

ORIGINAL ARTICLE

Open Access



Inclusion of MCQs written by radiology residents in their annual evaluation: innovative method to enhance resident's empowerment?

Nadia Amini^{1*}, Nicolas Michoux¹, Leticia Warnier², Emilie Malcourant², Emmanuel Coche¹ and Bruno Vande Berg¹

Abstract

Aim: We hypothesized that multiple-choice questions written by radiology residents (MCQ^{resident}) for their weekly case presentations during radiology staff meetings could be used along with multiple-choice questions written by radiology teachers (MCQ^{teacher}) for their annual evaluation. The current prospective study aimed at determining the educational characteristics of MCQ^{resident} and at comparing them with those of MCQ^{teacher}.

Methods: Fifty-one radiology residents of the first to the fifth year of training took the 2017 exam that contained 58 MCQ^{resident} and 63 of MCQ^{teacher}. The difficulty index, the discrimination power, and the distractor's quality were calculated in the two series of MCQs and were compared by using Student *t* test. Two radiologists classified each MCQ according to Bloom's taxonomy and frequencies of required skills of both MCQ series were compared.

Results: The mean \pm SD difficulty index of MCQ^{resident} was statistically significantly higher than that of MCQ^{teacher} (0.81 ± 0.1 vs 0.64 ± 0.2 ; $p < 0.0001$). The mean \pm SD discrimination index of MCQ^{resident} was statistically significantly higher than that of MCQ^{teacher} (0.34 ± 0.2 vs 0.23 ± 0.2 ; $p = 0.0007$). The mean number of non-functional distractors per MCQ^{resident} was statistically significantly higher than that per MCQ^{teacher} (1.36 ± 0.9 vs 0.86 ± 0.9 ; $p = 0.0031$). MCQ^{resident} required recalling skills more frequently than MCQ^{teacher} which required more advanced skills to obtain a correct answer.

Conclusions: Educational characteristics of MCQ^{resident} differ from those of MCQ^{teacher}. This study highlights the characteristics to optimize the writing of MCQs by radiology residents.

Keywords: Radiology training, Surveys and questionnaires, Quality, Learning, Internship and residency

Key points

- Scores obtained by PGY1-5 at their annual evaluation increase with education year whoever wrote the MCQs (radiology residents or teachers).
- MCQs written by radiology residents are easier and contain more nonfunctional distractors than MCQs written by radiology teachers, but their discriminant power is higher.

- Memory skills play a more important role in answering MCQs written by residents than by teachers.

Introduction

Training of radiology residents is generally based on supervised daily clinical practice during scheduled core rotations, autonomous activity during night calls, attendance to multi-disciplinary consultations and dedicated radiology-centered lectures, and presentations of clinical cases. Evaluation of their knowledge, skills, and attitudes is an important part of their training and depends on opinions of faculty staff and written or oral exams. In our institution, opinions of faculty staff and MCQ exams are used. Each year, all PG1-5

* Correspondence: amini.nadia@gmail.com

¹Department of Radiology, IREC, Cliniques Universitaires Saint-Luc UCLouvain, Avenue Hippocrate 10/2942, 1200 Brussels, Belgium
Full list of author information is available at the end of the article

residents perform a written examination containing 125 single best-option multiple-choice questions addressing knowledge in all fields of medical imaging as outlined in the ESR training curriculum [1].

Multiple-choice question (MCQ) is a validated educational tool for both formative and summative assessment because of its objective output, simplicity of utilization, and informative feedback to exam-takers and teachers [2–6]. For formative MCQ, immediate feedback to exam-takers promotes reflexion and further learning (catalytic effect). For summative assessment, the MCQ test provides an overall judgment about competence of the exam-taker. In addition, MCQ-based assessment could be more reproducible and objective than exams based on essays or open-ended type questions. There is no indication in the literature that open-questions are more reliable than closed-questions (selected response format) [7, 8].

Writing of MCQs is a difficult task, and teaching staff rarely have the time or incentive to develop high-quality formative questions, focusing instead on material for high-stakes assessments [9, 10]. In the current era of student empowerment, several educational teams proposed to engage students in their education by asking them to submit, review, and discuss MCQs items [3, 9, 11–14]. Collaborative or web-based question-writing is an interesting tool in learning enhancement [9–11, 15–17] because it may stimulate a deeper understanding of the taught subjects and self-monitoring. Although this approach is widely seen among medical students [9, 10, 15–19], there is no evidence in the literature that the same educational approach could be applied to medical residents in training. Therefore, we undertook this prospective pilot study to determine and compare the educational characteristics of MCQs written by radiology residents (MCQ^{resident}) and those written by teachers (MCQ^{teacher}), combined in a single computer-supported test.

Material and methods

Since October 2015, radiology residents from the third to the fifth year of residency (PGY3-5) rotating in our academic hospital are asked to integrate two single best-option multiple-choice questions (MCQs) into each of their presentations of clinical cases. MCQ templates including a stem and four-answer options consisting of one correct answer and three distractors were proposed [20–22]. Each MCQ^{resident} was prospectively classified by its author according to eight organs or anatomic systems, three estimated degrees of frequency (frequent, uncommon, rare), and of clinical relevance (important, moderate, less important). The residents were aware that some of their MCQs would be used for their annual evaluation (PGY1-5). In November 2016, 221 resident-authored MCQ^{resident} were available.

In January 2017, a first-year EDIR-certified fellow in radiology with interest in education and the staff member radiologist responsible for the resident's training (member of the French-speaking National Certification Board for Radiology) have in consensus selected 65 MCQ^{resident}, addressing items from all organ systems, with a frequent or an occasional occurrence and with a high or an intermediate degree of clinical relevance. Sixty-five MCQ^{teacher} written by nine radiologists after their lectures were also included in the annual evaluation. The writing of all MCQs was checked for spelling and format (four-option format with only one correct answer) [23]. In February 2017, 51 PGY1-5 residents in radiology took the examination on their PC. Anonymity of the residents was ensured by using pseudos enabling identification of the PGY. All the images contained in the MCQ fulfilled the usual criteria of confidentiality and anonymity for the patients. The participating residents received their personal score immediately at the end of the examination. Correct answers to all MCQs were presented to the residents by faculty staff during two meetings a few weeks after the examination.

Quantitative analysis

The internal consistency reliability was verified by the Kuder-Richardson formula 20 (KR-20) [24, 25]. Three quantitative parameters including the Difficulty Index, the Discrimination Index, and the number of nonfunctional distractors were calculated for each MCQ [25–29]. The Difficulty Index of an MCQ represents the ratio between the number of students who correctly answered the item and the total number of answering students [26, 27]. A high Difficulty Index (approaching 1.0) indicates an easy question. MCQs were classified as easy (Difficulty Index > 0.70), intermediate (Difficulty Index between 0.70 and 0.30), or difficult (Difficulty Index < 0.30). The ideal value for the degree of difficulty ranges between 0.50 and 0.70 [24]. The Discrimination Index of an MCQ assesses the relationship between how well students did on the item and their total exam score. It is most commonly referred to as the Pearson Point-Biserial correlation (r_{pbis}) [25, 26]. A high discriminant index indicates that the students who had high exam scores got the item correct, whereas students who had low exam scores got the item incorrect. An ideal range for the discrimination index is above 0.20 [23]. Usual working frame values are as follows: < 0.10 considered of a very poor discrimination power, 0.10–0.20 of a little discrimination power, and > 0.20 of a good discrimination power [26]. Finally, the degree of functionality of the distractors was calculated. A distractor is classified as non-functional distractor (NFD) if less than 5% of students have chosen it [20, 26]. Ideally, there must be no NFD at all, implying the educational power of a question.

In our test, there could be 0, 1, 2, or 3 NFD per item. The mean number of NFD was counted per series of MCQs.

Qualitative analysis

The two radiologists in charge of the evaluation separately classified each MCQ according to Bloom's cognitive taxonomy that indicates the most probable cognition process needed to correctly answer the item [30]. This hierarchical model of cognitive processes in solving problems includes four levels: remember, understand, apply, and analyze [27]. Based on previously published Blooming Anatomy and Histology tools [31, 32], we used Bloom's taxonomy type classification system to differentiate among different cognitive levels of radiology MCQs (Table 1).

Statistical analysis.

The exam scores of residents were expressed in mean value \pm standard deviation (SD) for both questionnaires and were plotted against year of training. Distribution of

scores was found to be normal according to the Kolmogorov-Smirnov test. Therefore, a two-sided *t* test (equivalently a Welch test when the equality of variance was not verified according to the *F* test) for independent samples was performed to assess statistical differences between scores from two groups: the MCQ^{resident} and MCQ^{teacher}. Due to the multiple comparisons that were performed, a Bonferroni correction of type $p < 0.05/n^{\text{comparisons}}$ was applied to the tests cited above, and the significance levels were adjusted accordingly.

The mean values \pm SD for Difficulty Index, Discrimination Index, and Distractor Functionality of MCQ^{resident} and MCQ^{teacher} were determined. Difficulty and Discrimination Index from residents' and teachers' MCQs were compared using the *t* test (after verifying the data distributions normality and the variance equality). The number of non-functional distractors of the two series was compared using the non-parametric Mann-Whitney test (*U* test) and the two-sided Fisher's exact test with the mid-P approach at $p < 0.05$. The *p* value less than

Table 1 Bloom's taxonomy type classification system to differentiate among different cognitive levels of radiology MCQs (Bloom's Taxonomy Radiology Tool)

Cognitive processes outlined in a radiology MCQ assessment			
Levels	General definition	Radiology application	Examples in liver cirrhosis
Remember	remembering facts, terms, concepts without necessarily understanding what they mean	memorizing clinical/physical facts, anatomy and variants, different radiologic patterns/signs recognition on a radiologic image	ultrasound signs of liver cirrhosis: parenchymal heterogeneous echotexture and surface nodularity, segmental hypertrophy/atrophy, signs of portal hypertension
Understand	capacity to organize, compare or translate facts or ideas	lesion characterization, artefact recognition, treatment monitoring, be able to give a relationship between a specific physiopathology and all related radiologic signs	signs to look for in order to conclude the portal hypertension: enlarged portal vein, slow/reversal portal venous flow, changes in sus hepatic spectral Doppler waveform, recanalization of paraumbilical vein, portosystemic collaterals, splenomegaly
Apply	using acquired knowledge to solve problems in new situations	detection some fine radiologic signs to make difference between different types of lesion showing similarities, discover a radiologic sign that is not present	detecting a suspicious lesion of hepatocarcinoma in the cirrhotic parenchyma and making difference with another type of lesion
Analyze	examining and breaking global information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations	make difference between different types of lesion using different imaging technics and taking into account clinical/biological data	answering a question if an HCC could be treated by interventional radiology (percutaneous ablation, chemo/radio-embolisation)

0.05 was considered to indicate statistically significant difference. The frequency of the highest levels of cognitive process involvement reached by MCQ^{resident} was compared with those of MCQ^{teacher} according to Bloom’s taxonomy [30].

The authorization of our ethical committee was not asked because our study did not involve patients. The project had been validated by resident representatives and faculty staff. All residents were aware of the projects, and they had signed a form for the use of their presentations.

Results

Study population

In February 2017, 51 radiology residents (31 men and 20 women, mean age 28 years, range 25–30) from the first to fifth year of training took the exam that initially contained 130 MCQs. Fifty-eight MCQ^{resident} and 63 MCQ^{teacher} were validated and 7 MCQ^{resident} and 2 MCQ^{teacher} were excluded due to technical difficulties during the examination (failure of video on some PCs) (Table 2). Ninety-two out of 121 MCQs included images (57 MCQ^{teacher} and 35 MCQ^{resident}). The mean scores (\pm SD) obtained at the MCQ^{resident} were statistically significantly higher than those at the MCQ^{teacher} for the residents of each year of residency ($p < 0.01$ for each year) (Fig. 1).

Quantitative analysis

The KR-20 value of the test was 0.905. There were less MCQ^{resident} than MCQ^{teacher} with an ideal difficulty index without statistically significant difference ($p = 0.94$), and their mean Difficulty Index (0.81 ± 0.1) was statistically significantly higher than that of MCQ^{teacher} (0.64 ± 0.2) ($p < 0.0001$) (Fig. 2). There were more MCQ^{resident} than MCQ^{teacher} with a good discrimination power ($p = 0.0002$), and the mean Discrimination Index \pm SD of MCQ^{resident} (0.34 ± 0.2) was statistically

significantly higher than that of MCQ^{teacher} (0.23 ± 0.2) ($p = 0.0007$) (Fig. 2). There were more non-functional distractors in MCQ^{resident} than MCQ^{teacher} ($p = 0.0022$), and the mean number of NFD per MCQ^{resident} (1.36 ± 0.9) was statistically significantly higher than that per MCQ^{teacher} (0.86 ± 0.9) ($p = 0.0031$) (Fig. 2). Examples of MCQ^{resident} and of MCQ^{teacher} with different educational characteristics are given in Fig. 3.

Qualitative analysis

The frequency of each cognitive process required for a correct answer is given in Table 3. For MCQ^{resident}, recalling-type cognitive process was more frequently required than for MCQ^{teacher} for both reviewers ($p = 0.004$ and 0.001). Application-type (for reviewer 1) and understanding-type (for reviewer 2) cognitive processes were less frequently needed for MCQ^{resident} than for MCQ^{teacher}.

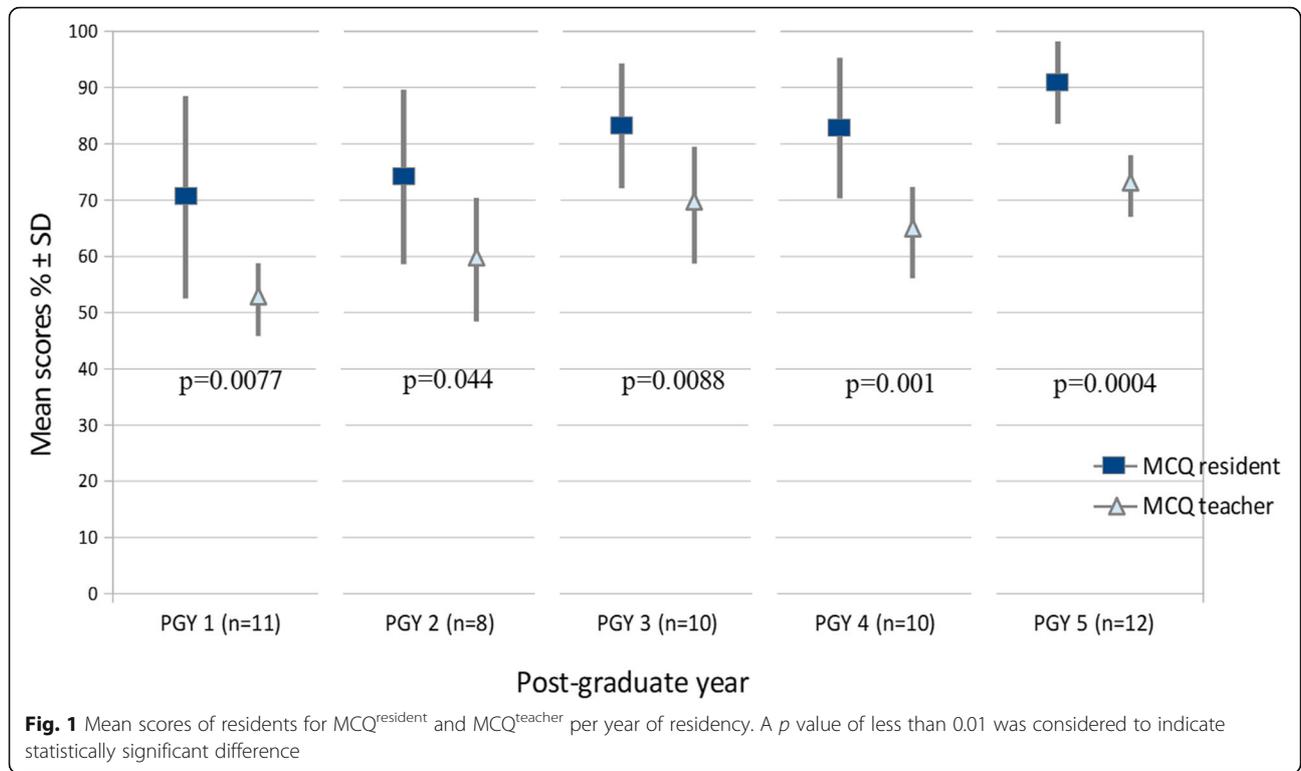
Discussion

In our institution, the annual evaluation task of the residents consists of three parts: a self-fulfilled logbook (clinical and scientific workload, radiology and multidisciplinary meeting attendance); a summary of the evaluation by the supervising faculty staff of their knowledge, skills, and attitudes; and our MCQ test addressing all radiology subspecialties. The MCQ test is performed on a yearly basis to provide the resident an insight on his/her learning curve throughout the 5 years of residency. The results of the PGY4 and PGY5 MCQs are validated by the National Accreditation Board and integrated in the qualification process, in the absence of a national board examination.

The current study demonstrated that the scores obtained by the residents varied according to their level of training in radiology with a non-exponential improvement throughout the 5 years of residency. Our learning gain

Table 2 Number of MCQ^{resident} and of MCQ^{teacher} per topic in the 2017 exam according to the degree of difficulty, frequency, and importance as indicated by their authors. Numbers are numbers of items for MCQ^{resident}/MCQ^{teacher}

Topic	No MCQ ^{resident} / MCQ ^{teacher}	Difficulty		Frequency		Importance	
		Basic	Advanced	Occasional	Frequent	Intermediate	High
Abdomen	12 / 12	6/6	6/7	10/5	2/8	3/3	9/10
Chest	8/10	3/5	5/5	6/3	2/7	4/3	4/7
Interventional	6/0	4/0	2/0	2/0	4/0	0/0	6/0
Musculo-skeletal	12/13	8/4	4/8	6/4	6/8	3/4	9/8
Head and neck	7/8	2/2	5/6	5/5	2/3	3/1	4/7
Pediatric	8/10	5/7	3/3	4/4	4/6	3/0	5/10
Women imaging	3/9	0/5	3/4	2/2	1/7	0/0	3/9
Technique	2/1	2/1	0/0	1/0	1/1	1/0	1/1
Total	58/63	30/30	28/33	36/23	22/40	17/11	41/52

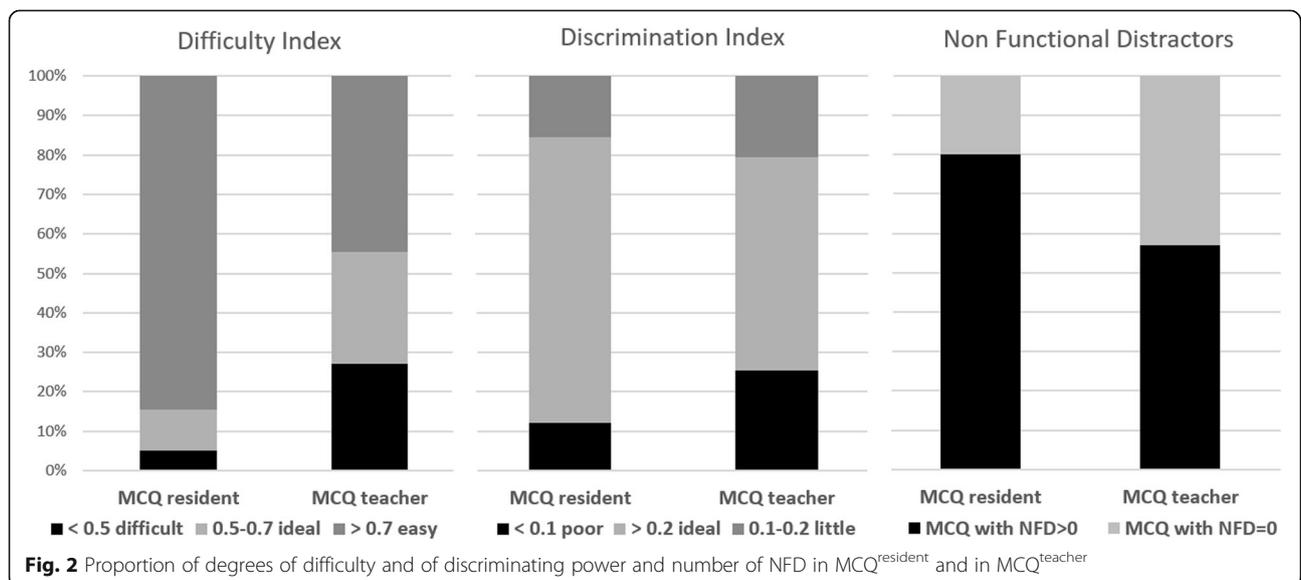


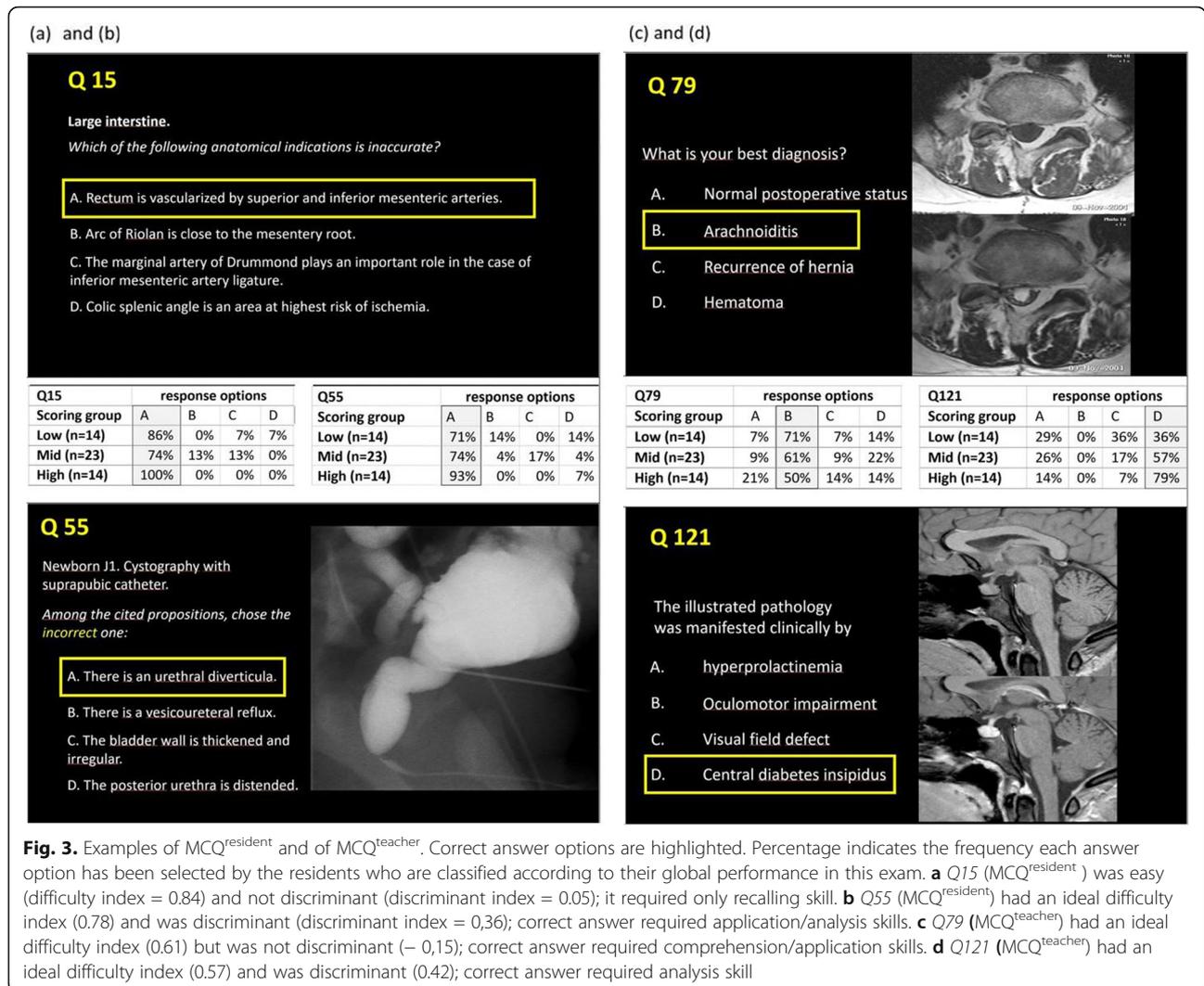
curve of radiology residents that seems to decelerate over time was similar to that observed by Ravesloot [28].

Second, the scores obtained in the MCQ^{resident} were statistically significantly higher than in the MCQ^{teacher} for all residents, independently of their post-graduate year. The most likely explanation was that the degree of difficulty of the MCQ^{resident} was lower than that of the MCQ^{teacher} and that the number of NFD was higher in

MCQ^{resident} than in MCQ^{teacher}. We cannot exclude the hypothesis that the residents deliberately lowered the difficulty level and included NFD because they knew that their MCQs would be used for their annual evaluation. It is most likely that these characteristics are inherent to the degree of qualification of the MCQ writer [33].

Third, the observation that the discrimination index of MCQ^{resident} was higher than that of MCQ^{teacher} warrants





further assessment as this feature is important when composing high-quality MCQs. A likely explanation was that MCQ^{resident} were written by PGY3–5 and not by PGY1–2 residents. Therefore, the PGY3–5 residents who, in overall, should obtain the highest scores obtain much better scores than the PGY1–2 residents in the

MCQs of their peers than in the MCQ^{teacher}, thus artificially increasing the discriminating index of the MCQ^{resident}.

Finally, the analysis of the MCQs according to Bloom’s taxonomy demonstrated that the MCQ^{resident} focused more on recalling skills than the MCQ^{teacher} that

Table 3 Comparative Bloom’s taxonomy analysis by two readers for MCQ^{resident} and MCQ^{teacher}. The *p* value < 0.05 was considered to indicate statistically significant difference

Bloom's taxonomy	reader 1				reader 2			
	MCQ (n=121)	MCQ ^{resident} (n=58)	MCQ ^{teacher} (n=63)	<i>p</i> -value	MCQ (n=121)	MCQ ^{resident} (n=58)	MCQ ^{teacher} (n=63)	<i>p</i> -value
Remember	29(24%)	22(38%)	7(11%)	0.004	51(42%)	33(57%)	18(29%)	0.001
Understand	26(21%)	15(26%)	11(17%)	0.57	32(26%)	10(17%)	22(35%)	0.01
Apply	48(40%)	14(24%)	34(53%)	0.001	21(17%)	9(16%)	12(19%)	0.38
Analyse	18(15%)	7(12%)	11(17%)	0.2	17(14%)	6(10%)	11(17%)	0.29

required the capacity to analyze and apply knowledge. This feature indicates the difficulty in writing high-quality MCQs that require more experience in solving problems [29–31]. However, although Bloom's taxonomy is a hierarchical model, the lowest levels of the hierarchical Bloom's taxonomy should not be disregarded as unimportant or unworthy of teaching [34]. Actually, while lesion detection may be considered as a (low) knowledge level (pattern recognition), there is a general agreement on the fact that most errors are detection errors rather than characterization errors. Furthermore, the distinction between the categories can be seen as artificial since any given cognitive task may entail a number of processes. Any attempt to nicely categorize cognitive processes into clean, cut-and-dried classifications undermines the holistic, highly connective and interrelated nature of cognition, a criticism that is directed at taxonomies of mental processes in general [35].

The effects of this collaborative approach for MCQ writing are controversial although it at least contributes to create questions that can support formal or summative evaluations [36]. Aflalo demonstrated the absence of statistically significant improvement in achievements, when comparing the examination grades before and after question generation in a group of 133 students generating questions [37]. Although students were able to write complex MCQs, they found some aspects of the writing process burdensome and tended not to trust the quality of each other's MCQs [10, 19]. The use of dedicated software like PeerWise which is a freely and globally available online platform allows students to write, share, answer, rate, and discuss peer-written MCQs. Studies demonstrated that PeerWise user students perform significantly better in end-of-course summative assessment than non-user students [16, 17, 38].

The effect of MCQ format on the resident's scores was not assessed as questions with videos were eliminated because of technical problems on certain personal computers. While taking the exam, residents were not able to scroll into images. The Clinically Orientated Reasoning Evaluation (CORE) computer-based format that replaced the oral examination in EDIR using DICOM viewer simulating the daily work of radiologists is most likely a better way to evaluate radiology residents [39].

The current study highlighted differences between MCQ^{resident} and MCQ^{teacher} that will be explained to current and future radiology residents in order to increase the quality of their MCQs. In addition, we plan to share this collaborative approach with other training centers to provide a broader supply of MCQ that would decrease the influence on the exam takers.

Our study had several limitations. First, it was a mono-center study with a limited number of MCQs, from residents and from teachers. Second, both MCQs were

selected by two radiologists to create a series of MCQs that would cover all fields of diagnostic and interventional radiology. To minimize selection bias, items were selected based on their characteristics indicated by the residents and the teachers and not by reading the MCQs. In addition, questions with a high degree of importance and a frequent occurrence in clinical practice have been privileged. Finally, our results were influenced by the facts that PGY3–5 residents composed the MCQs and that PGY1–5 residents took the examination. Residents were also aware of the fact that their MCQs would be used in the annual evaluation.

In conclusion, the current study demonstrated that the educational characteristics of MCQ^{resident} differ from those of the MCQ^{teacher} in many ways. The clear identification of these differences enabled us to indicate points of attention to address in MCQ writing guidance in order to achieve higher quality examinations with the collaboration of the teaching staff.

Abbreviations

CORE: Clinically Orientated Reasoning Evaluation; DICOM: Digital Imaging and Communications in Medicine; EDIR: European Diploma in Radiology; ESR: European Society of Radiology; MCQ: Multiple-choice question; MCQ^{resident}: Multiple-choice question written by residents; MCQ^{teacher}: Multiple choice-question written by teachers; NFD: Non-functional distractor; PGY: Post-graduate year; SD: Standard deviation

Acknowledgements

Special acknowledgment to Pavel Bakhmatov and Christine Algoet for the logistic support and help in data treatment.

Authors' contributions

NA contributed to the design, treatment of the data, analysis, and writing of the manuscript. BVB contributed to the design, analysis, and writing and editing of the manuscript. NM contributed to the data analysis and gave statistical support. LW and EM were professional consultants and specialized in educational topics such as student evaluations and MCQ. EC was the chief of the Radiology Department of Cliniques Universitaires Saint Luc and supported the study. All authors read and approved the final manuscript.

Funding

No funding.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable. The manuscript does not report on or involve the use of any animal or human data or tissue.

Consent for publication

Not applicable. The manuscript does not contain data from any individual person.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Radiology, IREC, Cliniques Universitaires Saint-Luc UCLouvain, Avenue Hippocrate 10/2942, 1200 Brussels, Belgium. ²Louvain Learning Lab, ADEF, Grand Rue 54/L1.06.01, 1348 Louvain-la-Neuve, Belgium.

Received: 15 May 2019 Accepted: 22 October 2019

Published online: 23 January 2020

References

- European Society of Radiology (ESR) (2018) European training curriculum for radiology. Available via <https://www.myesr.org/sites/default/files/ESR%20European%20Training%20Curriculum%20Level%20I%20-%20282018%29.pdf>
- Azer SA (2003) Assessment in a problem-based learning course: twelve tips for constructing multiple choice questions that test students' cognitive skills. *Biochem Mol Biol Educ* 31:428–434
- Draper SW (2009) Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *Bri J Educ Tech* 40:285–293
- Leclercq D, Gilles J-L (2003) Analyses psychométriques des questions des 10 check-up MOHICAN: vue d'ensemble. In: Leclercq D (Ed) *Diagnostic cognitif et métacognitif au seuil de l'université: le projet MOHICAN mené par les 9 universités de la Communauté française Wallonie-Bruxelles*. Presses universitaires de l'Université de Liège, Liège pp. 173–180
- Nicol DJ, Macfarlane-Dick D (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education* 31:199–218
- Kadiyala S, Gavini S, Kumar DS, Kiranmayi V, Rao PNS (2017) Applying blooms taxonomy in framing MCQs: an innovative method for formative assessment in medical students. *J Dr NTR University of Health Sciences* 6:86
- Hift RJ (2014) Should essays and other "open-ended"-type questions retain a place in written summative assessment in clinical medicine? *BMC Med Educ* 14:249
- Palmer EJ, Devitt PG (2007) Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ* 7:49
- Harris BHL, Walsh JL, Tayyaba S, Harris DA, Wilson DJ, Smith PE (2015) A novel student-led approach to multiple-choice question generation and online database creation, with targeted clinician input. *Teach Learn Med* 27:182–188
- Jobs A, Twesten C, Göbel A, Bonnemeier H, Lehnert H, Weitz G (2013) Question-writing as a learning tool for students—outcomes from curricular exams. *BMC Med Educ* 13:89
- The University of Auckland NZ (2014–2018) PeerWise. The University of Auckland, New Zealand. Available via <https://peerwise.cs.auckland.ac.nz/>
- Fellenz MR (2004) Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Evaluation in Higher Education* 29:703–719
- Arthur N (2006) Using student-generated assessment items to enhance teamwork, feedback and the learning process. *Synergy* 24:21–23
- Sharp A, Sutherland A (2007) Learning Gains... my (ARS) The impact of student empowerment using audience response systems technology on knowledge construction Student Engagement and Assessment REAP International Online Conference on Assessment Design for Learner Responsibility, 29th–31st May, pp 29–31
- Chandrasekar H, Gesundheit N, Nevins AB, Pompei P, Bruce J, Merrell SB (2018) Promoting student case creation to enhance instruction of clinical reasoning skills: a pilot feasibility study. *Adv Med Educ Pract* 9:249
- Kadir FA, Ansari RM, AbManan N, Abdullah MHN, Nor HM (2014) The impact of PeerWise approach on the academic performance of medical students. *Malays Online J Educ Tech* 2:37–49
- Walsh JL, Harris BH, Denny P, Smith P (2018) Formative student-authored question bank: perceptions, question quality and association with summative performance. *Postgrad Med J* 94:97–103
- Wagener S, Möltner A, Timbil S et al (2015) Development of a competency-based formative progress test with student-generated MCQs: results from a multi-centre pilot study. *GMS Z Med Ausbild* 32
- Grainger R, Dai W, Osborne E, Kenwright D (2018) Medical students create multiple-choice questions for learning in pathology education: a pilot study. *BMC Med Educ* 18:201
- Vegada B, Shukla A, Khilnani A, Charan J, Desai C (2016) Comparison between three option, four option and five option multiple choice question tests for quality parameters: a randomized study. *Indian J Pharmacol* 48:571
- Rodriguez MC (2005) Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. *Educ Measurement: Issues and Practice* 24:3–13
- Collins J (2006) Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics* 26:543–551
- Downing SM (2005) The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. *Adv Health Sci Educ Theory Pract* 10:133–143
- Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334
- Laveault D, Grégoire J (2014) Introduction aux théories des tests en psychologie et en sciences de l'éducation 3e éd. edn. De Boeck, Bruxelles
- Zimmaro DM (2016) Item analysis. Writing good multiple-choice exams. Faculty Innovation Center, University of Texas at Austin. Available via <https://facultyinnovate.utexas.edu/sites/default/files/writing-good-multiple-choice-exams-fic-120116.pdf>. Accessed 1 Jan 2017.
- Haladyna TM (2004) Developing and validating multiple-choice test items, Third edn. Lawrence Erlbaum Associates, USA
- Mehta G, Mokhasi V (2014) Item analysis of multiple choice questions: an assessment of the assessment tool. *Int J Health Sci Res* 4:197–202
- Braibant J-M Les examens QCM. Comment lire et interpréter les rapports d'analyse d'items (Contest) en vue d'améliorer la qualité de vos examens ? Service d'évaluation en appui à la qualité, UCL https://cdn.uclouvain.be/public/Exports%20reddot/adef/documents/EVA_QCM_version3.pdf. Accessed 5/05/2017
- Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of educational objectives, handbook I: the cognitive domain. New York: David McKay Co Inc
- Zaidi NB, Hwang C, Scott S, Stallard S, Purkiss J, Hortsch M (2017) Climbing Bloom's taxonomy pyramid: lessons from a graduate histology course. *Anat Sci Educ* 10:456–464
- Phillips AW, Smith SG, Straus CM (2013) Driving deeper learning by assessment: an adaptation of the Revised Bloom's Taxonomy for medical imaging in gross anatomy. *Acad Radiol* 20:784–789
- Bates SP, Galloway RK, Riise J, Homer D (2014) Assessing the quality of a student-generated question repository. *Phys Rev ST Phys Educ Res* 10: 020105
- Flannery MC (2007) Observations on biology. *Am Biol Teach* 69:561–565
- Krathwohl DR, Anderson LW (2009) A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman, New York
- Lam R (2014) Can student-generated test materials support learning? *Stud Educ Eval* 43:95–108
- Aflalo E (2018) Students generating questions as a way of learning. *Act Learn High Educ* 0:1469787418769120
- Hardy J, Bates SP, Casey MM et al (2014) Student-generated content: Enhancing learning through sharing multiple-choice questions. *Int J Sci Educ* 36: 2180–2194
- European Board of Radiology (EBR) (2018) The European Diploma in Radiology (EDIR): investing in the future of the new generations of radiologists. *Insights Imaging* 9:905–909

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)