

Analysis of the French national evaluation of radiology residents

Ursule Dumont, Stephane Sanchez, Benjamin Olivier, Jean-François Chateil, Luc Pellerin, Marie-Christine Beauvieux, Anne-Karine Bouzier-Sore, Hélène Roumes

▶ **To cite this version:**

Ursule Dumont, Stephane Sanchez, Benjamin Olivier, Jean-François Chateil, Luc Pellerin, et al.. Analysis of the French national evaluation of radiology residents. Diagnostic and Interventional Imaging, 2019, 100 (3), pp.185-193. 10.1016/j.diii.2018.11.006 . hal-02472400

HAL Id: hal-02472400
https://cnrs.hal.science/hal-02472400

Submitted on 22 Oct 2021

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 - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S2211568418302948
Manuscript_43d92dd91e7209a88dcaad0fea783e8f

Analysis of the French national evaluation of radiology residents

Gaël Dournes[a,b,c], Ivan Bricault[d,e], Jean-François Chateil[f,g]

[a]Bordeaux University, Centre de Recherche Cardio-Thoracique de Bordeaux, U1045, 33000 Bordeaux, France

[b]INSERM, Centre de Recherche Cardio-Thoracique de Bordeaux, U1045, 33000 Bordeaux, France

[c]Department of Cardiovascular and Thoracic Imaging, CHU de Bordeaux, 33600 Pessac, France

[d]Department of Medical Imaging, CHU de Grenoble, Hôpital Nord, 38043 Grenoble, France

[e]Université Grenoble-Alpes, TIMC-IMAG, 38000 Grenoble, France.

[f]Department of Pediatric imaging, CHU de Bordeaux, 33000 Bordeaux, France.

[g]Bordeaux University, Centre de Résonance Magnétique des Systèmes Biologiques, UMR 5536, 33076 Bordeaux, France

Corresponding author: Gaël Dournes, Centre de Recherche Cardio-thoracique de Bordeaux, INSERM U1045, Université Bordeaux Segalen, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.

E-mail: gael.dournes@chu-bordeaux.fr

© 2018 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract

Purpose: In France, a national evaluation is given annually to radiology residents. The aim of this study was to perform both a docimological analysis of the quality of the questionnaire and a statistical analysis of the results.

Materials and methods: This retrospective study, which included French radiology residents from Year 1 to Year 5 of residency, was performed from 2015 to 2017 across 25 medical universities in France. Both qualitative and quantitative docimological analyses were performed as assessed by the Cronbach alpha coefficient, the difficulty of question (PDI), and the coefficient of discrimination (Rir). Results to the questionnaire were compared between years of residency.

Results: The results of the analysis confirmed the quality of the questionnaire (Cronbach alpha coefficient=0.71, mean (PDI =0.40) though the majority of questions could be answered by memory rather than cognitive ability. The mean Rir was 0.02, indicating that students could not be certified using only the questionnaire. The results measuring resident level of knowledge were moderate, with mean results ranging from 9.2/20 at the first year to 11.3/20 at the fifth year of residency ($P<0.001$). There were no significant differences in results obtained between the third, fourth, and fifth year of residency but results were significantly different among university hospitals.

Conclusion: Even if close interactions exist between learning and pedagogic environment, our results suggest that it may be useful to further develop an evaluation process in relation with pedagogic instructions in order to provide more optimal training.

Keywords: Pedagogy; Docimology; Radiology, Learning

Introduction

Medical knowledge, visual skills such as recognition of patterns, and interpretation schemes are key components of developing radiological expertise [1–3]. Expert radiologists develop specific knowledge as well as advanced perception [4–6]. However, longer experience does not always lead to a high level of accuracy. Research shows that an experienced radiologist can occasionally miss or misunderstand lesions [2,7]. Additional training during residency should foster improved professional skills and, thus, minimize errors from certified

radiologists. Therefore, discovering how radiological expertise develops during residency and which factors may predict a higher level of accuracy may help optimize the teaching environment [2,8].

How radiological expertise develops remains largely unknown [8,9]. A previous study indicates that radiological skills improve from the first year to the third year of residency but may not improve beyond the third year [10]. Volume of examination could be a predictive factor of accuracy development as well as the level of expertise of the pedagogical environment [11]. In France, a voluntary national examination is administered annually to radiology residents from the first year to the fifth year of residency. The same examination is administered to the entire student body and covers 10 subspecialties. The benefit of the examination is to allow residents to annually auto-evaluate their own knowledge and provide the French College of Radiology Teachers (CERF) the opportunity to gather feedback regarding teaching methodology.

Interactions between the requests of residents, the teaching modalities, and how examinations are completed are necessary for improving learning and education. In pedagogy, “docimology” refers to the science related to the assessment of examinations and notation systems. To our knowledge, a docimological analysis of the French national examination of radiologists in training has not been reported to assess the quality of the questionnaire itself.

The purpose of this study was to perform a docimological analysis of the questionnaire administered in 2017 to radiologists in training in France during the national examination. The secondary objective was to analyze the results obtained in 2017, and retrospectively to study and compare the progression of the residents from 2015 to 2017.

Materials and methods

Study design

The study was retrospective and performed over 25 medical universities in France from 2015 to 2017.

Study population

Participants included French radiology residents from the first year to the fifth year of residency. In France, participation in the national evaluation is optional. All participants answered the same questionnaire regardless of their year of residency. In this study, residents

whose year of residency was unknown were excluded in order to accurately categorize the results.

National examination of radiology: structure and composition

The examination consisted of a questionnaire structured into 10 categories covering all subspecialties: neuroradiology, head and neck, thoracic, cardiovascular, gastrointestinal, urology, breast and gynecology, pediatrics, musculoskeletal, and oncology imaging. The questionnaire was built as follows: (1) the CERF board asked each subspecialty society to provide a set of questions; (2) the person in charge of the exam reviewed and formatted each set; (3) one or two subspecialists in each category then performed a second review. The questionnaire was then turned into a computer-based examination allowing online participation for each candidate (SIDES platform: Système Inter-universitaire Dématérialisé d'Evaluation en Santé) [12].

The questionnaire consisted of two question types: single-choice questions (SCQ) and multiple-choice questions (MCQ). Each MCQ was composed of five choices, and the number of correct answers allowed was from 1 to 5 per each question. Level of knowledge was assessed using the following classification, in accordance with the French training curriculum for radiology:

- Level 1: basic knowledge for routine purpose, required to be learned during the first three years of residency;
- Level 2: advanced knowledge, to be acquired during the fourth and fifth years of residency;
- Level 3: expert knowledge, more likely to be learned after the completion of radiology training during a post-certification subspecialization.

Each subspecialty's set of 10 questions followed pre-required instructions to authors to comply with 50% MCQ / 50% SCQ and seven Level 1 questions, two Level 2 questions, and one Level 3 question. The examination was administered once per year in October, on the same time and day across all 25 medical universities in France. All residents in each center are brought together in a single room with dedicated computers, under the responsibility of a regional teaching coordinator. The notation is done electronically through the SIDES platform. The scale used to score the final result is as follows: For SCQ, 2 points for a correct answer and 0 points for an incorrect answer. For MCQ, 2 points if all answers are correct, 1

point if one answer is erroneous, and 0 points in all other cases. Erroneous answers are either a missing answer or a false positive answer. Given that 10 questions are proposed per each module, the maximum score is 200 points.

Qualitative docimological analysis of the questionnaire

For study purposes, we arbitrarily divided questions into two categories: questions related to image analysis (IM) and questions related to pure factual knowledge (KL). For IM questions, we distinguished between those related to recognition and those requiring rendering. The analysis may relate to a synthesis of several data, or it may require some data that are not contained in the available images but in the clinical introduction provided in each question.

Quality of the national examination questionnaire was analyzed using two criteria: (1) structure, and (2) question cognition level. Two readers (one with 10 years and one with 30 years of imaging experience) analyzed the questionnaire together. Quality of structure was scored using binary scales (0=no; 1=yes) according to the following items: no double-meaning, no error, only one concept per question, affirmative statement, no modal adverb without appeal (never, always), no information on the number of correct answers, all snares are plausible.

The level of cognition was categorized from lowest to highest as follows: memory only (low), both memory and image analysis (medium), contextual interpretation (high). When several cognition levels were present within a single question, only the highest level was recorded.

Quantitative docimological analysis of the questionnaire

Three indexes were analyzed to quantify the quality of the questionnaire: the Cronbach alpha coefficient, the difficulty index P_{DI}, and the coefficient of discrimination Rir (item rest correlation) [13,14]. The definition and method of calculation of each index are extensively described in the Supplemental Methods section (Appendix). Briefly, the Cronbach alpha coefficient is an index which indicates the degree of internal consistency of a questionnaire. An alpha (α) coefficient > 0.70 was considered acceptable and satisfactory when > 0.80.

The P_{DI} index of difficulty is a proportion and calculated as the number of students who correctly answered a question divided by the total number of students. In this study, we

have considered that a question was easy when P_{DI} was > 0.80 and difficult when P_{DI} was < 0.20.

The item rest correlation (Rir) represents an index used to quantify the degree of discrimination allowed by a given question. It quantifies the relationship between the number of students who correctly answered a given question versus the final score of the full questionnaire. A positive Rir is considered discriminant when > 0.20. A negative Rir indicates an inconsistency because students who correctly answered the given question are those who had the lowest result for the full questionnaire. A Rir < 0.10 may indicate that the question should not be kept, unless the examination would be applied for certification purpose. Finally, the item advice report is a graphical representation where P_{DI} is plotted as a function of Rir.

Statistical analysis of questionnaire results

Multiple comparisons of means were performed using an ANOVA test. Comparisons between independent groups was assessed using Student *t*-test and comparison of paired means using Student *t*-test for paired data, with Bonferroni correction.

Results

Characteristics of the questionnaire

The question categories were heterogeneous across overspecialization, ranging from 100% (10/10) of IM questions regarding gastrointestinal imaging to 30% (3/10) of IM questions regarding oncology. However, both format and levels of questions were found similar between radiological subspecialties (Table 1).

Qualitative docimological analysis of the questionnaire

Table 2 shows that the overall, structural quality ranged from a minimum of 62/70 points (oncology questionnaire) to a maximum 69/70 points (gastrointestinal and breast questionnaires) out of a potential total score of 100 points. The most common errors were the occurrence of double-meanings in 6% (6/100) of questions, more than one concept per question in 18% (18/100) of questions, and modal adverbs in 7% (7/100) of questions.

In addition, the cognitive level of questions was mainly low, with 40% (40/100) of questions requiring memory only; of those, 32/40 (80%) were Level 1 questions (Table 3). This proportion was closely followed by medium-level cognition in 38% of questions

(38/100). High-level cognition questions, although represented by 22% (22/100) of questions, were proportionately heterogeneous (e.g., up to 70% [7/10] in neuroradiology and only 10% [1/10] in breast imaging/gynecology).

Quantitative docimological analysis of the questionnaire

The internal consistency of the questionnaire was acceptable with a Cronbach α coefficient of 0.71. The item advice report demonstrated that a substantial number of questions did not allow discrimination between students with high and low skill levels (Fig. 1). Indeed, 34% of questions (34/100) had a Rir > 0.10 and 56% (56/100) had a negative Rir. The mean level of difficulty was P_{DI} = 0.40 and the mean coefficient of discrimination was 0.02. The most difficult questionnaire concerned thoracic imaging (Table 4).

Only questions related to neuroradiology had a mean Rir of 0.20, which is the threshold to consider that students with high skill levels can be discriminated from students with low skill levels (Supplemental Methods). There were 5/10 subspecialties that had a negative mean Rir (*i.e.*, head and neck, thoracic, cardiovascular, breast-gynecology, and pediatric imaging).

Resident results obtained to the questionnaire in 2017

Characteristics of the residents

A total of 904 residents completed the questionnaire: 31 residents were excluded because the year of residency was missing. The proportion of residents varied according to year of residency (Fig. 2) and universities (Fig. 3). The majority of residents (904/1200; 75.33%) completed the questionnaire during the first year to the fourth year. However, 38.16% (79/207) of fifth-year residents participated. Because of data anonymization it was not possible to retrieve sex and age of participants retrospectively.

2017 examination scores

Fig. 4 shows the mean results obtained by the residents to the questionnaire in 2017, as expressed using a maximum score of 20 points. The repartition of values was Gaussian around the mean score (Fig. 5). The maximum progression slope (+1 point) was noticed between Year 1 and Year 2 of residency. There was a decrease of the slope to +0.6 between Year 2 and Year 3 and between Year 3 and Year 4 of residency. There were significant

differences in mean between Year 1 and Year 2 compared to all other years of residency ($P < 0.001$). Year 3 was significantly different from Year 4 ($p<0.001$) but not Year 5 ($P = 0.09$). There was no significant difference between Year 4 and Year 5 (11.4 vs. 11.3, respectively; $P =0.73$) (Fig. 4). Results were heterogeneous according to the subspecialties. The lowest proportion of results considered successful were found in thoracic, gastrointestinal, and head and neck imaging (Fig. 1, Table 4).

Longitudinal analysis of results obtained between 2015 and 2017

A retrospective analysis of results obtained between 2015 and 2017 was completed. For this, mean scores obtained in Year 1, Year 2, and Year 3 were followed up in 2016 and 2017. Results showed that there were significant differences between years of evaluation except for Year 3, between 2016 and 2017 (Fig. 6).

Discussion

In this study, we found that docimological analysis of the questionnaire showed that the quality of the structure was met owing to either no errors or minimal errors regarding double-meaning, formulation of questions, or snare plausibility. Rigorous internal reviewing of questions by at least two independent reviewers was done before final completion of the questionnaire. Nevertheless, descriptions of clinical cases were often very brief in this questionnaire related to imaging analyses; and thus, authors did not have to complete complex formulations. Interestingly, formulations were not homogeneous across radiological subspecialties. For instance, musculoskeletal imaging had a marked tendency to propose SCQ questions with IM category. However, recognition of a specific diagnosis based on images was not possible in other subspecialties such as pediatric radiology, where MCQ and both IM and KL questions were noticed. Regarding the level of cognition, it should be noticed that the major part of the questionnaire was composed of questions requiring memory only, which is the lowest level of cognition. Questions related to imaging analysis were mainly related to a visual recognition of a diagnosis. Additional more complex questions, with short cases—including successive questions or Clinically Oriented Reasoning Evaluation (CORE), as proposed in the European Diploma in Radiology [15-17]— may be useful in order to appreciate the level of competency of the residents.

The second point of interest is that quantitative analysis of the questionnaire demonstrated that it cannot be used for certification purposes. Despite an acceptable internal

degree of consistency, as assessed by a Cronbach alpha coefficient superior to 0.70, the mean coefficient of discrimination was only 0.02. Normally, a questionnaire resulting in certification should reach a Rir of 0.20 in order to allow discrimination between those students with a high level of skills and those students with a low level of expertise. In addition, 5 out of 10 subspecialties had a negative Rir *(i.e.*, head and neck, thoracic, cardiovascular, breast-gynecology, and pediatric imaging) indicating that those students who successfully passed these questions were not those who had the best results for the entire questionnaire. Although the majority of students learned neuroradiology, this result demonstrated that a substantial number of students may overspecialize during their residency, at the expense of a more generic set of skills and expertise. In this setting, it could be noted that, at the end of residency, independent evaluations of each subspecialty could be more relevant than a questionnaire dedicated to covering the entire 10 subspecialties. Another option would be to assess a general knowledge no later than the third or fourth year of residency.

The third point of interest of this study is that statistical analysis of the residents' results appeared to be in disagreement with the low level of cognition required, with mean results ranging from 9.2 at Year 1 to 11.3 at Year 5. In addition, the scoring system was prone to overestimate the results obtained by the resident since half of the points were still allocated whether a resident would have made a single mistake at MCQ. In addition, there were significant differences between Year 5 and Years 1 and 2 but not with Years 3 or 4. Several points can be discussed to explain this result. Even though nonsignificant, the absolute value of mean of Year 5 was inferior to that of Year 4. The national examination is a non-degree examination; and thus, no particular preparation was done before completing the questionnaire. As an optional test, a vast majority of residents of Year 5 were absent. Moreover, imaging analyses were done using single two-dimensional images, which is different from routine practice where all three-dimensional images over a full volume are available before analysis. Once again, new modalities (e.g., CORE examination with practical oriented cases) providing a full set of sonography, CT, or MRI images are very useful to better evaluate the level of knowledge and competence [17].

In most French university hospitals, radiology departments are divided into subspecialized units. Thus, to be able to answer the full questionnaire, a resident should complete a temporary internship in each subspecialized unit. Unfortunately, in 2017 the French system did not provide opportunities for residents to gain adequate work experience in

all subspecialized radiology units; allocation of internship was a resident's free choice. Some deliberately chose to overspecialize in some areas and not train in others. Thus, it is possible that any novel experience was acquired at the expense of previously acquired knowledge. Therefore, experience acquired during the later years of residency may not be fully representative of that also obtained during early residency. Fortunately, with the "new deal" reform launched in 2017, the third specialized cycle of medical teaching now requires a practical internship in each subspecialty for at least three months in order to give residents a better overview.

Another point of interest is that results were found heterogeneous across universities. Our analysis was unclear as to whether this was related to the residents themselves or the pedagogic environment to which they belonged. We did not have data on the residents' previous performance in high school or secondary school. Finally, interventional radiology was not part of the national examination; a substantial part of residents was probably dedicated to developing these therapeutic skills with specific competence, instead of pure diagnostic skills. These results may indicate that individual expertise largely develops during the first three years of residency. Detecting a lack of progression at the individual level is important during this period and may require special focus. Simply increasing experience may be insufficient as far as developing expertise; more specific training may be necessary to obtain feedback and improve accuracy.

Our study has some limitations. First, the level of cognition was sometimes ambiguous since radiological expertise can develop in a number of ways: frequent exposure to common diagnoses, direct recognition of lesions, or developing synthesis abilities that require more complex cognition exercise. Our analysis was not able to clearly determine a resident's strategy, exposure to diagnosis, or at what frequency. Because the national examination is optional, it does not represent the maximum level that a resident can reach but rather the basic, routine level of everyday practice. Furthermore, radiologists may glean additional information on their own by accessing data online. Therefore, the results of this evaluation may not be representative of the final quality of radiological reports that are provided to the patients in everyday practice.

In conclusion, our docimological analysis of the national examination revealed that the quality of structure and internal consistency was acceptable; whereas the cognition level required was relatively low, mainly related to pure memory. However, results obtained by the residents highlight that the questionnaire cannot be used for certification purposes. There was

a marked tendency of students to overspecialize precociously at the expense of a more generic knowledge, and results did not improve starting from the third year of residency. Therefore, the timing between an evaluation dedicated to general knowledge and then an overspecialized certification could be discussed. Further research could be worthwhile in order to better understand the predictive factors that would allow optimal training and improved pedagogical instructions.

Conflict of interest

The authors declare that they have no conflicts of interest concerning this article.

References

[1] Drew T, Evans K, Võ ML-H, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? Radiographics 2013;33:263–74.

[2] Taylor PM. A review of research into the development of radiologic expertise: implications for computer-based training. Acad Radiol 2007;14:1252–63.

[3] van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JCBM, Ravesloot CJ, van Schaik JPJ, et al. Interpretation of radiological images: towards a framework of knowledge and skills. Adv Health Sci Educ 2014;19:565–80.

[4] Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 2007;242:396–402.

[5] Mallett S, Phillips P, Fanshawe TR, Helbren E, Boone D, Gale A, et al. Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: visual perception of experienced and inexperienced readers. Radiology 2014;273:783–92.

[6] Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, et al. How experience and training influence mammography expertise. Acad Radiol 1999;6:575–85.

[7] Krupinski EA. Current perspectives in medical image perception. Atten Percept Psychophys 2010;72:1205–17.

[8] Gunderman R, Williamson K, Fraley R, Steele J. Expertise. Acad Radiol 2001;8:1252–6.

[9] Krupinski EA. The role of perception in imaging: past and future. Semin Nucl Med 2011;41:392–400.

[10] Ravesloot CJ, van der Schaaf MF, Kruitwagen CLJJ, van der Gijp A, Rutgers DR, Haaring C, et al. Predictors of knowledge and image interpretation skill development in radiology residents. Radiology 2017;284:758–65.

[11] Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. J Natl Cancer Inst 2002;94:369–75.

[12] Jolly D, Ambrosi P, Chaffanjon P, Dreyfuss D, Le Jeunne C, Lorette G, et al. Results of the 2015 national ranking in France. Press Med 2016;45:133–42.

[13] El Allaoui A, Rhazi Filali F, El Hadri EM, Fetteh K, Bouhadi M. Étude évaluative d'examen normalisé de sciences de la vie et de la terre au cycle secondaire collégial. Eur Sci J 2016;12:283-299.

[14] Zijlmans OE, Tijmstra J, Andries van A, Sijtsma K. Item-score reliability in empirical-data sets and its relationship with other item indices. Educ Psychol Meas 2018;6:998-1020.

[15] Breatnach E. Diploma (EDiR) of the European Society of Radiology. Radiologe 2011;51:1058–9.

[16] Tong E, Spooner M, Van Delden O, Uberoi R, Sheehan M, O'Neill DC, et al. The European Board of Interventional Radiology Examination: a cross-sectional web-based survey. Cardiovasc Intervent Radiol 2018;41:21–6.

[17] European Board of Radiology (EBR). The European Diploma in Radiology (EDiR): investing in the future of the new generations of radiologists. Insights Imaging 2018;doi: 10.1007/s13244-018-0665-7.

Figure legends

Figure 1. Item advice report of the questionnaire. The difficulty of question P_{DI} is represented in the vertical axis, as a function of the coefficient of discrimination Rir in the horizontal axis. Black dots represent the scores obtained by each question.

Figure 2. Participation of residents per year of residency in 2017. Percentages indicate the proportion of residents as compared to the total theoretical number of residents.

Figure 3: Resident participation per University Hospital in France in 2017. Blue bars represent the theoretical number of residents and orange bars the number of residents who actually participated to the examination.

Figure 4. Results to the questionnaire. Blue bars indicate the mean scores obtained by the residents to the questionnaire in 2017, expressed as a maximum score of 20. Year x corresponds the residents who belong to the x year of residency. Red lines indicate the standard deviation of scores.

Figure 5: Distribution of cores obtained by residents in 2017, expressed as a maximum score of 100. Blue bars indicate the percentage of residents who obtained a given score, as indicated in horizontal axis. Red line is a model of the distribution of scores.

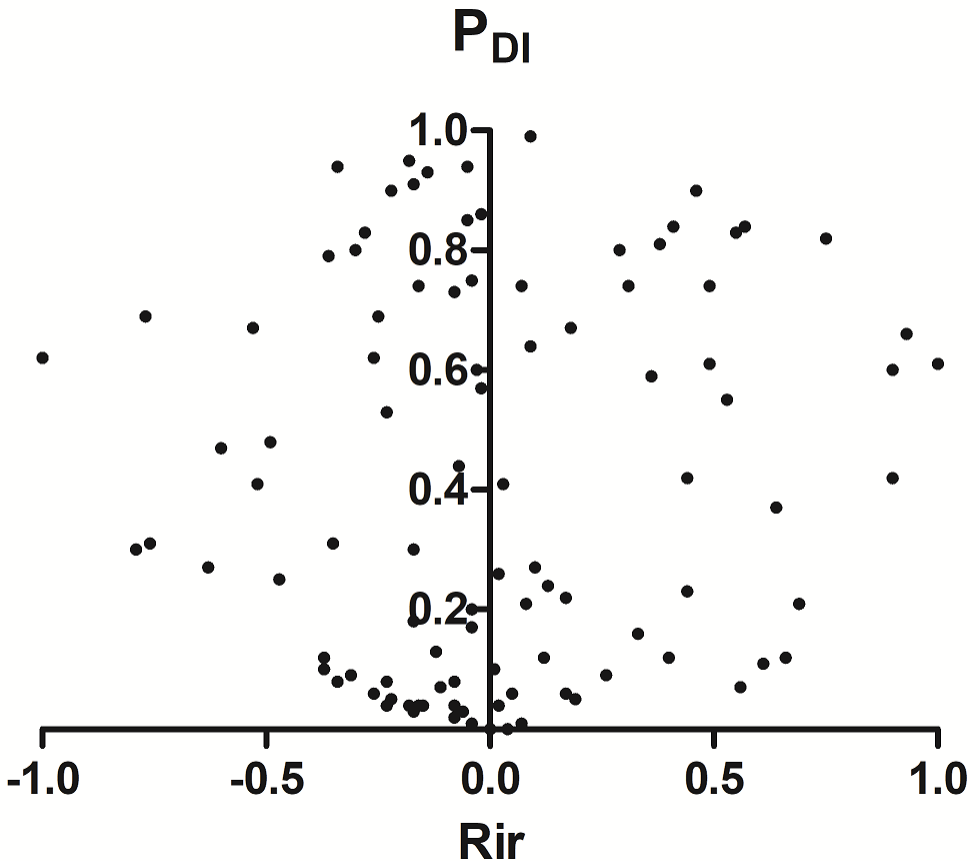
Figure 6. Year x correspond to the population of residents who were in their x year of residency in 2015. White bar, grey bar and black bar represent the mean score obtained by the resident at the examination in 2015, 2016 and 2017, respectively. Star sign (*) indicates a significant difference in paired mean scores. NS indicates no significant difference.

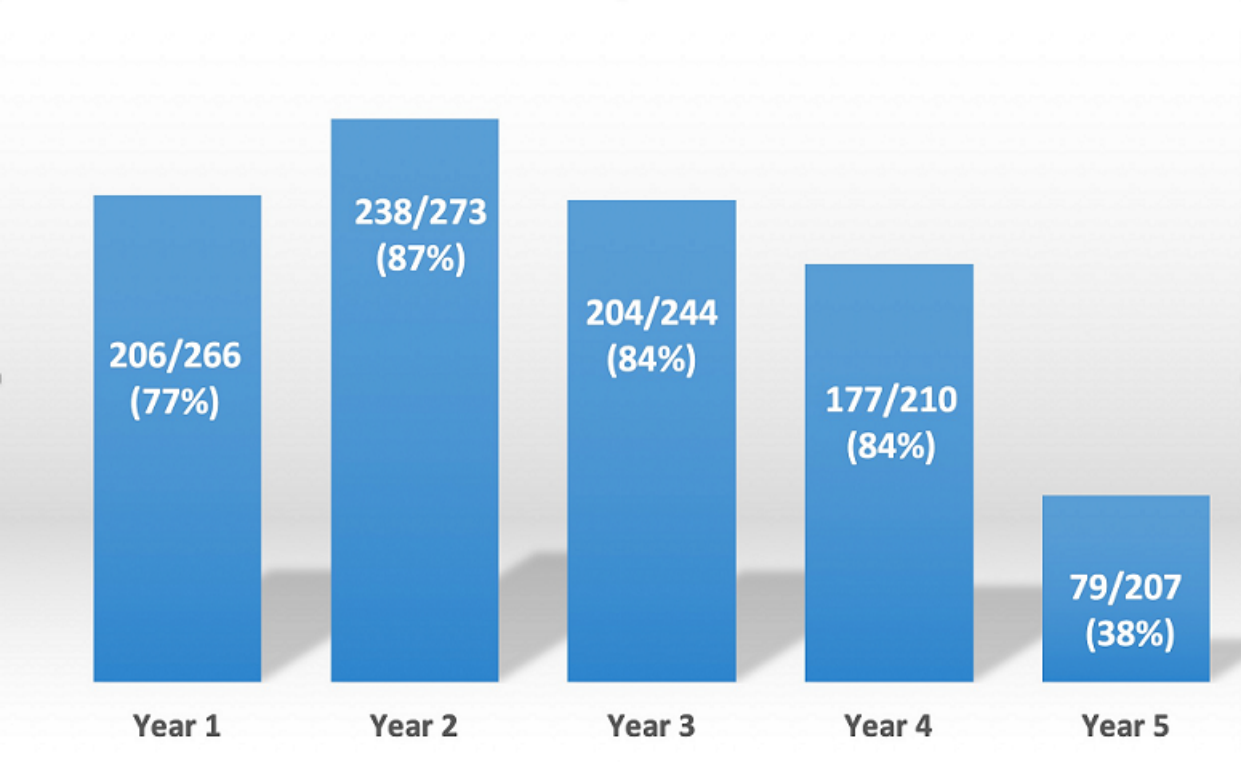
Table. 1. Question characteristics.

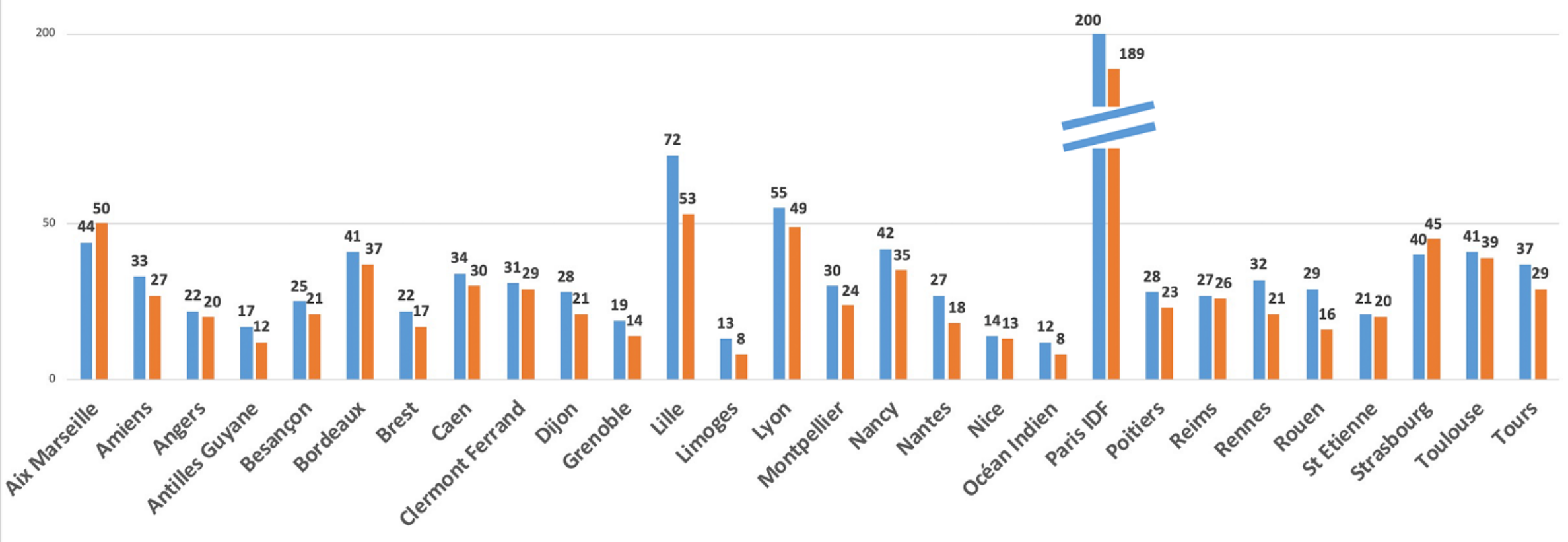
Table 2. Qualitative docimological analysis of the structural quality of the questionnaire.

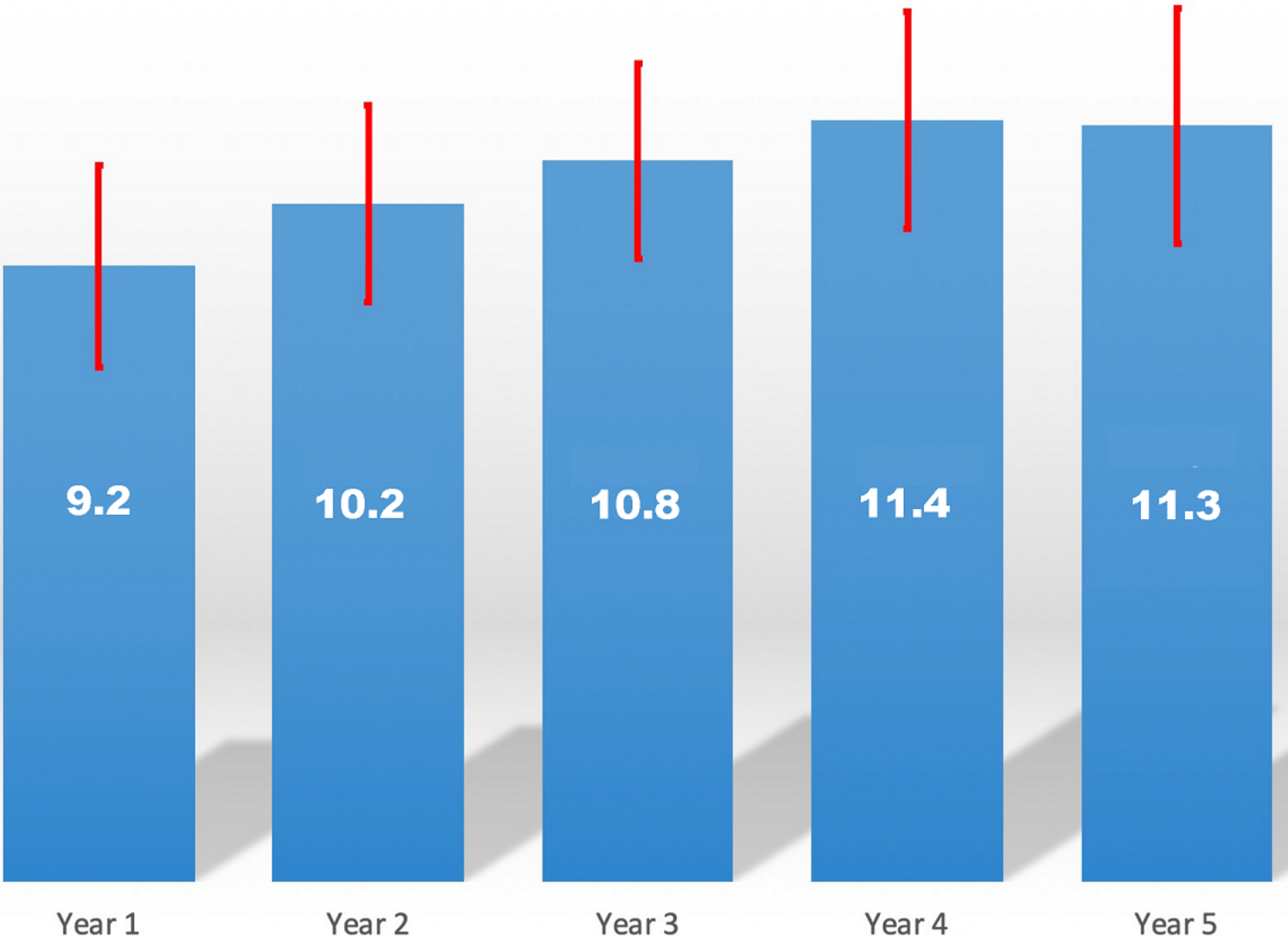
Table 3. Qualitative docimological analysis of cognitive level of the questionnaire.

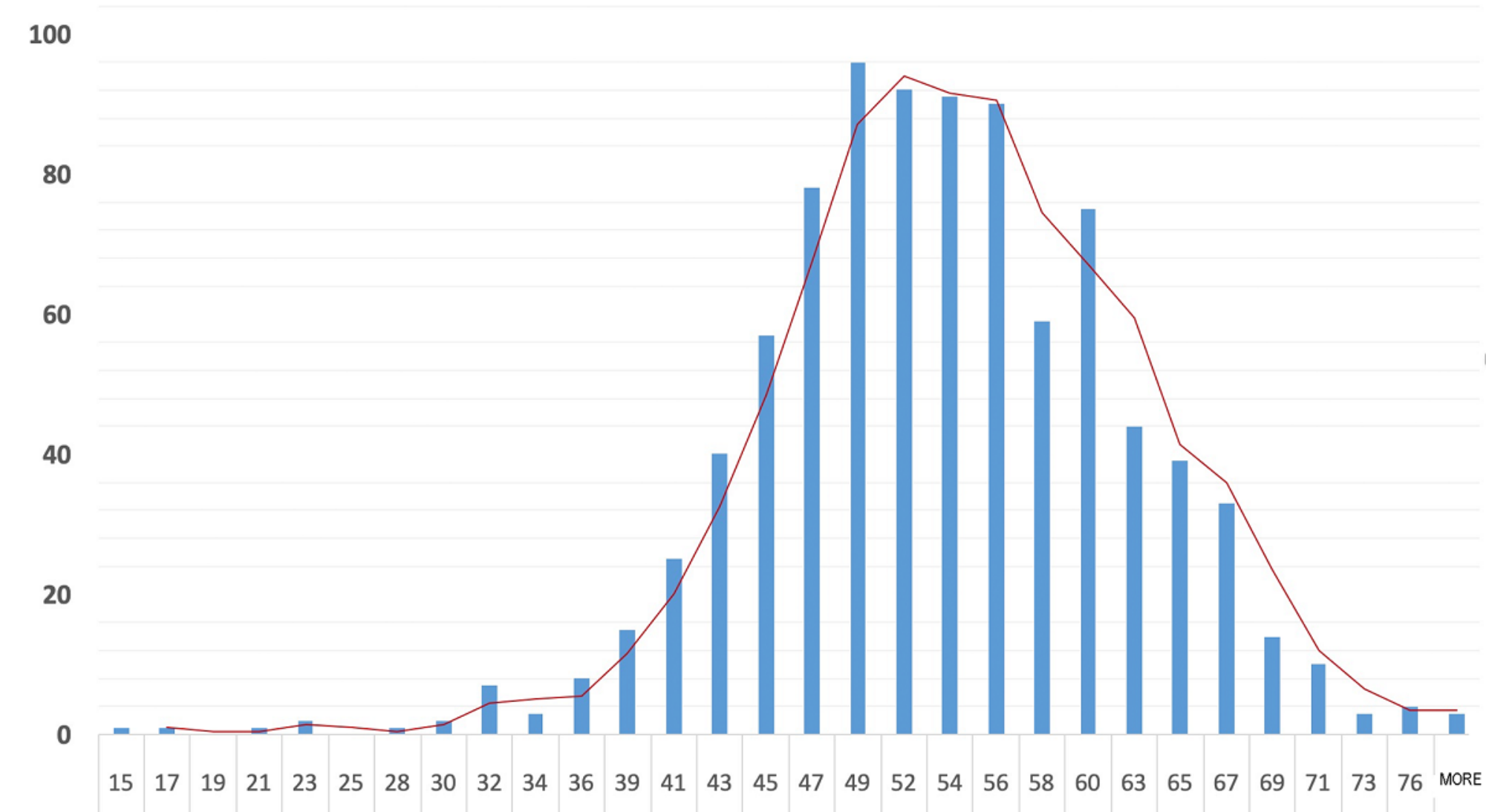
Table 4. Quantitative docimological analysis of the questionnaire per subspecialties.

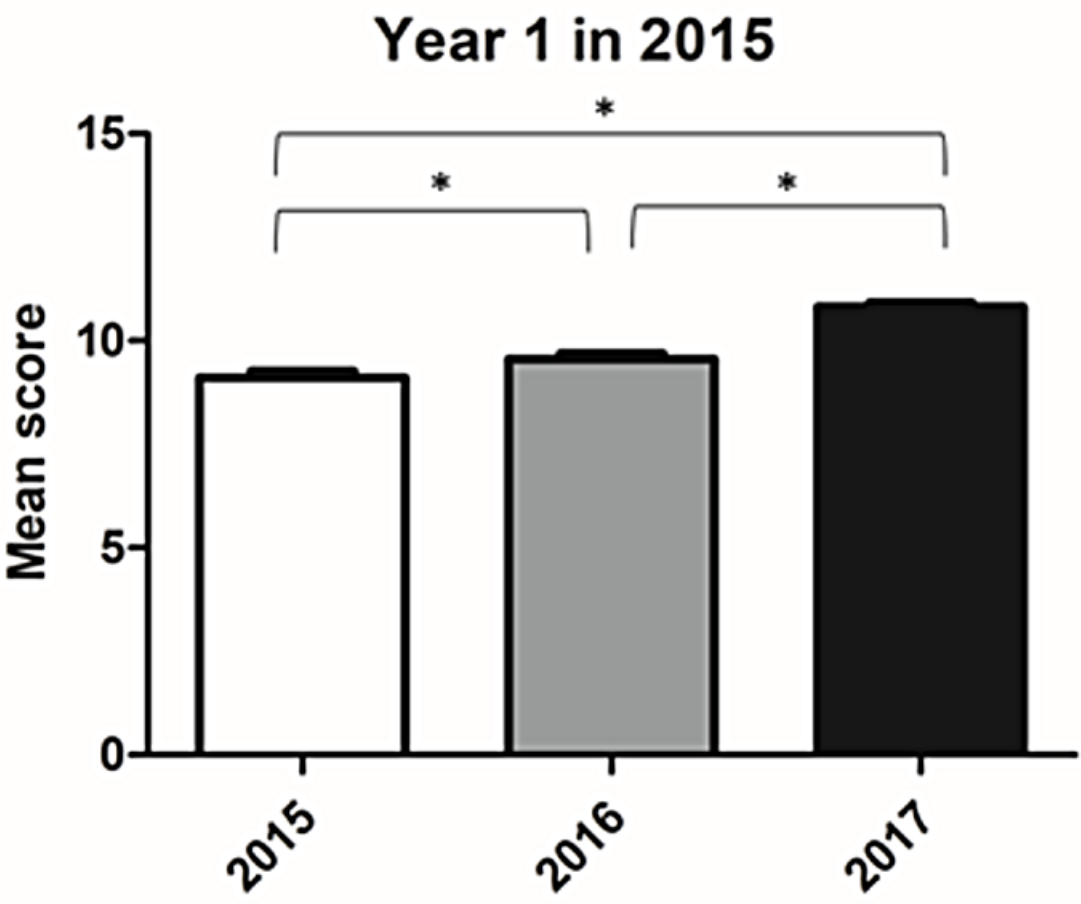

P_DI
1.0
0.8
0.6
0.4
0.2
-1.0
-0.5
0.0
0.5
1.0
Rir

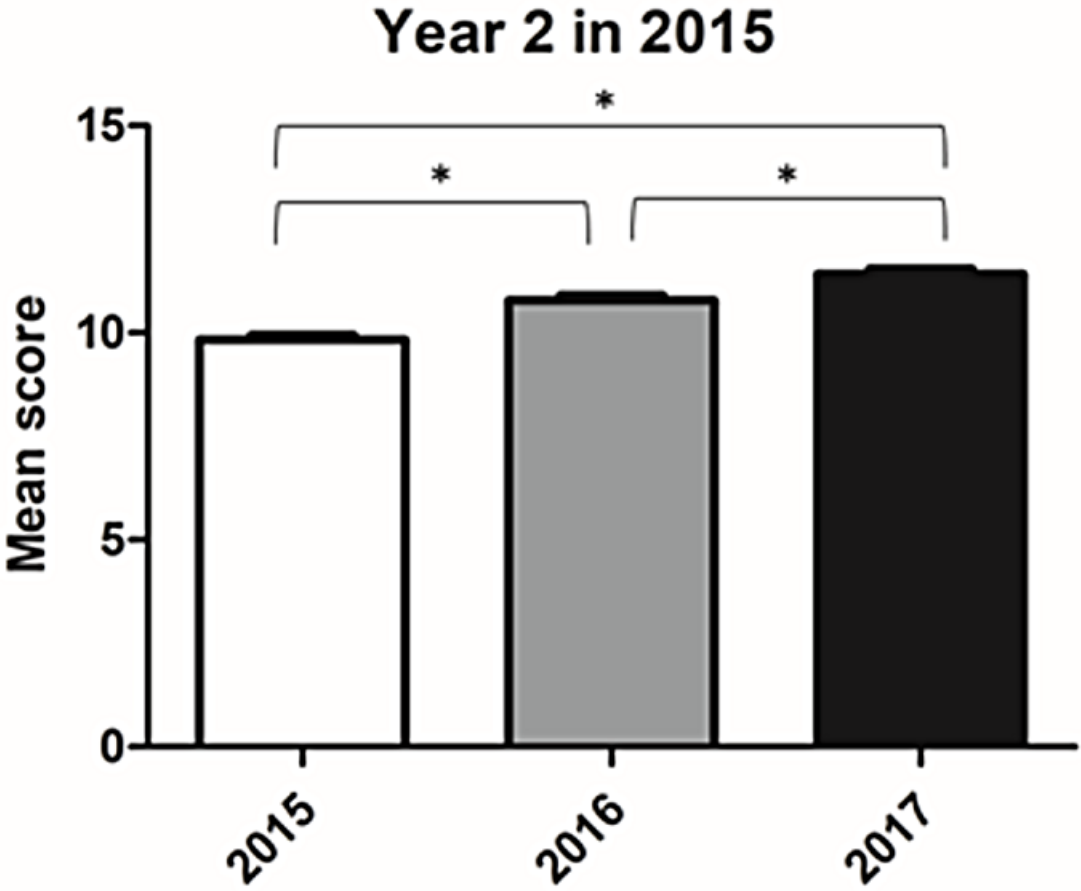

206/266
(77%)
238/273
(87%)
204/244
(84%)
177/210
(84%)
79/207
(38%)
Year 1
Year 2
Year 3
Year 4
Year 5

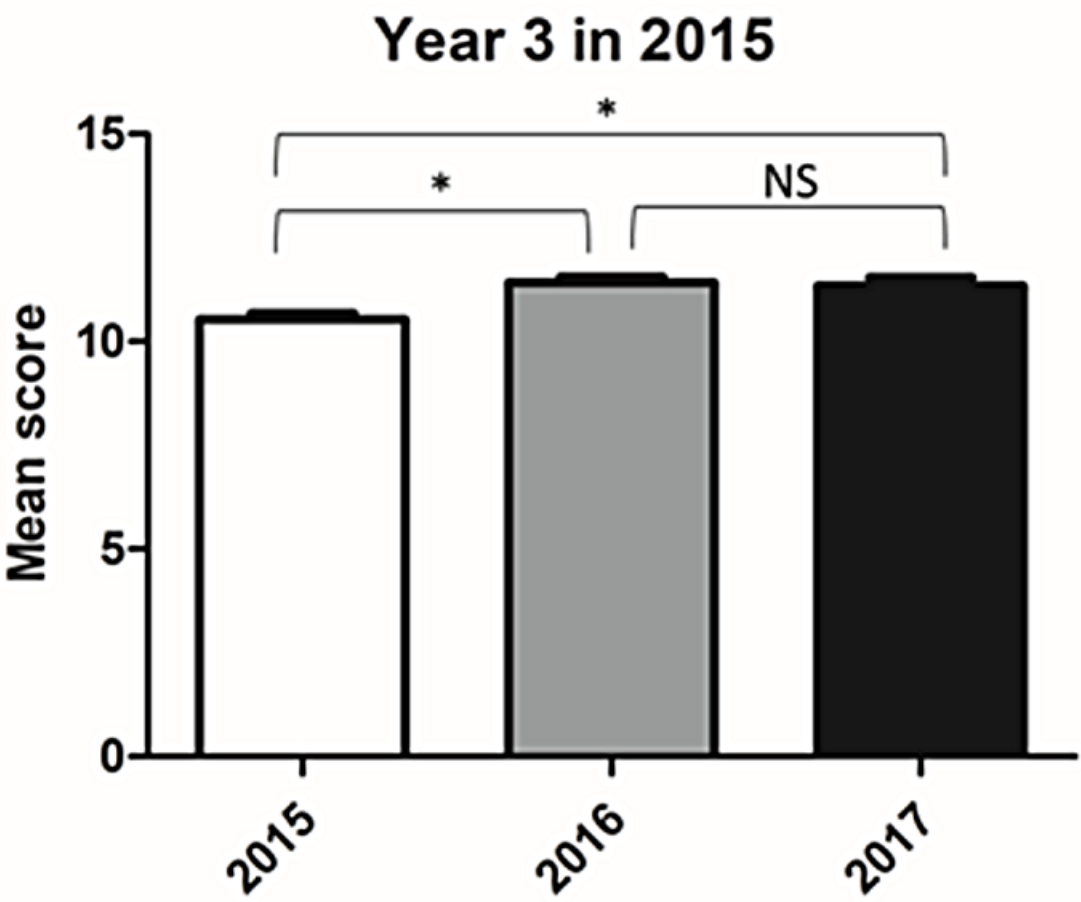

200
50
0
44 50
33 27
22 20
17 12
25 21
41 37
22 17
34 30
31 29
28 21
19 14
72 53
13 8
55 49
30 24
42 35
27 18
14 13
12 8
200 189
28 23
27 26
32 21
29 16
21 20
40 45
41 39
37 29
Aix Marseille
Amiens
Angers
Antilles Guyane
Besançon
Bordeaux
Brest
Caen
Clermont Ferrand
Dijon
Grenoble
Lille
Limoges
Lyon
Montpellier
Nancy
Nantes
Nice
Océan Indien
Paris IDF
Poitiers
Reims
Rennes
Rouen
St Etienne
Strasbourg
Toulouse
Tours


9.2
10.2
10.8
11.4
11.3
Year 1
Year 2
Year 3
Year 4
Year 5


100
80
60
40
20
0
15
17
19
21
23
25
28
30
32
34
36
39
41
43
45
47
49
52
54
56
58
60
63
65
67
69
71
73
76
MORE


Year 1 in 2015
Mean score
15
10
5
0
*
*
*
2015
2016
2017


Year 2 in 2015
Mean score
15
10
5
0
*
*
*
2015
2016
2017


Year 3 in 2015
Mean score
15
10
5
0
*
*
NS
2015
2016
2017

	Type		Format		Level		
Specialty	IM	KL	MCQ	SCQ	Level 1	Level 2	Level 3
Neuroradiology	9/10	1/10	5/10	5/10	7/10	2/10	1/10
Head & neck	8/10	2/10	5/10	5/10	7/10	2/10	1/10
Thoracic	9/10	1/10	5/10	5/10	7/10	2/10	1/10
Cardiovascular	3/10	7/10	5/10	5/10	7/10	2/10	1/10
Gastrointestinal	10/10	0/10	5/10	5/10	7/10	2/10	1/10
Urology	5/10	5/10	5/10	5/10	7/10	2/10	1/10
Breast/gynecology	8/10	2/10	5/10	5/10	7/10	2/10	1/10
Musculoskeletal	7/10	3/10	5/10	5/10	7/10	2/10	1/10
Pediatrics	8/10	2/10	5/10	5/10	7/10	2/10	1/10
Oncology	3/10	7/10	5/10	5/10	7/10	2/10	1/10
Total	70/100	30/100	50/100	50/100	70/100	20/100	10/100

Note. IM=imaging-analysis question; KL=knowledge question; MCQ=multiple choice question; SCQ=single choice question. Data are the proportion of questions.

	Neuroradiology	Head & neck	Thoracic	Cardiovascular	Gastrointestinal	Urology	Breast/gynecology	Musculoskeletal	Pediatrics	Oncology	Total
No double-meaning	10/10	10/10	10/10	9/10	10/10	8/10	9/10	10/10	9/10	9/10	94/100
No error	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	100/100
Single concept	10/10	6/10	10/10	5/10	10/10	9/10	10/10	10/10	10/10	3/10	82/100
Affirmative mode	9/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	95/100
No modal adverb	9/10	9/10	8/10	9/10	9/10	10/10	10/10	10/10	9/10	10/10	93/100
Number of answers	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	100/100
Snare plausibility	10/10	9/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	99/100
Total	68/70	65/70	67/70	63/70	69/70	67/70	69/70	70/70	68/70	62/70	

Note. Data are the proportions of correct questions.

Table 3. Qualitative docimological analysis of cognitive level of the questionnaire.

	LOW LEVEL (Memory alone)			MEDIUM LEVEL (Memory and imaging analysis)			HIGH LEVEL (Interpretation owing to context)		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Neuroradiology	2/10			1/10			4/10	2/10	1/10
Head & neck	2/10				7/10	1/10			
Thoracic	3/10	1/10		4/10				1/10	1/10
Cardiovascular	5/10	1/10	1/10	2/10	1/10				
Gastrointestinal	1/10			3/10	1/10		3/10	1/10	1/10
Urology	6/10			1/10	2/10				1/10
Breast/gynecology	4/10			3/10	1/10	1/10		1/10	
Musculoskeletal	3/10	1/10		4/10	1/10	1/10			
Pediatrics	1/10	1/10	1/10	2/10	1/10	1/10	4/10		
Oncology	5/10	2/10					2/10		
Total	32/100	6/100	2/100	20/100	14/100	4/100	13/100	5/100	4/100

Table 4. Quantitative docimological analysis of the questionnaire per subspecialties

	Mean P_{DI}	Mean Rir	Proportion of easy questions	Proportion of difficult questions
Neuroradiology	0.46	0.20	3/10	2/10
Head & neck	0.41	-0.05	2/10	4/10
Thoracic	0.26	-0.07	0/10	6/10
Cardiovascular	0.40	-0.01	1/10	4/10
Gastrointestinal	0.45	0.13	3/10	5/10
Urology	0.34	0.11	3/10	4/10
Breast/gynecology	0.34	-0.05	1/10	4/10
Musculoskeletal	0.41	0.00	1/10	3/10
Pediatrics	0.43	-0.06	3/10	4/10
Oncology	0.48	0.06	3/10	3/10
Total	0.40	0.02	20/100	39/100

Note P_{DI}=index of difficulty; Rir=coefficient of discrimination; an easy question was defined as $P_{DI} > 0.80$; a difficult question was defined as $P_{DI} < 0.20$.