School voor Educatieve Studies

Educatieve master in de gezondheidswetenschappen

Masterthesis

Certainty-Based Marking: Analyzing Socio-Cultural Factors and Inclusivity in Assessment Practices

Margot Cochez

Scriptie ingediend tot het behalen van de graad van Educatieve master in de gezondheidswetenschappen

PROMOTOR :
Prof. dr. Jean-Michel RIGO

COPROMOTOR :
Prof. dr. Jean-Michel RIGO

BEGELEIDER :
Prof. dr. Anouk AGTEN

School voor Educatieve Studies

Educatieve master in de
gezondheidswetenschappen

Masterthesis

Certainty-Based Marking: Analyzing Socio-Cultural Factors and Inclusivity in Assessment Practices

Margot Cochez
Scriptie ingediend tot het behalen van de graad van Educatieve master in de gezondheidswetenschappen

PROMOTOR :
Prof. dr. Jean-Michel RIGO

COPROMOTOR :
Prof. dr. Jean-Michel RIGO

BEGELEIDER :
Prof. dr. Anouk AGTEN

Preface

This master's thesis is the result of a collaborative effort characterized by open communication, complementary strengths, and shared commitment. All team members were involved in the process of data acquisition. The statistical analyses were mainly carried out by Margot and Dorien, while the creation of figures was primarily handled by Dorien and Brent. Alix played a leading role in the writing process, with contributions from all team members. This division of strengths enabled us to efficiently combine our individual talents into a coherent and comprehensive final product.

The research question was developed in close consultation with our promotor, Prof. Dr. Jean-Michel Rigo, and our supervisor, Prof. Dr. Anouk Agten. Their guidance helped us shape a relevant and timely study within the field of educational assessment.

Our research on the inclusivity of certainty-based marking is particularly significant in the context of ongoing efforts to develop more inclusive assessment strategies. As future professionals and educators, we are committed to continually improving evaluation methods to ensure that they are fair, transparent, and accessible to all students.

We would like to sincerely thank everyone who supported us during the completion of our master's thesis. Special thanks go to our promotor, Prof. Dr. Jean-Michel Rigo, for his expert guidance and valuable feedback, and to our copromotor, Prof. Dr. Anouk Agten, for her insightful input and support throughout the process. We are especially grateful to the students who participated in our study. Their openness and willingness to share their experiences were essential to our research. Thanks also to the faculty and staff of UHasselt for providing the resources and academic environment needed for this work. Finally, we thank our families, friends, and each other for the collaboration, motivation, and encouragement that carried us through this journey.

Abstract

Background: Certainty-Based Marking (CBM) is an assessment strategy that evaluates not only the correctness of student responses but also the confidence with which those responses are given. This dual-layer approach aims to enhance metacognitive engagement while discouraging random guessing. Despite its pedagogical potential, questions remain about CBM's inclusivity across diverse learner populations.

Aim: This study investigated whether CBM constitutes an inclusive assessment format by analyzing its interaction with socio-cultural background, academic preparation, confidence, stress and cognitive effort among first-year university students.

Method: A total of 316 students from three bachelor programs (Medicine, Economics and Rehabilitation Sciences) completed a structured digital questionnaire assessing perceptions and experiences with CBM. This self-report data was combined with anonymized academic performance scores, realism and level of centering. Statistical analyses were performed to explore associations between CBM outcomes and demographic, educational and affective variables.

Results: Significant effects were observed for academic field and secondary school background. Medicine students and those from the ASO track outperformed others across all three CBM parameters. In contrast, gender, home language and family background showed no significant effects, suggesting CBM's potential neutrality regarding socio-demographic variables. Self-confidence and ease of using certainty levels were positively associated with higher accuracy and performance, whereas elevated stress and processing difficulties correlated with poorer outcomes.

Conclusion: CBM appears to promote inclusive and reflective assessment, especially when combined with clear instruction and repeated exposure. However, its effectiveness remains influenced by prior academic experience and individual cognitive factors. To ensure equitable implementation, educators should embed CBM within differentiated instructional practices and provide support for learners with varied educational backgrounds.

1. Introduction

Assessment plays a pivotal role in the educational process, serving not only as a tool for grading but also as a mechanism for feedback, motivation, curriculum development and instructional improvement. Depending on its purpose, assessment can take several forms: diagnostic assessments identify students' prior knowledge and misconceptions to inform instructional planning (Onasanya & Ajamu, 2024); formative assessments are conducted during the learning process to provide feedback and support student growth (Onasanya & Ajamu, 2024; Muradova, 2025); summative assessments evaluate learning outcomes usually at the end of an instructional unit, often used for certification or high-stakes decisions (Suardana et al., 2023; Demir & Yalçın, 2022; Susanti & Hasanah, 2023; Petrovskaya et al., 2022). These types of assessment not only differ in timing and purpose but also in their influence on instructional design and student motivation (Mustamin, 2023). Formative assessment, in particular, fosters active learning and self-regulation, while summative assessment can drive performance-oriented behavior (Muradova, 2025). As educational goals have evolved from the mere transmission of knowledge to the cultivation of critical thinking, creativity and self-regulation, these criteria need to guide the methods used to evaluate student learning. Next, several evaluation methods are described, combined with their pros and cons regarding the evolved criteria.

The oral examination provides real-time, interactive assessment of a student's understanding. Oral exams are common in professional and medical education, where communication and critical thinking under pressure are essential. They allow examiners to probe student reasoning, adapt questions on the fly and clarify ambiguities. Nonetheless, oral assessments are inherently resource-intensive, require trained evaluators and are prone to examiner bias. Additionally, performance anxiety can significantly affect results, raising concerns about equity and reliability (Gharibyan, 2005).

The counterpart of oral examination is written evaluation, which is commonly divided into two categories: open-ended and closed-ended questions. Open-ended questions, for instance, allow students to articulate their reasoning, demonstrate problem-solving skills and minimize guessing (Schuman & Presser, 1979; Tan & Erdogan, 2004). These questions are considered beneficial for assessing higher-order thinking and individual expression, making them suitable for evaluating deeper conceptual understanding (Allen, 2002). However, open-ended formats are time-consuming to grade, susceptible to subjective evaluation biases and can limit content coverage due to the extended time required per response (Ilhan, 2016; Ozcelik, 2011).

Conversely, closed-ended questions (e.g., multiple-choice questions (MCQs)) are valued for their efficiency, objectivity and scalability. They are widely used in both formative and summative assessments, particularly in large-enrollment courses and standardized testing, as they are easy to administer and score (Schuman & Presser, 1979; Cotton, 1988). Despite their practicality, MCQs rely on predefined answer choices, which can constrain student expression and introduce response bias, especially when options fail to reflect students' actual understanding or perspectives (Schuman & Presser, 1979; Vinten, 1995). MCQs also tend to fall short in assessing higher-order cognitive skills, such as analysis, synthesis and evaluation. Moreover, they may encourage surface-level learning strategies and allow students to guess correct answers without genuine comprehension. These limitations have sparked growing criticism and prompted efforts to refine

assessment formats, including approaches that account for student certainty in scoring (Gharehbagh, 2022).

A further concern with traditional MCQ scoring lies in its binary structure, which classifies responses strictly as correct or incorrect. Consequently, educators miss opportunities to gain deeper insights into students' thought processes and metacognitive awareness, a limitation that ultimately affects instructional planning and decision-making at all levels of educational management (Curtis, 2013).

To discourage guessing, assessments can include questions with negative marking (QNM). However, this potential benefit also represents the approach's greatest drawback: QNM poses significant challenges in terms of inclusivity by failing to account for individual differences in confidence and risk tolerance. Research indicates that risk-averse students, particularly many female test-takers, may choose to leave questions unanswered to avoid penalties, leading to distorted assessment outcomes (Wu et al., 2021). This design introduces an inherent bias against cautious learners and ultimately fails to provide an equitable measure of knowledge. Such an approach may disproportionately affect learners from diverse educational backgrounds or those unfamiliar with the assessment format, thereby failing to reflect the complexity of student learning (Leclercq, 1982).

Because of these issues, there is growing interest in alternative evaluation methods, like Certainty-Based Marking (CBM). This system goes beyond simply grading answers as right or wrong. Instead, it asks students to indicate how confident they are in their answers, adding an extra dimension to the evaluation. CBM provides valuable insight into both students' understanding of the material and their confidence in their knowledge. It also encourages students to reflect on their learning process. Moreover, research suggests that repeated exposure to CBM can enhance students' understanding while gradually building their confidence in assessing their own knowledge (Remesal et al., 2024).

Another potential benefit of CBM is that it reduces random guessing, like QNM. However, with CBM, studies show that both male and female students benefit equally from CBM, with no significant gender differences in confidence progression after multiple trials (Remesal et al., 2024). This suggests that CBM can mitigate potential gender biases present in traditional assessments, where risk aversion in women, for instance, can skew results (Remesal et al., 2024; Gardner-Medwin & Gahan, 2003).

Although CBM offers pedagogical advantages, it can also be contextualized within the broader framework of confidence-based learning (CBL), where confidence serves not just as a measure of certainty but as a mechanism for internal feedback that enhances learning, even in the absence of external correction (Ptasczynski et al., 2022). Ptascynski et al. has found that students' prior academic background significantly influences their ability to accurately assess their own confidence. Therefore, it is recommended that CBM is implemented multiple times within a course to help students become familiar with the process, ensuring that less experienced students have the chance to develop their self-assessment skills over time (Remesal et al., 2024; De Finetti, 1965; Leclercq, 1982). This approach reduces initial disadvantages faced by students who are new to this evaluation method, thus making CBM more inclusive for learners at various stages of their academic journey.

There is still limited evidence on how inclusive the system is, which brings us to the following research question: Are CBM-based assessments inclusive? A possible hypothesis is that CBM promotes honesty and reduces guessing among almost all students. Moreover, students with more academic experience may benefit more from CBM than those with less experience.

2. Method

2.1 Participants

This study focused on first-year students from the bachelor's programs in Rehabilitation Sciences, Medicine and Economics, who recently completed exams using the Certainty-Based Marking (CBM) system. This diverse group of participants was chosen to generalize the results to various academic contexts. Each study area received information regarding the CBM system. Students were given a single explanation of the method and practiced with example questions using the CBM system prior to the exam, based on the scoring scale developed by Leclercq et al. (1993) (Figure 1). The study was approved by the Social and Societal Ethics Committee (SMEC) and all participants provided informed consent (Appendix 1).

Figure 1
Scoring table used in CBM

Certainty level	0% ↔ 25%	25% ↔ 50%	50% ↔ 70%	70% ↔ 85%	85% ↔ 95%	95% ↔ 100%
Points when correct	+0.65	+0.8	+0.85	+0.9	+0.95	+1
Points when incorrect	+0.2	+0.15	+0.1	0	-0.3	-1
Points when no answer	0					

Note. Figure 1 illustrates the scoring table used in the CBM exams. Scores were assigned based on the correctness of the response and the declared certainty level, with higher penalties for incorrect answers given with greater certainty.

2.2 Design of the student questionnaire

Data was collected through a digital questionnaire (Appendix 2) designed in Qualtrics, which consisted of six sections: general questions (demographic characteristics and academic background), understanding of CBM, experience with the system during exams, its impact on confidence and stress, general feedback about the system and questions related to the cognitive effort and expected achievement. This comprehensive design was intended to capture a well-rounded perspective on the students' experiences and perceptions.

The development of the questionnaire was informed by findings from previous literature. Wu et al. (2021) suggests that female students may be disadvantaged when working with QNM. In line with this, a question on gender was included to allow for analysis of possible gender-based differences.

Ptasczynski et al. (2022b) highlight that a student's academic background may affect their confidence in answering questions, so participants were asked to indicate their secondary school track. Based on the same study's findings that repeated exposure to CBM supports confidence-based learning (CBL), the questionnaire also included items about students'

first encounter with CBM, how often and clearly instructions were given and whether they understood them before the exams.

Beyond the variables identified in previous research, we aimed to explore additional factors that could potentially influence students' results and self-assessment. These included questions about the students' home situation, native language, perceived stress and self-confidence, perceived value of the CBM system and if the students utilized facilities.

2.3 Data acquisition

The data collection occurred during lectures in the classrooms of the included study areas (Figure 2). Before completing the questionnaire, students were provided with an informed consent form, which explained the study's purpose, procedures and the confidentiality of the data, as well as emphasizing that participation was voluntary. Students who agreed to participate signed the form and were given access to the questionnaire. The questionnaire was completed digitally via a personal link, with participants using their own devices. On average, completing the questionnaire took 10 to 15 minutes.

The perceived data from the questionnaire was anonymized by the principal investigators (PI's). Additionally, the PI's completed the data set with information concerning students' academic performance (i.e., exam scores), realism (i.e., self-assessment accuracy, ranging from 0 to 1; closer to 1 indicates better self-assessment ability) and the level of centering (i.e., self-assessment bias indicated with a score between -100 to +100; negative values indicating underestimation, between the -15 and +15 is considered neutral, positive values indicating overestimation) corresponding to the anonymized students. These additional data referred to the first course taken by each student during the current academic year and consisted of CBM-related content.

Figure 2

*Timeline of exam dates and questionnaire administration*
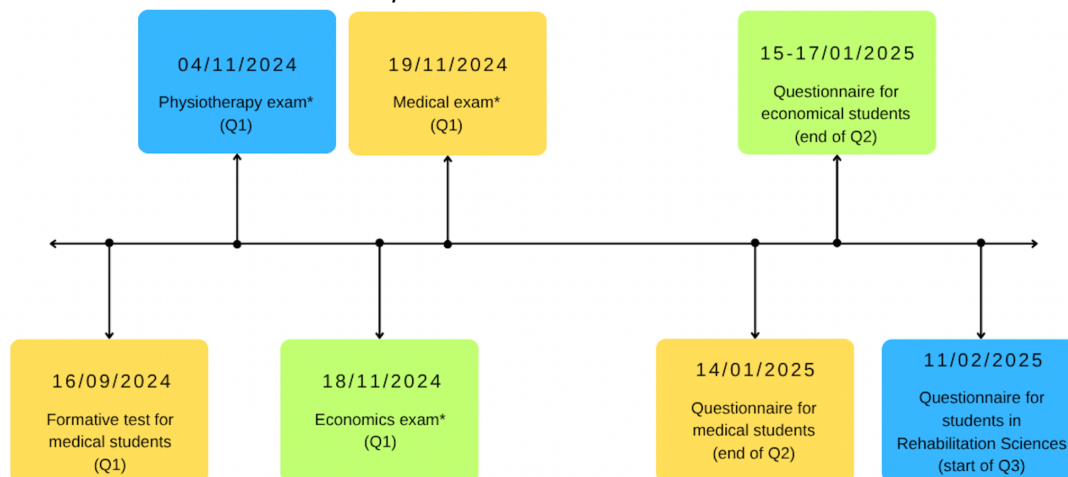
Note. *Figure 2 illustrates the timeline of exam dates and questionnaire administration across the three study programs. (*) indicate data points for which academic performance, realism and level of centering were collected. Q = quartile.*

2.4. Statistical analysis

The data ordening and statistical analysis were performed in JMP for Windows Pro version 17. The data ordening first involved calculating the average scores for academic performance, centering and realism data for the study areas of Medicine and Economics. This approach was chosen because these programs included multiple CBM-related evaluation moments. Not all evaluation moments involved CBM; therefore, scores from non-CBM-related exams were excluded from the analysis. For the study area of Rehabilitation Sciences, the regular data for academic performance, centering and realism were used instead of average scores, as there was only one CBM-related evaluation moment. Thereafter, the academic performance presented as ranges in the questionnaire were excluded from the analysis, as access to the participants' anonymized exam scores were obtained. Lastly, individuals with missing data were excluded from the analysis (Figure 3).

Figure 3

Participant flow diagram.
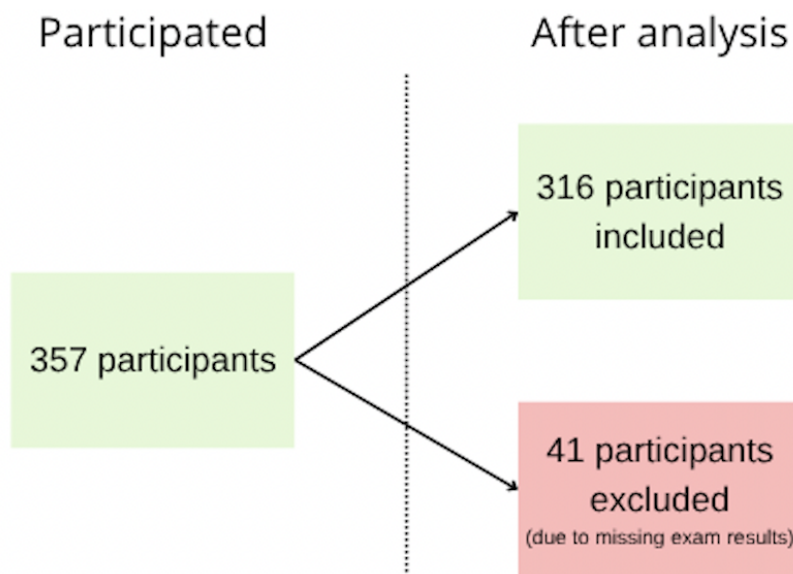
Note. *Figure 3: Out of 357 participants, 41 were excluded due to missing data, resulting in a final sample of 316 participants included in the analysis.*

To compare differences in CBM-related content across more than two independent groups, statistical tests were selected based on assumption checks. Normality of the residual data was assessed using the Shapiro–Wilk test and homogeneity of variances was tested using Levene's test. If both assumptions were met, a one-way ANOVA was conducted. In cases where the assumption of normality was violated or variances were unequal, alternative methods were used. Specifically, Welch's ANOVA was applied when variances were unequal but normality was retained and the Kruskal-Wallis test was used when both assumptions were violated. The significance level was set at 0.05 for both tests. A nonparametric post hoc test (Comparisons for Each Pair Using the Wilcoxon Method) was conducted to determine between which groups significant differences were present.

For comparisons across two independent groups, independent samples t-tests were conducted for the CBM-related content. When group sizes were sufficiently large (n > 30), the assumption of normality was considered satisfied based on the central limit theorem. For smaller groups, normality was assessed using the Shapiro–Wilk test. Equality of variances was examined using Levene's test. If both assumptions were met, a standard t-test was used. In cases of unequal variances (p < 0.05), Welch's t-test was applied. In cases where the assumption of normality was violated in one or both groups the Mann-Whitney U test was conducted instead. The significance level was set at 0.05 for both tests.

3. Results

The results of the study are summarized in Table 1, which provides a comprehensive overview of the effects of various demographic and experiential factors on academic performance (exam scores), realism (accuracy of self-assessment) and level centering (self-assessment bias).

3.1 General questions

3.1.1 Study Area in higher education

A significant effect of the current bachelor's program was observed for all three parameters: academic performance (p < 0.0001), realism (p < 0.0001) and level of centering (p < 0.0001) (Table 1). Medical students achieved the most favorable results across all dimensions. Compared to students in Economics and Rehabilitation Sciences, they scored higher, assessed their certainty more accurately and showed less deviation in their level of centering. Economics students followed, with moderate scores on all three outcomes. Students in Rehabilitation Sciences obtained the lowest results in academic performance, realism and level of centering. No significant difference was found in realism between Economics students and students in Rehabilitation Sciences (Figure 4).

Figure 4

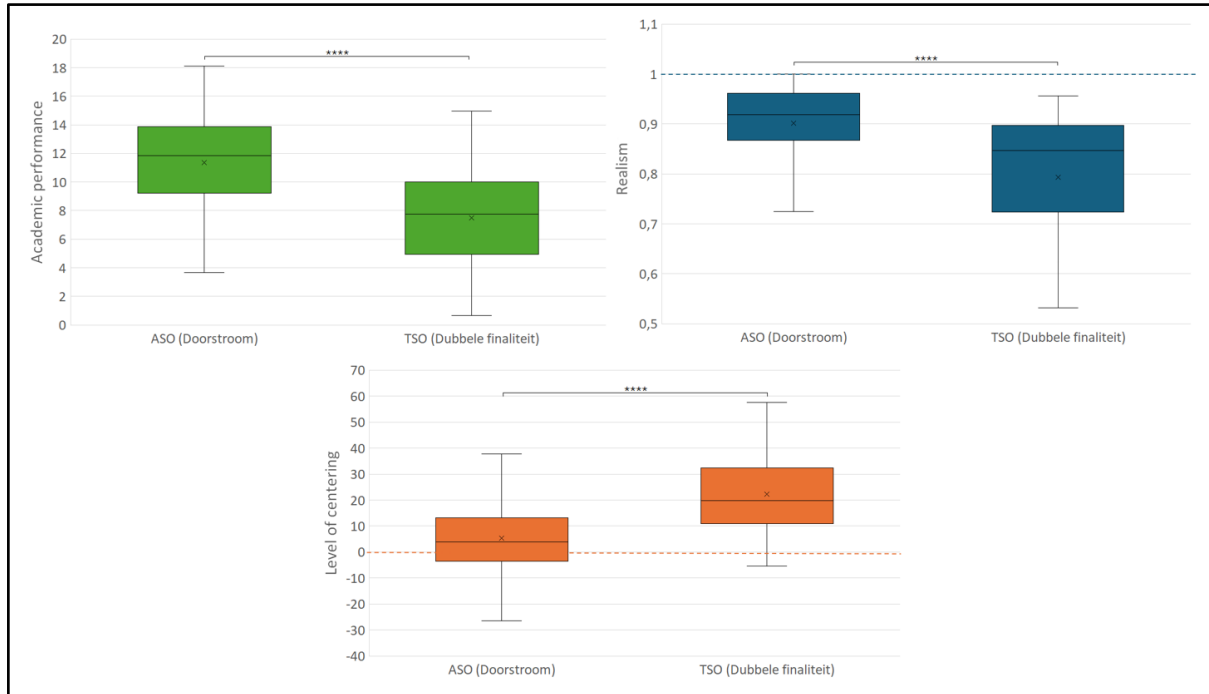*Academic performance, realism and level of centering of study area higher education*
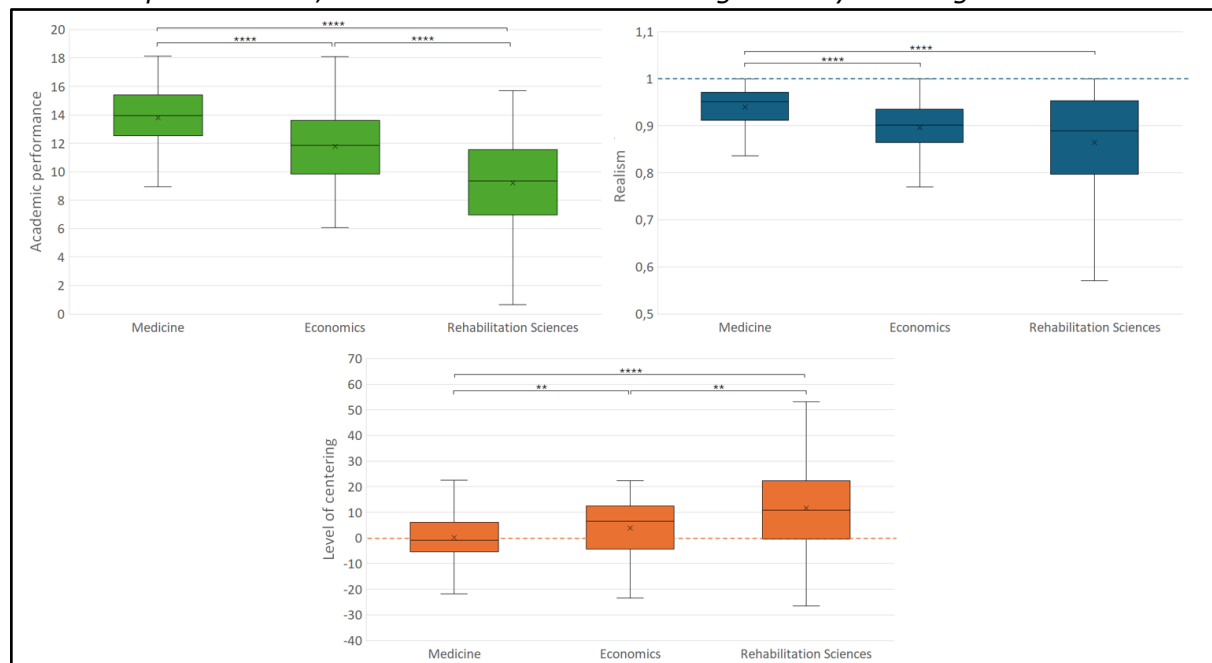
Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

A similar effect was found for the type of secondary education students completed prior to university. Students who studied ASO (doorstroomfinaliteit) scored significantly higher on academic performance ($p < 0.0001$), realism ($p < 0.0001$) and level of centering ($p < 0.0001$) compared to those from a TSO background (dubbele finaliteit) (Figure 5).

Figure 5

*Academic performance, realism and level of centering of study area secondary school*

Note. *(*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.*

3.1.2 Gender and demographic factors

Gender showed no significant effects on any of the dimensions (Figure 6). Additionally, demographic factors including parental language, living situation, family composition, parental employment status also showed no significant influence on the realism, level of centering, or academic outcomes (Figures not included).

Figure 6

Academic performance, realism and level of centering of gender
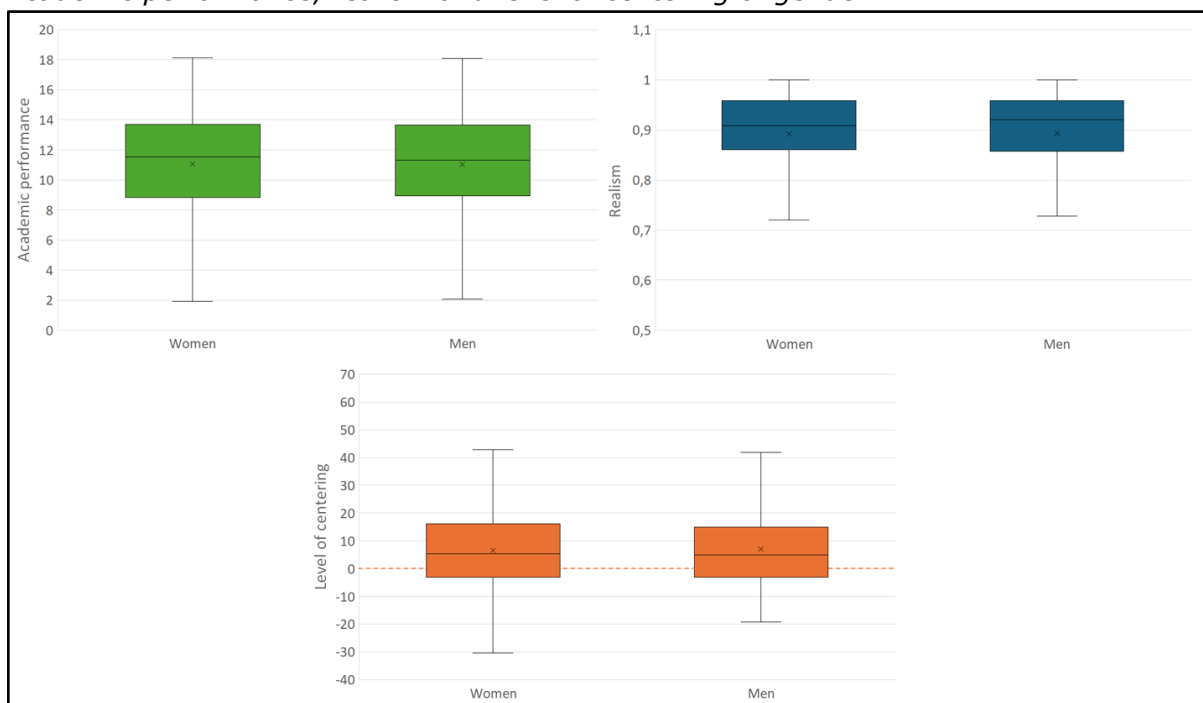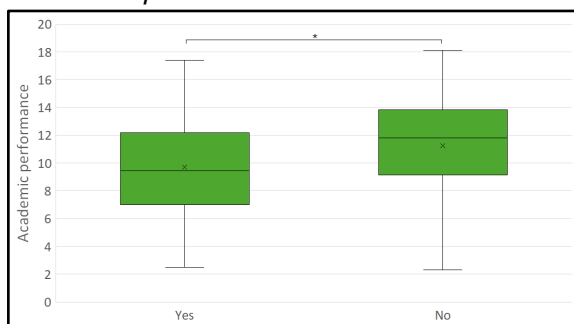

Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

3.1.3 Exam facilities

Access to exam facilities (such as additional time, separate rooms, or other accommodations) was significantly associated with lower academic performance (p = 0.0195), though no significant effects were observed for realism or level of centering (Figure 7; Table 1).

Figure 7

Academic performance of the utilization of facilities

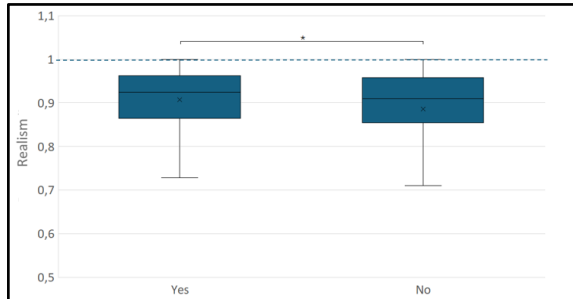**Note.** (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

3.2 Understanding of CBM

3.2.1 Instruction CBM and received advice from others

While the instruction received on CBM did not result in significant differences (Figures not included). As seen in Figure 8, receiving advice from others was significantly associated with better realism outcomes ($p = 0.0352$), though not with level of centering or academic performance (Table 1).

Figure 8

Realism of received advice from others
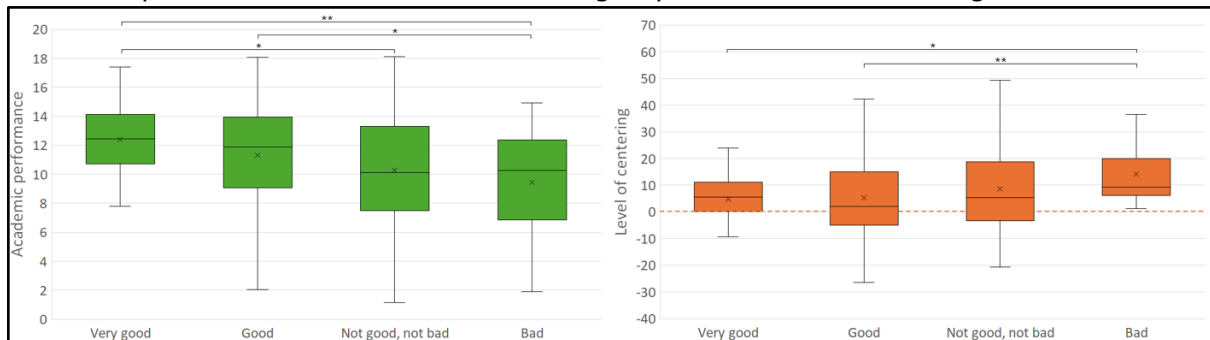

Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

3.2.2 Perceived understanding of CBM

Several subjective perceptions of CBM were found to have statistically significant effects. Perceived understanding of how CBM works has a significant effect on level of centering ($p = 0.0234$) and academic performance ($p = 0.0019$) (Figure 9 and 16; Table 1).

Figure 9

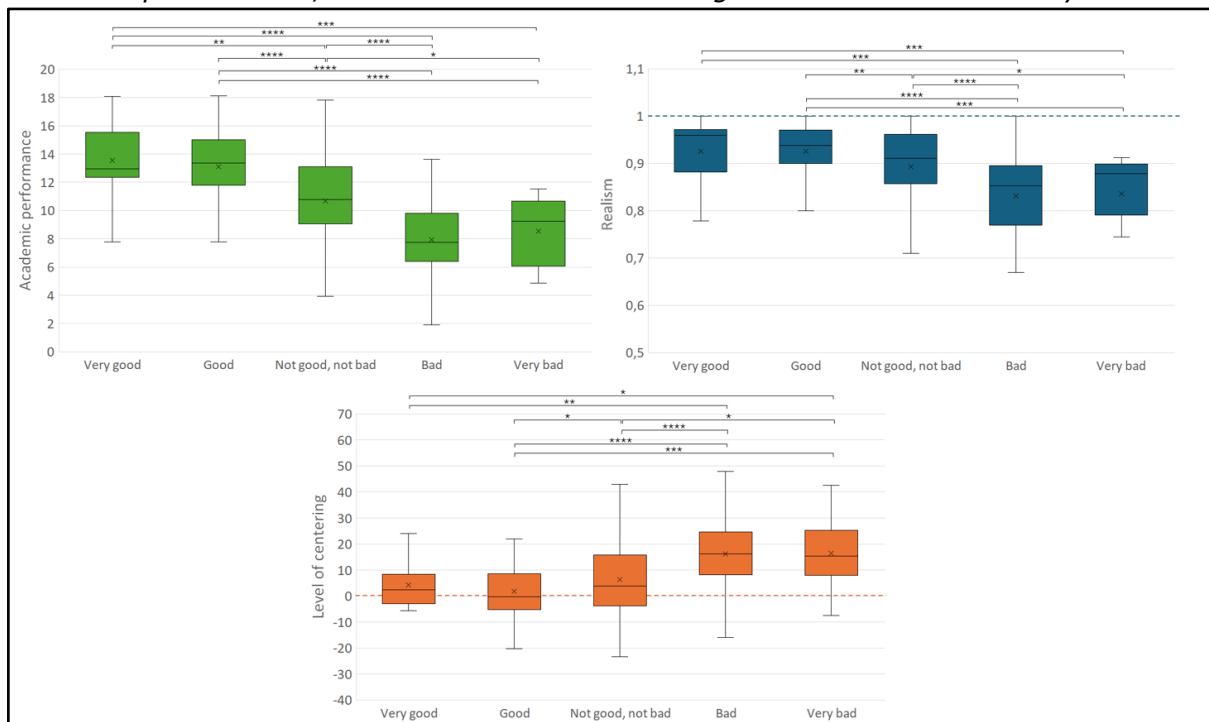*Academic performance and level of centering of perceived understanding of CBM*

Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

3.3 Experience with the system during exams

The self-assessment ability, which indicates their self-assessment of how well they were able to assess their own confidence, was significantly associated with academic performance, realism and level of centering ($p < 0.0001$ for each) (Table 1). Students who reported difficulties in accurately assessing themselves tend to have poorer outcomes (Figure 10).

Figure 10

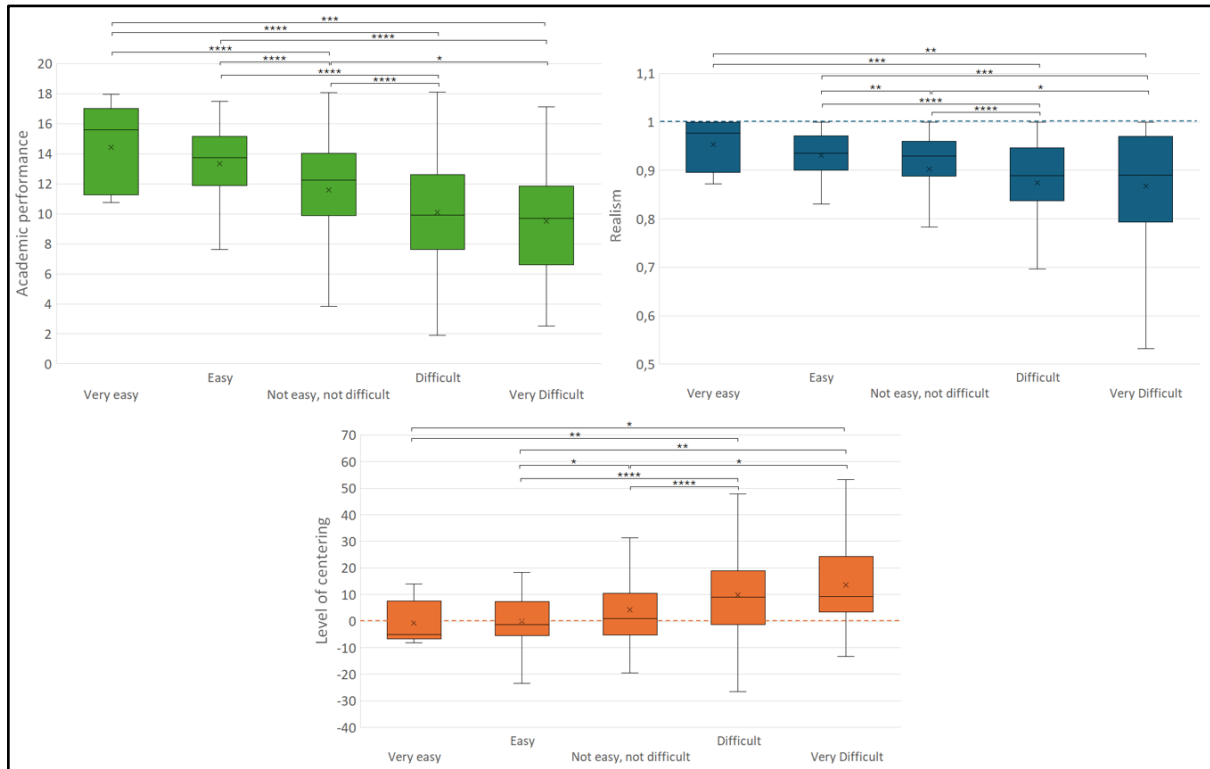*Academic performance, realism and level of centering of self-assessment ability*

Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

Ease of certainty indication, which is the perceived difficulty to indicate a level of confidence for each question, levels demonstrated significant effects on academic performance (p < 0.0001), realism (p = 0.0004) and level of centering (p < 0.0001) (Figure 11; Table 1). In contrast, prior exposure to CBM did not show a statistically significant effect on any of the three outcome measures (Figures not included).

Figure 11

Academic performance, realism and level of centering of certainty indication

Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

3.4 Stress and self-confidence

3.4.1 Stress

Stress levels were significantly associated with level of centering (p = 0.0059) and academic performance (p < 0.0001), but not with realism (Table 1). Students who reported stress using CBM during the exam tend to have poorer outcomes (Figure 12).

Figure 12

Academic performance and level of centering of stress
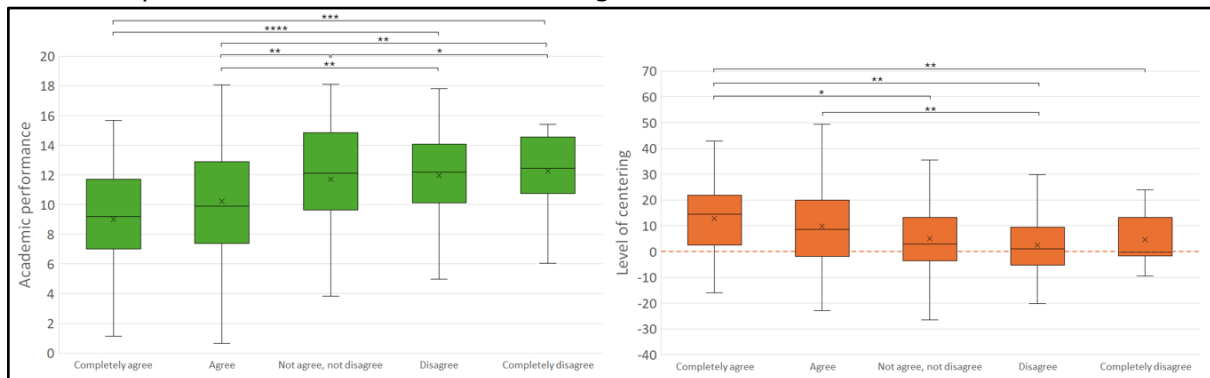
Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

In contrast, self-confidence exhibited a statistically significant effect on all three areas, most notably academic performance ($p < 0.0001$), with additional significance for realism ($p = 0.0382$) and level of centering ($p = 0.0004$) (Table 1). Students who reported to be more self-confident tend to have better outcomes (Figure 13).

Figure 13

Academic performance, realism and level of centering of self-confidence
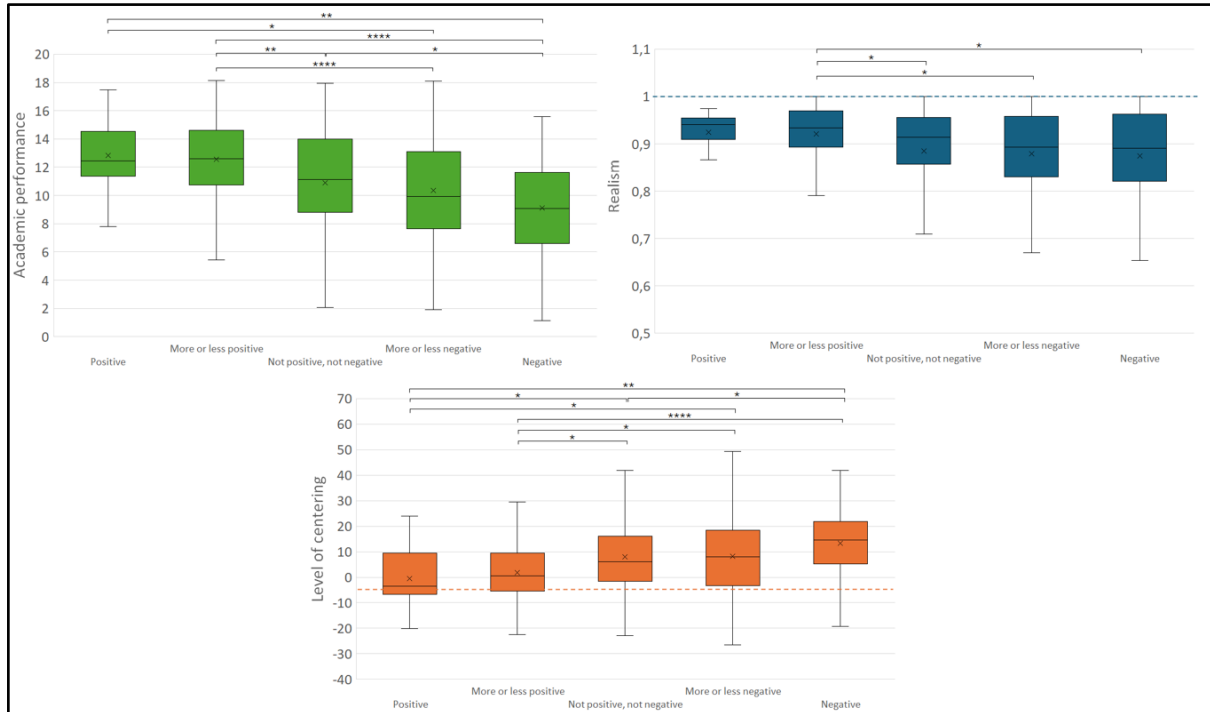
Note. *(*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.*

3.5 Cognitive effort and expected achievement

Students' experience with processing learning material (categorized as difficult, moderate, or fluent) significantly influenced all three domains, with $p = 0.0120$ for realism, $p = 0.0005$ for level of centering and $p < 0.0001$ for academic performance (Table 1). Students who reported that they processed the learning material more effectively exhibited more favorable outcomes (Figure 14).

Figure 14

Academic performance, realism and level of centering of processing learning material
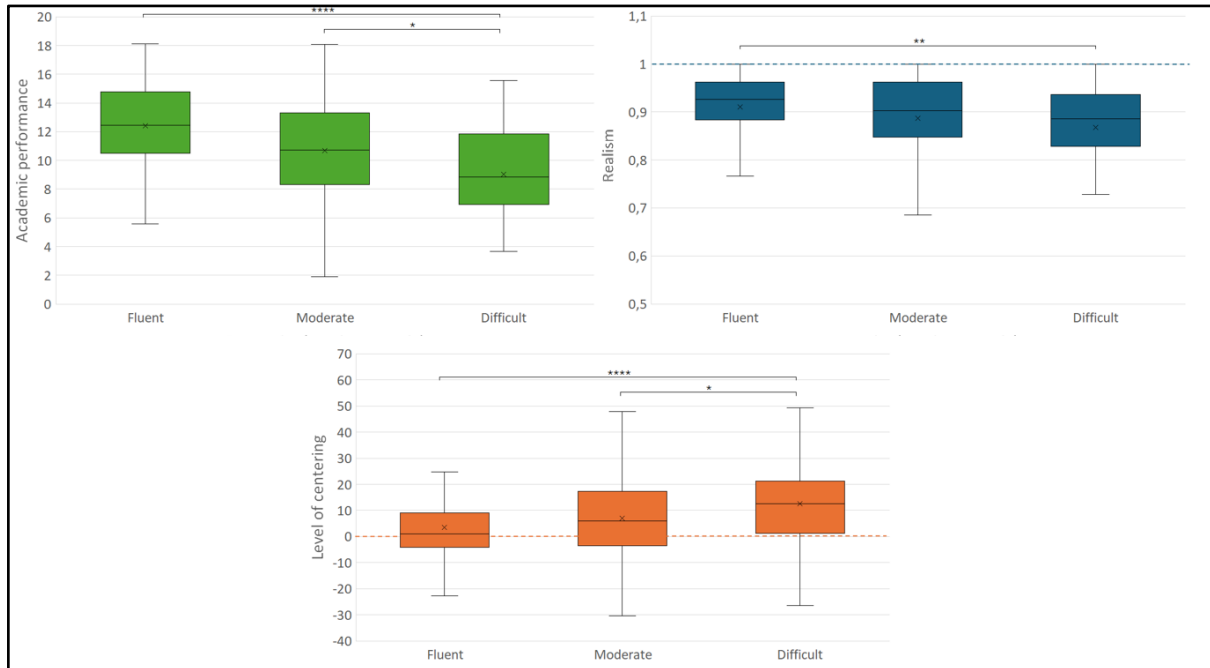
Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

Students who felt that their exam results aligned with their expectations demonstrated significantly better outcomes in academic performance, realism and level of centering (all p < 0.0001) (Figure 15; Table 1).

Figure 15

Academic performance, realism and level of centering of perceived result accuracy
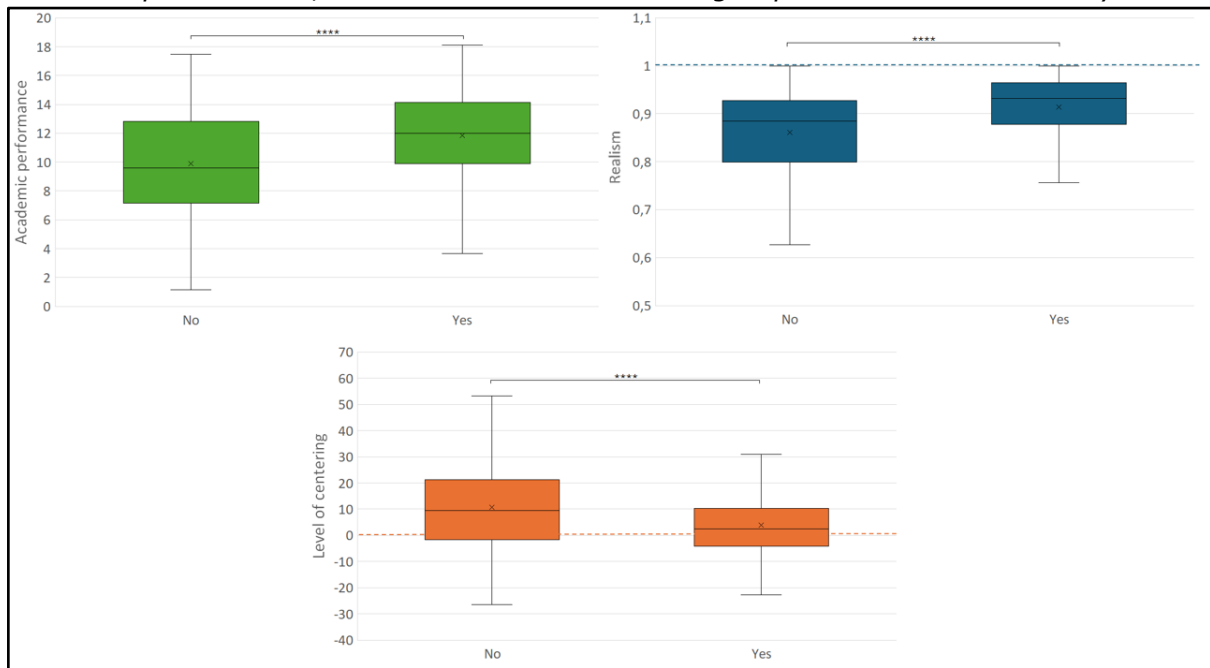
Note. (*) p-value <0.05, (**) p-value <0.01, (***) p-value <0.001, (****) p-value <0.0001.

Table 1

Results academic performance, realism and level of centering

Questioned	Academic performance	Realism	Centering
General questions			
Study area	****	****	****
Medicine	(W)	(KW)	(KW)
Economics			
Rehabilitation sciences			
Study area secondary school	****	****	****
ASO/doorstroomfinaliteit	(MW)	(MW)	(MW)
TSO/dubbele finaliteit			
Gender	-	-	-
Male	(MW)	(MW)	(MW)
Female			
Language parents	-	-	-
Both parents speak Dutch	(KW)	(KW)	(KW)
One parent speaks Dutch			
No parent speaks Dutch			
First contact with CBM	-	-	-
Yes	(MW)	(MW)	(MW)
No			
Living situation	-	-	-
Dormitory	(KW)	(KW)	(KW)
With parents			
Own house			
Family situation: parents are	-	-	-
Single	(KW)	(KW)	(KW)
Married			
Divorced			
Others			
Work situation parents	-	-	-
Both working	(KW)	(KW)	(KW)
One working			
None working			
Others			
Facilities during exam	*	-	-
Yes	(T)	(T)	(T)
No			

Understanding of CBM

Received advice from others			
Yes	-	*	-
No	(T)	(W)	(T)
Instruction CBM	-	-	-
Too much	(KW)	(KW)	(KW)
Much			
Not too much, not too little Little			
Too little			
Perceived understanding of CBM	**	-	*
Very good	(KW)	(KW)	(KW)
Good			
Not good, not bad			
Bad			
Very bad			

Experience with the system during exams

Self-assessment ability			
Very good	****	****	****
Good	(KW)	(KW)	(KW)
Not good, not bad			
Bad			
Very bad			
Ease of certainty indication	****	***	****
Very easy	(KW)	(KW)	(KW)
Easy			
Not easy, not difficult			
Difficult			
Very difficult			
Perceived result accuracy	****	****	****
Yes	(W)	(W)	(W)
No			

Cognitive effort and expected achievement

Stress			
Completely agree	****	-	**
Agree	(KW)	(KW)	(KW)
Not agree, not disagree			
Disagree			
Completely disagree			
Self-confidence	****	*	***
Positive	(W)	(KW)	(KW)
More or less positive			
Not positive, not negative			
More or less negative			
Negative			

Note. KW = kruskal wallis test, W = Welch test, T = T-test, MW = Mann-whitney U -test. (-) = no effect, (*) = significant effect with p-value <0.05, (**) = significant effect with p-value <0.01, (***) = significant effect with p-value <0.001, (****) = significant effect with p-value <0.0001.

4. Discussion

This study examined the inclusivity and perceived effectiveness of CBM in assessment practices among first-year university students. The results suggest that while CBM has the potential to promote equitable assessment, its implementation and interpretation are subject to several contextual and methodological considerations.

4.1 Interpretation of Key Findings

4.1.1 General questions
4.1.1.1 Study area

Significant differences were found between study areas in both secondary and higher education. Students with an ASO background (doorstroomfinaliteit) outperformed those from TSO (dubbele finaliteit) across all three parameters: academic performance, realism, level of centering. These results suggest that ASO students are not only academically stronger but also more capable of accurately estimating their certainty when answering questions. In contrast, TSO students tend to overestimate their knowledge, often assigning high certainty to incorrect answers. This is consistent with educational theory, as ASO students generally receive more cognitively oriented instruction, which may better prepare them for tasks involving metacognitive judgment and self-assessment (Schoendorfer & Emmett, 2012).

At the university level, students in Medicine outperformed those in Economics and Rehabilitation Sciences (Physiotherapy) across all three parameters: academic performance, realism, level of centering. This may reflect variations in curriculum demands and prior academic preparation. Medical students, due to the selection process and nature of training, may develop higher-order reasoning skills more quickly (Schoendorfer & Emmett, 2012). Additionally, students in the Medicine program must pass a competitive entrance examination, which may result in a more academically homogeneous and high-performing student group. What stands out is that students in Rehabilitation Sciences more often come from TSO compared to students in Medicine and Economics. Specifically, 15.48% of students in Rehabilitation Sciences have a TSO background, compared to 1.41% in Economics and 1.18% in Medicine. These factors possibly contribute to the stronger performance of medical students, not only in terms of academic results but also in their ability to realistically assess their own certainty.

4.1.1.2 Gender and Demografic Factors

No significant differences were found based on gender, home situation, or parental employment. This is important because traditional negative marking systems like QNM often disadvantage certain groups, especially risk-averse students, such as females (Schoendorfer & Emmett, 2012). In contrast, CBM appears to be more neutral and inclusive, reducing these sociocultural biases. For example, gender did not have an effect on students' performance or their ability to assess their own certainty, which supports previous research suggesting CBM helps minimize gender related disparities common in conventional testing. Similarly, differences in social and cultural backgrounds, such as family situation, the language spoken at home, or parents' employment status, did not influence results. This suggests that CBM offers a fairer assessment environment where such factors have less impact on outcomes.

It should be noted, however, that this study only examined the type of living arrangement and did not explore other aspects of students' living environments. Ramli and Zain (2019) highlight that student housing plays a significant role in academic performance. Importantly, their findings suggest that it is not simply the type of living arrangement whether living with parents, in a dormitory, or alone that matters most, but rather the quality and suitability of the living environment. Comfortable, quiet and well equipped housing conditions provide students with better study environments, reducing stress and enabling greater focus and study time. Thus, positive academic outcomes are closely tied to living conditions that support effective learning rather than the form of accommodation itself.

4.1.1.3 Exam facilities

Students with access to exam facilities performed significantly worse in terms of academic performance. However, no significant effects were observed on realism or level of centering. This suggests that students with access to exam facilities did not have a more negative or unrealistic self-perception compared to those who did not have access. In conclusion, CBM appears to be inclusive for students who receive support (e.g. more exam time). Nevertheless, students with facilities tend to perform worse academically. This finding is surprising, as literature shows that adequate support and facilities typically lead to more equal academic performance (Ramli & Zain, 2019).

4.1.2 Understanding of CBM and the experience with the system during exams

Students' perceived understanding of CBM, self-assessment ability and the ease of indication certainty were all significantly associated with better outcomes. These findings align with theoretical frameworks suggesting that informed and confident use of CBM tools enhances performance and metacognitive calibration (De Finetti, 1965). Notably, prior exposure to CBM did not lead to a statistically significant difference in outcomes. However, this finding should be interpreted with caution, as only 12 students had prior exposure. Nevertheless, the results suggest that clear instruction and preparation may play a more decisive role than familiarity alone. As De Finetti (1965) emphasizes, the scoring method must be clearly understood by test-takers to allow for meaningful engagement.

Students who indicated that their exam results aligned with their expectations showed significantly higher scores in academic performance, realism and level of centering. This suggests that perceived result accuracy may reflect stronger metacognitive awareness, where students are able to accurately predict their performance and assign appropriate certainty levels.

4.1.3 Cognitive effort and expected achievement

Self-confidence appears to be positively associated with all three parameters: academic performance, realism and level of centering. Students who reported higher levels of self-confidence generally achieved better scores, showed less bias in their self-assessment and demonstrated a higher degree of accuracy in estimating whether their answers were correct. These results are consistent with earlier research showing that confidence and self-efficacy are positively linked to academic performance and metacognitive accuracy (Grazziotin-Soares et al., 2021).

Stress was significantly associated with both level of centering and academic performance. Students who experienced stress of using CBM during the exam tended to overestimate their performance, indicating a statistically greater discrepancy between their expected and actual performance. Nevertheless, the centering scores remained within a range of +15 to -15, which is generally interpreted as neutral, suggesting that while stress had an effect, the extent of it was limited.

These findings highlight the role of emotional factors in the application of CBM. Although the system is designed to support more accurate and reflective assessment, elevated stress may reduce its effectiveness for some students. However, clear instructions can help reduce stress, which in turn supports better performance. This aligns with previous observations that students must not only understand the scoring mechanism but also be able to apply it consistently under pressure, in line with decision-theoretical principles (De Finetti, 1965).

4.1.4 Learning process

Students which indicated that they had increased difficulty (moderate to difficult) processing the learning materials significantly scored worse on respectively two or three parameters with regards to the students who indicated they fluently processed the material. These findings emphasize the importance of cognitive accessibility and learner-centered material design in supporting academic success. They also point to the necessity for educators to implement differentiated instructional strategies and provide scaffolding to support learners with varying levels of processing ability, thereby promoting more equitable learning outcomes. This is in line with the findings of Qorib (2024), who underscores the critical role of differentiated instruction in fostering Critical Diversity Literacy (CDL) within inclusive education. According to Qorib, tailored instructional approaches not only enhance engagement among diverse learners but also help bridge performance gaps by accommodating varied cognitive processing speeds and learning needs.

4.2. Methodological Considerations

It is important to recognize that a pre-selection effect may have influenced the results. All participants were already enrolled at Hasselt University, where a minimum level of Dutch language proficiency is required.

Another limitation of this study is that the questionnaire was not formally validated. Although it was based on existing literature and expert input, no psychometric testing was carried out to assess aspects such as construct validity, internal consistency, or factor structure. For future use, a pilot phase with validation measures is recommended.

Additionally, the questionnaire data collection was conducted during the second trimester for Medicine and Economics and during the third trimester for Rehabilitation Sciences. This timing may have led to less optimal results, as the questionnaire focused on courses from the first trimester. This introduces a potential recall bias, as students were asked to evaluate first-trimester courses several weeks or months later, possibly affecting the accuracy of their responses.

Furthermore, the study specifically targeted first-year students, who had limited to no prior exposure to CBM. This increases the reliability of the findings concerning first-time CBM use. Nevertheless, some participants may have previously completed the course (e.g., due to retaking the subject), potentially introducing variability in CBM familiarity. In this sample, 20 participants indicated that this was not their first encounter with CBM, suggesting that a small subset may have had prior experience. Future studies should control for this factor more explicitly to ensure a clearer interpretation of first-time use effects.

Finally, the limited representation of international students in the sample restricts the generalizability of the findings to multilingual or non-Dutch-speaking populations. Since only 28 students indicated that one or both of their parents speak a language other than Dutch at home, the conclusions about the linguistic inclusiveness of CBM should be interpreted with caution, as the limited representation makes it difficult to draw robust conclusions.

4.3. Broader Implications

The findings support the claim that CBM can enhance assessment practices by engaging students in higher-order thinking processes and promoting deeper learning (Schoendorfer & Emmett, 2012). By requiring learners to reflect on their certainty, CBM fosters greater awareness of personal knowledge gaps and encourages caution in decision-making, a crucial skill in high-stakes fields such as healthcare. However, the full inclusivity of CBM can only be assured when all students are adequately trained to understand and apply its logic effectively.

5. Conclusion

This study aimed to explore the inclusivity of CBM in higher education by examining its relationship with socio-cultural, academic, cognitive and emotional factors. The findings indicate that CBM is a viable alternative to traditional assessment methods. It requires students to reflect on their certainty in addition to providing correct answers, promoting metacognitive engagement.

No significant differences were found in academic performance, realism, or level of centering based on gender, language spoken at home or family background, supporting the inclusive potential of CBM. In contrast, students with an ASO background and those studying Medicine outperformed peers in all three measured parameters, suggesting a role of prior academic preparation. Similarly, lower outcomes were observed in students who experienced higher stress, had difficulty processing learning materials, or received exam facilities, although the latter group did not differ in self-assessment accuracy.

The effectiveness of CBM was also linked to students' understanding of the method and their confidence in using it. Prior exposure alone did not improve outcomes, underlining the importance of explicit instruction and engagement. These results highlight the need for differentiated instructional strategies and cognitive accessibility to ensure CBM benefits all learners equally.

In sum, CBM offers an efficient and reflective assessment format that supports equitable evaluation, provided that its implementation is supported by adequate training and inclusive teaching practices.

To further enhance CBM, future research should develop validated instruments, include more diverse student populations and better align data collection with course timing. Longitudinal and qualitative studies, especially involving students with learning difficulties or accommodations, are essential to improve its inclusivity and educational value.

Bibliography

Abdolreza Gharehbagh, Z., Mansourzadeh, A., Montazeri Khadem, A., & Saeidi, M. (2022). Reflections on using open-ended questions. *Medical Education Bulletin, 3*(2), 475–482. https://doi.org/10.22034/meb.2022.333518.1054

Consortium for Educational Development, Evaluation, and Research, Texas A&M University–Corpus Christi College of Education, Hu, B. Y., Pate, R. S., Gauthier, L. R., Schorzman, E. M., Culbertson, D., ... Lucido, F. (2012). *Educational research and innovations 2012 CEDER yearbook* (E. Ortlieb et al., Eds.). CEDER, Texas A&M University–Corpus Christi.

De Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology, 18*(1), 87–123. https://doi.org/10.1111/j.2044-8317.1965.tb00695.x

Fabre, J. (1978). Docimologie expérimentale et évaluation par questionnaires: Étude du jugement multiple et de l'autopondération. *Bulletin de Psychologie, 32*(338), 97–99. https://doi.org/10.3406/bupsy.1978.11569

Gardner-Medwin, A., & Curtin, N. (2007). Certainty-Based Marking (CBM) for reflective learning and proper knowledge assessment. *Creative Commons*. http://discovery.ucl.ac.uk/178612/

Gardner-Medwin, T., Curtin, N., University College London, & Imperial College London. (n.d.). *Certainty-Based Marking (CBM) for reflective learning and proper knowledge assessment*. https://www.ucl.ac.uk/lapt/REAP_cbm.pdf

Gharibyan, H. (2005). Assessing students' knowledge. *ACM SIGCSE Bulletin, 37*(3), 143–147. https://doi.org/10.1145/1151954.1067487

Grazziotin-Soares, R., Blue, C., Feraro, R., Tochor, K., Ardenghi, T. M., Curtis, D., & Ardenghi, D. M. (2021). The interrelationship between confidence and correctness in a multiple-choice assessment: Pointing out misconceptions and assuring valuable questions. *BDJ Open, 7*(1). https://doi.org/10.1038/s41405-021-00067-4

Hendriks, W. J. A. J., Bakker, N., Pluk, H., De Brouwer, A., Wieringa, B., Cambi, A., ... Klaren, P. H. M. (2019). Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. *BMC Medical Education, 19*(1). https://doi.org/10.1186/s12909-019-1610-2

Karadag, N., Yuksekdag, B. B., Akyildiz, M., & Ibileme, A. I. (2020). Assessment and evaluation in open education system: Students' opinions about open-ended question (OEQ) practice. *Turkish Online Journal of Distance Education, 22*(1), 179–193. https://doi.org/10.17718/tojde.849903

Kolbitsch, J., Ebner, M., Nagler, W., & Scerbakov, N. (2008). Can confidence assessment enhance traditional multiple-choice testing. In *Interactive Computer Aided Learning, ICL2008*.

Leclercq, D. (1982). Confidence marking: Its use in testing. *Evaluation in Education: An International Review Series, 6*(2), 182–271.

Mustamin, R. (2024). The role of formative and summative assessment in improving learning quality and student learning outcomes. *Journal of Strategy and Transformation in Educational Management*, 17–21.

Muradova, M. U. (2025). The role of formative and summative assessment in shaping instruction. *Экономика и социум, (1-1 (128))*, 198–201.

Onasanya, S. A., & Ajamu, A. O. (2024). Formative and summative assessment using technology: A critical review. *Mimbar Ilmu, 29*(1), 1–8. https://doi.org/10.23887/mi.v29i1.72042

Prosperi, O. (2016). Le réalisme avec degrés de certitude. *Mesure et Évaluation en Éducation, 38*(1), 121–140. https://doi.org/10.7202/1036553ar

Ptasczynski, L. E., Steinecker, I., Sterzer, P., & Guggenmos, M. (2022). The value of confidence: Confidence prediction errors drive value-based learning in the absence of external feedback. *PLoS Computational Biology, 18*(10), e1010580. https://doi.org/10.1371/journal.pcbi.1010580

Qorib, M. (2024). Analysis the impact of differentiated instruction on critical diversity literacy in inclusive education. *Aksaqila International Humanities and Social Sciences [AIHSS] Journal, 3*(1), 1–19.

Ramli, A., & Zain, R. M. (2019). The impact of facilities on students' academic achievement. *Scientific International (Lahore), 30*(2), 299–311. https://www.researchgate.net/publication/337590619

Remesal, A., García-Mínguez, P., Domínguez, J., & Corral, M. J. (2024). Certainty-based self-assessment in higher education. *International Journal of Advanced Corporate Learning (iJAC), 17*(3), 58–67. https://doi.org/10.3991/ijac.v17i3.45675

Sanderson, P. J. (2010). Multiple-choice questions: A linguistic investigation of difficulty for first-language and second-language students. *University of South Africa* [Thesis].

Schoendorfer, N., & Emmett, N. (2012). Use of certainty-based marking in a second-year medical student cohort: A pilot study. *Advances in Medical Education and Practice*, 139. https://doi.org/10.2147/amep.s35972

Schuman, H., & Presser, S. (1979). The open and closed question. *American Sociological Review, 44*(5), 692. https://doi.org/10.2307/2094521

Smrkolj, Š., Bančov, E., & Smrkolj, V. (2022). The reliability and medical students' appreciation of certainty-based marking. *International Journal of Environmental Research and Public Health, 19*(3), 1706. https://doi.org/10.3390/ijerph19031706

Wu, Q., Vanerum, M., Agten, A., Christiansen, A., Vandenabeele, F., Rigo, J., & Janssen, R. (2021). Certainty-based marking on multiple-choice items: Psychometrics meets decision theory. *Psychometrika, 86*(2), 518–543. https://doi.org/10.1007/s11336-021-09759-0

Appendix 1

Informed Consent Formulier

Titel van het Onderzoek:
Hoe inclusief zijn zekerheidsgraden?

Onderzoekers:

Alix Lumbeeck
alix.lumbeeck@student.uhasselt.be

Dorien Boeckmans
dorien.boeckmans@student.uhasselt.be

Margot Cochez
margot.cochez@student.uhasselt.be

Brent Van Ballaer
brent.vanballaer@student.uhasselt.be

Doel van het Onderzoek:
Het identificeren van belemmerende factoren bij inclusie tijdens evaluaties met zekerheidsgraden.

Wat houdt deelname in?
Als u besluit deel te nemen, wordt u gevraagd om de volgende vragenlijst waarheidsgetrouw in te vullen. Uw deelname zal ongeveer 10 minuten in beslag nemen.

Vrijwilligheid en Recht om te Stoppen:
Deelname aan dit onderzoek is volledig vrijwillig. U kunt op elk moment besluiten om te stoppen zonder dat dit gevolgen heeft. U hoeft geen reden op te geven.

Vertrouwelijkheid en Anonimiteit:
Alle gegevens die tijdens dit onderzoek worden verzameld, worden strikt vertrouwelijk behandeld. Uw antwoorden worden geanonimiseerd door de supervisors en zullen niet aan uw identiteit worden gekoppeld in rapportages of publicaties.

Mogelijke Risico's en Voordelen:
Er zijn geen bekende risico's verbonden aan uw deelname. Deelname kan echter bijdragen aan een beter inzicht in het gebruik van zekerheidsgraden en eventuele tekortkomingen, wat op termijn voordelen kan opleveren voor alle studenten die gebruik maken van dit systeem.

Contactinformatie:
Als u vragen heeft over dit onderzoek, kunt u contact opnemen met de onderzoekers. Dit project werd goedgekeurd door de ethische commissie van de UHasselt. Voor vragen over uw rechten als deelnemer kunt u contact opnemen met de UHasselt.

Toestemming:
Door hieronder te ondertekenen, verklaart u dat:

- U geïnformeerd bent over het doel en de aard van het onderzoek.
- U begrijpt dat uw deelname vrijwillig is en dat u op elk moment kunt stoppen.
- U akkoord gaat met het verzamelen en verwerken van uw gegevens zoals hierboven beschreven.

Appendix 2

Enquête Masterthesis

Q37 Ik ga akkoord met de voorwaarden. Ik heb het document, ter beschikking via volgende link, gelezen en begrepen en ga akkoord met de voorwaarden: https://docs.google.com/document/d/1vYfBM5X9qlnXI7IR9BElem4d475UagMKj8noBCGzI4Q/edit?usp=sharing.

○ Ja (1)

○ Nee (2)

Q40 Studentennummer

_____

Was het in 2024 uw eerste kennismaking met zekerheidsgraden (Ans/docimo)?

○ Ja (1)

○ Nee (2)

Geslacht?

○ Man (1)

○ Vrouw (2)

○ Niet-binair/derde geslacht (3)

○ Ik zeg dat liever niet (4)

Studierichting middelbaar?

○ ASO (Doorstroom) (1)

○ TSO (Dubbele finaliteit) (2)

○ BSO (Arbeid) (3)

Gezinssituatie: zijn uw ouders

○ Alleenstaande (1)

○ Gehuwd/samenwonend (2)

○ Gescheiden (3)

○ Anders (4)

Deze vraag weergeven:
  If Gezinssituatie: zijn uw ouders = Anders

Q32 Indien anders:

_____

Pagina-einde ——

Spreken uw ouders Nederlands thuis?

◯ Beide ouders spreken Nederlands thuis (1)

◯ Eén ouder spreekt Nederlands thuis (2)

◯ Geen enkele ouder spreekt Nederlands (3)

Q38 Is Nederlands uw moedertaal

◯ Ja (1)

◯ Nee (2)

Q39 Indien nee, wat is uw moedertaal?

_____

Pagina-einde

Q34 Woonsituatie tijdens de schoolweek

○ Op kot (1)

○ Bij de ouders (2)

○ In eigen woning (3)

Werksituatie ouders

○ Beide ouders werken (1)

○ 1 ouder thuis (2)

○ Beide ouders thuis (3)

○ Anders (4)

Q35 Indien anders:

_____

Pagina-einde

Heeft u recht op faciliteiten tijdens het maken van een examen?

○ Ja (1)

○ Nee (2)

Q36 Heeft er iemand in uw omgeving, die al bekend is met zekerheidsgraden, u tips of advies gegeven?

○ Ja (1)

○ Nee (2)

Einde blok: Standaard vragenblok

Start van blok: Begrip van het zekerheidsgradensysteem

Q12 Hebt u voldoende instructies gekregen over het zekerheidsgradensysteem?

○ Veel te veel (1)

○ Enigszins te veel (2)

○ Noch te veel, noch te weinig (3)

○ Enigszins te weinig (4)

○ Veel te weinig (5)

Q13 Hoe goed dacht u het zekerheidsgradensysteem te begrijpen voor u aan het examen begon?

○ Zeer goed (1)

○ Goed (2)

○ Noch goed, noch slecht (3)

○ Slecht (4)

○ Zeer slecht (5)

Hoe goed heeft u uw eigen zekerheid kunnen inschatten tijdens het examen?   Hierbij vragen we naar zelfinschatting, niet naar kennis. Bijvoorbeeld: wanneer u vaak de laagste zekerheid heeft aangeduid en dan ook veel fouten heeft gemaakt, heeft u een goede inschatting gemaakt.

○ Zeer goed (1)

○ Goed (2)

○ Noch goed, noch slecht (3)

○ Slecht (4)

○ Zeer slecht (5)

Q15 Was het moeilijk om voor iedere vraag een zekerheidsgraad aan te geven?

○ Zeer moeilijk (1)

○ Eerder moeilijk (2)

○ Noch gemakkelijk, noch moeilijk (3)

○ Eerder gemakkelijk (4)

○ Zeer gemakkelijk (5)

Q17 In hoeverre beïnvloedde het zekerheidsgradensysteem uw zelfvertrouwen tijdens het examen?

○ Positief, ik voelde me zelfverzekerder (1)

○ Enigszins positief (2)

○ Noch positief, noch negatief (3)

○ Enigszins negatief (4)

○ Negatief, ik voelde me minder zelfzeker (5)

Q18 Verhoogde het moeten aangeven van zekerheid uw stressniveau tijdens het examen?

    ○ Helemaal niet mee eens (1)

    ○ Enigszins mee oneens (2)

    ○ Noch eens, noch oneens (3)

    ○ Enigszins mee eens (4)

    ○ Helemaal mee eens (5)

Einde blok: Zelfvertrouwen en stress

Start van blok: Algemeen oordeel over het zekerheidsgradensysteem

Q19 Hoe waardevol vond u het zekerheidsgradensysteem in dit examen?

    ○ Zeer waardevol (1)

    ○ Waardevol (2)

    ○ Neutraal (3)

    ○ Waardeloos (4)

    ○ Zeer waardeloos (5)

Q20 Zou u dit zekerheidsgradensysteem aanbevelen voor andere examens?

    ○ Nee (1)

    ○ Onwaarschijnlijk (2)

    ○ Neutraal (3)

    ○ Waarschijnlijk (4)

    ○ Ja (5)

Q21 Welke van de volgende verbeteringen zou u willen zien in het zekerheidsgradensysteem? (Meerdere antwoorden mogelijk)

- ☐ Betere uitleg vooraf (1)

- ☐ Voorbeelden of oefenvragen vooraf (2)

- ☐ Minder zekerheidsintervallen (3)

- ☐ Meer tijd per vraag (4)

- ☐ Andere (5)

Q22 Indien andere op de vorige vraag, welke verbeteringen?

_____

Einde blok: Algemeen oordeel over het zekerheidsgradensysteem

Start van blok: Vakspecifiek

Q22 Welke richting volgt u?

- ○ Revalidatiewetenschappen (1)

- ○ Geneeskunde (2)

- ○ Handelsingenieur of toegepaste economische wetenschappen (3)

Q41 Het behaalde cijfer voor het examen macro-economie in welk interval?

○ <8 (1)

○ 8-9 (2)

○ 10-13 (3)

○ 14-16 (4)

○ >16 (5)

Q42 Hoe verliep het verwerken van de leerstof macro-economie?

○ Vlot (1)

○ Matig (2)

○ Moeizaam (3)

○ Ik heb de leerstof niet verwerkt (4)

Q43 Komt het behaalde resultaat voor macro-economie overeen met uw verwachtingen?

○ Ja (1)

○ Nee (2)

Q23 Het behaalde cijfer voor het examen organisatieniveaus/ basisprincipes kinesiologie in welk interval?

○ <8 (1)

○ 8-9 (2)

○ 10-13 (3)

○ 14-16 (4)

○ >16 (5)

Q24 Hoe verliep het verwerken van de leerstof organisatieniveaus/ basisprincipes kinesiologie?

○ Vlot (1)

○ Matig (2)

○ Moeizaam (3)

○ Ik heb de leerstof niet verwerkt (4)

Q25 Komt het behaalde resultaat voor organisatieniveaus/ basisprincipes kinesiologie overeen met uw verwachtingen?

○ Ja (1)

○ Nee (2)

Q26 Het behaalde cijfer voor het examen histologie en genetica welk interval?

○ <8 (1)

○ 8-9 (2)

○ 10-13 (3)

○ 14-16 (4)

○ >16 (5)

Q27 Hoe verliep het verwerken van de leerstof histologie en genetica?

○ Vlot (1)

○ Matig (2)

○ Moeizaam (3)

○ Ik heb de leerstof niet verwerkt (4)

Q28 Komt het behaalde resultaat voor histologie en genetica overeen met uw verwachtingen?

○ Ja (1)

○ Nee (2)

Q29 Het behaalde cijfer voor het examen zorg en gezondheid welk interval?

○ <8 (1)

○ 8-9 (2)

○ 10-13 (3)

○ 14-16 (4)

○ >16 (5)

Q30 Hoe verliep het verwerken van de leerstof zorg en gezondheid?

○ Vlot (1)

○ Matig (2)

○ Moeizaam (3)

○ Ik heb de leerstof niet verwerkt (4)

Q31 Komt het behaalde resultaat voor zorg en gezondheid overeen met uw verwachtingen?

○ Ja (1)

○ Nee (2)

Einde blok: Vakspecifiek