



**HAL**  
open science

## Formative Assessment of Diagnostic Testing in Family Medicine with Comprehensive MCQ Followed by Certainty-Based Mark.

Charles Herbaux, Aurélie Dupré, Wendy Rénier, Ludovic Gabellier, Emmanuel Chazard, Philippe Lambert, Vincent Sobanski, Didier Gosset, Dominique Lacroix, Patrick Truffert

### ► To cite this version:

Charles Herbaux, Aurélie Dupré, Wendy Rénier, Ludovic Gabellier, Emmanuel Chazard, et al.. Formative Assessment of Diagnostic Testing in Family Medicine with Comprehensive MCQ Followed by Certainty-Based Mark.. Healthcare (Basel), 2022, Healthcare (Basel), 10, 10.3390/healthcare10081558 . hal-04396921

**HAL Id: hal-04396921**

**<https://hal.univ-lille.fr/hal-04396921>**

Submitted on 16 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Article

# Formative Assessment of Diagnostic Testing in Family Medicine with Comprehensive MCQ Followed by Certainty-Based Mark

Charles Herbaux <sup>1,2,\*</sup>, Aurélie Dupré <sup>3</sup>, Wendy Rénier <sup>2</sup>, Ludovic Gabellier <sup>2</sup>, Emmanuel Chazard <sup>4</sup>, Philippe Lambert <sup>5</sup>, Vincent Sobanski <sup>6</sup>, Didier Gosset <sup>7</sup>, Dominique Lacroix <sup>7</sup> and Patrick Truffert <sup>8</sup>

- <sup>1</sup> Clinical Hematologic Department, University of Lille, CHU Lille, ULR 2694 Metrics, F-59000 Lille, France  
<sup>2</sup> Clinical Hematologic Department, University of Montpellier, CHU Montpellier, F-34000 Montpellier, France  
<sup>3</sup> Laboratoire CIREL (EA 4354), Service Conseil et Accompagnement à la Pédagogie (DIP-CAPE), University of Lille, F-59000 Lille, France  
<sup>4</sup> CERIM Public Health Department, University of Lille, CHU Lille, ULR 2694 Metrics, F-59000 Lille, France  
<sup>5</sup> General Medicine Department, University Montpellier, CHU Montpellier, F-34000 Montpellier, France  
<sup>6</sup> Internal Medicine Department, University of Lille, CHU Lille, ULR 2694 Metrics, F-59000 Lille, France  
<sup>7</sup> Lille University School of Medicine, University of Lille, CHU Lille, ULR 2694 Metrics, F-59000 Lille, France  
<sup>8</sup> Pediatric Department, University of Lille, CHU Lille, ULR 2694 Metrics, F-59000 Lille, France  
\* Correspondence: c-herbaux@chu-montpellier.fr



**Citation:** Herbaux, C.; Dupré, A.; Rénier, W.; Gabellier, L.; Chazard, E.; Lambert, P.; Sobanski, V.; Gosset, D.; Lacroix, D.; Truffert, P. Formative Assessment of Diagnostic Testing in Family Medicine with Comprehensive MCQ Followed by Certainty-Based Mark. *Healthcare* **2022**, *10*, 1558. <https://doi.org/10.3390/healthcare10081558>

Academic Editor: Marco Dettori

Received: 26 June 2022

Accepted: 14 August 2022

Published: 17 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Introduction: The choice of diagnostic tests in front of a given clinical case is a major part of medical reasoning. Failure to prescribe the right test can lead to serious diagnostic errors. Furthermore, unnecessary medical tests are a waste of money and could possibly generate injuries to patients, especially in family medicine. Methods: In an effort to improve the training of our students to the choice of laboratory and imaging studies, we implemented a specific multiple-choice questions (MCQ), called comprehensive MCQ (cMCQ), with a fixed and high number of options matching various basic medical tests, followed by a certainty-based mark (CBM). This tool was used in the assessment of diagnostic test choice in various clinical cases of general practice in 456 sixth-year medical students. Results: The scores were significantly correlated with the traditional exams (standard MCQ), with matched themes. The proportion of “cMCQ/CBM score” variance explained by “standard MCQ score” was 21.3%. The cMCQ placed students in a situation closer to practice reality than standard MCQ. In addition to its usefulness as an assessment tool, those tests had a formative value and allowed students to work on their ability to measure their doubt/certainty in order to develop a reflexive approach, required for their future professional practice. Conclusion: cMCQ followed by CBM is a feasible and reliable evaluation method for the assessment of diagnostic testing.

**Keywords:** education; diagnostics

## 1. Introduction

The choice of diagnostic tests in front of a given clinical case is a major part of medical reasoning. Failure to prescribe the right test can lead to serious diagnosis delays and errors, especially in family medicine. Furthermore, unnecessary medical tests are a waste of money and could possibly generate injuries to patients [1,2]. Using open questions or modified essay questions [3] to evaluate the choice of laboratory and imaging studies allows for an unlimited range of answers but is time consuming to mark and present considerable variation in standards of marking [4]. New methods can consistently assess clinical reasoning, for example, with think-aloud and concept-mapping protocols [5]. However, given their complexity, these evaluations are difficult to apply to the regular assessment of large numbers of students. Thus, multiple-choice questions (MCQs) are frequently used during university tests and commonly in medical studies examinations in France. They often require less time to administer for a given amount of material than would tests requiring

written responses and they have a high level of reliability. Moreover, they display high potential time–economy advantages for the correctors. In contrast, MCQs allow sight recognition or random guessing [6–8]. The majority of MCQs test lower-order thinking skills (recall and comprehension) rather than higher-order skills, such as application and analysis. Nevertheless, it has been shown that a well-structured MCQ has the ability to assess higher-ordered thinking [9], as described by Bloom’s taxonomy [10,11]. Creating high-quality MCQ is time consuming and the work of MCQ authors needs to be constantly and carefully assessed [12–14].

Certainty-based marking (CBM) is another tool that can be used in the educational field. CBM is also known as confidence-based marking. A question is added for each answer, asking the student to assess the level of confidence in the given answer (i.e., “are you sure at 60%, 80% or 100%?”) [15,16]. This technique has been used in medical education to encourage reflection on reasoning prior to making clinical-based decisions [17,18] and it has already been associated with MCQs [19]. It is thought to enhance deeper levels of learning at the expense of common learning practices [20,21].

In an effort to improve our students’ training to diagnostic testing, we implemented a specific type of MCQ, with a high number (107) of options matching various basic medical tests. We called this assessment method “comprehensive” MCQ (cMCQ), because all possible options reasonably needed by a general practitioner for a given outpatient situation are given. cMCQ was then followed by a CBM. The aim of the present study was to explore the feasibility and reliability of cMCQ/CBM to assess the choice of laboratory and imaging studies in front of various clinical cases.

## 2. Methods

### 2.1. Participants

This prospective study was conducted in October 2017, at Lille University School of Medicine. The study was approved by the local Institutional Review Board (“*Conseil de Faculté*”). Students were informed orally and with a written protocol about study proceedings, on a voluntary basis. cMCQ/CBM scores did not account for official faculty assessment. The experimental questions were submitted to students right before the standard tests. All data were anonymously analyzed. Overall, out of the 498 students attending their sixth year of medical school, 456 students agreed to participate (91%) and were included in this study. In the whole year group, twenty-three students were not attending the faculty test and nineteen denied participation. The results of this study were presented to students at the end of the academic year.

### 2.2. Procedure and Data Collection Method

Each question in the experimental assessing method was composed of three different parts. The first one was a short clinical scenario, ending with one of the two following questions: “In this situation, which diagnostic test(s) is (are) indicated?” or “What are the “x” diagnostic tests that can confirm your main hypothesis?”, where “x” was a number set by the author of the cMCQ. The second part was the list of possible diagnostic tests to prescribe (Appendix A). This list included 108 proposals, corresponding to 107 basic medical tests, as well as the following proposal “No further examination is necessary”. They were written to cover all possible options reasonably needed in front of a consultation in family medicine. The third part consisted of a request for degrees of certainty regarding the previous answer (CBM). The student could choose one out of six different degrees of certainty allocated to their response (i.e., 0%, 20%, 40%, 60%, 80% and 100%). This numerical method is described to be more relevant in the CBM [21]. From a teaching point of view, asking students to assess their degree of confidence allows us to distinguish 4 types of answers [22]: a/ “serious mistake”: error made with high degree of certainty; b/ “mastered answer”: correct answer with high degree of certainty; c/ “ignorance”: error made with low degree of certainty; d/ “weak knowledge”: correct answer given with low

degree of certainty. Responses falling into the two last categories are strongly influenced by randomness.

Only the first part of the cMCQ was modified from question to question, parts 2 and 3 remained unchanged. Two examples of cMCQ/CBM are presented in Appendix B. The terms and conditions of this experimental assessing method were explained to students one month before the tests. At that moment, they also received the list of laboratory and imaging tests (Appendix A), which allowed them to familiarize themselves with the list in order not to waste time discovering all 107 proposals. Three minutes were allowed for each question. Two sessions of fifteen minutes each (five cMCQ/CBM) were submitted to students, right before two standard faculty assessments.

### 2.3. Scoring Scale

The scoring scale of cMCQ depended on the number of correct answers and was empirically chosen. In case of a single correct answer, if this answer was chosen, 1 point was awarded. Knowing that the score could not be negative, 0.2 points were removed for any incorrect answer ticked. In case of multiple correct answers, 1 point divided by the number of expected answers was awarded for each good answer (for example, when there were 4 correct answers, each one valued 0.25 points). Knowing that the score could not be negative, 0.05 points were removed for any incorrect answer ticked. The score for the total of 10 cMCQ was converted out of 100. In addition to this main mark, the CBM could lead to a maximum of 2 “bonus” points. The average of all the certainty levels for *correct* answers was calculated and corresponded to a maximum of 1 bonus point. For example, if the average of all the certainty levels for correct answers was 90%, the bonus was 0.9 points ( $90\% = 0.9$ ). Similarly, the average of all the certainty levels for *incorrect* answers was calculated and subtracted from 1, which corresponded to a maximum of 1 bonus point. For example, if the average of all the certainty levels for incorrect answers was 30%, the bonus was 0.7 points ( $30\% = 0.3; 1 - 0.3 = 0.7$ ). Of course, this scoring scale values the right answers with high degrees of certainty (“mastered answer”). Nonetheless, and this is more unusual and, therefore, one of the originalities of this work, the wrong answers given with low degrees of certainty (“ignorance”) were also valued. It is assumed that in a real professional situation, the student would have tried to gather additional information or asked for a peer’s opinion. This scale was designed to minimize the influence of the assessment of CBM on the student’s final score. It mainly aimed to enlarge score distribution and to separate students with equal or very close scores.

### 2.4. Faculty Tests and Context

Standard exams consisted of three-hour sessions with 120 MCQs in the format imposed for the ECN (*Épreuves Classantes Nationales*), the national medical examination which takes place at the end of the sixth year of medical school in France. Each MCQ included a description of a short clinical scenario followed by five sentences. One to four of these sentences were correct and one to four were incorrect. Two examples of these MCQs, which will be called “standard MCQ”, are presented in Appendix C. The scoring scale was as follows: no mistake: 1 point; 1 mistake: 0.5 points; 2 mistakes: 0.2 points; 3 mistakes or more: 0 points. Both tests, experimental and standard, were conducted on a tablet computer (iPad®, Apple Inc., Victor Valley, CA, USA). According to the rules of the ECN, the evaluation by MCQ on tablet computer is mandatory in France. This allowed us to use those tablet computers for the cMCQ and CBM. The themes of cMCQ were identical to those of the faculty tests (rheumatology and infectious diseases).

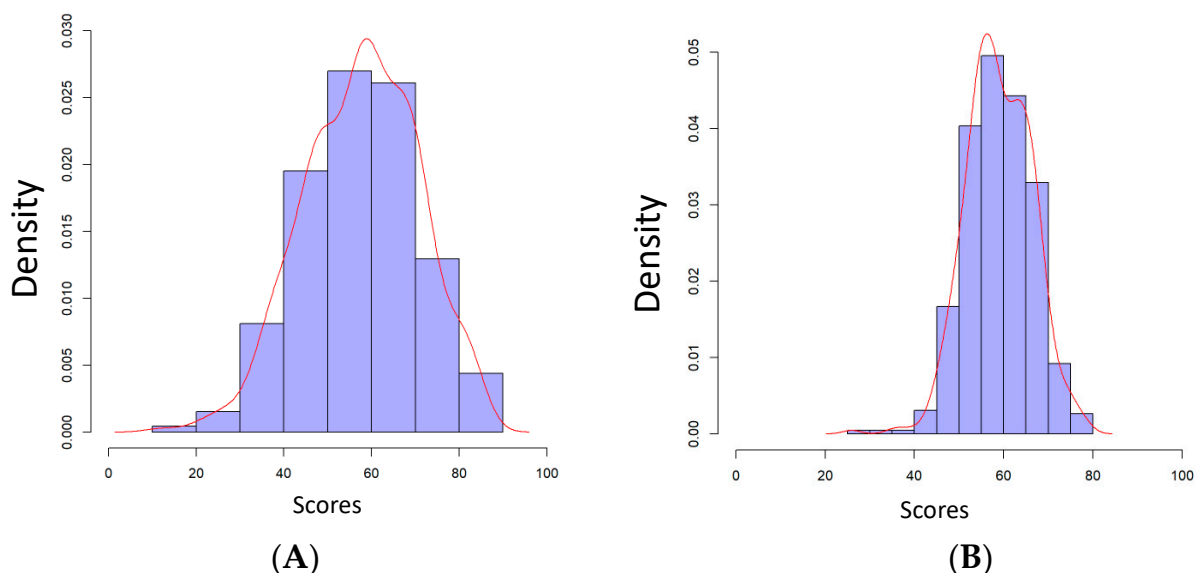
### 2.5. Statistical Analysis

We first performed a descriptive analysis on variables of interest. Qualitative variables were quoted as the frequency (percentage). Quantitative variables were quoted as mean  $\pm$  standard deviation (SD) when normally distributed or as the median (interquartile range (IQR)) otherwise. The 95% confidence interval (CI) was calculated using a normal

distribution. Independence between categorical variables was tested using a chi-squared test or Fisher's exact test. Independence between categorical and quantitative variables was tested using Student's *t*-test or analysis of variance. Independence between quantitative variables was tested using the Pearson correlation coefficient nullity test. Their relationship was modeled using linear regression. There were no missing data. All tests were two-sided and the threshold for statistical significance was set to  $p < 0.05$ . Analyses were performed using R software (version 3.3.1, R Foundation for Statistical Computing, Vienna, Austria).

### 3. Results

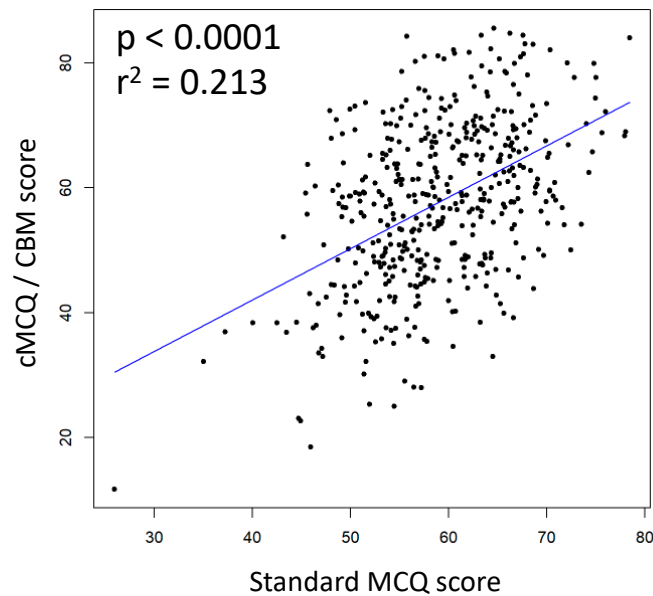
The mean age of the 456 students was 24.5 ( $\pm 2.0$ ) years. During the first session of 15 min, the average time spent on the five experimental questions was 9 min 26 s ( $\pm 2$  min 30 s), with a minimum of 4 min 6 s and a maximum of 15 min. Through the second session, the average time was 8 min 27 s ( $\pm 2$  min 23 s), with a minimum of 3 min 2 s and a maximum of 15 min. The mean score of the two cMCQ/CBM sessions was 57.7 points ( $\pm 13.1$ ); the distribution of the scores is presented in Figure 1A. In comparison, mean score of the two standard MCQ sessions was 58.9 points ( $\pm 7.3$ ); the distribution of the scores is presented in Figure 1B.



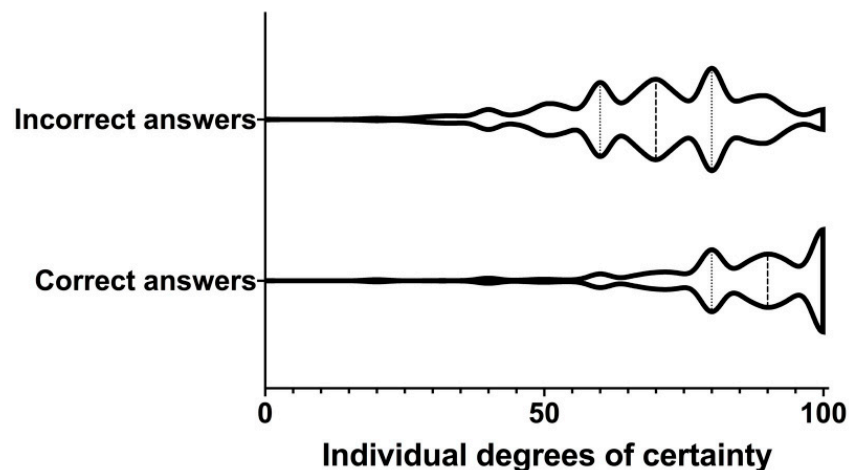
**Figure 1.** Distribution of the scores obtained with cMCQ/CBM (A) and with standard MCQ (B). Scores are more spread out with cMCQ/CBM, with a standard deviation of 13.1 versus 7.3 for standard MCQ.

Figure 2 shows a scatter plot and regression line of experimental and standard scores. We found a significant correlation between these two parameters ( $p < 0.0001$ ) and the proportion of “cMCQ/CBM score” variance explained by “standard MCQ score” was 21.3% ( $r^2 = 0.213$ ).

Regarding the CBM, the degrees of certainty were significantly higher for correct answers (median and quartiles: 90% (80%; 100%)) vs. incorrect answers (70% (60%; 80%)) ( $p < 0.0001$ ; Figure 3). “Serious mistakes” were easily identified, corresponding to false answers with high level of confidence. For example, 5.2% of the whole cohort answered the first question incorrectly with a certainty level of 100%.



**Figure 2.** Scatter plot with regression line of individual scores obtained with cMCQ/CBM vs. standard MCQ. R-squared value is denoted by  $r^2$ . The two scores are significantly correlated.



**Figure 3.** Violin plots of individual levels of certainty for correct and incorrect answers. The dashed lines show the medians and the dotted lines the lower and upper quartiles. Although the levels of confidence appear high for incorrect answers, they are significantly higher for correct answers ( $p < 0.0001$ ).

#### 4. Discussion

We described the usage of a specific type of MCQ with a high number of options matching a comprehensive list of diagnostic tests, followed by a certainty-based marking. This assessment tool was feasible in sixth-year medical students to assess the choice of laboratory and imaging studies in front of a clinical case of general practice. Our prospective study was performed on a large number of students, which allowed us to show that cMCQ and CBM can be easily implemented in a medical school using MCQs on a regular basis. As the first part of the cMCQ (the short clinical scenario) is the only one to change from question to question, the framing of a well-structured cMCQ is easy to learn and to perform for teachers. Given the large number of proposed responses, sight recognition or random guessing were minimized [23]. Finally, thanks to the CBM, “serious mistake”, “mastered answer”, “ignorance” and “weak knowledge” could be identified among the students.

Score distribution was larger with experimental evaluation than with standard faculty exam. This would facilitate student ranking by reducing the rate of close or identical scores. Furthermore, cMCQs are very similar to diagnostic test choices in real medical practice, since most of test prescription is now computer based, offering similar lists of laboratory and imaging studies. We observed a significant correlation between the results of standard and experimental methods, presumably because of the impact of the skills and knowledge of students to correctly answer both MCQ types. Themes in both assessment methods were matched to minimize the variability that could have been brought by unequal theme knowledge. The proportion of “cMCQ/CBM score” variance explained by “standard MCQ score” was 21.3%. The proportion of unexplained variance stayed, therefore, relatively high, showing that these two methods are different and do not explore the same knowledge. This can also be explained by differences in skills that are assessed by the two methods. In addition to its usefulness as an examination, the results of cMCQ/CBM can be used for other teaching purposes. Mainly, “serious mistakes” feedback could be recognized individually, in order to deconstruct false knowledge. Of course, a high frequency of 100% certainty level for an incorrect answer may only indicate a “serious mistake” once a poorly asked question or erroneous correction was discarded. More broadly, the CBM can also help in developing a reflexive attitude, taking doubt into account, which is essential in medical practice [24]. In that regard, the CBM exercise—by valuing “ignorance” in the scoring scale—induces another relationship to error and can, therefore, facilitate the right attitude to adopt in real practice: gathering literature or seeking advice from a peer. On the other hand, being sure of a correct answer indicates potentially lasting knowledge, which is also valued in our scoring scale. We were surprised by the high overall levels of certainty. This could be explained by insufficient explanations about CBM from our part to the students. Furthermore, this was the first time students were confronted with this CBM exercise. Students could have been overconfident, which is not uncommon among young doctors. Finally, the thorough preparation of medical students for this major testing can also be an explanation, as this sixth-year exam is the most important in France, basically orienting the rest of the students’ professional lives by the ECN final national ranking [25].

Our study has some limitations. Firstly, there is not a specific tool that is fundamentally objective to assess whether one evaluation method is better than another. Some studies claim that the optimum number of options for an MCQ is three [26,27], while our study used more. However, these conclusions should not be of concern, because those studies assessed the single best response in MCQ, with two distractors and one correct answer. It seems reductive to limit the problematic nature of the choice of paraclinical exams to one correct answer to pick up in three propositions. Secondly, as mentioned above, our scoring system was empirically chosen. The number of points withdrawn per false proposal selected could have been different, as well as the number of bonus points for the CBM or the level of certainty to identify serious mistakes. Other scales could have been chosen to give greater weight to CBM [28]. However, it can be easily adapted for faculty assessment goals. Then, the optional nature of this experimental evaluation may have influenced the results but including an experimental evaluation in the official faculty evaluation did not seem appropriate to us. Finally, the results of the CBM would certainly have been different if the students had been able to practice this new exercise, which is difficult to apprehend given the usual university context, where the goal is always to “find the right answer”. From a purely teaching point of view, it would have also been interesting to ask the students about their experience during the exam. Future studies will include such an assessment, such as: “Were you unsettled by the large number of proposals?”; “Were you unsettled by the CBM?”; and “Did you find this mode of evaluation relevant?”.

To better understand the high overall levels of certainty, students from other fields should also be evaluated with similar methods. Thus, as future scientists or biology teachers, the levels of confidence should vary and perhaps subgroups would be identified, based on the type of careers aimed at by students. Indeed, we expect to uncover significant discrepancies and probably even hallmarks of the different university subjects when testing

other students of other series or even medical students undertaking other examinations. The results we obtained could be, in part, explained by self-confidence. Hence, marked differences, indeed, appeared in regards to self-esteem between students of health sciences in another country [29]. Due to the impact of the ranking in “ECN” tests on French medical students’ future lives and careers, other students could also be less prepared when taking their own examinations. Moreover, medical doctors are worldwide supposed to be self-assured, so as to inspire trust in their diagnoses and prescribed treatments [30]. Even if self-esteem seems to be impacted by gender, being significantly lower in female students [31], the ability of the future medical doctors to convince their future patients remains essential. However, as clinical decisions require insight and foresight, when both are lacking, overconfidence and error might appear [19]. When self-assurance can be beneficial, overconfidence might also be problematic.

Overall, we believe that this work is of interest in three ways. First, the “comprehensive” MCQ placed students in authentic situations, close to professional life. Then, CBM brought an important dimension to the test and can also be a learning tool. Indeed, doubt management is a skill that is particularly necessary in the field of medicine, where students (and doctors) must constantly update their knowledge. Finally, this type of examination test is easy to set up and retains the advantages of the MCQ format. The use of this relatively innovative method was found to be a feasible and reliable tool to assess the choice of laboratory and imaging studies in family medicine. Nevertheless, this method of evaluation, using a comprehensive list of choices followed by a CBM, can be easily applied to other fields of medicine. Here, we chose 107 general medical tests, tailored to the practice of general practice, which is thought to be appropriate to sixth-year medical students in France. Hence, we can definitely create a list of tests adapted to each medical specialty (hematology, neurology, etc.). Furthermore, such a methodology could also be applied to the choice of clinical examinations to perform; diagnosis to state; or therapies to propose.

**Author Contributions:** Conceptualization, C.H., A.D. and P.T.; methodology, C.H. and P.T.; validation, L.G., E.C., P.L., V.S., D.G. and D.L.; writing—original draft preparation, C.H., A.D. and P.T.; writing—review and editing, W.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board *CONSEIL DE FACULTÉ* Lille University School of Medicine (protocol code 201712-20-003 approved on 04 December 2017).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. No individual can be identified in this publication.

**Data Availability Statement:** Raw data supporting reported results can be requested from the corresponding author by email at [c-herbaux@chu-montpellier.fr](mailto:c-herbaux@chu-montpellier.fr).

**Acknowledgments:** The authors would like to thank Frédéric Garcia (Psychology and Anthropology Degree), Céline Maurice and Magali Lance for their support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. List of Laboratory and Imaging Studies

“No further test is necessary”

Standard biology (26)

- Complete blood count
- Reticulocytes
- Blood sugar level
- Urea
- Serum creatinine
- Basic serum electrolytes (Na, K, Cl)



- Calcium
- Phosphorus
- Albumin
- Serum protein electrophoresis
- NT-proBNP
- CPK
- Troponin
- PT, PTT
- INR
- D-dimer
- Arterial blood gas test total CO<sub>2</sub>
- Lactate concentration
- CRP
- Erythrocyte sedimentation rate
- Liver function tests (ALT, AST, GGT, ALP)
- Total bilirubin, unconjugated bilirubin
- Lipid profile (cholesterol, triglycerides)
- Lipase
- LDH
- Haptoglobin

Biology related to infectiology (20)

- PCT
- Urine culture
- Sputum culture
- Blood culture
- Sputum culture for Mycobacterium tuberculosis
- Stool culture
- Lyme serology
- HIV serology
- TPHA-VDRL
- Anti-HBc Ab
- Anti-HBs Ab
- Antigen HBs
- HAV serology
- HCV serology
- HDV serology
- Chlamydia trachomatis PCR
- Mantoux test
- Microscopic examination of blood films for malaria
- Rapid strep test
- Neisseria gonorrhoeae test (swab)

Biology of other specialties (24)

- Lumbar puncture
- TSH
- T<sub>3</sub>, T<sub>4</sub>
- Prolactin
- Glycated hemoglobin test
- Bone marrow examination
- Ferritin
- Serum iron, total iron-binding capacity, transferrin saturation
- ABO and Rh (D) typing
- Irregular antibodies
- Vitamins B<sub>9</sub> and B<sub>12</sub>

- Immunoglobulin quantitation
- Blood immunofixation
- Fibrinogen
- Factor II and Factor V quantification
- Anti-nuclear antibodies
- Rheumatoid factor
- Cryoglobulin
- Complement (CH50, C3, C4)
- ANCA
- Urine electrolytes
- Total quantity of protein in a 24 h urine collection test
- Proteinuria/creatininuria index
- Anti-transglutaminase IgA

#### Medical imaging (25)

- Chest X-ray
- Abdominal X-ray
- Bone X-ray of a limb
- Cervical spine X-ray
- X-rays of the thoracic spine
- X-rays of the lumbosacral spine
- Radiography of the pelvis
- CT scan of the chest without contrast
- CT angiography of the chest
- CT of abdomen and pelvis with contrast
- CT of abdomen and pelvis without contrast
- CT of the neck without contrast
- Abdominal ultrasonography
- Abdominal MRI
- Arterial Doppler of lower extremities
- Venous Doppler of lower extremities
- Supra-aortic trunk Echo-Doppler
- Whole body FDG—PET /CT
- Cranial CT scan
- Head MRI with Gadolinium contrast
- Ventilation/perfusion lung scan
- Bone densitometry
- Bone scintigraphy
- CT scan of facial bones
- Transfontanelle ultrasonography

#### Specialized imaging (9)

- Dilated fundus examination
- Slit lamp examination
- Cardiac ultrasound
- Bronchial fibroscopy
- Colonoscopy
- Oesophagogastroduodenoscopy
- Capsule endoscopy
- Panendoscopy (of upper aerodigestive tract)
- Cystoscopy

#### Others (3)

- Urine test strip
- 12-lead ECG

- Pulmonary function testing

## Appendix B

### Appendix B.1. cMCQ #1

During a general practitioner consultation, you see a 58-year-old patient with lower-back pain. The pain occurred abruptly during an effort 3 days ago. It is accompanied by major functional impotence. Coughing and defecation efforts are at the origin of pain paroxysms. The rest of the clinical examination is normal.

Which diagnostic test(s) is (are) indicated?

Answer: No further examination is necessary.

Comment: Typical uncomplicated lumbago, without any of the so-called “red flags” signs, no exam indicated.

### Appendix B.2. cMCQ #2

You are taking care of a 31-year-old female patient in the emergency room complaining of pain on passing urine during the past five days with episodes of shivering. At the clinical examination you find an abdominal pain that radiates along the flank towards the back. The body temperature is 39.3 °C and the blood pressure is 92/43 mmHg.

Which diagnostic test(s) is(are) indicated?

Answer: Urine test strip, urine culture, blood culture, complete blood count, CRP, urea, creatinine, abdominal ultrasonography.

Comment: Complementary exams indicated in accordance with the infectious diseases reference.

## Appendix C

### Appendix C.1. Standard MCQ #1

You suspect that a 28-year-old man has inflammatory rachialgia, potentially associated with axial spondyloarthritis.

Which of the following supports the diagnosis of spondyloarthritis?

- Presence of the HLAB 27 antigen
- Biological inflammatory syndrome with high CRP
- Family or personal history of psoriasis
- Presence of marginal erosions on the heads of the 5th metatarsals (radiographs)
- Inefficiency of nonsteroidal anti-inflammatory drugs on pain

Answer: A, B and C.

### Appendix C.2. Standard MCQ #2

Mr. D, 92 years old, is referred to the emergency room for behavioral disorders associated with hyperthermia. He fell down in undetermined circumstances this morning. He is on warfarin and bisoprolol. On admission: BP 158/72 mmHg, CF = 80/mn, SaO<sub>2</sub> = 94%, temperature 39.2 °C. Medical interview was noncontributory. Clinical examination was poor except for marked temporospatial disorientation and fluctuating agitation. There's no stiff neck, no vomiting. No specific neurological sign. Glasgow score of 13. The abdomen is tender and generally defenseless with perceived hydro-aerial sounds. Bladder scan: 680 mL, urine test strip: leukocytes 2+.

Which of the following proposals for additional tests you request in an emergency are (is) correct?

- In the absence of nitrite in the urine test strip, a urine culture is not necessary
- A lumbar puncture must be performed after brain imaging
- A lumbar puncture must be performed after a coagulation test
- Fall on Vitamin K antagonist in a patient on delirium requires brain imaging
- Performing blood cultures is not necessary

Answer: C and D.

## References

- Lehnert, B.E.; Bree, R.L. Analysis of appropriateness of outpatient CT and MRI referred from primary care clinics at an academic medical center: How critical is the need for improved decision support? *J. Am. Coll. Radiol.* **2010**, *7*, 192–197. [[CrossRef](#)] [[PubMed](#)]
- Chan, P.S.; Patel, M.R.; Klein, L.W.; Krone, R.J.; Dehmer, G.J.; Kennedy, K.; Nallamotheu, B.K.; Weaver, W.D.; Masoudi, F.A.; Rumsfeld, J.S.; et al. Appropriateness of percutaneous coronary intervention. *JAMA* **2011**, *306*, 53–61. [[CrossRef](#)] [[PubMed](#)]
- Rabinowitz, H.K. The modified essay question: An evaluation of its use in a family medicine clerkship. *Med. Educ.* **1987**, *21*, 114–118. [[CrossRef](#)]
- Wood, E.J. What are Extended Matching Sets Questions? *Biosci. Educ.* **2003**, *1*, 1–8. [[CrossRef](#)]
- Pottier, P.; Hardouin, J.-B.; Hodges, B.D.; Pistorius, M.-A.; Connault, J.; Durant, C.; Clairand, R.; Sébille, V.; Barrier, J.-H.; Planchon, B. Exploring how students think: A new method combining think-aloud and concept mapping protocols. *Med. Educ.* **2010**, *44*, 926–935. [[CrossRef](#)]
- Schuwirth, L.W.T.; van der Vleuten, C.P.M. Different written assessment methods: What can be said about their strengths and weaknesses? *Med. Educ.* **2004**, *38*, 974–979. [[CrossRef](#)]
- Veloski, J.J.; Rabinowitz, H.K.; Robeson, M.R.; Young, P.R. Patients don't present with five choices: An alternative to multiple-choice tests in assessing physicians' competence. *Acad. Med.* **1999**, *74*, 539–546. [[CrossRef](#)]
- Palmer, E.J.; Devitt, P.G. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. *BMC Med. Educ.* **2007**, *7*, 49. [[CrossRef](#)]
- Javaeed, A. Assessment of Higher Ordered Thinking in Medical Education: Multiple Choice Questions and Modified Essay Questions. *MedEdPublish* **2018**, *7*, 128. [[CrossRef](#)]
- Bloom, B.S. Taxonomy of educational objectives. In *Cognitive Domain*; McKay: New York, NY, USA, 1956; Volume 1, pp. 20–24.
- Adams, N.E. Bloom's taxonomy of cognitive learning objectives. *J. Med. Libr. Assoc.* **2015**, *103*, 152–153. [[CrossRef](#)]
- Chaudhary, N.; Bhatia, B.D.; Mahato, S.K. Framing a Well-Structured Single Best Response Multiple Choice Questions (MCQs)-An Art to be Learned by a Teacher. *J. Univers. Coll. Med. Sci.* **2014**, *2*, 60–64. [[CrossRef](#)]
- Mehta, G.; Mokhasi, V. Item Analysis of Multiple Choice Questions—An Assessment of the Assessment Tool. *Int. J. Health Sci. Res. (IJHSR)* **2014**, *4*, 197–202.
- Mehta, M.; Banode, S.; Adwal, S. Analysis of multiple choice questions (MCQ): Important part of assessment of medical students. *Int. J. Med. Res. Rev.* **2016**, *4*, 199–204. [[CrossRef](#)]
- Leclercq, D. Validity, Reliability, and Acuity of Self-Assessment in Educational Testing. In *Item Banking: Interactive Testing and Self-Assessment*; NATO ASI Series; Leclercq, D.A., Bruno, J.E., Eds.; Springer: Berlin/Heidelberg, Germany, 1993; pp. 114–131.
- Gardner-Medwin, A.R. *Confidence-Based Marking-towards Deeper Learning and Better Exams*, 1st ed.; Routledge: London, UK, 2006.
- Schoendorfer, N.; Emmett, D. Use of certainty-based marking in a second-year medical student cohort: A pilot study. *Adv. Med. Educ. Pract.* **2012**, *3*, 139–143. [[CrossRef](#)] [[PubMed](#)]
- Barr, D.A.; Burke, J.R. Using confidence-based marking in a laboratory setting: A tool for student self-assessment and learning. *J. Chiropr. Educ.* **2013**, *27*, 21–26. [[CrossRef](#)]
- Tweed, M.; Thompson-Fawcett, M.; Schwartz, P.; Wilkinson, T.J. Determining measures of insight and foresight from responses to multiple choice questions. *Med. Technol.* **2013**, *35*, 127–133. [[CrossRef](#)]
- Brown, T.A.; Shuford, E.H. *Quantifying Uncertainty into Numerical Probabilities for the Reporting of Intelligence*; RAND Corporation: Santa Monica, CA, USA, 1973.
- Shuford, E.H.; Albert, A.; Edward Massengill, H. Admissible probability measurement procedures. *Psychometrika* **1966**, *31*, 125–145. [[CrossRef](#)]
- Castaigne, J.L.; Abry, S.; Bailly, B.; Sylvestre, E. Aller Plus Loin Avec les QCM Grâce aux Degrés de Certitude. PENSERA Network, Grenoble, France. Available online: [https://www.slideshare.net/JeanLoup\\_Castaigne/aller-plus-loin-avec-les-qcm-grce-aux-degrs-de-certitude](https://www.slideshare.net/JeanLoup_Castaigne/aller-plus-loin-avec-les-qcm-grce-aux-degrs-de-certitude) (accessed on 13 December 2015).
- Draaijer, S.; Jordan, S.; Ogden, H. Calculating the Random Guess Score of Multiple-Response and Matching Test Items. In *Technology Enhanced Assessment*; Communications in Computer and Information Science; Ras, E., Guerrero Roldán, A.E., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 210–222.
- Charlin, B.; Lubarsky, S.; Millette, B.; Crevier, F.; Audétat, M.-C.; Charbonneau, A.; Fon, N.C.; Hoff, L.; Bourdy, C. Clinical reasoning processes: Unravelling complexity through graphical representation. *Med. Educ.* **2012**, *46*, 454–463. [[CrossRef](#)]
- CNG, Établissement Public Administratif. Épreuves Classantes Nationales (ECN). Paris, France. Available online: <https://www.cng.sante.fr/concours-examens/epreuves-classantes-nationales-ecn> (accessed on 25 March 2022).
- Nwadinigwe, P.I.; Naibi, L. The Number of Options in a Multiple-Choice Test Item and the Psychometric Characteristics. *J. Educ. Pract.* **2013**, *4*, 189–196.
- Vyas, R.; Supe, A.N. Multiple choice questions: A literature review on the optimal number of options. *Natl. Med. J. India* **2008**, *21*, 130–133.
- Leclercq, D.; Poumay, M. Trois nouveaux indices de réalisme dans l'auto-évaluation des performances. *Cah. Du Serv. De Pédagogie Expérimentale* **2003**, *15*, 189–196.

29. Abdulghani, A.H.; Almelhem, M.; Basmaih, G.; Alhumud, A.; Alotaibi, R.; Wali, A.; Abdulghani, H.M. Does Self-Esteem leads to high achievement of the science college's students? A study from the six health science colleges. *Saudi J. Biol. Sci.* **2019**, *27*, 636–642. [[CrossRef](#)]
30. Huang, L.; Thai, J.; Zhong, Y.; Peng, H.; Koran, J.; Zhao, X.-D. The Positive Association Between Empathy and Self-Esteem in Chinese Medical Students: A Multi-Institutional Study. *Front. Psychol.* **2019**, *10*, 1921. [[CrossRef](#)] [[PubMed](#)]
31. Spoto-Cannons, A.C.; Isom, D.M.; Feldman, M.; Zwygart, K.K.; Mhaskar, R.; Greenberg, M.R. Differences in medical student self-evaluations of clinical and professional skills. *AMEP* **2019**, *10*, 835–840. [[CrossRef](#)] [[PubMed](#)]