Students' preference on examination format and the relation with learning strategies

Hage Niggendijker

S4784081

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group: 2324_2a_08

Supervisor: prof. dr. Rob R. Meijer

Second evaluator: dr. Stacey M. Donofrio

In collaboration with: Silke Kok, Matsen Piepers, Annika Selesnew, Juraj Takács, and Ziying 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

Acknowledgements

I acknowledge the use of ChatGPT to only improve the papers structure and writing style. No output was used to generate information about the topic of this paper.

Abstract

This research investigated the preferences of students' on examination format and the relationship between learning strategies. The sample consisted of 128 psychology students, predominantly first-year students. A questionnaire was constructed that measured 4 different attitudes: difficulty, potential to showcase performance, valence, and objectivity. These attitudes combined measured overall preference for examination formats. Furthermore, two learning styles were measured, namely the surface learning strategy and deep learning strategy. Results showed that students prefer multiple-choice exams over open-question exams and performance tasks. Students perceived multiple-choice exams the least difficult, rated them higher in valence, and thought they were more objective. Students perceived open-question exams to be the most difficult and had more potential to showcase performance. Moreover, students perceived performance tasks to be the least objective, higher in valence than open-question exams, and had more potential to showcase performance than multiple-choice exams. The findings only supported the hypothesis that students find multiple-choice exams to be better suited for surface learning than for performance tasks. Future research should focus on what influence the perception of potential to showcase performance has on learning effort and perceptions that teachers have about these examination formats when constructing an exam.

Keywords: multiple-choice, open-question, essays, performance tasks, portfolios, learning strategies, deep learning strategy, surface learning strategy, preferences, perceptions, assessment, students

Students' preference on examination format and the relation with learning strategies

There has been a longstanding debate regarding the use of multiple-choice exams (MC exams) versus open-question exams (OQ exams). Popular beliefs are, for example, that OQ exams are more capable to test higher-order thinking compared to MC exams (Meijer, 2021). Additionally, it is argued that OQ exams are more suited to test a reasoning process that is more representative of real-life problem-solving than MC exams would examine (Hift, 2014). However, Hift (2014) contends that many general assumptions about the advantages of OQ exams over MC exams are not substantiated by the literature. Instead, Hift (2014) demonstrated that MC exams are equally well suitable to measure higher-order cognitive thinking than OQ exams. On top of that, MC exams are in general more reliable (Hift, 2014) and more efficient than OQ exams (Lukhele et al., 1994; Elstein, 1993).

Despite these findings, there remains a discrepancy between test-takers' attitudes and perceptions of MC exams and OQ exams and what is documented in the literature. From the students' perspective, studies reported that students preferred MC exams over OQ exams in general (Holzinger et al., 2020; Zeidner, 1987; Yang et al, 1985). They perceived MC exams as more convenient, less time taking to prepare, less anxiety evoking, and less difficult than OQ exams. They also viewed MC exams as more objective than OQ exams. On the other hand, students found OQ exams were more appropriate to obtain long-term knowledge (Holzinger et al., 2020), more reflective of students' knowledge, and perceived as fairer (Zeidner, 1987). Some of these perceptions contrast with findings by Hift (2014) that MC exams are equally capable at measuring higher-order cognitive processes as OQ exams. Students perceived OQ exams as better reflecting their knowledge than MC exams. This may reflect the idea that OQ exams are better suited to measure higher-order cognitive thinking than MC exams.

Regarding traditional MC and OQ exam formats, studies indicated that these forms of assessment had negative effects on their learning processes (Struyven et al., 2005; Sambell et al, 1997). Instead, alternative forms of assessment were seen as more conducive for the quality of learning. For instance, students reported that portfolio assessments enhanced their learning more than traditional formats (Slater, 1996). Van de Watering et al. (2008) found that students in New Learning Environments (NLE) preferred alternative assessment forms such as writing papers and portfolios. These alternative forms of assessment provide a new way to assess students' problem-solving abilities, critical thinking, and practical application of knowledge (Darling-Hammond & Adamson, 2010). These alternative assessments, known as performance tasks (PTs), provide a practical learning format, such as writing essays, portfolios, or research papers, and conducting research (Braun, 2019). Assessors have recognized the potential of alternative assessment forms (Dochy & McDowell, 1997).

Although PTs appear promising for application in the educational context, little research has been conducted on the attitudes of students toward PTs. The present research aims to integrate perceptions of PTs by comparing students' attitudes regarding their preferred examination format and the reasons for their preferences.

Learning strategies and preferences

Research on students' perceptions of exam format (and assessment in general) has found that perceptions about examination formats can influence learning strategies (Entwistle, 1991). In the late 1970s, studies on learning styles were initiated by research in cognitive psychology (Entwistle, 1998). Biggs (1976) developed the *Study Behaviour Questionnaire* (SBQ), revealing a distinction between two approaches to learning – namely the *deep approach* and the *surface approach*. This deep approach to learning defines itself by seeking personal understanding of the study material and the surface approach defines itself by memorizing the study material for recall and reproducing aspects of it. Marton and Säljö

(1976) associated these learning approaches with specific learning intentions: understanding ideas for oneself (deep approach) versus meeting course requirements (surface approach). (The original SBQ also measured a third learning approach called *achieving*, in which students try to get the highest grade possible. This bachelor thesis is restricted to the deep strategy and surface strategy).

According to Entwistle (1987), a student's learning approach is influenced by their perception of the learning environment. Furthermore, the assessment procedure can alter the learning strategy used by students (Entwistle, 1991). This implies that students adjust their learning strategies based on the assessment form used to measure their performance. Students who prefer surface learning may be reinforced by teachers who provide "ready-to-learn" material (Entwistle and Tait, 1990). Thus, the perception that students have of learning and assessment forms determines the learning strategy that is being used (Entwistle, 1991).

A literature review by Struyven et al. (2005) made important conclusions about the association between learning strategies and assessment forms. Firstly, OQ exams invoked deeper approaches to learning than MC exams. On the other hand, MC-exams invoked more surface approaches (Entwistle & Entwistle, 1991). Secondly, if students perceived that the assessment form was inappropriate for testing course material, this would elicit more of a surface approach to learning rather than a deep approach to learning. Thirdly, there was a mismatch between students their general preference for MC exams and the perceived benefits of OQ exams. Although students prefer MC exams over OQ exams, the latter are perceived as more appropriate for testing course materials. Therefore, their learning strategies tends to be more focused on a deep style of learning instead of a more surface style of learning.

From the literature, it is evident that students' preferences and attitudes influences their learning styles. However, there is limited research on how students perceive PTs in relation to learning strategies. Previous studies have shown that MC and OQ exams negatively

impact students' learning processes (Struyven et al., 2005; Sambell et al., 1997), while alternative assessment forms are considered more beneficial for learning quality (Slater, 1996). It is crucial that any assessment format tests and evaluates a combination of cognitive processes, such as reproducing, comprehending, problem-solving, explaining, drawing conclusions, critical thinking, and applying knowledge (Van de Watering et al., 2008). If an exam format assesses these important cognitive processes, students are likely to perceive it as appropriate to test course material. This perception can encourage a deeper approach to learning for the course material, leading to higher quality learning compared to the surface learning style associated with inappropriate exam formats.

The present research aims to investigate this idea by exploring the relationship between learning strategies and examination formats.

Research question and hypothesis

This study compares students' general preference regarding MC exams, OQ exams and PTs. The first research question examines which type of examination (MC exams, OQ exams or PTs) students prefer. Based on the literature, it is hypothesized that students prefer MC exams over OQ exams (Holzinger et al., 2020; Zeidner, 1987; Yang et al, 1985). Due to the limited research on PTs, it is unclear how they compare to MC and OQ exams. Given that OQ exams require the creation of responses, similar to PTs, it is anticipated that students may prefer MC exams over PTs. This leads to the first hypothesis:

Hypothesis 1: students prefer MC exams over OQ exams and PTs. There is no directional hypothesis with respect to which students prefer OQ exams or PTs more.

To further analyse the general preferences of students, different aspects of these examination formats will be explored and compared: difficulty, potential to showcase performance, valence (positive emotion), and objectivity. These aspects summarize previous research on students' preferences and perceptions of examination formats. Given the limited

research on PTs compared to MC and OQ exams, no directional hypothesis are given for the relationship between these three examination formats. Therefore, an exploratory analysis will be conducted on these four aspects.

Lastly, the relationship between examination format and learning strategies will be examined. The study will compare deep and surface learning strategies for each exam format. The research question examines whether there is a difference in learning strategies (deep versus surface) across examination formats. Based on the literature, two hypotheses are proposed:

Hypothesis 2: Students consider MC exams less suited to measure deep learning strategies than OQ exams and PTs.

Hypothesis 3: Students consider MC exams better suited for measuring surface processing than for OQ exams and PTs.

Method

Sample

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

There were 37.5% Psychology (EN) students, 61.7% Psychology (NL) and 0.8% Sociology. For practical reasons, we limited our sample to students of this faculty. 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. 35.2 percent of students were in their second year of studying or higher,

while there were 64.8% first-year students. A higher number of first-year students was expected since the sampling was mostly done through Sona Systems (https://www.sona-systems.com).

Incentive

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

Procedure

Participants were recruited through convenience sampling. The survey 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 second-year students or higher 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, 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 shortly. 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 and levels of trait test anxiety. Participants were then asked to indicate their preferences for different exam formats: open-question exams, MC exams, and performance-based tasks (both individual and group). Subsequently, prior experience and learning strategies for the different formats were assessed. Next, respondents answered questions regarding their levels of state evaluation anxiety for each examination method. The order in which these different sections were shown was randomized for each participant. After completing the questionnaire, participants were directed to the Sona Systems website to receive credits, if applicable.

Ethics

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

Materials

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

Overall preference

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

The inventory consisted of four scales, each consisting of two to four items, where participants indicated their level of agreement with displayed statements on a five-point Likert scale. Three scales were based on an existing inventory by Lindner et al. (2019): 1) potential

to show performance, 2) objectivity, and 3) valence. The scale measuring potential to show performance was directly adopted from Lindner's research, where it consisted of four items. An example item of this scale is "Open question exams/ MC exams/ Performance tasks allow me to express my knowledge precisely." Perceived objectivity was assessed through two items, one adopted directly from Lindner (2019) 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.

For every subscale, means were calculated that are derived from the summed item scores. After calculating mean scores for every subscale, the different subscales were combined to form the general preference score. Reliability was estimated using Cronbach's alpha and proved adequate for all scales. Moreover, the reliability of all scales combined as a measure of preference was also sufficient (see Table 1).

Table 1

Cronbach's Alpha for the different subscales and total scale of the EPI

Subscales	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

Learning strategies

To measure the learning strategies of students, the Learning Preference Inventory was used by Gordon et al. (2007). The LPI consists of five subscales, but only two of the subscales were taken into account. The subscales that have been used from the LPI were *Shallow Processing* and *Deep Learning*. The subscale Shallow Processing measures surface learning and Deep Learning measures deep learning strategies. The scale Shallow Processing contained four items and Deep Learning contained seven items. Participants indicated their level of agreement with displayed statements on a five-point Likert scale. An example item of Deep Learning is “I work practice problems to check my understanding of new concepts or rules”. An example item of Shallow Processing is “I try to memorize the steps for solving problems presented in the text or in class”. Important to note, is that both scales were rephrased for measuring learning strategies for PTs (see Appendix B).

Again, reliability was estimated using Cronbach's alpha. Alpha fluctuated between .68 and .71 for shallow processing and between .64 and .74 for deep learning across conditions, which is similar across the three conditions, which is similar for the original LPI where an Alpha score of .71 was found.

Results

General Preference

Following data collection, descriptive statistics were analysed to determine students' general preference for each examination format, based on the EPI scores. As was hypothesized, a higher preference score was found for MC exams (M = 3.45, SD = .54) in comparison to OQ exams (M = 3.01, SD = .46), and PTs (M = 3.00, SD = 0.48). The difference between the means of PTs and OQ exams was negligible and therefore significance testing was not conducted for this pair. Given the amount of comparisons required to test all

hypotheses, Bonferroni corrections were conducted. This resulted in a specified $\alpha = .003$ level.

The paired one-sided sample t-test between MC exams and PTs showed that preference for MC exams was significantly higher, $t(127) = 6.16$, $p < .001$, $d = 0.54$, CI [0.23, 0.67][1] Similarly, the paired one-sided t-test between MC exams and OQ exams indicated that preference was significantly higher for MC exams compared to OQ exams, $t(127) = 6.21$, $p < .001$, $d = 0.55$, CI [0.22, 0.65]. The effect sizes were calculated using Cohen's *d* and showed moderate effects ($d > 0.54$). Therefore, the results support the first hypothesis, demonstrating that MC exams are favoured in comparison with OQ exams and PTs.

Subscales of General Preference

The subscale scores were explored without directional hypotheses, providing specific insights into how different aspects relate to exam formats The means of the four subscales for each examination format can be found in Table 2. Each subscale was compared using two-sided paired t-tests and will be individually analysed.

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

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

Difficulty

Students perceived OQ exams as significantly more difficult than MC exams, $t(127) = 13.559$, $p < .001$, $d = 1.12$, CI [0.88, 1.38] with a very high effect size. Additionally, students perceived PTs significantly more difficult than MC exams, $t(127) = 4.204$, $p < .001$, $d = .37$, CI [0.09, 0.55], with a moderate effect size and significantly less difficult than OQ exams, $t(127) = 13.715$, $p < .001$, $d = 1.21$, CI [0.63, 0.99], with a very high effect size.

Potential to showcase performance

Students perceived that PTs have as much potential to showcase performance than OQ exams, $t(127) = 2.762$, $p = .003$, $d = .24$, CI [-0.02, 0.49] with a moderate effect size. Students perceived PTs to have significantly more potential to showcase performance than MC exams, $t(127) = 5.427$, $p < .001$, $d = .48$, CI [0.27, 0.94], with a moderate effect size. Additionally, students also perceived OQ exams to have significantly more potential to showcase performance than MC exams, $t(127) = 7.366$, $p < .001$, $d = .65$, CI [0.49, 1.18], with moderate effect size.

Valence

Students rated PTs significantly higher in valence than OQ exams, $t(127) = 2.986$, $p < .002$, $d = .26$, CI [0.00, 0.53], with a small effect size. PTs are rated lower in valence than MC exams, $t(127) = 4.455$, $p < .001$, $d = .39$, CI [0.17, 0.93], with a moderate effect size. Furthermore, results showed that MC exams were significantly higher rated in valence than OQ exams, $t(127) = 6.179$, $p < .001$, $d = .55$, CI [0.41, 1.19], with a moderate effect size.

Objectivity

Students perceived MC exams as more objective than OQ exams, $t(127) = 13.870$, $p < .001$, $d = 1.23$, CI [1.09, 1.70] and PT's, $t(127) = 16.562$, $p < .001$, $d = .1.46$, CI [1.49, 2.15], with very high effect sizes. Furthermore, results showed that students significantly rated OQ

exams higher in objectivity than PTs, $t(127) = 4.151$, $p < .001$, $d = .37$, CI [0.12, 0.74], with a moderate effect size.

Learning strategies

Descriptive statistics for each learning strategy per exam format are provided in Table 3. Notably, the means for MC exams and OQ exams were similar, with mean differences observed between MC exams and PTs. Thus, two matched t-tests were conducted[2]. There was partial support for hypothesis 2. Students considered MC exams less suited to measure deep learning strategies than PTs, $t(127) = 3.425$, $p < .001$, $d = .30$, CI [0.02, 0.38], with a small effect size. For surface learning, scores on PTs were significantly lower than for MC exams, $t(127) = 3.091$, $p = .001$, $d = .27$, CI [0.004, 0.46] with a small to moderate effect size. The hypothesis that students perceive MC exams as less suited for measuring deep learning strategies compared to PTs and OQ exams was not supported. However, a note of caution must be added for interpreting these results because of the small effect sizes.

Table 3
Means and Standard Deviations of learning strategies per exam format

Measure	Multiple Choice		Open Question		Performance Tasks	
	M	*SD*	*M*	*SD*	*M*	*SD*
Surface learning	3.91	.69	3.92	.70	3.68	.78
Deep learning	3.61	.62	3.67	.65	3.41	.73

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

Discussion

Preferences and perceptions on exam format

In this study, we aimed to investigate students' preferences for different exam formats and to explore the aspects on which these formats might differ. Furthermore, this research investigated the link between exam format and learning strategies, given the established link in prior research (Struyven et al., 2005). Our findings indicate a general preference for MC exams over OQ exams and PTs. These results are consistent with our hypothesis and previous studies on MC and OQ exams (Holzinger et al., 2020; Zeidner, 1987; Yang et al, 1985). Adding to the limited amount of research on PTs as an exam format, no significant differences were found in general preference between PTs and OQ exams. Thus indicating that students do not generally prefer one format over the other when MC exams are excluded.

From our findings, we can conclude that students prefer MC exams and that they found them the least difficult, rated them higher in valence, and found them the most objective. Conversely, students found OQ exams to be the most difficult and have more potential to showcase performance. Students perceived PTs to be the least objective, were rated higher in valence than OQ exams, and had more potential to showcase performance than MC exams. Importantly, there is a mismatch between the exam format students prefer and their perceptions of these formats. While students prefer MC exams, they do not believe these exams have the most potential to showcase their performance or are the most challenging. Despite arguments by Hift (2014) that MC exams can assess higher cognitive abilities as effectively as OQ exams, students do not necessarily share this perception.

Learning strategies

It was hypothesized that students would consider MC exams less suited to measure deep learning strategies than OQ exams and PTs. Furthermore, it was hypothesized that students consider MC exams better suited for measuring surface processing than for OQ

exams and PTs. The findings only partially support these hypotheses. There was support for the hypothesis that students perceive MC exams as better suited to use surface learning compared to PTs.

While these results do not align with conclusions made by Struyven et al. (2005), they do support findings by Hift (2014). Hift's review of the literature discusses the notion that linking exam format to a specific learning strategy is overly simplistic and that evidence for such claims is likely misinterpreted (Joughin, 2010). Students' learning styles seem to depend on multiple interacting variables, such as innate learning motivation and preferred learning strategies. Thus, focusing solely on exam format to promote a specific learning strategy may not be a profound approach to benefiting students' learning.

These findings might have been influenced by the sample. The sample primarily consisted of psychology students who have extensive experience with MC exams in the current curriculum of the University of Groningen. It seems plausible that different results might emerge from a curriculum with a more diverse use of examination formats. Another potential confounding variable is the questionnaire used, which was not originally designed to measure learning strategies for PTs. A more appropriate questionnaire might have yielded different results for PTs.

Implications and future recommendations

Our findings have several implications. First, while students generally prefer MC exams, their perceptions do not necessarily align with their preferences. When evaluating courses and assessments, evaluators and assessors should consider this discrepancy. For instance, a student might express positive feedback about an MC exam but also believe they could showcase better performance with an OQ exam or PT. Although no link was found between learning strategies and exam format, future research should investigate how perceptions of showcasing performance influence learning effort for different assessment

formats. Perhaps, the perception of needing to showcase performance impacts students' learning effort.

Interestingly, MC exams can potentially assess higher cognitive abilities as effectively as OQ exams, provided that the MC exams are well-constructed. MC exams should than be as difficult as OQ exams. Future research should also explore the perceptions of teachers and assessors who create MC and OQ exams. Teachers' attitudes towards exam formats may influence the quality of the exams they produce. For example, a teacher who believes that MC exams can only measure factual knowledge may create less effective MC exams that only test factual knowledge, reinforcing the perception that OQ exams are better for assessing higher-order cognitive abilities and showcasing performance.

Limitations current research

There are several limitations in the current research that should be acknowledged. First, the sample mainly consisted of first-year psychology students who received an incentive to participate. While the research topic is most likely of interest to them, caution should remain about generalizing to other student populations. Additionally, it is uncertain whether students from different faculties share the same preferences. Psychology students at the University of Groningen are very accustomed to MC exams, which could influence their preferences. Moreover, the analyses could have been more sophisticated by employing a repeated measurement design with contrasts.

References

Biggs, J. B. (1976). Dimensions of study behaviour: another look at a.t.i. *British Journal of Educational Psychology, 46,* 68-80. https://doi.org/10.1111/bjep.12274

Braun, H. (2019). Performance assessment and standardization in higher education: A problematic conjunction? *British Journal of Educational Psychology, 89*(3), 429 – 440

Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: the role of performance assessment in achieving 21st century standards of learning*. Standford University, Stanford Center for Opportunity Policy in Education

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

Elstein, A. S. (1993). Beyond multiple-choice questions and essays: The need for a new way to assess clinical competence. *Academic Medicine, 68*(4), 244–249. https://doi-org.proxy-ub.rug.nl/10.1097/00001888-199304000-00002

Entwistle, N. J. (1987). A model of the teaching-learning processes. In J. T. E. Richardson, M. W. Eysenck & D. Warren Piper (eds.). *Student Learning: Research in Education and Cognitive Psychology* (pp. 13-28). Open University Press.

Entwistle, N. J. (1991). Approaches to Learning and Perceptions of the Learning Environment: Introduction to the Special Issue. *Higher education, 22*(3), 201-204

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education, 22,* 205-277

Entwistle, N. J. & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education, 19,* 169-194

Entwistle, N. J. (1998). Approaches to learning and forms of understanding. In B. Dart & G. Boulton-Lewis (Eds.). *Teaching and learning in higher education: From theory to practice*. Australian Council for Educational Research.

Gordon, S.C., Dembo, M.H., & Hocevar, D. (2007). Do teachers' own learning behaviors influence their classroom goal orientation and control ideology? *Teaching and Teacher Education, 23*, 36 – 46

Hift, R.J. (2014). Should essays and other open-ended-type questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education, 14*(249).

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*(312).

Joughin, G. (2010). The hidden curriculum revisited: a critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education, 35*(3), 335 – 345.

Lindner, M.A., Mayntz, S.M., & Schult, J. (2019). Studentische Bewertung und Präferenz von Hochschulprüfungen mit Aufgaben im offenen und geschlossen Antwortformat. *Zeitschrift für Pädagogische Psychologie, 32*(4), 239 – 248. https://doi.org/10.1024/1010-0652/a000229

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement, 31*(3), 234–250. http://www.jstor.org/stable/1435268

Marton, F., & Säljö (1976). On qualitative differences in learning. I – outcome and process. *British Journal of Educational Psychology, 46*, 4-11

Meijer, R. R. (2021). Meerkeuzetentamens zijn betrouwbaarder en efficiënter. *Ukrant.* Geraadpleegd op 14 maart 2023, van https://ukrant.nl/meerkeuzetentamens-zijn-betrouwbaarder-en-efficienter/

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

Slater, T. F. (1996). Portfolio assessment strategies for grading first-year university physics students in the USA. *Physics Education, 31*(5), 329-333

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

Yang, J., Verheul, I. C., Verhelst, N., & Van Essen, E. (1985). Opinions about multiple choice- and open questions-tests and differences in study activities as a result of test expectation: Report of a questionnaire. *Tijdschrift Voor Onderwijsresearch*, *10*(4), 179–188.

Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. In B. Nevo & R. S. Jäger (Eds.), *Educational and psychological testing: The test taker's outlook.* (pp. 67–82). Hogrefe & Huber Publishers.

Appendix A

Examination Preference Inventory

This appendix comprises the five scales and their corresponding items of the Examination Preference Inventory. 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.

Appendix B

Rephrased items from LPI used for performance tasks

This appendix comprises the items that were rephrased from the LPI to measure surface learning and deep learning strategies for PT's.

Surface learning

1) I find reviewing previously papers/presentations to be a good way to study for these assignments
2) When I study for making papers/presentations, I review my class notes and look at examples
3) I try to memorize the steps for making papers/presentations presented in the text or in class
4) When I study I use example papers/presentations and my notes from class or a book to help me memorize the steps involved in making papers/presentations

Deep learning

1) I examine example papers/presentations that have already been worked to help me figure out how to do similar papers/presentations on my own
2) When I work a problem for an essay, I analyze it to see if there is more than one way to get the right answer
3) I draw pictues or diagrams to help me solve some problems
5) I work through example papers/presentations to check my understanding of new concepts or rules
6) I work through several examples of the same type of papers/presentations when studying so I can understand problems better

7) When studying, I try to combine different pieces of information from course material in new ways