

The Role of Performance Tasks in Higher Education Assessment: A Study of Students'

Preferences

Matsen Piepers

S4795490

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group 8

Supervisor: Prof. Dr. Rob Meijer

Second evaluator: Drs. S.M. Stacey Donofrio

In collaboration with: Annika Selesnew, Hage Niggendijker, Juraj Takács, Silke Kok, and Ziyang Zhang.

June 21, 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.

Abstract

This study aims to gain a deeper understanding of students' perceptions of performance tasks as an assessment format in higher education, an area less researched than traditional formats such as multiple-choice and open-question exams. By investigating and comparing general preferences for multiple-choice exams, open-question exams, and performance tasks, we sought to determine where performance tasks rank among these formats. Our findings indicate a clear preference for multiple-choice exams, while there was no preference for performance tasks or open-question exams. Furthermore, a comparative analysis of students' opinions on individual versus group assignments revealed a higher preference score for individual assignments, contrary to our predictions. Additionally, qualitative feedback highlighted the need for more performance tasks with better guidance, feedback, and objective grading. It seems that though students acknowledge the value of performance tasks, the benefits of traditional assessment formats such as objectivity and ease outweigh this when it comes to preference. The study underscores the importance of continually measuring student opinions and incorporating students' feedback into assessment strategies to foster an engaging and effective learning environment. Future research should explore students' perceptions of performance tasks across different disciplines and investigate the impact of these preferences on their learning outcomes and academic performance.

Keywords: assessment formats, performance tasks, student preferences, group assignments

The Role of Performance Tasks in Higher Education Assessment: A Study of Students' Preferences

In the evolving landscape of higher education, assessment remains a cornerstone of the learning process. It not only evaluates students' understanding of course material but also helps to improve teaching methods and fosters student development (Boud & Falchikov, 2007). While traditional methods like multiple-choice and open-question exams have long been the standard in the academic world, educators and researchers have increasingly recognized their limitations in measuring higher-order thinking skills and real-world application of knowledge (Freeman et al., 2014; Krathwohl, 2002).

This recognition has sparked a shift from a testing culture to an assessment culture (Birenbaum, 1994). Exams are now used as tools for learning, not just for assigning grades, leading to a growing interest in alternative and holistic evaluation methods (Dochy & McDowell, 1997). Performance tasks, which involve activities like experiments, essays, or presentations, have emerged as a promising method (Braun, 2019; Darling-Hammond & Adamson, 2010). This study investigates students' opinions on performance tasks in higher education, specifically compared to multiple-choice and open-question exams.

Benefits and Barriers of the Implementation of Performance Tasks

The appeal of performance tasks lies in their ability to assess students' skills more thoroughly, going beyond mere recall of information to evaluate their thinking, communication, and teamwork abilities (Pellegrino et al., 2001). Performance tasks thus give students the opportunity to show their academic performances in a more comprehensive manner. Moreover, these tasks potentially allow students to not only demonstrate their understanding of course concepts but also to develop skills important for their future work field (Darling-Hammond & Adamson, 2010).

Despite their potential benefits, however, the use of performance tasks in higher education remains relatively limited, with many instructors sticking to more traditional assessment formats (Torrance & Pryor, 2001). This hesitancy might stem from multiple factors, including concerns about how complicated and resource-heavy it might be to design and implement performance-based assessments (Wren, 2019). Moreover, some benefits of traditional assessment formats are lost with performance tasks, such as the objective nature and simple grading process of a multiple-choice exam. Additionally, teachers may face institutional obstacles and lack the support or training needed to integrate performance tasks effectively (Moss et al., 2006).

The Importance of Students' Preferences

Besides the educational benefits and barriers, students' preferences and perceptions of assessment formats should also play a crucial role in whether and how performance tasks are implemented in higher education (Smith et al., 2013). Students' attitudes towards different assessment formats can significantly impact their motivation, engagement, and overall learning experience (Dochy et al., 1999; Entwistle, 1991). A study by Lizzio et al. (2002) supported this sentiment, finding that students' perceptions of their learning environment were a more accurate predictor of university performance than prior school achievements. It seems that understanding students' preferences and perceptions is essential for designing assessments that not only accurately measure learning outcomes, but also encourage meaningful engagement and support student success (Brown, 2004).

Students' Opinions on Different Exam Formats

The body of research on students' opinions of performance tasks is not as wide as that of opinions on multiple-choice and open-question exams. Research on the latter formats consistently shows that students prefer multiple-choice exams over open-question ones. Often mentioned reasons for this preference are that multiple-choice exams are graded more

objectively and are easier to prepare for and pass (Holzinger et al., 2020; Sambell et al., 1997; Zeidner, 1987). While students' preference thus lies with multiple-choice exams, some benefits of open-question exams are acknowledged. Students often indicate that the open-ended format leads to more engagement with and better long-term retention of the material (Holzinger et al., 2020). In line with this, Zeidner (1987) found that high school students perceived open-question exams as more valuable to their education than multiple-choice exams.

The research on performance tasks is scarcer, but students generally highlight benefits like better long-term retention, engagement, and comprehension (Sambell et al., 1997; Struyven et al., 2005). A review by Struyven et al. (2005) found that students perceive traditional assessments to be arbitrary and irrelevant, whereas alternative assessments like performance tasks were seen as meaningful and leading to high-quality learning.

The fact that students acknowledge these benefits does not necessarily mean they prefer performance tasks over other assessment formats, however. As can be seen in the multiple-choice versus open questions debate, students prefer formats they perceive to be easier (Holzinger et al., 2020; Sambell et al., 1997; Zeidner, 1987). In a qualitative study by Sambell et al. (1997), some students suggested that they simply did not have the time or motivation to put in the effort required for alternative assessments. Moreover, opinions on performance tasks are not free of criticism, with students indicating a need for clearer guidelines and expectations, fair and consistent grading, and more timely and constructive feedback (Carless, 2006; Jonsson & Svingby, 2007; Smith & Colby, 2007; Williams, 1992).

All things considered, there is no clear answer to the question of what students prefer when alternative assessments are considered. Though we know students generally prefer multiple-choice exams over open-question exams, research directly investigating where performance tasks are placed in this ranking is lacking. Since performance tasks show great

potential to enhance students' learning experiences, it is of interest to gain more insight into students' perceptions of this assessment format.

Individual and Group Assignments

While the concept of multiple-choice and open-question exams is more straightforward, performance tasks can be implemented in a variety of ways. Not only are there different types of tasks, but another decision is whether students perform these tasks alone or in collaboration with others. In previous research, students often mention specific benefits of group assignments, like learning from others, improving communication skills, sharing the workload, and developing critical thinking skills (Ford & Morice, 2003; Hassanien, 2007). Most studies conclude that group assignments are generally a positive learning experience for students (Bentley & Warwick, 2013; Hassanien, 2007). Dijkstra and Zwaal (2020) found that in both of their two samples of hotel school students, around 70% preferred a group assignment over an individual written exam. Moreover, when asked about the balance between group and individual assignments, half of the students indicated that they preferred more group work (51%), while 22% emphasized they preferred working individually, and 27% preferred an equal distribution. There are also studies that yield less conclusive results, however. In Bentley and Warwick's (2013) sample of students at a UK university, 51% of students preferred group assignments, while 49% preferred individual assignments, indicating no clear overall preference in this sample. In another study, 79% of students initially preferred a group marketing project over an individual one. This percentage dropped to 53 at the end of the semester, however, mainly due to logistical issues. This means that ultimately, only a very small majority preferred the group project (Ryan & Ogilvie, 2005).

Moreover, students are also in agreement about the downsides of group projects. For one, while group assignments are by definition worked on collaboratively, students might be

graded either individually or as a group. Fairness is not ensured for group grades, since some students might let others do the work for them. This tendency is called “free riding,” and it is the issue that students most often cite concerning group assignments (Benning, 2022; Ford & Morice, 2003; LaBeouf et al., 2016). Other often-mentioned problems are poor communication, poor attendance at group meetings, and inequality of members’ skills, among others (Ford & Morice, 2003; Hassanien, 2007; LaBeouf et al., 2016). These issues may outweigh the benefits of group work; that is, transform it from a positive experience into an anxiety-inducing one (LaBeouf et al., 2016; Strauss et al., 2011). Group work can thus be preferred or avoided, and this seems to depend on the balance between its inherent positive and negative aspects.

Current Study

As has been discussed, insight into students’ assessment preferences is crucial to making education as valuable and enjoyable as possible. Still, research into opinions on performance tasks is lacking when compared to the literature on multiple-choice and open-question exams. The present study seeks to address this gap in the literature by directly comparing students’ preferences for multiple-choice exams, open-question exams, and performance tasks as assessment formats in higher education. As was mentioned, previous research shows a general preference for multiple-choice exams over open-question exams, as multiple-choice exams are perceived to be less difficult and more objective (Holzinger et al., 2020; Sambell et al., 1997; Zeidner, 1987). Given this seemingly robust preference for multiple-choice exams and considering the higher effort and lower objectivity associated with performance tasks, we expect multiple-choice exams to be preferred. This leads to our first hypothesis:

Hypothesis 1. Students prefer multiple-choice exams over both open-question exams and performance tasks.

Due to a lack of supporting literature, we do not hypothesize which format is preferred between performance tasks and open-question exams. The constructs of objectivity, potential to show performance, difficulty, and valence (i.e., overall liking) are explored to gain insight into students' preferences, as all these themes have previously been shown to influence preference (Darling-Hammond & Adamson, 2010; Holzinger et al., 2020; Lindner et al., 2018; Sambell et al., 1997; Zeidner, 1987).

In addition to assessing preferences for assessment formats, our study also examines students' attitudes towards individual and group assignments, in hopes of gaining a deeper understanding of students' attitudes towards different types of performance tasks. Previous research suggests students are generally positive about group assignments (Bentley & Warwick, 2013; Ford & Morice, 2003; Hassanien, 2007), and studies directly measuring preferences found either a preference for group assignments or no clear preference (Bentley & Warwick, 2013; Dijkstra & Zwaal, 2020; Ryan & Ogilvie, 2005). Based on this, our second hypothesis is as follows:

Hypothesis 2. Students prefer group assignments over individual assignments.

Specific constructs are explored for this hypothesis as well, including fairness, valence, value, and anxiety, as in the literature, each of these constructs has been suggested to affect students' opinions of assignment formats (Benning, 2022; Ford & Morice, 2003; Hassanien, 2007; Labeouf et al., 2016; Strauss et al., 2011).

Furthermore, our study includes an open-ended question inviting students to provide suggestions for improving the implementation of performance tasks in higher education, to gather students' views on the practical issues and challenges associated with performance-based assessment. The aim of this question is to explore options rather than to support a certain hypothesis, but based on previously mentioned findings, we expect students to mention factors like more guidance from faculty, a fair grading procedure, and more

feedback, among others (Carless, 2006; Jonsson & Svingby, 2007; Smith & Colby, 2007; Williams, 1992).

Method

Sample

The study involved students from the faculty of Behavioral and Social Sciences at the University of Groningen. Initially, 143 students were sampled, but 12 did not complete the survey. Throughout the survey, five attention checks were included. Three participants failed two or more checks. Those who did not complete the survey or failed the attention checks were excluded, resulting in a final sample of 128 students for analysis.

The sample mainly consisted of Psychology students (EN: 37.5%, NL: 61.7%), plus a small number of Sociology students (0.8%). The sample was predominantly female (72.7%), followed by male (23.4%), non-binary / third gender (3.1%), and those preferring not to say (0.8%). Most students were in their first year of studying (64.8%), aligning with expectations due to recruitment primarily via Sona Systems (<https://www.sona-systems.com/>).

Incentive

First-year Psychology students received 0.6 course credits via Sona Systems for participating in the study, which count towards passing the first-year course "A Practical Introduction to Research Methods." Higher-year students did not receive participation incentives.

Procedure

Participants were recruited via convenience sampling and accessed the survey through Sona Systems. Additionally, the study link was shared in WhatsApp group chats to obtain participants in higher years of study. Researchers were not present during survey completion; participants were asked to complete the survey online independently. The survey ran from April 16th to April 24th, 2024.

Before starting, participants were briefed on the research's purpose, their voluntary participation, incentives offered, and data confidentiality. Contact information for researchers was provided for any possible questions. Participants then gave informed consent to participate in the study and allow their data to be processed.

The questionnaire started with demographic questions, including gender, year of study, and subject of study, 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 each individual collaborator.

Following demographics, participants answered questions on personality type, trait test anxiety levels, and exam format preferences (open-question exams, multiple-choice exams, performance tasks in general, and group / individual performance tasks specifically). Experience with and learning strategies for each format were assessed, along with state evaluation anxiety levels specific to each exam type. The order of these sections was randomized for each participant. Upon completion, participants were redirected to Sona Systems to receive credits, if eligible.

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

An Examination Preference Inventory (EPI) was developed to assess preferences for various examination methods. Presented separately for each type, it comprised four scales with two to four items each. Participants indicated their level of agreement with displayed

items on a five-point Likert scale, ranging from *strongly disagree* to *strongly agree*. Three scales were adapted from an existing exam preference inventory by Lindner et al. (2018): potential to show performance, objectivity, and valence. The fourth scale measured difficulty and was adapted from Zeidner's (1987) exam preference inventory. Most scales were slightly modified, either by adding items to improve construct validity or by revising existing items to prevent misinterpretations. An example item of the potential to show performance scale is "Open-question exams / Multiple-choice exams / Performance tasks allow me to express my knowledge precisely." Appendix A displays the complete inventory.

Subscale means were combined to form the general preference score. Reliability, assessed using Cronbach's alpha, was deemed acceptable ($\alpha > .70$), good ($\alpha > .80$), or excellent ($\alpha > .90$) for most scales, though the reliability of the valence scale for performance tasks was questionable ($\alpha > .60$; Cronbach, 1951). The reliability of all scales combined as a measure of preference was acceptable (see Table 1).

Table 1

Cronbach's Alpha Scores for the EPI Scale and Subscales

Scale	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

Individual Versus Group Assignments

A self-developed Assignment Preference Inventory (API) was used to assess preferences for individual and group assignments, as no existing inventory was available. Students completed the API twice, once for each assignment type (individual and group).

The API included four constructs: fairness, valence, value, and anxiety, each with three Likert-scale items measuring agreement from *strongly disagree* to *strongly agree*. For instance, a valence scale item was: "Individual / Group assignments are enjoyable to work on." Appendix B contains the complete inventory.

Cronbach's alpha evaluated scale reliability. Most alphas were acceptable ($\alpha > .70$), good ($\alpha > .80$), or excellent ($\alpha > .90$), except for the fairness scale for group assignments, which showed questionable reliability (Cronbach, 1951). Results dependent on this scale should thus be interpreted with caution. The reliability of the total preference scale was good (see Table 2).

Finally, an open question was added in which students were asked to explain their ideas on how the implementation of performance tasks in higher education can be improved. Answers were limited to 400 characters.

Table 2

Cronbach's Alpha Scores for Reliability of the API

Scale	Individual assignments	Group assignments
	α	α
Fairness	.70	.66
Valence	.90	.92
Value	.75	.81
Anxiety	.87	.86
Total	.86	.86

Results

General Preference

In line with our first hypothesis, a significantly higher preference score was found for multiple-choice exams ($M = 3.45$, $SD = .54$) in comparison to both open-question exams ($M = 3.01$, $SD = .46$), $t(127) = 6.21$, $p < .001$, $d = 0.55$, 99.5% CI [0.24, 0.63] and performance tasks ($M = 3.00$, $SD = 0.48$), $t(127) = 6.16$, $p < .001$, $d = 0.54$, 99.5% CI [0.25, 0.65].¹² Both effect sizes were medium ($0.50 \geq d \leq 0.80$; Cohen, 1988).

The difference between the means of performance tasks ($M = 3.00$, $SD = 0.48$) and open-question exams ($M = 3.01$, $SD = .46$) was negligible and significance testing was thus not conducted. These results suggest no clear preference between open-question exams and performance tasks.

Subscales of General Preference

Subscale scores were explored to gain insight into the general preference findings. The means of the four subscales for each examination format can be found in Table 3.

Respondents placed performance tasks (PTs) in between open-question (OQ) exams and multiple-choice (MC) exams on all scales except objectivity, where they scored lowest.

¹Although a repeated-measures ANOVA might have been more suitable for this research, the researchers' curriculum did not cover this method adequately. Paired sample t-tests were thus ran. The assumptions of independence between subjects, same subject paired measurements, and normal distribution of differences were met, but some outliers were found. While a Kruskal-Wallis test might thus have been more appropriate, it was also not covered in the curriculum. The results of the paired sample t-tests should thus be interpreted with caution.

² To reduce the risk of a type I error due to multiple tests, the original alpha value of .05 was adjusted to .005 using the Bonferroni correction, also considering the t-tests for the subscales.

Table 3*Means and Standard Deviations of the Different Scales of the EPI*

Subscale	Multiple-choice		Open question		Performance tasks	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
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.

To test for significance, two-sided paired sample t-tests were performed for the differences in subscale scores between PTs and MC exams and PTs and OQ exams. Differences between MC and OQ were not tested, as this was not relevant to the current research.

PTs vs MC Exams

The results of the tests comparing PTs and MC exams are shown in Table 4. PTs were rated more positively than MC exams on just one subscale: potential to show performance. This difference was found to be significant with a small effect size. PTs were rated less positively than MC exams on all other subscales (i.e., lower on valence and objectivity, higher on difficulty). These differences were also significant, with small effect sizes for difficulty and valence, but a very large effect size for objectivity. Overall, PTs were thus rated as having significantly greater potential to show performance, but also as being more difficult, less liked (valence), and especially less objective than MC exams.

Table 4*Paired Sample T-Test Results for Subscale Differences Between PT and MC Exams*

Subscale	Mean difference	SD	<i>t</i> (127)	Two-sided <i>p</i>	Cohen's <i>d</i>	99.5% CI of the difference	
						Lower	Upper
Difficulty	0.32	0.86	4.20	<.001	0.37	0.10	0.53
PSP	0.60	1.26	5.43	<.001	0.48	0.28	0.92
Objectivity	-1.82	1.24	-16.56	<.001	-1.46	-2.13	-1.51
Valence	-0.55	1.41	-4.46	<.001	-0.39	-0.91	-0.20

Note. The subscale Potential to show performance was abbreviated to PSP to improve the table's readability.

PTs vs OQ Exams

Different results were found for the paired sample t-tests between the subscales scores of PTs and OQ exams, as can be seen in Table 5. PTs were seen as significantly less difficult than OQ exams, with a large effect size. Moreover, PTs were rated more positively on valence, this effect being small. OQ exams were given a higher score on objectivity and potential to show performance, though only the difference in objectivity was significant, with a small effect size. All in all, PTs were thus rated as being liked more (valence), as well as being less difficult than OQ exams, but were also found to be less objective.

Table 5*Paired Sample T-Test Results for Subscale Differences Between PT and OQ Exams*

Subscale	Mean difference	SD	<i>t</i> (127)	Two-sided <i>p</i>	Cohen's <i>d</i>	99.5% CI of the difference	
						Lower	Upper
Difficulty	-0.81	0.67	-13.72	<.001	-1.21	-1.54	-0.89
PSP	-0.23	0.96	-2.76	.007	-0.24	-0.50	0.008
Objectivity	-0.43	1.16	-4.15	<.001	-0.37	-0.62	-0.11
Valence	0.25	0.94	2.99	.003	-0.26	0.01	0.52

Note. The subscale Potential to show performance was abbreviated to PSP to improve the table's readability.

Individual and Group Assignments

General Preference

General preference scores for individual and group assignments were calculated based on the API. The mean score for group assignments was lower ($M = 3.08$, $SD = .68$) than the mean score for individual assignments ($M = 3.41$, $SD = .62$). Since these preliminary results already contradicted the hypothesis that group assignments would be rated more favorably than individual assignments, further significance testing was not conducted.

Subscales of General Preference

Since general preference was not tested on significance, neither were the differences in subscale scores between the two assignment formats, as these were meant to support the main hypothesis. However, means and standard deviations for each subscale are still reported for further exploration (see Table 6). Individual assignments were given a more favorable score on fairness, valence, and value, while the higher score on anxiety was unfavorable.

Table 6*Means and Standard Deviations of the Different Scales of the API*

Subscale	Individual assignments		Group assignments	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Fairness	3.74	0.71	2.67	0.79
Valence	3.38	0.98	3.22	1.04
Value	4.02	0.70	3.75	0.90
Anxiety	3.52	0.95	3.31	1.03

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

Ideas on the Improvement of PTs

The open question on improving PT implementation in higher education was analyzed using thematic analysis, largely according to the guidelines of Braun and Clarke (2006). Responses were read and initially coded according to key ideas and words. The codes were then put into themes. What was determined to be a theme was based on how frequently a certain idea was mentioned. The cohesion of the responses within each theme was checked, and potential subthemes were identified.

There were 35 participants who either did not respond or gave nonsensical answers. Another 21 responses did not fit into a main theme and were deemed not relevant or not frequently mentioned enough to solicit a separate theme. This gave us a total of 56 unclassified responses.

Ultimately, responses from 72 participants were categorized into seven themes, which can be found in Table 7, along with example items. The biggest themes were frequency ($n =$

23), guidance and feedback ($n = 17$), grading ($n = 15$), and diversity ($n = 12$). Frequency included general calls for more PTs ($n = 9$), more individual assignments ($n = 10$), and more group assignments ($n = 4$). Guidance and feedback responses emphasized the need for clearer guidelines and expectations and more chances to receive help or feedback. For instance, one participant stated, “Performance assignments work as the basis of a lot of skills. Better instructions, feedback as well as advice on how to approach the procedure could prove to be highly effective.” Grading responses highlighted the need for objective grading criteria, suggesting stricter criteria ($n = 7$) and individual grading in group assignments ($n = 8$). The following quote is an example of a response put into the latter subtheme.

A lot of stress comes from the fact that oftentimes group members do not put an equal amount of effort into the project. It would be a good idea to maybe grade the group members individually, to make sure everyone pulls their weight and to avoid social loafing.

The diversity theme contained responses that mentioned using multiple forms of assessment. The main sentiment in this theme was the need for a better balance of the different types of PTs used.

Smaller themes included creativity ($n = 6$), practicality ($n = 5$), and reducing anxiety ($n = 4$). Students indicated wanting more creative freedom, such as choosing topics or formats. One participant said, “More focus on creativity rather than following strict guidelines would motivate students and make the work more enjoyable.” Suggestions for practicality included making tasks more applicable to real-life (work) situations. Lastly, some participants highlighted the need to reduce anxiety around PTs, though specific solutions were not always mentioned.

Table 7*Themes and Examples of Ideas to Improve PTs*

Theme	Example quotes
Frequency	<p>“I think including more shorter performance tasks could be a good idea.”</p> <p>“Doing more group projects.”</p>
Guidance and feedback	<p>“Maybe giving us more instructions or more time to work on it.”</p> <p>“Providing clear guidance on what is expected, and opportunities for getting help / asking questions to other students (like a group) or teachers (before or during the assessment, depending on if it’s an exam or a longer form assignment).”</p>
Grading	<p>“It can be improved by integrating individual component assessment into group performance assessments.”</p> <p>“Very strict rules about the grading, so the chances of it being subjective or biased are very small.”</p>
Diversity	<p>“I think you should implement a more equal use of the different performance tasks.”</p> <p>“Having variations in the types of assessment used”</p>
Creativity	<p>“By letting students pick a subject of their interest to give them the opportunity to do more research and gain more information about it. This will also help them in finding what they are interested in to work with in the future.”</p>

Note. Table 7 continues on the next page.

Table 7 (Continued)*Themes and Examples of Ideas to Improve PTs*

Theme	Example quotes
Creativity	“Doing it in a fun way about topics that are actually interesting.”
Practicality	“I think it's important to use a variety of assessment formats, especially practical and theoretical. We learn a lot of theory which is of course important, but a lot of the time people don't get to learn how to put that theory into practice before starting on a job.”
Reducing anxiety	“More practical tasks, less theoretical.” “Help students reduce anxiety about presentation and help them be confident in front of a group of people.” “Reduce the relevance for assessment to reduce stress and anxiety. Focus on skills rather than grading and give more advise / supervision.”

Discussion

The objective of this study was to gain a better understanding of students' perceptions of PTs as an assessment format within higher education, as research on this topic is scarce in comparison to that on traditional formats such as MC and OQ exams.

General Preference

Firstly, we investigated and compared general preferences for MC exams, OQ exams, and PTs to determine where PTs rank. Our findings showed a higher preference score for MC exams over OQ exams and PTs, supporting our hypothesis that MC exams would be preferred

over both formats. The difference in preference scores between PTs and OQ exams was negligible, indicating no overall preference between these formats among our sample.

Analysis of the subscales of general preference gives us more insight into our results. It seems that the biggest factor adding to the preference for MC exams over PTs is objectivity since this difference had a very large effect size. MC exams were by far found to be the most objective, besides also scoring lowest on difficulty and highest on valence. PTs scored higher on potential to show performance. These results align with existing literature, where MC exams are preferred for their objectivity and ease (Holzinger et al., 2020; Sambell et al., 1997; Zeidner, 1987), and PTs are valued for showcasing performance and engagement (Sambell et al., 1997; Struyven et al., 2005). However, the prevailing overall preference for MC exams in our study suggests that PTs' higher potential to show performance does not weigh up to the perceived benefits of MC exams.

When comparing PTs with OQ exams, the benefits of both formats seem to weigh each other out: while PTs were found to be less difficult and more liked overall, OQ exams were found to be more objective. There was no significant difference in potential to show performance. These results were somewhat unexpected, given the emphasis put on PTs' potential to show performance in the literature (Sambell et al., 1997; Struyven et al., 2005). However, students have previously indicated similar benefits of OQ exams, though this was specifically compared to MC exams (Holzinger et al., 2020). Furthermore, students may fail to recognize the superior potential of PTs in showcasing performance over OQ exams if this potential is not effectively realized in practice through proper task design.

Moreover, it is interesting that the largest effect size was for the difference in difficulty. Students seem to find PTs less difficult than OQ exams. Initially, this might seem counterintuitive, as PTs often require keeping up with the study material and in total might require more work and thought from students than simply studying for an exam. However, the

perceived difference in difficulty might stem from the fact that if students do put in the required work and follow the guidelines, success is often more guaranteed for PTs than for OQ exams, as not knowing the exam questions beforehand adds an element of uncertainty.

Individual Versus Group Assignments

Secondly, we compared students' opinions on individual and group assignments, as whether a PT is executed individually or collaboratively might influence students' perception of the format. Counter to our hypothesis that students would prefer group assignments, we found a higher general preference score for individual assignments. Moreover, individual assignments were rated more positively on the subscales of fairness, valence, and value, while only receiving a less favorable score on anxiety. Although we did not further test these results for significance, it is clear that our sample of students did not show a preference for group assignments. This result was surprising, as we concluded from our literature review that students are generally positive about group assignments, with the few studies directly comparing preference between individual and group assignments either finding students to prefer group assignments or finding no clear overall preference (Bentley & Warwick, 2013; Dijkstra & Zwaal, 2020; Ryan & Ogilvie, 2005). It could be that our literature search was not extensive enough, and that research underlining a preference for individual assignments has been published but was simply not found. Upon further inspection, however, scientific research indicating a clear preference for individual assignments could still not be found. Another explanation for our findings might be our sample. It could very well be that different types of students prefer different types of assignments. We based our hypothesis on research such as that of Dijkstra and Zwaal (2020) and Ryan and Ogilvie (2005), who respectively found that hotel school and marketing students showed a preference for group tasks. Assignments might take a different shape for students of the social sciences, and this could

influence the suitability and enjoyability of group assignments. Moreover, statistical limitations such as subscales with questionable reliability could have influenced our results.

Ideas on the Improvement of PTs

Lastly, our study included a qualitative component, inviting students to share their opinions on how the implementation of PTs in higher education can be improved. The main themes determined were frequency, guidance and feedback, grading, and diversity. Students indicated generally wanting more PTs, but needing better guidance and feedback, as well as an objective form of grading and a good balance of the types of PTs used. A smaller number of students articulated a desire for PTs to exhibit greater creativity and practical relevance, aligning more closely with their future work fields. Additionally, some participants expressed the need to reduce the anxiety surrounding PTs. However, it should be noted that, due to the small sample size, even the largest categories consisted of a relatively small number of students, and these results should thus be interpreted with caution.

The determined themes have all been mentioned in the literature previously, and it thus seems student opinions on the limitations of PTs are quite consistent (Carless, 2006; Jonsson & Svingby, 2007; LaBeouf et al., 2016; Smith & Colby, 2007; Strauss et al., 2011; Williams, 1992). It is interesting that, although our sample of students showed a general preference for MC exams, the biggest theme that emerged here was the indication that PTs should be used more frequently. It seems that students recognize the value of using multiple forms of assessment, rather than solely employing their format of preference.

Another parallel can be drawn with regard to our second hypothesis. When prompted by the open question, a greater number of students expressed a preference for more individual assignments compared to those advocating for more group assignments. This strengthens our finding that, in our sample, most students did not prefer group assignments over individual assignments.

Limitations and Strengths

This study has several limitations that should be noted. Firstly, due to limited statistical expertise, we opted to use a statistical method that might not have been the most suitable. Additionally, while most scales showed good reliability, some exhibited only acceptable or questionable reliability. These statistical limitations suggest the results should be interpreted with caution. Moreover, although subscales were chosen based on existing literature, they remain arbitrary; different subscales might have yielded different results. Specifically, the valence subscale could be problematic, because it is not a distinct construct inherent to exam formats that might independently influence overall preference. Instead, it closely resembles the concept of preference itself and thus overlaps with it, which might limit its suitability for further explaining preference. Furthermore, the sample was limited to Psychology and Sociology students, primarily first-years, restricting the generalizability of our findings. Lastly, 56 students did not provide classifiable answers to the open question, which might have influenced the results.

Despite these limitations, the study has notable strengths, including its comprehensive approach that combines quantitative and qualitative methods to provide a thorough understanding of students' perceptions and preferences. Additionally, our study has an innovative focus. By investigating PTs, an area that is less researched compared to traditional assessment formats, we contribute valuable insights. Furthermore, our analysis of perceptions on group versus individual assignments offers a deeper understanding of opinions on the specific types within this broad category.

Implications

The results of this study contribute to the existing literature on assessment preferences in higher education, particularly by addressing the gap in research on PTs. Educators and researchers mainly praise PTs for their inherent engagement with the material and their ability

to assess students' performance comprehensively (Darling-Hammond & Adamson, 2010; Pellegrino et al., 2001). Our results suggest, however, that this quality does not largely influence student's preference for the format, as this benefit was acknowledged, but MC exams were still preferred overall. This implies that students' preferences are heavily influenced by the immediate benefits of traditional assessments, such as simplicity and clear-cut grading criteria. Moreover, previous research indicates PTs have greater potential to show performance than traditional assessment formats, including open-question exams (Darling-Hammond & Adamson, 2010; Struyven et al., 2005), while our sample of students did not. This finding underscores the importance of educators focusing on designing PTs effectively to realize their full potential.

These insights can help educators reach the goal of designing assessment formats that not only measure learning outcomes accurately but also align with student's preferences and motivations. The strong preference for MC exams indicates that any shift towards incorporating more PTs should be accompanied by efforts to enhance their perceived objectivity and manageability. This could include providing clear guidelines, fair and consistent grading rubrics, and timely, constructive feedback, as was echoed in students' ideas of how PTs can be improved.

Furthermore, while we cannot definitively say if students prefer individual assignments over group assignments, the surprising finding that group assignments were not preferred supports the idea that different samples of students prefer different types of assessment. This highlights the importance of continually measuring student preferences, as they may be unexpected, but likely impact students' learning quality and enjoyment (Dochy et al., 1999; Entwistle, 1991).

The qualitative feedback highlights specific areas for improvement in the implementation of PTs. Our found themes were in line with the existing literature,

emphasizing the prevalence of these themes. Students' calls for more frequent use of PTs, coupled with better guidance, feedback, and fair grading, underline the need for robust support systems. Providing diverse types of PTs that are creative and aligned with students' future professional contexts could also enhance their practical relevance and acceptance. It is crucial to listen to students not only in matters of exam preference but also in practical considerations like these, as they are the ones performing the tasks and thus the experts on how to enhance them.

Conclusions and Recommendations

Overall, the study underscores the importance of understanding and addressing students' preferences and perceptions when designing assessment strategies. The preference for traditional assessment formats seems deeply ingrained in students. While we do not completely discourage the use of formats such as MC and OQ exams, as they provide a fast, efficient, and mostly objective way to measure the academic achievements of large numbers of students, we believe that PTs add value to higher education by capturing skills that are important for students' futures. A shift in the balance between the different exam formats could thus be recommended, with PTs being implemented more. In this process, educators should focus on incorporating feedback on performance assessments to satisfy students' needs. This approach will foster a more engaging and effective learning environment that not only meets educational standards but also supports students' academic success and professional development.

Future research should explore the use of PTs across a wider range of disciplines to enhance the generalizability of the findings, and to explore to what extent different types of students have different preferences. Additionally, different types of PTs besides group and individual assignments could be examined, as opinions are likely to differ between specific PTs. Finally, more research on how exactly students' preferences influence their learning and

academic performance would provide valuable insights and possibly strengthen the view that student opinions matter.

Acknowledgments

During the preparation of this work, the author used ChatGPT (OpenAI, 2024) to refine the quality and readability of the final manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Prompts and generated outputs can be found in Appendix C.

Furthermore, it is important to acknowledge that while most sections of this thesis were written individually, a portion of the work was conducted collaboratively. Therefore, the terms "we" and "our" are employed throughout the text for the sake of consistency.

References

- Benning, T. M. (2022). Reducing free-riding in group projects in line with students' preferences: Does it matter if there is more at stake? *Active Learning in Higher Education*, 146978742211188. <https://doi.org/10.1177/14697874221118864>
- Bentley, Y., & Warwick, S. (2013). Students' experience and perceptions of group assignments. *Journal of Pedagogic Development*, 3(3), 11-19. <https://www.beds.ac.uk/jpd/volume-3-issue-3/an-investigation-into-students-perceptions-of-group-assignments>
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge. <https://doi.org/10.4324/9780203964309>
- 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>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318. <https://doi.org/10.1080/0969594042000304609>
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/bf02310555>

- Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.
https://globaled.gse.harvard.edu/files/geii/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-report_0.pdf
- Dijkstra, R., & Zwaal, W. (2020). Students motives for choosing between an individual test or a group assignment. *Research in Hospitality Management*, 10(2), 75–83.
<https://doi.org/10.1080/22243534.2020.1869464>
- 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](https://doi.org/10.1016/s0191-491x(97)86211-6)
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331–350.
<https://doi.org/10.1080/03075079912331379935>
- Entwistle, N. (1991). Approaches to learning and perceptions of the learning environment. *Higher Education*, 22(3), 201–204. <https://doi.org/10.1007/bf00132287>
- Ford, M., & Morice, J. (2003). How Fair are Group Assignments? A Survey of Students and Faculty and a Modest Proposal. *Journal of Information Technology Education*, 2, 367–378. <https://doi.org/10.28945/335>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8410–8415. <http://dx.doi.org.proxy-ub.rug.nl/10.1073/pnas.1319030111>

- Hassanien, A. (2007). A Qualitative student evaluation of group learning in higher education. *Higher Education in Europe*, 32(2–3), 135–150.
<https://doi.org/10.1080/03797720701840633>
- Holzinger, A., Lettner, S., Steiner-Hofbauer, V., & Melsner, 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>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
<https://doi.org/10.1016/j.edurev.2007.05.002>
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- LaBeouf, J. P., Griffith, J. C., & Roberts, D. (2016). Faculty and Student Issues with Group Work: What is Problematic with College Group Assignments and Why? *Journal of Education and Human Development*, 5(1). <https://doi.org/10.15640/jehd.v5n1a2>
- Lindner, M. A., Mayntz, S. M., & Schult, J. (2018). Studentische Bewertung und Präferenz von Hochschulprüfungen mit Aufgaben im offenen und geschlossenen Antwortformat. *Zeitschrift Für Pädagogische Psychologie*, 32(4), 239–248.
<https://doi.org/10.1024/1010-0652/a000229>
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Chapter 4: Validity in educational assessment. *Review of Research in Education*, 30(1), 109–162.
<https://doi.org/10.3102/0091732x030001109>
- OpenAI. (2023). *ChatGPT* (June 2023 version) [Large language model].
<https://openai.com/chatgpt/>

- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.
<https://doi.org/10.17226/10019>
- Ryan, M. M., & Ogilvie, M. (2005). The preference for group work - not always the case: A case study. In E. Manalo & G. Wong-Toi (Eds.), *Communication skills in university education: The international dimension*. (pp. 150–158). Pearson Education New Zealand. <https://ro.ecu.edu.au/ecuworks/2456/>
- Sambell, K., McDowell, L., & Brown, S. (1997). “But is it fair?”: An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349–371. [https://doi.org/10.1016/s0191-491x\(97\)86215-3](https://doi.org/10.1016/s0191-491x(97)86215-3)
- Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students ‘assessment literacy.’ *Assessment & Evaluation in Higher Education*, 38(1), 44–60.
<https://doi.org/10.1080/02602938.2011.598636>
- Smith, T. W., & Colby, S. A. (2007). Teaching for Deep Learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 80(5), 205–210.
<https://doi.org/10.3200/TCHS.80.5.205-210>
- Strauss, P., Alice, U., & Young, S. (2011). ‘I know the type of people I work well with’: student anxiety in multicultural group projects. *Studies in Higher Education*, 36(7), 815–829. <https://doi.org/10.1080/03075079.2010.488720>
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students’ perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30(4), 325–341. <https://doi.org/10.1080/02602930500099102>

- Torrance, H., & Pryor, J. (2001). Developing Formative Assessment in the Classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615–631. <https://doi.org/10.1080/01411920120095780>
- Williams, E. (1992). STUDENT ATTITUDES TOWARDS APPROACHES TO LEARNING AND ASSESSMENT. *Assessment & Evaluation in Higher Education*, 17(1), 45–58. <https://doi.org/10.1080/0260293920170105>
- Wren, D. G. (2019). *Assessing Deeper Learning: Developing, Implementing, and Scoring Performance Tasks*. Rowman & Littlefield. <https://doi-org.proxy-ub.rug.nl/https://rowman.com/ISBN/9781475845778#>
- 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 and their corresponding items. Participants rate each statement on a five-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Difficulty

- 1) Open-question exams / Multiple-choice exams / Performance tasks are complex.
- 2) Open-question exams / Multiple-choice exams / Performance tasks are easy.
- 3) Open-question exams / Multiple-choice exams / Performance tasks are challenging.

Potential to Show Performance

- 4) Open-question exams / Multiple-choice exams / Performance tasks give me the opportunity to show that I have understood the subject matter very well.
- 5) Open-question exams / Multiple-choice exams / Performance tasks give me the opportunity to show that I know more than other students.
- 6) Open-question exams / Multiple-choice exams / Performance tasks allow me to express my knowledge precisely.
- 7) Open-question exams / Multiple-choice exams / Performance tasks are an appropriate examination format for important exams.

Objectivity

- 8) Open-question exams / Multiple-choice exams / Performance tasks are evaluated objectively.
- 9) Open-question exams / Multiple-choice exams / Performance tasks are graded without bias.

Valence

- 10) Open-question exams / Multiple-choice exams / Performance tasks should be the main method of examination.
- 11) Open-question exams / Multiple-choice exams / Performance tasks are interesting.
- 12) Open-question exams / Multiple-choice exams / Performance tasks are liked by me.

Appendix B

Assignment Preference Inventory

This appendix comprises the four scales of the Assignment Preference Inventory and their corresponding items. Participants rate each statement on a five-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Fairness

- 1) Individual / Group assignments lead to a grade representative of effort.
- 2) Individual / Group assignments are a fair form of assessment.
- 3) Individual / Group assignments are evaluated objectively.

Valence

- 4) Individual / Group assignments are enjoyable to work on.
- 5) Individual / Group assignments are interesting to do.
- 6) Individual / Group assignments are liked by me.

Value

- 7) Individual / Group assignments add value to a course.
- 8) Individual / Group assignments are able to teach skills not covered by other forms of assessment.
- 9) Individual / Group assignments are indispensable (necessary) in higher education.

Anxiety

- 10) Individual / Group assignments are anxiety inducing to work on.
- 11) Individual / Group assignments lead to high stress levels.
- 12) Individual / Group assignments lead to feelings of dread.

Appendix C

ChatGPT Prompts and Outputs

This appendix includes ChatGPT (OpenAI, 2024) prompts and their corresponding outputs, sorted by section of the thesis. Material taken from the output is marked in gray. Besides the shown prompts, ChatGPT was also used to check for spelling and grammar mistakes in most of the thesis. Moreover, the specific prompt “Is this sentence/paragraph correct formal English?” was used frequently. The output of these prompts is not shown, as it is quite straightforward, and including the output would make the appendix excessively long.

Introduction

Prompt A

Can you rewrite this paragraph to make it a bit more academic? Though higher education is ever-evolving, assessment remains an integral part of the learning process. It not only assesses students’ understanding of the course material, but also helps improve teaching methods and support student growth (Boud & Falchikov, 2007). Traditionally, multiple-choice and open-question exams have been the standard in the academic world. However, educators and researchers are increasingly aware that these methods fall short in measuring higher-level thinking skills and real-world application of knowledge (Freeman et al., 2014; Krathwohl, 2002).

Output A

In the evolving landscape of higher education, assessment remains a cornerstone of the learning process. It not only evaluates students' understanding of course material but also helps to improve teaching methods and fosters student development (Boud & Falchikov, 2007). While traditional methods like multiple-choice and open-question exams have long been staples in academia, educators and researchers have increasingly recognized their

limitations in measuring higher-order thinking skills and real-world application of knowledge (Freeman et al., 2014; Krathwohl, 2002).

Prompt B

Rewrite this sentence so that it doesn't use the word nuanced: Performance tasks thus give students the opportunity to show their academic performances in a more nuanced manner.

Output B

Performance tasks thus give students the opportunity to demonstrate their academic abilities in a more comprehensive manner.

Prompt C

Can you rewrite this sentence to make it a bit more academic? This hesitancy might come from multiple factors, such as worries about how complicated and demanding it could be to design and implement performance-based assessments (Wren, 2019).

Output C

This hesitancy might stem from multiple factors, including concerns about how complicated and resource-heavy it might be to design and implement performance-based assessments (Wren, 2019).

Prompt D

How can I say “students want feedback more often and want it to be helpful” in one academic sentence?

Output D

Students emphasize the need for more timely and constructive feedback.

Prompt E

How can I write this so that it flows nicer and doesn't use as many commas: Another study had students choose between an individual or group marketing project and found that while initially, around 79% of students preferred the group task, at the end of the semester, this

percentage had dropped to 53, mainly due to logistical issues, meaning that ultimately, only a very small majority preferred the group project.

Output E

Here's a revised version of your text for better flow and fewer commas:

"Another study asked students to choose between an individual or group marketing project. Initially, around 79% of students preferred the group task. However, by the end of the semester, this percentage had dropped to 53%, mainly due to logistical issues. Ultimately, only a very small majority preferred the group project."

This revision breaks down the information into shorter, more manageable sentences and reduces the number of commas for smoother readability.

Prompt F

Is this sentence correct and can it be improved: Since the preference for multiple-choice exams seems quite strong, and performance tasks are likely to require more effort from students and are not as objective, we expect multiple-choice exams to be preferred over performance tasks as well.

Output F

The sentence is generally correct, but it can be improved for clarity and conciseness. Here's a revised version:

"Given the strong preference for multiple-choice exams and the higher effort and lower objectivity associated with performance tasks, we expect multiple-choice exams to be preferred."

Prompt G

Can you edit this sentence to make it better: We do not hypothesize which format is preferred between performance tasks and open-question exams, due to the lack of literature to base a hypothesis on.

Output G

Certainly! Here's an improved version of the sentence:

"Due to a lack of supporting literature, we do not hypothesize a preference between performance tasks and open-question exams."

Method***Prompt A***

Can you rewrite this to make it flow better and more concise: 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.

Output A

Before starting, participants were briefed on the research's purpose, their voluntary participation, incentives offered, and data confidentiality.

Prompt B

Can you rewrite this so it doesn't use adjusting twice: Most scales were adjusted slightly, either by adding items to improve construct validity or by adjusting existing items to prevent misinterpretations.

Output B

Most scales were slightly modified, either by adding items to improve construct validity or by revising existing items to prevent misinterpretations.

Results***Prompt A***

Can you rewrite this so that it flows better and sounds academic? Taken together, PTs were thus rated to have significantly more potential to show performance, but also to be more difficult, less liked (valence), and especially less objective than MC exams.

Output A

Overall, PTs were rated as having significantly greater potential to demonstrate performance, but they were also perceived as more difficult, less favored (in terms of valence), and particularly less objective compared to MC exams.

Prompt B

Can you rewrite this to make it flow better: Since these preliminary results were already in contradiction with the hypothesis that group assignments would be rated more favorably than individual assignments, these differences were not tested further for significance.

Output B

Since these preliminary results contradicted the hypothesis that group assignments would be rated more favorably than individual assignments, further significance testing was not conducted.

Prompt C

Can you change the word order or rewrite this so that it doesn't start with a number: 35 participants did not respond or gave nonsensical answers.

Output C

There were 35 participants who either did not respond or gave nonsensical answers.

Prompt D

Can you change the word order or rewrite this so that it doesn't start with a number: Seventy-two responses were categorized into themes.

Output D

Responses from 72 participants were categorized into themes.

Discussion

Prompt A

Can you give me an alternative for the phrase "finding out" in this sentence that sounds a bit more academic? Firstly, we investigated and compared general preferences for multiple-choice exams, open-ended exams, and performance tasks, with the purpose of finding out where performance tasks fall within this ranking.

Output A

Sure, you can use the phrase "determining" instead. Here is the revised sentence:

"Firstly, we investigated and compared general preferences for multiple-choice exams, open-ended exams, and performance tasks, with the purpose of determining where performance tasks fall within this ranking."

Prompt B

Can you rewrite this sentence to make it better: However, the found overall preference for multiple-choice exams suggests that the fact that performance tasks might be a better way to engage with the material and show performance seems to not weigh up to the perceived benefits of multiple-choice exams.

Output B

However, the prevailing preference for multiple-choice exams suggests that the perceived benefits of such assessments outweigh the potential advantages of performance tasks in engaging with the material and demonstrating proficiency.

Prompt C

Can you adjust this a bit for better flow and clarity? Moreover, it could be that students don't recognize performance tasks' superior potential to show performance over open-question exams if this potential is not realized in practice. After all, this potential can only come forward if the tasks are effectively designed.

Output C

Furthermore, students may fail to recognize the superior potential of performance tasks in showcasing performance over open-question exams if this potential is not effectively realized in practice through proper task design.

Prompt D

Can you rewrite this sentence to make it more academic: These results are a bit surprising, given the emphasis put on performance tasks' potential to show performance in the literature (Sambell et al., 1997; Struyven et al., 2005).

Output D

These results are somewhat unexpected, considering the significant emphasis placed on the potential of performance tasks to demonstrate students' abilities in existing literature (Sambell et al., 1997; Struyven et al., 2005).

Prompt E

Can you rewrite this to make it more academic: Moreover, individual assignments were rated more positively on the subscales of fairness, valence, and value, and only given a more unfavorable score on anxiety.

Output E

Moreover, individual assignments received more favorable ratings on the subscales of fairness, valence, and value, while only receiving a less favorable score on anxiety.

Prompt F

Can you rewrite this to make it more academic? Additionally, a smaller number of students indicated wanting performance tasks to be more creative and more practical, meaning more in line with their future workfield, and some expressed the need to reduce the anxiety surrounding performance tasks.

Output F

A minority of respondents articulated a desire for performance tasks to exhibit greater creativity and practical relevance, aligning more closely with their anticipated professional domains. Additionally, some participants underscored the necessity of mitigating anxiety associated with performance tasks.

Prompt G

Can you rewrite the following sentence to make it make more sense and maybe a bit more academic? It thus seems that students do see the value of using multiple forms of assessment, and not just the one they prefer.

Output G

Hence, it is evident that students recognize the efficacy of employing diverse assessment methodologies, rather than solely favoring their personal preferences.

Prompt H

Can you rewrite this to make it more academic: Another parallel can be drawn in regard to our second hypothesis. When prompted by the open question, there were more students who indicated wanting more individual assignments than students indicating they want more group assignments. This strengthens our finding that in our sample, most students did not prefer group assignments over individual assignments.

Output H

Our second hypothesis is further supported by another observation. When asked openly, a greater number of students expressed a preference for additional individual assignments compared to those advocating for more group assignments. This reinforces our finding that, within our sample, the majority of students favor individual assignments over group tasks.

Prompt I

Rewrite this to make it clearer what I mean: Specifically the valence subscale might be problematic, since this is not actually a separate construct inherent to exam formats that might influence general preference, but is very similar to the concept of preference in itself.

Output I

Certainly! Here's a clearer version of your sentence:

"The valence subscale **could** be problematic because it does not represent a **distinct** construct inherent to exam formats that might **independently** influence overall preference. **Instead, it closely resembles the concept of preference itself.**"

This revision clarifies that the valence subscale is not a separate factor affecting preference but rather overlaps significantly with the general concept of preference.

Prompt J

Can you rewrite this to make it flow a bit better: Moreover, our study has an innovative focus. We not only investigate performance tasks, an under-researched area compared to traditional assessment formats, but the analysis of opinions on group and individual assignments also provides more insight into more specific types of this broad category.

Output J

Additionally, our study has an innovative focus. By investigating performance tasks, an area that is less researched compared to traditional assessment formats, we contribute valuable insights. Furthermore, our analysis of opinions on group versus individual assignments offers a deeper understanding of the specific types within this broad category.

Prompt K

Can you rewrite this to make it more clear and flow better: Moreover, the fact that while previous research suggest performance tasks to have more potential to show performance than traditional assessment formats including open question exams, but our students did not,

indicates that educators should focus on designing performance tasks effectively to make this potential come forward.

Output K

Furthermore, the discrepancy between previous research suggesting that performance tasks have greater potential to assess performance compared to traditional formats such as open question exams, and our students' lack of recognition of this potential, **underscores the importance for educators** to prioritize effective design of performance tasks **to realize their full potential.**

Prompt L

Can you change this sentence so that it doesn't use the word emphasize? Students' calls for more frequent use of performance tasks, coupled with better guidance, feedback, and fair grading, **emphasize** the need for robust support systems.

Output L

Students' calls for more frequent use of performance tasks, coupled with better guidance, feedback, and fair grading, **underline** the need for robust support systems.

Prompt M

Can you change the last sentence a bit to make it flow better? Not only in exam preference, but also in practical considerations like this, it is important to listen to students; suggestions, as they are the ones performing the tasks, making them the experts on how to improve them.

Output M

It is crucial to listen to students not only in matters of exam preference but also in practical considerations like these, as they are the ones performing the tasks, making them the experts on how to **enhance** them.

Prompt N

Can you rewrite this so that the sentences are a bit shorter: Overall, the study underscores the importance of understanding and addressing students' preferences and perceptions when designing assessment strategies. It seems that preference for traditional assessment formats is deeply engrained in students. Educators should thus focus on incorporating students' feedback on performance assessment in order to satisfy student needs, while also fostering a more engaging and effective learning environment that not only meets educational standards, but also supports students' academic success and professional development.

Output N

Overall, the study underscores the importance of understanding and addressing students' preferences and perceptions when designing assessment strategies. The preference for traditional assessment formats seems deeply ingrained in students. Therefore, educators should focus on incorporating students' feedback on performance assessments to satisfy their needs. This approach will foster a more engaging and effective learning environment that meets educational standards and supports students' academic success and professional development.