multiple-choice exams. Our results supported this further, as open-mindedness significantly predicted a preference for multiple-choice assessment. This result supports our second hypothesis and is consistent with Chamorro-Premuzic et al. Other personality traits were not significant predictors of assessment preference, thus providing no support for our third and fourth hypotheses.

Our partial replication of Chamorro-Premuzic et al. (2005) suggests that open-mindedness has the most predictive ability in assessment preference, raising questions about how to best utilize this finding. We hypothesize that individuals who are open-minded, prone to empirical thinking, and open to challenging their thoughts (King et al., 1996) may prefer more open ways to interpret findings and thus dislike the limited expression opportunities in multiple-choice examinations. This supports the notion that individuals who score higher on the trait of open-mindedness enjoy (and perform well in) the humanities, arts, and other creative fields because they are fundamentally poetic and artistic (Costa & McCrae, 1992). Therefore, a multiple-choice examination may not be well-suited for them. Despite the fact that the effect was not substantial, this discovery is worthy of further investigation.

Additionally, our findings suggest that personality traits are less significantly associated with preferences for assessment than prior experience. Thus, our hypothesis concerning the role of familiarity as a moderator in the relationship between personality traits and assessment preference should be reevaluated and investigated as a direct correlational relationship. Further

examination of our current results and replication findings is required due to the overall disparity between our findings and the research of Chamorro-Premuzic et al. (2005). The efficacy of personality in an academic environment should be evaluated with caution, given the careful consideration of methodology in the present study.

**Limitations, Strengths and Future Directions**

This study had several limitations. Firstly, the low reliability of familiarity scales. Secondly, due to the questionnaire's limited size and collaboration with other researchers, a shorter variant of the personality inventory was implemented. Thirdly, the sample was homogeneous, consisting primarily of first-year University of Groningen students from the Faculty of Behavioral and Social Sciences in the Netherlands. With this in mind, future studies should use different personality measures and explore familiarity as an independent link to assessment preferences. To further generalize these findings, it is advisable to use a more diverse student sample.

An additional factor to consider is the lack of analysis regarding personality facets. In particular, the distinction between openness to experience and its narrower traits has been previously investigated, and the results support this distinction by demonstrating differential validity for organizational outcomes (Woo et al., 2013). Furthermore, additional research on the facets of conscientiousness (Chernyshenko, 2002) indicates that the predictive validity of conscientiousness is enhanced when facets are taken into account (MacCann et al., 2009). Following our results, it would be beneficial to replicate the current personality findings on the level of facets.

We acknowledge the strengths of the study, particularly the larger sample size of university students. Secondly, the use of valid and reliable measurements with the exception of familiarity. Thirdly, the research adhered to the ethical guidelines, specifically the inclusion of informed consent, ensuring the confidentiality and safety of the participants. In addition, the study retained a low drop-out rate of participants.

We trust that these results will provide additional insight into the academic environment and its assessment formats. The perceived objectivity of multiple-choice exams is a strong indicator that they will continue to remain popular. Personality appears to play only a minor role in this relationship, if at all. We suggest that when assessment formats are designed to evaluate inventive and empirical thinking, it is important to consider the capacity to demonstrate performance, as evidenced by the lower value in multiple-choice exams.

**Conclusion**

In summary, university students exhibit a greater preference for multiple-choice examinations than for performance tasks and open-question examinations. This preference is linked to lower perceived difficulty, positive valence, and higher perceived objectivity. The current data indicates that personality has a negligible influence on assessment preferences; however, the trait of open-mindedness may warrant additional investigation. The objective of this study was to improve the utility of academic assessment by identifying robust predictors of academic preference. Despite not supporting all initial hypotheses, prior experience shows promise in predicting academic assessment preferences. Additionally, further exploration of personality is warranted. Student perspectives are aligned with multiple-choice examinations, which facilitate efficient learning. We discovered that the potential to demonstrate performance is more pronounced in performance tasks and open-question assessments than in multiple-choice examinations. Consequently, there is support for a distinction between assessments that require different types of knowledge. In courses that require extensive factual knowledge, a multiple-choice format may be more viable. In contrast, courses that assess the quality of students' knowledge, particularly in subjects where the demonstration of creativity is evaluated, may benefit more from performance tasks or open-ended question formats. Aligning students' perspectives with practice is expected to initiate more promising academic learning.

## References

Ackerman, P. L. (1999). Traits and knowledge as determinants of learning and individual differences: Putting it all together. In *American Psychological Association eBooks* (pp. 437–462). https://doi.org/10.1037/10315-019

Birenbaum, M. (1994). Toward adaptive assessment — The student's angle. *Studies in Educational Evaluation*, *20*(2), 239–255. https://doi.org/10.1016/0191-491x(94)90011-6

Bjork, R. A. (1975). Retrieval as a memory modifer. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Lawrence Erlbaum Associates.

Chamorro-Premuzic, T., Furnham, A., Dissou, G., & Heaven, P. C. L. (2005). Personality and preference for academic assessment: A study with Australian University students. *Learning and Individual Differences*, *15*(4), 247–256. https://doi.org/10.1016/j.lindif.2005.02.002

Chernyshenko, O. S. (2002). Applications of ideal point approaches to scale construction and scoring in personal ity measurement: The development of a six-faceted measure of conscientiousness. Champaign: University of Illinois at Urbana-Champaign.

Costa, P. T., & McCrea, R. B. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. http://ci.nii.ac.jp/ncid/BA20569924

De Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist :/Educational Psychologist*, *31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, *23*(4), 279–298. https://doi.org/10.1016/s0191-491x(97)86211-6

Furnham, A., & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences*, *37*(5), 943–955. https://doi.org/10.1016/j.paid.2003.10.016

Gasser, C. E., Larson, L. M., & Borgen, F. H. (2004). Contributions of personality and interests to explaining the educational aspirations of college students. *Journal of Career Assessment*, *12*(4), 347–365. https://doi.org/10.1177/1069072704266644

Gronlund, N. E. (1976). *Measurement and evaluation in teaching* (3rd ed.). New York: Macmillan.

King, L. A., Walker, L. M., & Broyles, S. J. (1996). Creativity and the Five-Factor model. *Journal of Research in Personality*, *30*(2), 189–203. https://doi.org/10.1006/jrpe.1996.0013

Liao, H. I., Yeh, S. L., & Shimojo, S. (2011). Novelty vs. Familiarity Principles in Preference Decisions: Task-Context of Past Experience Matters. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00043

Lindner, M. A., Mayntz, S. M., & Schult, J. (2018). Studentische Bewertung und Präferenz von Hochschulprüfungen mit Aufgaben im offenen und geschlossenen Antwortformat. *Zeitschrift Fur Padagogische Psychologie*, *32*(4), 239–248. https://doi.org/10.1024/1010-0652/a000229

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of Conscientiousness. *Learning and Individual Differences*, *19*(4), 451–458. https://doi.org/10.1016/j.lindif.2009.03.007

Mammadov, S. (2021). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255. https://doi.org/10.1111/jopy.12663

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The value of using tests in education as tools for Learning—Not just for assessment. *Educational Psychology Review*, *35*(3). https://doi.org/10.1007/s10648-023-09808-3

Murphy, K. R. (2021). In praise of Table 1: The importance of making better use of descriptive statistics. *Industrial and Organizational Psychology*, *14*(4), 461–477. https://doi.org/10.1017/iop.2021.90

Nevo, B. (1985). FACE VALIDITY REVISITED. *Journal of Educational Measurement*, *22*(4), 287–293. https://doi.org/10.1111/j.1745-3984.1985.tb01065.x

Pan, W., & Hu, T. (2024). More familiar, more credible? Distinguishing two types of familiarity on the truth effect using the drift-diffusion model. *the Journal of Social Psychology/Journal of Social Psychology*, 1–19. https://doi.org/10.1080/00224545.2024.2363366

Piedmont, R. L., Hill, D. C., & Blanco, S. (1999). Predicting athletic performance using the five-factormodel of personality. *Personality and Individual Differences*, *27*(4), 769–777. https://doi.org/10.1016/s0191-8869(98)00280-3

Schuwirth, L., Van Der Vleuten, C., & Donkers, J. (1996). A closer look at cueing effects in multiple-choice questions. *Medical Education*, *30*(1), 44–49. https://doi.org/10.1111/j.1365-2923.1996.tb00716.x

Sona Systems (n.d.). *Sona Systems: Cloud-based Participant Management Software* [Computer software]. Sona Systems, Ltd. https://www.sona-systems.com/

Soto, C. J., & John, O. P. (n.d.). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Star, J. R., & Stylianides, G. J. (2013). Procedural and Conceptual knowledge: Exploring the gap between knowledge type and knowledge quality. *Canadian Journal of Science, Mathematic and Technology Education/Canadian Journal of Science, Mathematics and Technology Education*, *13*(2), 169–181. https://doi.org/10.1080/14926156.2013.784828

Teigen, K. H. (1987). Intrinsic interest and the novelty-familiarity interaction. *Scandinavian Journal of Psychology*, *28*(3), 199–210. https://doi.org/10.1111/j.1467-9450.1987.tb00756.x

Whittle, R. J., Benson, A., Ullah, S., & Telford, A. (2018). Investigating the influence of question type and cognitive process on academic performance in VCE Physical Education: a secondary data analysis. *Educational Research and Evaluation*, *24*(8), 504–522. https://doi.org/10.1080/13803611.2019.1612256

Woo, S. E., Chernyshenko, O. S., Stark, S. E., & Conz, G. (2013). Validity of Six openness Facets in Predicting Work Behaviors: A Meta-Analysis. *Journal of Personality Assessment*, *96*(1), 76–86. https://doi.org/10.1080/00223891.2013.806329

You, M., Laborde, S., Vaughan, R. S., & Salinas, A. (2021). Influence of personality and emotional competences on academic performance: direct and indirect pathways mediated by perceived stress. *Current Issues in Personality Psychology*, *10*(1), 61–70. https://doi.org/10.5114/cipp.2021.111423

Zeidner, M. (1987). Essay versus Multiple-Choice Type Classroom Exams: The Student's Perspective. *The Journal of Educational Research*, *80*(6), 352–358. https://doi.org/10.1080/00220671.1987.10885782

# Appendix A

## Examination Preference Inventory

This appendix comprises the four scales of the Examination Preference Inventory, Familiarity and their corresponding items. Participants rate each statement on a five-point Likert scale, ranging from “Strongly disagree” to “Strongly agree”.

***Difficulty***

1. Open-question exams/MC exams/performance tasks are complex.
2. Open-question exams/MC exams/performance tasks are easy.
3. Open-question exams/MC exams/performance tasks are challenging.

***Potential to show performance***

1. Open-question exams/MC exams/performance tasks give me the opportunity to show that I have understood the subject matter very well.
2. Open-question exams/MC exams/performance tasks give me the opportunity to show that I know more than other students.
3. Open-question exams/MC exams/performance tasks allow me to express my knowledge precisely.
4. Open-question exams/MC exams/performance tasks are an appropriate examination format for important exams.

***Objectivity***

1. Open-question exams/MC exams/performance tasks are evaluated objectively.
2. Open-question exams/MC exams/performance tasks are graded without bias.

***Valence***

1. Open-question exams/MC exams/performance tasks should be the main method of examination.
2. Open-question exams/MC exams/performance tasks are interesting.
3. Open-question exams/MC exams/performance tasks are liked by me.

***Familiarity***

1) Open-question exams/MC exams/performance tasks are familiar to me.
2) Open question exams/MC exams/performance tasks are a common form of assessment to me.
3) Open-question exams/MC exams/performance tasks are rarely used as a form of assessment in my studies.