Students' Preference for Assessment Methods

Silke Nienke Kok

s3735141

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group Number 8

Supervisor: prof. dr. R.R. Meijer

Second evaluator: dr. S.M. Donofrio

In collaboration with: Annika Selesnew, Hage Niggendijker, Juraj Takács, Matsen Piepers

Ziying Zhang

June 2nd, 2024

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.

The generative AI tool "ChatGPT" was used for background information and for the introduction and discussion section of this thesis to improve grammar, spelling, and wording (see Appendix B) and make it in line with APA guidelines.

Due to the collaborative nature 'we' is used to describe the research, even though some parts are written individually.

Abstract

This study investigated students' preference for different assessment methods, namely multiple-choice (MC) exams, open-question (OQ) exams, and performance tasks (PT). We explored the factors that influence the preference for both differences in the exam format and individual differences, including familiarity with the exam format and personality. We addressed this by looking at data from 128 students from the Faculty of Behavioral and Social Sciences at the University of Groningen, who were sampled through a convenience sample. A higher preference was found for multiple-choice exams in comparison to open-question exams and performance tasks. MC exams were perceived to be less difficult, more objective, and higher in valence than OQ exams and PT, while having a lower potential to show performance. Higher familiarity with MC exams and PT was positively correlated with the general preference for those exam formats. Contrary to our hypothesis, no relation was found between familiarity with OQ exams and overall preference for OQ exams. Contrary to our expectations and prior research, we did not find evidence that extraversion and neuroticism was associated with preference for specific assessment methods. Only openness to experience had a small negative relation with a preference for MC exams. The differences with existing literature might be explained through possible moderators or mediators. In future research, it would be interesting to investigate the individual factors further.

Keywords: students' preference, multiple-choice examinations, open-question examinations, performance tasks, familiarity, personality

Students' Preference for Assessment Methods

A wide variety of assessment methods have been developed. The Assessment Preference Inventory by Birenbaum (1994) alone recognizes 18 different (sub)types of assessment, ranging from written tests without supporting materials to making a portfolio. But how do we select the most suitable method? One approach could be to consider the student's preference. This is especially crucial since students who are satisfied with their assessment method tend to achieve better outcomes (Neto et al., 2022). Additionally, undergoing an exam in a preferred format might lead to higher motivation in students and differences in preparation (Holzinger et al., 2020; Liu et al., 2023). An essay assignment might be more likely to encourage students to use deep learning strategies compared to multiple-choice exams (Scouller, 1998).

In this thesis, we will further explore students' preference. This research aims to continue previous research, by examining the preference for various assessment methods, considering factors that affect these preferences.

Preference of Assessment Methods

In this research, the focus was on three often-used assessment methods: *multiple-choice examinations* (MC exams), *open-question examinations* (OQ exams), and *performance tasks* (PT). MC exams include forms of assessment where the student needs to decide on the best answer from different options. This is also referred to as the selected response format (Hift, 2014). Examination methods in which students generate their own answers are called constructed response or OQ exams (Hift, 2014). Lastly, in PT, students create a product to demonstrate their knowledge, skills, or abilities, in response to a real-world task (Braun, 2019).

Several studies found that multiple-choice exams are more popular among students than open-question exams (Furnham et al., 2011; Zeidner, 1987); but the relation between MC, OQ

exams and performance tasks is largly unknown. In this thesis, we will examine different factors that determine preferences, considering both structural elements of the exam format and personal differences among students. For structural elements of the exam formats, we will direct our focus to four subscales of preference: difficulty, objectivity, potential to show performance, and valence. These subscales were chosen as they provide a multi-dimensional view of general preference and were embedded in previous literature (Lindner et al., 2018; Zeidner, 1987). Research showed that MC exams are perceived as less difficult, more objective, but lower in potential to show performance (Lindner et al., 2018; Zeidner, 1987). Furthermore, earlier studies did not provide a conclusive answer to which assessment method scores higher on valence (Lindner et al.,2018). The upcoming section will explain the differences between exam formats more in-depth.

Differences in the Perception between Exam Formats

Zeidner (1987) found that MC exams were perceived by students to be less difficult, caused less anxiety, have higher success expectancy, and make students feel more at ease compared to OQ exams. OQ exams, on the other hand, were observed as slightly more valuable and more sufficient to assess a student's knowledge. Similar to Zeidner (1987), Lindner et al. (2018) found that OQ exams were perceived as having a greater potential to show performance. MC exams were associated with a lower workload and higher objectivity. This higher objectivity of MC exams has also been identified in later studies (Holzinger et al., 2020). There were similar strong scores on valence for both multiple-choice and open-question exams (Lindner et al., 2018), meaning both assessment methods were equally liked.

The above-mentioned research mostly focused on the comparison between MC exams and OQ exams. There is limited research comparing aspects of MC exams with other exam

formats, such as performance tasks. There are, however, some findings on the perception of performance tasks. Performance tasks seem to be popular because they provide students with the opportunity to generate their own ideas (Hether, 2023). Additionally, there are also some studies on the preference for exam formats that allow supporting material, which includes performance tasks. The use of supporting materials could reduce anxiety, give students the perception of a higher chance of success, and allow students to engage with the material more in-depth (Ben-Chaim & Zoller, 1997; Van de Watering et al., 2008).

Current Study: Differences in Exam Format Perception

In this thesis, we will examine the preference for various assessment methods, exploring factors that affect these preferences. In particular, we compare three types of exam formats: multiple-choice examinations, open-question examinations, and performance tasks. These assessment methods were selected due to their general popularity in academia (Holzinger et al., 2020; Neto et al., 2022). Previous studies are often limited to comparing MC and OQ exams, neglecting other types of exams. Therefore, much remains unclear about the preference for different examination forms, particularly regarding the differences between MC exams and PT. We will thus explore the following hypotheses:

Hypothesis 1a. There is an overall preference for MC exams in comparison to OQ exams and PT.

Hypothesis 1b. Students perceive MC exams as less difficult than OQ exams and performance tasks.

Hypothesis 1c. Students perceive MC exams as more objective than OQ exams and performance tasks.

Hypothesis 1d. Students perceive MC exams as having a lower potential to show their performance than OQ examinations and performance tasks.

Hypothesis 1e. There will be differences in valence among the exam formats. However, given the lack of conclusive research, no specific hypotheses were formulated.

Individual Differences

In addition to exploring the influence of various aspects of the test format on preference, we will also investigate individual differences in *familiarity* and the relationship with some *personality* traits. Investigating these differences could provide theoretical insights into the topic. Naturally, it will not be possible to cater exams to every individual in practice. However, research on individual differences could improve the design of universities' curricula. For instance, if the most commonly chosen exam format is more suited for a certain type of student, more variety in assessment methods could be crucial to improve equality between students.

Familiarity. An important aspect that may affect preference for an exam type is the level of experience with a particular test format, also called familiarity (Lindner et al., 2018). When confronted with an assessment type that was not previously known, students can experience anxiety or fear of the unexpected (Lynam & Cachia, 2017). This can lead to negative experiences, panic or sleeplessness, which influences student performance. The impact of familiarity was also illustrated by Dang et al. (2022). This study found that written assignments were favored more than oral assessments. Even though familiarity was not studied directly, Dang et al. (2022) noted that this preference could be due to written assignments being a more common type of exam for students. Furthermore, the role of familiarity is more pronounced when the stakes are high (Parkes & Stefanou, 2010; Stefanou & Parkes, 2003), for instance when a test is of greater importance to a student. When students consider receiving a good grade

important, they would rather avoid the uncertainty, and choose a familiar testing format to help them achieve a desired grade (Parkes & Stefanou, 2010).

Personality. Various dimensions of the Big Five have been associated with preferences for specific assessment methods (Neto et al., 2022). The Big Five Model depicts five dimensions of personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (McCrae & John, 1992). Neto et al. (2022) revealed that extraversion was positively related to a preference for open-question examinations. This study also identified a negative correlation between neuroticism and a preference for engaging in dissertation research, which may be considered a type of performance task: highly neurotic students may struggle to cope with the stress associated with a final project (Neto et al., 2022). Furthermore, lower neuroticism was associated with a preference for essay-type exams (Furnham & Chamorro-Premuzic, 2005). However, this relation was not consistently observed (Chamorro-Premuzic et al., 2005). Lastly, openness to experience has been positively linked to performance tasks, such as essays and dissertations, but negatively to multiple-choice exams (Furnham et al., 2008; Furnham et al., 2011). People who score high on openness tend to be more creative and original and, therefore, prefer an exam format that allows them to use this creativity (Neto et al., 2022).

Current Study: Individual Differences

In conclusion, although familiarity and several personality traits are linked to preference of certain types of exams relations are not always clear and some results from earlier studies are contradictory. Therefore, in this study we will further investigate the relationship between exam preference and level of familiarity, and personality traits. To assess familiarity, we will first examine whether there is a significant difference in familiarity among MC exams, OQ exams, and PT. As will be described in the Method section, in thus study we used students from the

behavioral and social sciences. Based on inspection of the curriculum we formulated the following hypothesis for this population:

Hypothesis 2. There is a higher familiarity with MC exams, compared to OQ exams and PT for students of the faculty of Behavioral and Social Sciences at the University of Groningen.

Even though familiarity has been considered in earlier studies, research on the relationship with general preference is scarce. In general, we see that uncertainty about a type of exam can result in anxiety or fear, which can lead to lower student performance (Lynam & Cachia, 2017). Familiarity could combat this effect, making familiar assessment methods more appealing to students. Therefore, we have formulated the following hypotheses:

Hypothesis 3a. Familiarity with MC exams is positively related to preference for MC exams.

Hypothesis 3b. Familiarity with OQ exams is positively related to preference for OQ exams.

Hypothesis 3c. Familiarity with PT is positively related to preference for PT.

Furthermore, this study aims to investigate the relationship between the dimensions of the Big Five and general preference. Even though earlier studies have explored this relationship, this study hopes to strengthen earlier findings in a sample of psychology students. To explore these relationships, we formed the following hypotheses:

Hypothesis 5. Extraversion is positively related to a preference for OQ exams.

Hypothesis 6. Neuroticism is negatively related to preference for PT and negatively related to OQ exams.

Hypothesis 7. Openness to experience is positively related to preference for PT, but negatively related to preference for MC exams.

Method

Sample

The present study was conducted among 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 two or more attention checks. If a participant failed to complete the survey or the attention checks, that participant was removed from the final data set. This resulted in a final sample size of 128 students, which was used for the data analysis.

Of the participants, 37.5% was a Psychology (EN) student, 61.7% Psychology (NL) and 0.8% Sociology. For practical reasons, we limited our sample to students of the faculty of Behavioral and Social sciences. Of the sample 72.7% was female, 23.4% was male, 3.1% was 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 64.8% were 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 was accessible 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 accessible from April 16th, 2024 to April 24th, 2024.

Before filling in 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 in handling the data. The researchers' contact information was provided in case of any questions about the research. Subsequently, participants were required to fill in the informed consent about participating in the study and processing their data.

The questionnaire started with demographic questions, namely gender, year of study, and subject of study. This was 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

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. Participants indicated their level of agreement with displayed statements on a five-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). Three scales were based on an existing inventory by Lindner et al. (2018): potential to show performance, objectivity, and 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 item scores of the different subscales were combined to form the general preference score. Reliability was estimated using Cronbach's alpha and proved sufficient (>.70), for all scales except one (Bland & Altman, 1997). Moreover, the reliability of all scales combined as a measure of preference was also sufficient (see Table 1).

Table 1

Cronbach's Alpha Scores for the EPI scale and subscales

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

Familiarity

Familiarity was measured with an adapted version of Lindner's (2018) familiarity scale. The scale encompassed three items, two were rephrased to align better with the current study (e.g. "...are a common form of assessment to me") (see Appendix A). The participants reported their agreement with each statement on a 5-point Likert scale (1 = *strongly disagree*, 5 =

strongly agree). The reliability was estimated with Cronbach's alpha resulting in alpha levels of .63 for multiple-choice exams, .63 for open-question exams, and .75 for performance tasks. Due to the relatively low reliability, the item-test correlations were inspected, which showed low scores for the third question of the familiarity scale (e.g., "... are rarely used as a form of assessment in my studies."). However, the item-total correlations for this question were not below 0.2 or negative, indicating acceptable item-test correlations (Everitt & Skrondal, 2010). Therefore, the complete scale was used, but the results should be interpreted with care.

Personality

Personality was measured using The Big Five Inventory-2 Short Form (BFI-2-S; Soto and John, 2017). This inventory included six questions for all five domains: extraversion (e.g., "... is full of energy."), agreeableness (e.g., "... assumes the best about people.") conscientiousness (e.g., "...is reliable, can always be counted on."), negative emotionality/neuroticism (e.g., "...worries a lot.") and open-mindedness/openness to experience (e.g., "...is original, comes up with new ideas."). This inventory can be found in Appendix A. The participants reported their agreement with each statement on a 5-point Likert scale (1 = *strongly disagree,* 5 = *strongly agree*). The reliability was estimated with Cronbach's alpha resulting in alpha levels of .79 for extraversion, .71 for agreeableness, .78 for conscientiousness, .84 for negative emotionality/neuroticism, and .74 for open-mindedness. These reliability scores are acceptable for comparing groups (Bland & Altman, 1997).

Results

Differences in the Exam Format

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 open-question exams ($M = 3.01$, $SD = .46$), and performance tasks ($M = 3.00$, $SD = 0.48$). Supporting the first hypothesis, a paired sample t-test between MC exams and PTs showed that preference for MC exams was significantly higher than for PTs $t(127) = 6.16$, $p < .001$, $d = 0.54$, 99.5% CI [0.24, 0.66]. Also, students prefer MC exams over OQ exams, $t(127) = 6.21$, $p < .001$, $d = 0.55$ [0.23, 0.64].[1,2] The effect sizes showed moderate effects. The difference between the means of performance tasks ($M = 3.00$, $SD = 0.48$) and OQ exams ($M = 3.01$, $SD = .46$) was negligible and significance testing was thus not conducted.

Subscales of General Preference

To further explore students' preferences, we analysed the subscales of the EPI (see Footnotes 1 and 2). The descriptive statistics of the four subscales for each examination format can be found in Table 2. Based on our hypotheses, we conducted paired-sample t-tests for objectivity, difficulty, and potential to show performance and valence, between selected exam formats. The effect sizes of these t-tests showed medium to large effects.

---

[1] The assumptions of independence between subjects, same subject paired measurements, normal distribution of differences between pairwise comparisons were met. Some outliers were found. In such a case, 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 conducted. 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. As such, the critical p-value was corrected to 0.0045 when taking into account the t-tests for the subscales later in this section.

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

Note. Here, a higher score on difficulty means the format was found to be more difficult. The coding of this variable was reversed for the calculation of the general preference score.

Difficulty. The differences in difficulty between MC exams and performance tasks and MC exams and OQ exams were significant, $t(127) = 4.204$, $p < .001$, $d = .37$, [0.10, 0.54] and $t(127) = 13.559$, $p < .001$, $d = 1.20$, [0.90, 1.36], respectively. Performance tasks and OQ exams were rated as more difficult than MC exams, which aligns with hypothesis 1b.

Potential to Show Performance. The next paired sample t-tests tested the differences in potential to show performance for MC exams and OQ exams, and MC exams and performance tasks. OQ exams and performance tasks were rated significantly higher on that subscale compared to MC exams, $t(127) = 7.366$, $p < .001$, $d = .65$, [0.51, 1.17] and $t(127) = 5.427$, $p < .001$, $d = 0.48$, [0.28, 0.93], respectively, which supports hypothesis 1d.

Objectivity. Supporting hypothesis 1c, the objectivity scores for MC exams were significantly higher compared to performance tasks, $t(127) = 16.562$, $p < .001$, $d = 1.46$, 99.5% CI [1.50, 2.14], and compared to OQ exams. $t(127) = 13.870$, $p < .001$, $d = 1.23$, [1.10, 1.69].

Valence. For valence, a significant difference was found between MC exams and OQ exams ($t(127) = 6.179$, $p < .001$, $d = .55$, [0.43, 1.18], MC exams and PT ($t(127) = 4.455$, $p < .001$, $d = .54$, [0.19, 0.91]), and OQ exams and PT ($t(127) = 2.986$, $p = 0.003$, $d = .26$, [0.01, 0.49]). This means that MC exams had the highest valence, followed by PT and OQ exams.

Individual Differences

Familiarity

We found a significant difference between the familiarity of MC exams (M = 4.52, SD = 0.59) and OQ exams (M = 3.13, SD = 0.84; $t(127) = 14.46$, $p < .001$, $d = 1.28$, 97,5% CI [1.18, 1.61]), as well as between the familiarity of MC exams and PT (M = 3.17, SD = 0.94; $t(127) = 13.87$, $p < .001$, $d = 1.23$, 97,5% CI [1.12, 1.56]) with large effect sizes. [3,4] Thus, MC exams were more familiar for the students in our sample than OQ exams and PT, which supports our second hypothesis.

To investigate whether familiarity is positively related to general preference for MC exams, OQ exams, and performance tasks, respectively, we inspected the correlations between the variables. Supporting hypothesis 3a, general preference for MC exams correlated positively with familiarity with MC exams ($r(126) = .21$, $p < .019$, 95% CI [.04, .37]). There was also a positive correlation between general preference for PT and familiarity with PT, supporting hypothesis 3c ($r(126) = .26$, $p < .003$, [.09, .42]). Both correlations were small, however. We did not find evidence that general preference for OQ exams correlated with familiarity with OQ exams, ($r(126) = .07$, $p = .459$, [-.11, .24]). Therefore, hypothesis 3b was not supported.

---

[3] The assumptions of independence between subjects, same subject paired measurements, normal distribution of differences between pairwise comparisons were met. One outlier was found, but this outlier was not far out. Because of this, the paired sample t-test was still run, but the results need to be interpreted with caution.

[4] 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 0.025.

Personality

The descriptive statistics for the different personality domains can be found in Table 3.

Extraversion. Contrary to the fifth hypothesis, we did not find evidence of a positive correlation between extraversion and preference for open-question exams ($r(126) = .01$, $p = .96$, 95% CI [-.17, .18]).

Neuroticism. There was a lack of support for hypotheses 6a and 6b, with no evidence that neuroticism correlated with general preference for OQ exams ($r(126) = -.09$, $p = .30$, [-.26, .08]) or general preference for PT ($r(126) = -.04$, $p = .68$, [-.21, .14]).

Openness to experience. Supporting hypothesis 7b, a small negative correlation was found between openness to experience and preference for MC exams ($r(126) = -.19$, $p = .03$, [-.35, -.02]). Hypothesis 7a was not supported; there was no positive significant correlation between openness to experience and a general preference for PT ($r(126) = .05$, $p = .61$, [-.13, .22]).

Table 3

Means and Standard Deviations of the Different Domains of Personality

	M	SD
Extraversion	3.13	0.76
Agreeableness	3.81	0.62
Conscientiousness	3.18	0.79
Neuroticism/negative emotionality	3.23	0.86
Open-mindedness	3.66	0.73

Discussion

This study investigated students' preference for different types of exams. Previous studies have been limited to comparisons between MC and OQ exams, overlooking other assessment methods. Contributing to previous research, this thesis also examined the differences in preference between MC exams and PT. Furthermore, we explored factors that might determine preference, considering both structural elements of the exam and personal differences.

Overall Preference

We found that students have a higher general preference for MC exams compared to the other two assessment methods. In line with our first hypothesis, students perceive MC exams to be less difficult, more objective, and higher in valence than OQ exams. MC exams are perceived as having less potential to show performance. This pattern of results builds on the findings of previous literature (Lindner et al., 2018; Zeidner, 1987). Contributing to prior research, our results show that performance tasks are less preferred and are seen as more difficult, less objectively graded, and liked less compared to MC exams. However, they are considered an exam format that allows students a higher potential to show performance.

Interestingly, many students presume that OQ exams and PT are better suited to demonstrate their level of comprehension than MC exams. This is especially compelling since research shows that MC exams have the potential to allow students to exhibit higher-order thinking skills through context-rich questions (Hift, 2014; Liu et al., 2023). A possible explanation for this discrepancy between theory and the perception of students is that in practice context-rich questions only occur in a small segment of the exam, or are not incorporated at all (Liu et al., 2023; Zheng et al., 2008). Therefore, students could perceive MC exams as having a lower potential to show performance, without that being what research suggests.

Individual Differences

For the current research, we explored the familiarity with the test format and students' personality traits as they may be related to preference for types of exams. We found that for students in the behavioral and social sciences, familiarity with MC exams and PT is positively correlated with the general preference for those exam formats. Even though earlier research on this subject is scarce, these results align with expectations of findings that uncertainty and unfamiliarity can lead to anxiety, which can result in lower student performance (Lynam & Cachia, 2017). Contrary to our hypothesis, no relation was found between familiarity with OQ exams and overall preference for OQ exams.

Contrary to our expectations, we did not find evidence that extraversion and neuroticism were associated with a preference for specific assessment methods. As expected, openness to experience had a small negative relationship with a preference for MC exams. Except for this correlation, the results were not in line with earlier research where associations were found for the different personality domains (Furnham & Chamorro-Premuzic, 2005; Furnham et al., 2008; Furnham et al., 2011; Neto et al., 2022). One explanation may be that in the present study we used relatively few items to measure each of the personality traits. As a result, important facets that together define an individual trait may have not have been covered enough. Future research may look into which facets determine the relation between personality and preference for exam types.

Limitations and Strengths

This study has some limitations that should be addressed. First, the generalizability of the results should be further investigated, because the current research was limited to the Faculty of Behavioral and Social Sciences at the University of Groningen. Additionally, most participants

were first-year psychology students who participated in this study through the Sona Systems. Different contexts could therefore yield different results. Future research could, for example, focus on the replication of these results among master's students or in different cultural settings. Second, although in contrast to earlier research, we used different subscales and were able to obtain a more detailed picture of general preference, not all subscales were reliable enough, which might have affected our results. Furthermore, general preference could be improved by adding other subscales, such as the degree of anxiety provoked or learning effort, which was also mentioned in previous literature (Lindner et al., 2018; Zeidner, 1987).

Limitations aside, there are also several strengths. First, this study contributed to previous research by not only comparing MC exams and OQ exams but also MC exams with PT. Earlier research on this topic has been scarce and this research leads to interesting insights into how PT are perceived. Secondly, as mentioned above we used a measure for preference consisting of different subscales. While in the past, preference was limited to three or four items per assessment method (see Neto et al., 2022), this research included a multi-dimensional scale with 12 items per assessment method, which provides us with a broader overview.

Theoretical and Practical Implications

Despite the limitations, the results have theoretical and practical implications. The current study contributes to the body of evidence that MC exams are more favored by students in comparison to OQ exams or PT. This study focused on the current scarcity of research regarding the position of PT in contrast to MC exams. Furthermore, the relationship of familiarity and personality with general preference was examined. Even though some research on familiarity existed, the relationship between familiarity and preference was mostly overlooked in the past.

This study therefore provides an important perspective on how familiarity can impact general preference.

Furthermore, faculty and instructors may consider students' preferences when selecting an assessment method. After the corona crisis, some voices emerged stating that open-question exams should be more commonly used (Fens, 2021). The current research, however, shows that students prefer MC exams, even though OQ exams and PT contain a higher potential to show performance. Therefore, we might reconsider whether replacing multiple-choice exams with open-question exams is the optimal choice. Especially, since students' preferences can have an impact on their motivation, learning strategies, and student successful outcomes (Holzinger et al., 2020; Liu et al., 2023; Neto et al., 2022; Scouller, 1998).

Future Research

In terms of future research, it would be useful to broaden the current findings by examining the relation between personality and general preference in more detail. The current study did not find positive or negative corrleations between different personality domains and general preference, with the exception of a small negative correlation between openness to experience and preference for MC exams. Future research could, for example, explore the relationships between preference and the NEO PI-R or as discussed foucus more on the different facets in a particular personality domain.

Moreover, future research could explore the relationship between individual difference variables through moderator or mediator analyses. For instance, the correlation between familiarity and the level of anxiety. Past research has suggested that being unfamiliar with an assessment method can cause anxiety, which might lead to negative experiences (Lynam & Cachia, 2017). Therefore, the influence of familiarity may be related to anxiety.

Conclusion

In summary, our research replicates earlier findings that MC exams are preferred over other assessment methods (Lindner et al., 2018; Zeidner, 1987). The findings contribute to the body of evidence that MC exams are perceived to be less difficult, more objective, and higher in valence than OQ exams and performance tasks while having a lower potential to show performance. This study addressed the lack of research on the relationship between familiarity and preference, finding correlations between familiarity for MC exams and PT and general preference for these exam formats. Other individual differences were also studied, focusing on the Big Five personality domains. Extraversion and neuroticism were not associated with preference for any specific exam formats. The results did indicate a small negative correlation between openness to experience and preference for MC exams. We hope that the current research will stimulate further investigation of these different factors.

References

Ben-Chaim, D., Zoller, U. Examination-type preferences of secondary school students and their teachers in the science disciplines. *Instructional Science 25*, 347–367 (1997). https://doi.org/10.1023/A:1002919422429

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ. British Medical Journal*, *314*(7080), 572. https://doi.org/10.1136/bmj.314.7080.572

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

Braun, H. (2019). Performance assessment and standardization in higher education: A problematic conjunction? *British Journal Of Educational Psychology*, *89*(3), 429–440. https://doi.org/10.1111/bjep.12274

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

Dang, B. Y. Y., Ho, E., & Tsang, A. (2022). Learner's Assessment Preferences in Higher Education: A Comparison Study of High-Achievers and Low-Achievers. *The Asia-Pacific Education Researcher*, *32*(5), 595–604. https://doi.org/10.1007/s40299-022-00679-w

Everitt, B.S. & Skrondal, A. (2010). *The Cambridge Dictionary of Statistics* (4th edition). Cambridge University Press. https://doi.org/10.1017/cbo9780511779633

Fens, D. T. (2021, 16 november). Tentamens na corona: gaan we terug naar meerkeuze?

UKrant.nl. https://ukrant.nl/magazine/tentamens-na-corona-gaan-we-weer-terug-naar-meerkeuze/

Furnham, A., Batey, M., & Martin, G. T. (2011). How would you like to be evaluated? The correlates of students' preferences for assessment methods. *Personality And Individual Differences*, *50*(2), 259–263. https://doi.org/10.1016/j.paid.2010.09.040

Furnham, A., & Chamorro-Premuzic, T. (2005). Individual Differences and Beliefs Concerning Preference for University Assessment Methods. *Journal Of Applied Social Psychology*, *35*(9), 1968–1994. https://doi.org/10.1111/j.1559-1816.2005.tb02205.x

Furnham, A., Christopher, A., Garwood, J., & Martin, N. G. (2008). Ability, demography, learning style, and personality trait correlates of student preference for assessment method. *Educational Psychology*, *28*(1), 15–27. https://doi.org/10.1080/01443410701369138

Hether, H. J. (2023). Cultivating Creative Problem-Solving Skills in a Strategic Communication Class: Student Perceptions of a Collaborative Assignment. *Journal Of Creative Behavior*, *57*(4), 495–502. https://doi.org/10.1002/jocb.617

Hift, R. (2014). Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education*, *14*(1). https://doi.org/10.1186/s12909-014-0249-2

Holzinger, A., Lettner, S., Steiner-Hofbauer, V., & Melser, M. C. (2020). How to assess? Perceptions and preferences of undergraduate medical students concerning traditional assessment methods. *BMC Medical Education*, *20*(1). https://doi.org/10.1186/s12909-020-02239-6

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

Liu, Q., Wald, N., Daskon, C., & Harland, T. (2023). Multiple-choice questions (MCQs) for higher-order cognition: Perspectives of university teachers. *Innovations in Education And Teaching International*, 1–13. https://doi.org/10.1080/14703297.2023.2222715

Lynam, S., & Cachia, M. (2017). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 223–234. https://doi.org/10.1080/02602938.2017.1329928

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. Journal Of Personality, 60(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Neto, J., Neto, F., & Furnham, A. (2022). Predictors of students' preferences for assessment methods. *Assessment And Evaluation in Higher Education/Assessment & Evaluation in Higher Education*, *48*(4), 556–565. https://doi.org/10.1080/02602938.2022.2087860

Parkes, J., & Stefanou, C. (2010). Does pragmatism trump motivation in college students' preferences for examination formats? *Learning Environments Research*, *13*(3), 225–241. https://doi.org/10.1007/s10984-010-9077-4

Scouller, K. (1998). The influence of assessment method on students' learning approaches: multiple-choice question examination versus assignment essay. *Higher Education, 35*(4), 453–472. https://doi.org/10.1023/A:1003196224280

Soto, C. J., & John, O. P. (2017). 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

Stefanou, C., & Parkes, J. (2003). Effects of Classroom Assessment on Student Motivation in

Fifth-Grade Science. *The Journal Of Educational Research*, *96*(3), 152–162.

https://doi.org/10.1080/00220670309598803

Van de Watering, G., Gijbels, D., Dochy, F., & Van Der Rijt, J. (2008). Students' assessment

preferences, perceptions of assessment and their relationships to study results. *Higher*

Education, *56*(6), 645–658. https://doi.org/10.1007/s10734-008-9116-6

Zeidner, M. (1987). Essay versus Multiple-Choice Type Classroom Exams: The Student's

Perspective. *The Journal of Educational Research, 80*(6), 352–358.

http://www.jstor.org/stable/27540265

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's

Taxonomy Debunks the "MCAT Myth". *Science*, *319*(5862), 414–415.

https://doi.org/10.1126/science.1147852

Appendix A

Examination Scales

Examination Preference Inventory

This appendix comprises the four scales of the Examination Preference Inventory 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

4) Open-question exams / MC exams / performance tasks give me the opportunity to show
   that I have understood the subject matter very well.

5) Open-question exams / MC exams / performance tasks give me the opportunity to show
   that I know more than other students.

6) Open-question exams / MC exams / performance tasks allow me to express my
   knowledge precisely.

7) Open-question exams / MC exams / performance tasks are an appropriate examination
   format for important exams.

Objectivity

8) Open-question exams / MC exams / performance tasks are evaluated objectively.

9) Open-question exams / MC exams / performance tasks are graded without bias.

Valence

10) Open-question exams / MC exams / performance tasks should be the main method of examination.

11) Open-question exams / MC exams / performance tasks are interesting.

12) Open-question exams / MC exams / performance tasks are liked by me.

Familiarity

The items used for the scale familiarity were as follows. Participants rate each statement on a five-point Likert scale, ranging from "*Strongly disagree*" to "*Strongly agree*".

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.

Personality

Familiarity was measured using The Big Five Inventory-2 Short Form (BFI-2-S) by Soto and John (2017). Participants rated how much they agreed with each statement.

1) Tends to be quiet …

2) Is compassionate, has a soft heart …

3) Tends to be disorganized …

4) Worries a lot …

5) Is fascinated by art, music, or literature …

6) Is dominant, acts as a leader …

7) Is sometimes rude to others …

8) Has difficulty getting started on tasks …

9) Tends to feel depressed, blue …

10) Has little interest in abstract ideas …

11) Is full of energy …

12) Assumes the best about people …

13) Is reliable, can always be counted on …

14) Is emotionally stable, not easily upset …

15) Is original, comes up with new ideas …

16) Is outgoing, sociable …

17) Can be cold and uncaring …

18) Keeps things neat and tidy …

19) Is relaxed, handles stress well …

20) Has few artistic interests …

21) Prefers to have others take charge …

22) Is respectful, treats others with respect …

23) Is persistent, works until the task is finished …

24) Feels secure, comfortable with self …

25) Is complex, a deep thinker …

26) Is less active than other people …

27) Tends to find fault with others …

28) Can be somewhat careless …

29) Is temperamental, gets emotional easily …

30) Has little creativity …

Appendix B

Use of ChatGPT

The following prompts were put into ChatGPT: "Can you check the following text on grammar and APA guidelines?". Both the introduction and the discussion of the thesis were added. The following text is the input. I have marked sections that were changed in my final thesis with the help of the ChatGPT output.

A wide variety of assessment methods have been developed. The Assessment Preference Inventory by Birenbaum (1994) alone recognizes 18 different (sub)types of assessment, ranging from written tests without supporting materials to making a portfolio. But how do we select the most suitable method? One approach could be to consider the student's preference. This is especially crucial since students who are satisfied with their assessment method, tend to achieve more better outcomes (Neto et al., 2022). Additionally, undergoing an exam in a preferred format might lead to higher motivation in students and differences in preparations (Holzinger et al., 2020; Liu et al., 2023). An essay assignment might be more likely to encourage students to use deep learning strategies compared to multiple-choice exams (Scouller, 1998).

In this thesis, we will further explore students' preference. This research aims to continue previous research, by examining the preference for various assessment methods, considering factors that affect these preferences.

Preference of Assessment Methods

In this research, the focus was on three often-used assessment methods: *multiple-choice examinations* (MC exams), *open-question examinations* (OQ exams), and *performance tasks* (PT). MC exams include forms of assessment where the student needs to decide on the best

answer from different options. This is also referred to as the selected response format (Hift, 2014). Examination methods in which students generate their own answers are called constructed response or OQ exams (Hift, 2014). Lastly, in PT, students create a product to demonstrate their knowledge, skills, or abilities, in response to a real-world task (Braun, 2019).

Several studies found that multiple-choice exams are more popular among students than open-question exams (Furnham et al., 2011; Zeidner, 1987); but the relation between MC, OQ exams and performance tasks is largly unknown. In this thesis we will examine different factors that determine preferences, considering both structural elements of the exam format and personal differences among students. For structural elements of the exam formats, we will direct our focus to four subscales of preference: difficulty, objectivity, potential to show performance, and valence. These subscales were chosen as they provide a multi-dimensional view of general preference and were embedded in previous literature (Lindner et al., 2018; Zeidner, 1987). Research showed that MC exams are perceived as less difficult, more objective, but lower in potential to show performance (Lindner et al., 2018; Zeidner et al., 1987). Furthermore, earlier studies did not provide a conclusive answer to which assessment method scores higher on valence (Lindner et al., 2018). The upcoming section will explain the differences between exam formats more in-depth.

Differences in the Perception between Exam Formats

Zeidner (1987) found that MC exams were perceived by students to be less difficult, cause less anxiety, have higher success expectancy, and make students feel more at ease compared to OQ exams. OQ exams, on the other hand, were observed as slightly more valuable and more sufficient to assess a student's knowledge. Similar to Zeidner (1987), Lindner et al. (2018) found that OQ exams were perceived as having a greater potential to show performance.

MC exams were associated with a lower workload and higher objectivity. This higher objectivity of MC exams has also been identified in later studies (Holzinger et al., 2020). There were similar strong scores on valence for both multiple-choice and open-question exams (Lindner et al., 2018), meaning both assessment methods were equally liked.

The above-mentioned research mostly focused on the comparison between MC exams and OQ exams. There is limited research comparing aspects of MC exams and other exam formats, such as performance tasks. There are, however, some findings on the perception of performance tasks. Performance tasks seem to be popular because they provide students with the opportunity to generate their own ideas (Hether, 2023). There are also some studies on the preference for exam formats that allow supporting material, which includes performance tasks. The use of supporting materials could reduce anxiety, give students the perception of a higher chance of success, and allow students to engage with the material more in-depth (Ben-Chaim & Zoller, 1997; Van de Watering et al., 2008).

Current Study: Differences in Exam Format Perception

In this thesis, we will examine the preference for various assessment methods, exploring factors that affect these preferences. In particular, we compare three types of exam formats: multiple-choice examinations, open-question examinations, and performance tasks. These assessment methods were selected due to their general popularity in academia (Holzinger et al., 2020; Neto et al., 2022). Previous studies are often limited to comparing MC and OQ exams, neglecting other types of exams. Therefore, much remains unclear about the preference for different examination forms, particularly regarding the differences between MC exams and PT. We will thus explore the following hypotheses:

Hypothesis 1a. There is an overall preference for MC exams in comparison to OQ exams and PT.

Hypothesis 1b. Students perceive MC exams as less difficult than OQ exams and performance tasks.

Hypothesis 1c. Students perceive MC exams as more objective than OQ exams and performance tasks.

Hypothesis 1d. Students perceive MC exams as having a lower potential to show their performance than OQ examinations and performance tasks.

Hypothesis 1e. There will be differences in valence among the exam formats. However, given the lack of conclusive research, no specific hypotheses were formulated.

Individual Differences

In addition to exploring the influence on preference of various aspects of the test format, we will also investigate individual differences in *familiarity* and the relation with some *personality* traits. Investigating these differences could provide theoretical insights into the topic. Naturally, it will not be possible to cater exams to every individual in practice. However, research on individual differences could improve the design of universities' curricula. For instance, if the most commonly chosen exam format is more suited for a certain type of student, more variety in assessment methods could be crucial to improve equality between students.

Familiarity. An important aspect that may affect preference for an exam type is the level of experience with a particular test format, also called familiarity (Lindner et al., 2018). When confronted with an assessment type that was not previously known, students can experience anxiety or fear of the unexpected (Lynam & Cachia, 2017). This can lead to negative experiences, panic or sleeplessness, which influences student performance. The impact of

familiarity was also illustrated by Dang et al. (2022). This study found that written assignments were favored more than oral assessments. Even though familiarity was not studied directly, Dang et al. (2022) noted that this preference could be due to written assignments being a more common type of exam for students. Furthermore, the role of familiarity is more pronounced when the stakes are high (Parkes & Stefanou, 2010; Stefanou & Parkes, 2003), for instance when a test is of greater importance to a student. When students consider receiving a good grade important, they would rather avoid the uncertainty, and choose a familiar testing format to help them achieve a desired grade (Parkes & Stefanou, 2010).

Personality. Various dimensions of the Big Five have been associated with preferences for specific assessment methods (Neto et. al, 2022). The Big Five Model depicts five dimensions of personality: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (McCrae & John, 1992). Neto et al. (2022) revealed that extraversion was positively related to a preference for open-question examinations. This study also identified a negative correlation between neuroticism and a preference for engaging in dissertation research, which may be considered a type of performance task: highly neurotic students may struggle to cope with the stress associated with a final project (Neto et al., 2022). Furthermore, lower neuroticism was associated with a preference for essay-type exams (Furnham & Chamorro-Premuzic, 2005). However, this relation was not consistently observed (Chamorro-Premuzic et al., 2005). Lastly, openness to experience has been positively linked to performance tasks, such as essays and dissertations, but negatively to multiple-choice exams (Furnham et al., 2008; Furnham et al., 2011). People who score high on openness tend to be more creative and original and, therefore, prefer an exam format that allows them to use this creativity (Neto et al., 2022).

Current Study: Individual Differences

In conclusion, although familiarity and several personality traits are linked to preference of certain types of exams relations are not always clear and some results from earlier studies are contradictory. Therefore, in this study we will further investigate the relationship between exam preference and level of familiarity, and personality traits.

To assess familiarity, we will first examine whether there is a significant difference in familiarity among MC exams, OQ exams, and PT. As will be described in the Method section, in thus study we used students from the behavioral and social sciences. Based on inspection of the curriculum we formulated the following hypothesis for this population:

Hypothesis 2. There is a higher familiarity with MC exams, compared to OQ exams and PT for students of the faculty of Behavioral and Social Sciences at the University of Groningen.

Even though familiarity has been considered in earlier studies, research on the relation with general preference is scarce. In general, we see that uncertainty about a type of exam can result in anxiety or fear, which can lead to lower student performance (Lynam & Cachia, 2017). Familiarity could combat this effect, making familiar assessment methods more appealing to students. Therefore, we have formulated the following hypotheses:

Hypothesis 3a. Familiarity with MC exams s is positively related to preference for MC exams.

Hypothesis 3b. Familiarity with OQ exams is positively related to preference for OQ exams.

Hypothesis 3c. Familiarity with PT is positively related to preference for PT.

Furthermore, this study aims to investigate the relationship between the dimensions of the Big Five and general preference. Even though earlier studies have explored this relation, this

study hopes to strengthen earlier findings in a sample of psychology students. To explore these relationships, we formed the following hypotheses:

Hypothesis 5. Extraversion is positively related to a preference for OQ exams.

Hypothesis 6. Neuroticism is negatively related to preference for PT and negatively related to OQ exams.

Hypothesis 7. Openness to experience is positively related to preference for PT, but negatively related to preference for MC exams.

Discussion

This study investigated students' preference for different types of exams. Previous studies have been limited to comparisons between MC and OQ exams, overlooking other assessment methods. Contributing to previous research, this thesis also inspected the differences in preference between MC exams and PT. Furthermore, we explored factors that might determine preference, considering both structural elements of the exam and personal differences.

Overall Preference

We found that students have a higher general preference for MC exams compared to the other two assessment methods. In line with our first hypothesis, students perceive MC exams to be less difficult, more objective, and higher in valence than OQ exams. MC exams are perceived as having less potential to show performance. This pattern of results builds on the findings of previous literature (Lindner et al., 2018; Zeidner, 1987). Contributing to prior research, our results show that performance tasks are less preferred and are seen as more difficult, less objectively graded, and liked less compared to MC exams. However, they are considered an exam format that allows students a higher potential to show performance.

Interestingly, many students presume that OQ exams and PT are better suited to demonstrate their level of comprehension than MC exams. This is especially compelling since research shows that MC exams have the potential to let students show higher-order thinking skills through context-rich questions (Hift, 2014; Liu et al., 2023). A possible explanation for this discrepancy between theory and the perception of students is that in practice context-rich questions only occur in a small segment of the exam, or are not incorporated at all (Liu et al., 2023; Zheng et al., 2008). Therefore, students could perceive MC exams as having a lower potential to show performance, without that being what research suggests.

Individual Differences

For the current research, we explored the familiarity with the test format and students' personality as they may be related to preference for types of exams. We found that familiarity with MC exams and PT is positively correlated with the general preference for those exam formats. Even though earlier research on this subject is scarce, these results fit with expectations of findings that uncertainty and unfamiliarity could lead to anxiety, which can lead to lower student performance (Lynam & Cachia, 2017). Contrary to our hypothesis, no relation was found between familiarity with OQ exams and overall preference for OQ exams.

Contrary to our expectations, we did not find evidence that extraversion and neuroticism were associated with a preference for specific assessment methods. As expected, openness to experience had a small negative relation with a preference for MC exams. Except for this correlation, the results were not in line with earlier research in which associations were found for the different personality domains (Furnham & Chamorro-Premuzic, 2005; Furnham et al., 2008; Furnham et al., 2011; Neto et al., 2022). One explanation may be that in the present study we used relatively few items to measure each of the personality traits. As a result, important facets

that together define an individual trait may have not have been covered enough. Future research may look into which facets determine the relation between personality and preference for exam types.

Limitations and Strengths

This study has some limitations that should be addressed. First, the generalizability of the results should be further investigated, because the current research was limited to the Faculty of Behavioral and Social Sciences at the University of Groningen. Additionally, most participants were first-year psychology students who participated in this study through the Sona Systems. Different contexts could therefore yield different results. Future research could, for example, focus on the replication of these results among master's students or in different cultural settings. Second, although in contrast to earlier research, we used different subscales and were able to get a more detailed picture of general preference, not all subscales were reliable enough, which might have affect our results. Furthermore, general preference could be improved by adding other subscales, such as the degree of anxiety provoked or learning effort, which was also mentioned in previous literature (Lindner et al., 2018; Zeidner, 1987).

Limitations aside, there are also several strengths. First, this study contributed to previous research by not only comparing MC exams and OQ exams but also MC exams with PT. Earlier research on this topic has been scarce and this research leads to interesting insights into how PT are perceived. Secondly, as mentioned above we used a measure for preference consisting of different subscales. While in the past, preference was limited to three or four items per assessment method (see Neto et al., 2022), this research included a multi-dimensional scale with 12 items per assessment method, which provides us with a broader overview.

Theoretical and Practical Implications

Despite the limitations, the results have theoretical and practical implications. The current study contributes to the body of evidence that MC exams are more favored by students in comparison to OQ exams or PT. This study focused on the current scarcity of research regarding the position of PT in contrast to MC exams. Furthermore, the relationship of familiarity and personality with general preference was examined. Even though some research on familiarity existed, the relationship between familiarity and preference was mostly overlooked in the past. This study therefore provides an important perspective on how familiarity can impact general preference.

Furthermore, faculty and instructors may consider students' preferences when selecting an assessment method. After the corona crisis, some voices emerged stating that open-question exams should be more commonly used (Fens, 2021). The current research, however, shows that students prefer MC exams, even though OQ exams and PT contain a higher potential to show performance. Therefore, we might reconsider whether replacing multiple-choice exams with open-question exams is the optimal choice. Especially, since students' preferences can have an impact on their motivation, learning strategies, and student successful outcomes (Holzinger et al., 2020; Liu et al., 2023; Neto et al., 2022; Scouller, 1998).

Future Research

In terms of future research, it would be useful to broaden the current findings by examining the relation between personality and general preference in more detail. The current study did not find positive or negative correlations between different personality domains and general preference, with the exception of a small negative correlation between openness to experience and preference for MC exams. Future research could, for example, explore the

relationships between preference and the NEO PI-R r as discussed focus more on the different facets in a particular personality domain.

Moreover, future research could explore the relationship between individual difference variables through moderator or mediator analyses. For instance, the correlation between familiarity and the level of anxiety. Past research has suggested that being unfamiliar with an assessment method can cause anxiety, which might lead to negative experiences (Lynam & Cachia, 2017). Therefore, the influence of familiarity may be related to anxiety.

Conclusion

In summary, our research replicates earlier findings that MC exams have a higher preference score than other assessment methods (Lindner et al., 2018; Zeidner, 1987). The findings contribute to the body of evidence that MC exams are perceived to be less difficult, more objective, and higher in valence than OQ exams and performance tasks while having a lower potential to show performance. This study addressed the lack of research on the relationship between familiarity and preference, finding correlations between familiarity for MC exams and PT and general preference for these exam formats. Other individual differences were also studied, focusing on the Big Five personality domains. Extraversion and neuroticism were not associated with preference for any specific exam formats. The results did indicate a small negative correlation between openness to experience and preference for MC exams. We hope that the current research will stimulate further investigation of these different factors.