Providing training on critical appraisal of medical literature by junior and senior teacher : an equivalence randomized trial in Algeria

Enseignement de la lecture critique des publications médicales par un enseignant junior versus un enseignant sénior : un essai randomisé d'équivalence

Abdeldjellil Bezzaoucha[1], Abderrezak Bouamra[2], Benyoucef Sidoumou[3], Ahmed Ben Abdelaziz[4]

1 : University Saad Dahlab of Blida. Country: Algeria
2: University Saad Dahlab of Blida. Country: Algeria.
3: Teaching hospital of Blida. Country: Algeria.
4: Teaching hospital of Sousse. Country: Tunisia Research Unit UR 12 SP 36

RÉSUMÉ

Contexte: La pénurie d'enseignants formés à la lecture critique des publications médicales est un obstacle majeur au développement de cet apprentissage dans les pays au revenu bas et intermédiaire.

Objectif: L'objectif de l'étude était de montrer qu'un enseignant junior en épidémiologie était aussi performant qu'un enseignant sénior en épidémiologie pour enseigner la lecture critique.

Méthodes: Des apprenants, enseignants de toutes spécialités à la Faculté de Médecine de Blida (Algérie) et volontaires, ont été répartis par un tirage au sort équilibré entre les deux enseignants dans un essai randomisé d'équivalence. Un test de Fresno adapté et noté sur 200 points a été utilisé pour mesurer, en insu, la performance des apprenants.

Résultats: Les deux groupes, constitués chacun de 33 apprenants ayant subi le post-test, étaient fondamentalement identiques au départ. Aucune différence significative n'a été mise en évidence entre les scores moyens, attribués au post-test, du groupe de l'enseignant junior et du groupe de l'enseignant sénior, respectivement 87.83 ± 27.44 et 84.31 ± 20.45 (intervalle de confiance de la différence des deux moyennes : -15.41 ; 8.38). Cet intervalle de confiance, situé à l'intérieur de l'intervalle d'équivalence (-20 ; +20), a montré que les deux enseignants étaient aussi performants l'un que l'autre pour enseigner la lecture critique des publications médicales.

Conclusion : Un enseignant junior en épidémiologie était finalement au moins aussi capable des mêmes performances qu'un enseignant sénior en épidémiologie pour enseigner la lecture critique des publications médicales. Cela pourrait ouvrir la voie à d'autres initiatives similaires par l'implication d'autres spécialités pour constituer une masse critique de jeunes enseignants capables d'enseigner la lecture critique.

Mots-clés

Ecoles de médecine – Enseignants – Prise de décision – Essai d'équivalence - Médecine factuelle - Algérie

SUMMARY

Background: The scarcity of trained teachers in charge of critical appraisal of medical literature is a major obstacle for the development of this learning in low-middle-income countries.

Objectives: The aim of the study was to show equivalence in providing training on critical appraisal by either junior or senior teacher in epidemiology.

Methods: Learners, volunteer teachers of all specialties at the medical school of Blida (Algeria), were randomized between the two teachers in an equivalence randomized controlled trial. An adapted Fresno test scored out of 200 points was used to blindly measure learners' performance.

Results: Main characteristics of the two learners' groups, of 33 each, who took the post-test, appeared to be similar at baseline. No significant difference was found between the two average scores attributed to the post-test, respectively 87.83 ± 27.44 and 84.31 ± 20.45 (95% confidence interval of the difference between these two means: -15.41; 8.38). That confidence interval, inside the equivalence interval of (-20, +20), revealed that junior teacher was sufficiently efficient to teach critical appraisal in comparison with a senior teacher.

Conclusion: Equivalence in providing training on critical appraisal by either junior or senior teacher could pave the way for other similar initiatives among all specialities.

Key-words

Schools, Medical - School Teachers - Equivalence Trial - Decision Making - Evidence-Based Medicine- Algeria

توفير التدريب على القراءة النقدية للمقالات الطبية
من قبل مدرس مبتدئ وأخر معلم : تجربة معادلة عشوائية

عبد الجليل بزاوشة ، عبد الرزاق بو عمرة ، بنيوسف سيدمو، احمد بن عبد العزيز

الخلفية: تمثل ندرة المعلمين المدربين المكلفين بالتدريب على القراءة النقدية للمقالات الطبية، عقبة رئيسية أمام تطور هذا التعلم في البلدان ذات الدخل المنخفض والمتوسط.

الهدف: هدف هذه الدراسة هو إظهار التكافؤ في توفير التدريب على التقييم النقدي للمقال، من قبل معلم مبتدئ و آخر متمرس في علم الأوبئة.

الطريقة: تم الاختبار العشوائي و التطوعي لأساتذة من جميع التخصصات في كلية الطب في البليدة (الجزائر) ، لتجربة مكافئة عشوائية. تم استخدام اختبار فريسنو المعدل من 200 نقطة، لقياس أداء المتعلمين بشكل أعمى.

النتائج: يبدو أن الخصائص الرئيسية لمجموعتي المتعلمين، 33 في كليهما، بعد الاختبار ، مماثلة في خط الأساس. لم يتم العثور على فروق ذات دلالة إحصائية لمتوسط النتائج المنسوبة إلى ما بعد الاختبار، بين المجموعتين على التوالي 87.83 ± 27.44 و 84.31 ± 20.45 (فاصل الثقة ٪95 من الفرق بين هاتين الطريقتين: 15.41- ؛ 8.38). كشفت هذه الفترة الفاصلة ضمن فترة التكافؤ (20-، 20+) أن المعلم المبتدئ كان كفؤًا بشكل كاف لتدريس التقييم النقدي للمقال بالمقارنة مع معلم متمرس.

الاستنتاج: يمكن أن يساعد مبدأ التكافؤ في توفير التدريب على القراءة النقدية للمقال من قبل معلم مبتدئ مقارنة بمدرب متمرس متمرسين في تمهيد الطريق لمبادرات أخرى مماثلة بين جميع التخصصات.

الكلمات المفاتيح: المدارس الطبية ـ المدرسون ـ تجربة التكافؤ ـ صنع القرار ـ الطب المؤسس على القرائن ـ الجزائر

First, corresponding and submitting author:

Pr BEZZAOUCHA Abdeldjellil.

Professional affiliation: University Saad Dahlab of Blida. Country: Algeria

E-mail : bezzaoucha_djellil@yahoo.fr

Tel-fax: 213 25 209027

Postal address: BEZZAOUCHA Abdeldjellil

Head of the Department of Epidemiology

Hospital Frantz Fanon

Teaching hospital (CHU) of Blida

09200 Zabana

Blida – Algeria

INTRODUCTION

Teaching critical appraisal of medical scientific literature has been subjected to evaluation ever since the early 1990s, Audet N critically reviewed 10 reports of teaching critical appraisal published between 1980 and 1989 [1]. Another systematic review, in 2000, incorporated 8 of the previous reports and expanded on it by adding two articles published between 1992 and 1994 [2]. The authors reported on the poor methodological quality of most of these reports, therefore making it difficult to assess the effectiveness of teaching critical appraisal; only one out the 12 studies carried out [3] was a randomised trial. Later in 2003, the small proportion (2/17 = 11.8%) of the randomized trials was also highlighted by Coomarasamy A in a systematic review of postgraduate teaching in Evidence-Based Medicine (EBM) which is a broader concept including its principal component: critical appraisal [4]. In a similar review aimed at surgical residents, in 2012, Ahmadi N also reported the poor methodological quality and the small percentage of randomized trials [5]. Ahmadi SF, in 2014, found 9 randomised trials out of 27 studies included in a review that dealt with teaching EBM to undergraduate medical studies i.e. a much high percentage (9/27=33.3%). Out of the 9 trials, 5 had a good or acceptable quality. The majority of the other studies were of poor methodological quality [6]. The benefits of teaching EBM on patient care have also not been demonstrated [5, 7] even if other authors expect a benefit in this field [8].

Nobody, however, seems to question the relevance of teaching critical appraisal of medical scientific literature. Over the past two decades, EBM has become an essential component of the medical school curriculum despite the challenges facing this learning [9]. Critical appraisal of medical scientific literature has not yet been taught in Algeria, either at graduate level or post-graduate level.. The Algerian medical teachers' research output was very low over the past decade. One of the solutions that has been put forward to remedy this situation is the integration of critical appraisal of scientific literature in medical curricula at graduate and post-graduate levels [10]. But the challenge lies in training a critical mass of well-trained and specialized medical teachers, regardless of speciality, to undertake such teaching duties. No specialty can claim critical appraisal of literature for its own, but epidemiology could set an example. EBM was actually launched by epidemiologists from McMaster University (Canada) in the early 1990s [11].

Hypotheses tested by randomized trials in teaching EBM were varied (short courses, workshops or seminars, medical journal clubs, e-learning...). None of the published randomized trials assessed the teacher aptitude in this field. The need to pay attention to EBM teachers, as well as pedagogical interventions and outcomes, has been integrated into a conceptual framework proposed in the mid-2000s [12]. Moreover, the scarcity of EBM-trained teachers was a major obstacle for the development of this learning in low-middle-income countries [13].

We set out to demonstrate that a junior lecturer in epidemiology would be able achieve satisfactory outcomes, in teaching critical appraisal of literature to medical teachers, in comparison with a senior lecturer.

METHODS

Study design

The introductory course on critical appraisal of literature for teachers of the Medicine School of Blida – Algeria (which had around 300 clinical teachers) was animated by a junior teacher in epidemiology with less than one-year of teaching experience and a senior teacher in epidemiology with 35 years' experience who authored a book on critical appraisal of scientific literature [14]. The junior teacher was thus compared to the senior teacher through a single-center equivalence randomized trial design. Volunteering candidate teachers were randomized into an intervention group (junior teacher) or a control group (senior teacher). The two teachers were blinded to the group allocation.

A balanced draw technique has been used to ensure numerical equality of both groups among the total group of learners attending the pre-test [15]. A random number table was used to assign learners, who were absent in the pre-test, to each group. The course took place at Blida University Hospital under the aegis of Medicine School of Blida between February 28 and May 30, 2016. Learners from both groups filled a questionnaire four weeks following the start of the training to judge their comparability.

Organization and content of the course

The course was made of 10 weekly sessions of three hours each. Both groups took the course on the same day at the same hour in two adjoining classrooms. Group change was not permitted. Residents were in charge of keeping track of the attendance record of each learner. Learners were given the senior teacher's authored book and were asked to read, each week, a published article related to a specific study design in the following order: cross-sectional study, case-control study, case series, systematic review of literature, diagnostic accuracy study, diagnostic reliability study, randomized trial, prognostic study, cohort study and descriptive study. Questions and suggested answers to the questions were in the book [14]. The raised questions along with their subsequent

answers were meant to enable learners, at the end of the session, to fill the reading grid for each article and form an opinion about the quality of the study, based on the set criteria developed by Salmi LR [16]. The criteria were mainly related to the clear formulation of the objective, the study design befitting the objective, patients' inclusion and exclusion criteria, required sample size, withdrawals, reliability and validity of key variables as well as results analysis.

A pre-test was administered to learners one week before the start of the course (February 28, 2016) and the post-test was administered to teachers one week after the end of the course (May 30, 2016). The pre-test and the post-test were also an article to be criticized. During the tests, teachers had access to any document they wanted and to the Internet.

At the beginning of the course, the pre-test correction was the opportunity to introduce for all participants, during a three-hour session, the main epidemiological and statistical tools to read an article: structure of a scientific medical article, main statistical tests and epidemiological association measures. At the request of voluntary learners of both groups, these tools and other developments (confidence intervals, sample sizes, power, analysis of tables with three variables to detect confusion and / or an interaction, survival analysis) were the subject of supplemental five sessions. These sessions were taught exclusively by the junior teacher for two hours per session; examples to illustrate these developments were contained in the senior teacher's handbook.

No manual calculation for these tools should be carried out by learners. All calculations for the exposed tools were performed by two free software packages: OpenEpi (Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version. www. OpenEpi.com, updated 2014/09/22) and BiostaTgv (https://marne.u707.jussieu.fr/biostatgv/) that teachers had to master gradually throughout the course. The pre-test has also been an opportunity to emphasize to learners that the post-test would be administered according to the same modalities used for the pre-test.

Pre-test and post-test performance measurement

Open questions, essentially based on the proposed articles, were asked in the pre-test and post-test by adapting them to the twelve components of the Fresno test and to the corresponding notation proposed [17]. A total of 23 questions were taken from the pre-test article (a published prognostic study) for the first eight components of the Fresno test. An independent exercise of the pre-test article with four questions was used to score the

ninth and tenth component of the Fresno test while two short questions were used to score the eleventh and the twelfth component of the same test. For the post-test, 26 questions were taken from the post-test article (a published randomized trial) for the first seven components of the Fresno test. A first independent exercise with two questions was dedicated to the eighth component of the test, while a second independent exercise concerned the ninth and tenth component of the test. Two other short questions were used to grade the eleventh and twelfth component of the Fresno test. The pre-test and the post-test were scored out of 200 points.

Learners had three hours to read the article of the tests and try to answer the asked questions. Articles and asked questions were only accessible to learners at the beginning of the tests. The pre-test was anonymously corrected by the senior teacher. Both the senior and the junior teachers scored anonymously the post-test. The analysis of pre-test and post-test performance used the scores attributed by the senior teacher.

Sample size and statistical analysis

The sample size was determined to show, at α-risk of 2.5% with a power of 90%, that the mean post-test score obtained by trainees after one course on critical appraisal provided by a junior teacher was of the same order of magnitude as that obtained with a senior teacher with an equivalence interval of 20 points out of 200 and a standard deviation of the score of 20 points. The sample size required per group had to be 26 each, a total sample size of 52 learners (the unit of analysis was learners). In other words, if there was truly no difference between the mean scores obtained by the two teachers, then 52 learners were required to be 90% sure that the limits of a two-sided 95% confidence interval of the mean scores difference would not be outside the interval [-20; +20] (Sealed Envelope Ltd. 2012. Power calculator for continuous outcome equivalence trial. [Online] Available from: https://www.sealedenvelope.com/power/continuous-equivalence/).

Comparison of the two groups was carried out, at α-risk of 5%, using the Chi-square test or Fisher's exact test when the response variable was qualitative. When the response variable was quantitative, the Kolmogorov-Smirnov test was systematically used to assess the normality of the distribution. Both means were either then compared by the Student's t-test or by the Mann-Whitney U-test. The Mantel-Haenszel χ^2 was used to test the relationship between two variables taking into account a third variable. Furthermore, multiple linear regression analysis was performed: the dependent variable was the overall score obtained in the post-test and the independent variables

were the variables for which the two groups differed at p ≤ 20 %. The agreement between the scores attributed to the post-test by the two teachers was assessed by the intra-class correlation coefficient. Factorial analysis of variance allowed identifying the sources of variability on the post-test score. Cronbach's alpha coefficient was used to assess internal consistency of scores Fresno's post-test in all its components. All corresponding calculations were made with the SPSS software in its 17th version. Equivalence test between the two mean scores was performed by the Minitab software in its 17th version (https://www.minitab.com). This test consisted of performing a two-sample equivalence test (two one – sided test), which gave two p-values. Equivalence could be claimed when the greater of the two p-values was ≤ 0.025.

Ethical approval and written consent
The trial was approved by the ethics committee of the medical school of Blida. Teachers participating in the trial were informed of the formation of the two groups by random drawing. They expressed their written consent by signing the questionnaire that has been administered to them.

RESULTS

Learners' recruitment process
Figure 1 illustrates the diagram flow that has marked the recruitment of candidates' teachers for the training course of critical appraisal. The analysis of the results concerned only the 66 learners who took the post-test, divided into two groups of 33 learners each. No subject has switched between groups during the experiment.

Comparison of the two learner groups at baseline
Table 1 provides a comparison of the two groups of learners recruited with regard to the variables from the questionnaire that the 66 teachers completed and signed. The two groups did not differ with regard to any of the variables except for the incorporation into clinical practice of the results of an original article, the ability to read a whole article in English, and the number of scientific articles read per term. This number was not in fact related to the group of learners when the number of hours per week engaged in reading articles was neutralized (p = 0.11). Similarly, the incorporation into practice of the results of an original article was no longer linked to the group when the number of scientific articles read per term was taken into consideration (p = 0.14). But the ability to read a whole article in English narrowly remained linked to the group when this number of articles was taken into account (p = 0.05).
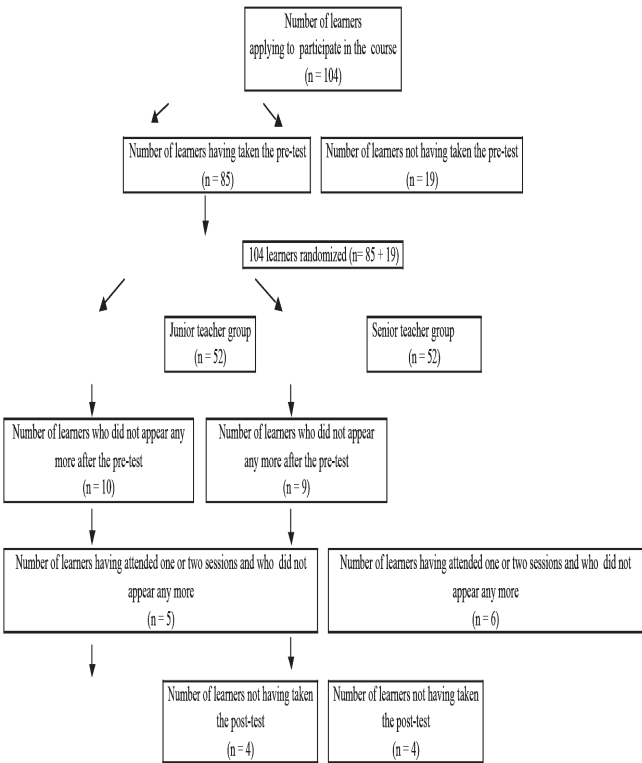
Figure 1: Diagram flow of the learner recruitment Medicine School of Blida (Algeria) – February to May 2016

The learners who requested the additional training were distributed without significant difference between the groups (69.7% for the junior teacher and 54.5% for the senior teacher, p = 0.20). Information pertaining to the assertion of eight learners that claimed to be authors or co-authors of one or more articles indexed in Medline, was checked by consulting this bibliographic database. Three names did not appear in the database, one in the "junior teacher" group and two in the "senior teacher" group. Two other names in this second group were authors or co-authors of an abstract for a conference and a symposium and not of an original article.

Internal consistency of the post-test and agreement between the scores attributed by the two teachers
Cronbach's alpha coefficient of the Fresno test components for the post-test was 0.73 indicating acceptable internal consistency. The agreement between the scores attributed to the post-test by the two teachers was significant for the overall scores and scores for each component of the Fresno test (Table 2). Factorial analysis of variance

Table 1 : Baseline characteristics of the learning groups at baseline Medicine School of Blida (Algeria) – February to May 2016

Characteristic	Junior teacher (n = 33)		Senior teacher (n = 33)		p
	Number	%	Number	%	
Female gender	21	**63,6**	27	**81,8**	0.10[*]
Projection at Blida University Hospital	26	**78 ,8**	23	**69 ,7**	0.40[*]
Medical-surgical speciality	25	**75,8**	20	**60,6**	0.19[*]
Senior lecturer rank	28	**93,3**	28	**93,3**	0.66[†]
Leadership position in department	13	**39,4**	14	**42,4**	0.80[*]
Defended or registered doctoral dissertation	10	**30.3**	13	**39.4**	0.44[*]
Member of registered research project	4	**12.1**	8	**24.2**	0.20[*]
Author or co-author of articles indexed in Medline	2	**6.1**	6	**18.8**	0.20[†]
Foreign collaboration in thesis, project or article	3	**9.1**	5	**15,2**	0.71[*]
Collaboration with a biostatistician and/or an epidemiologist in thesis, project or article	11	**33.3**	15	**45.5**	0.31[a]
Prior training in biostatistics and/or epidemiology outside medical curriculum	9	**27.3**	10	**30.3**	0.79[*]
Duration of prior training in biostatistics and/or in epidemiology over 30 days	3	**33.3**	6	**60.0**	0.37[*]
Prior training in critical reading of publications and/or in medical writing outside medical curriculum	9	**27.3**	14	**42.3**	0.20[*]
Duration of prior training in critical reading of publications and /or in medical writing over 8 days	9	**100**	11	**78.6**	0.25[*]
Number of scientific articles (≥ 3) read in average per term	16	**48.5**	24	**72.7**	0.04[*]
Number of hours (≥ 3) engaged in average per week in reading articles	14	**42.4**	17	**51.5**	0.46[*]
Critical reading practice in department meetings	4	**12.1**	7	**21.3**	0.32[*]
Incorporate into clinical practice of scientific results from original publication	14	**42.4**	23	**69.7**	0.05[†]
Sufficient English proficiency to read a whole article	12	**36.3**	23	**69.7**	0.02[*]
Sufficient English proficiency to write a whole article	3	**9.1**	4	**12.1**	1.00[†]
Personal commitment at the end of this course to develop critical reading in department	28	**84.8**	29	**87.9**	1.00[†]
Took pre-test	26	**78.8**	29	**87.9**	0.32[*]
Total pre-test score(out of 200) : (n) m ± s	(26) 24.26 ± 10.65		(29) 23.00 ± 8.06		0.91[¶]
Average number of absences over the course: (n) m ± s	(33) 3.73 ± 2.00		(33) 3.52 ± 2.83		0.72[‡]
Age (years) : (n) m ± s	(33) 41.55 ± 7.20		(33) 42.27 ± 7.86		0.70[‡]
Average number of supplemental instruction sessions: (n) m ± s	(23) 2.26 ± 1.38		(18) 2.56 ± 0.85		0.43[‡]
Seniority as a teacher since senior lecturer (years) : (n) m ± s	(33) 5.76 ± 5.25		(33) 7.18 ± 6.83		0.34[‡]

[*] Chi-square test   [†] Fisher test   [‡] Student t test   [¶] Mann-Whitney U test

allowed highlighting a "learner" factor (p <0.001), but not a "teacher" factor (p=0.40) in the attribution of the overall post-test score.

Comparison of the two groups with regard to attributed scores

Multiple linear regression analysis highlighted that none of the variables presented in Table 1 was related to the overall post-test score. The number of learners in the "junior teacher" group who scored 100 points or more on the post-test was 11, or 33.3%. This number was 10 for learners in the "senior teacher", or 30.3%. The difference between these two proportions was not significant (p = 0.79). The proportion of learners with a score equal to or higher than 80 points was identical in both groups (60.6%). No significant difference was found between the mean scores of the two learner groups within the different components of the Fresno test (pre-test and post-test) as well as for the overall scores, except for the twelfth section of the Fresno's pre-test scored out of 4 (Table 3).

When adjusting for baseline scores to measure the effectiveness of the training (for 55 learners who attended pre-test and post-test) by determining the difference of scores obtained by learners in pre-test and post-test (per learner difference between the scores), the average difference for junior group (n=26; 64.17 ± 21.03) was not significantly different from that of senior group (n = 29, 59.86 ± 19.30): t = 0.79; p = 0.43; 95% confidence interval

of the difference between these two means: -6.60; 15.22. As regards the equivalence test between teachers for the two overall average scores (equal variances were assumed for the analysis), p value for the first null hypothesis (difference ≤ 20) was less than 0.001. For the second null hypothesis (difference ≥ 20), p value was 0.004. Claiming equivalence between the two teachers became possible. The method using the 95% (100% - 2 * 2.5%) confidence interval around the difference of average scores naturally led to the same conclusion since this interval (-15.41; 8.38) was within the equivalence interval of (-20, +20).

Table 2: Intra-class correlation coefficient between the scores attributed to the post-test by the two teachers according to the different components of Fresno test Medicine School of Blida (Algeria) – February to May 2016

Area of knowledge tested	Intra-class correlation coefficient	95% confidence interval	p
1. Formulate a question	0.65	0.48 ; 0.77	$< 10^{-3}$
2. Sources of information	0.82	0.72 ; 0.88	$<10^{-3}$
3. Searching strategies	0.76	0.64 ; 0.85	$<10^{-3}$
4. Study design	0.79	0.67 ; 0.86	$<10^{-3}$
5. Relevance	0.71	0.56 ; 0.81	$<10^{-3}$
6. Internal validity	0.81	0.71 ; 0.88	$<10^{-3}$
7. Magnitude of effect	0.91	0.85 ; 0.94	$<10^{-3}$
8. Statistical values of diagnostic study*	0.69	0.54 ; 0.80	$<10^{-3}$
9. Impact assessment†	0.82	0.73 ; 0.89	$<10^{-3}$
10. Confidence interval	0.38	0.12 ; 0.57	$<10^{-2}$
11. Best study design, diagnostic	0.82	0.72 ; 0.88	$<10^{-3}$
12. Best study design, prognostic	0.84	0.91 ; 0.96	$<10^{-3}$
Overall score	0.94	0.91 ; 0.97	$<10^{-3}$

* Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio
† Risk difference (or absolute risk reduction), relative risk reduction (or etiologic fraction in exposed), number needed to treat

Table 3 : Average scores obtained by learners of both groups in pre-test and post-test according to the different components of the Fresno test, Medicine School of Blida (Algeria) – February to May 2016

Area of knowledge tested	Score max	Pre-test				Post-test			
		Junior teacher (n = 26) m ± s	Senior teacher (n = 29) m ± s	95% CI*	p	Junior teacher (n = 33) m ± s	Senior teacher (n = 33) m ± s	IC 95%*	p
1. Formulate a question	12	4 .71 ± 2.0	4.51 ± 1.71	-1.21 ; 0.82	0.91†	7.95 ± 1.71	7 .68 ± 1.68	- 1.10 ; 0.56	0.47†
2. Sources of information	24	5.9 ± 2.53	6.17 ± 1.92	-0.94 ; 1.48	0.66	8.36 ± 4.64	8.15 ± 4.54	-2.47 ; 2.04	0.84†
3. Searching strategies	24	0.73 ± 2.18	0.68 ± 1.25	- 0.99 ; 0.90	0.93	7.10 ± 5.19	5.27 ± 4.47	-4.21 ; 0.55	0.13
4. Study design	24	1.94 ± 2.69	2.08 ± 2.31	-1.21 ; 1.50	0.83	9.62 ± 5.39	8.75 ± 6.16	-3.71 ; 1.98	0.45†
5. Relevance	24	8.82 ± 4.11	8.39 ± 4.27	-2.70 ; 1.84	0.70	13.71 ± 3.4	14 .25 ± 2.4	-0.91 ; 2.00	0.45
6. Internal validity	24	0.23 ± 0.71	0.17 ± 0.60	-0.41 ; 0.29	0.74	8.60 ± 5.1	8.78 ± 4 .1	- 2.10 ; 2.47	0.87
7. Magnitude of effect	24	0. 65 ± 0.79	0.65 ± 0.72	-0 .40 ; 0.41	0.99	8.30 ± 8.10	6.21 ± 5.23	- 5.44 ; 1.26	0.22
8. Statistical values of diagnostic study	20	0 ± 0.0	0 ± 0.0	-	-	9.25 ± 3.54	8.96 ± 4.41	- 2.25 ; 1.68	0.77
9. Impact assessment	12	0.30 ± 1.22	0 ± 0.0	-0.76 ; 0.14	0.18	6.80 ± 3.42	6.72 ± 3.36	-1.74 ; 1.59	0.92
10. Confidence interval	4	0.38 ± 1.13	0.20 ± 0.81	-0.70 ; 0.35	0.50	1.86 ± 1.49	2.27 ± 1.44	-0.31 ; 1.13	0.26
11. Best study design, diagnostic	4	0.30 ± 1.08	0.79 ± 2.32	-0.51 ; 1.48	0.33	3.15 ± 1 .66	3.63 ± 1.16	-0.22 ; 1.19	0.17
12. Best study design, prognostic	4	0.61 ± 1.41	0 ± 0.0	-1.16 ; 0.60	0.02	3.15 ± 1.66	3.33 ± 1.47	-0 .58 ; 0.95	0.63
Overall score	200	24.26 ± 10.65	23.00 ± 8.06	-6 .34 ; 3.81	0.91†	87.83 ± 27.44	84 .31 ±20.45	-15.41 ; 8.38	0.56†

* Confidence interval around the difference of average scores   † Mann- Whitney test (score is not normally distributed)

DISCUSSION

Main findings

The study revealed that both senior and junior teachers in epidemiology performed equally well in teaching critical appraisal of literature. The proper timing for integrating EBM instruction across the spectrum of medical education has been addressed [18] but it does not appear to contest the benefit received by learners who are medical teachers. Regardless of the country's level of development, these teachers must not only write medical articles but also keep strictly to the international standards of writing [19]. It is true, however, that doctors have a problem with the figures and statistic interpretation, three-quarters of Connecticut medical residents, for example, acknowledged that they did not understand all of the statistics they encountered in journal articles [20]. In any case, if a junior teacher has been able to do as well as a senior teacher in teaching critical appraisal, it may reasonably be thought that other teachers of the same specialty and even others specialties could be capable of the same performance .

Explanation of the study's findings

Randomization to assign learners to the two teachers and the efforts made to maintain this comparability until the end of the trial (no change affecting the groups) were not the only explanation of the study's findings. This achievement could potentially become possible because the same pedagogical support, developed by the senior teacher [14], was used. This eliminated the source of variability due to the learning instrument. The use of the same pedagogical support, whatever it is, helps to standardize learning objectives despite nuances introduced from one teacher to another. Moreover, if a "learner" factor was naturally highlighted (probably explained by a different motivation and receptivity of the learners), there was no "teacher" factor influencing the scores obtained.

As regards performance measurement, skills required for critical appraisal in medicine were drafted even before the 2000s [21]. The Fresno test, based on the student's ability to solve a clinical problem as represented in a hypothetical scenario, is now widely used and is considered a valid instrument for measuring EBM competence [8, 22, 23]. It has been necessary to adapt this test according to the articles (of the pre-test and the post-test) submitted to the learners, providing material from a pre-selected article seemed compatible with such an assessment approach. The adaptation of the Fresno test, which can lead to different versions, is a widespread practice [7, 8, 23-26].

Comparisons with other studies

To our knowledge, no study published has compared teachers when it comes to teaching EBM, which necessarily limits the possibilities for comparisons. However, EBM teaching has gained much appeal and is constantly subject of publications. In a cross-sectional survey, three-quarters of Canadian residents in gynecology-obstetrics had little/ no confidence interpreting research statistics despite the training provided during their course. Nevertheless, confidence was raised significantly when associated with increased seniority, prior publications and prior epidemiology/statistics course [27]. These characteristics were obviously taken into account in our questionnaire to judge the comparability of both groups. Gender was also taken into account, with women identified as having a lower publication rate [28].

Post-graduate residents in Algeria, regardless of their specialty, as well as graduate students, did not benefit from any structured training in epidemiology/statistics during their course. When some of these residents became teachers, they had the same weaknesses explaining the very low scores of the pre-test. Only one-third of our learners had the average (overall score ≥ 100) in the post-test at the end of the training. The mean scores obtained at the pre-test were comparable to the mean score of a pre-test obtained by Jordanian graduation students (26.7 ± 16.1 out of 200) but they obtained a better mean score in the post-test (119.5 ± 28.5) [29]. But average difference between pre-test and post-test scores was less than 50 (out of 212 points) among residents in Catalonia [24], whereas it was about 60 (out of 200 points) for each of our two groups. This average difference was only 21 (out of 156 points) in both studies conducted by Mc Cluskey [7, 23]. It is true that average scores above 100 (out of 212 points), even before the pedagogical intervention, and an average difference of 15 points, after the intervention, were also observed [8]. Difficulty level or clarity of asked questions, previous training received could be at the origin of the observed disparities.

The purpose of the EBM is not in the assignment of scores but in the acquisition of knowledge that might influence practices when faced with a patient or a population (better care or better preventive measures). The need to shift EBM teaching from classrooms to clinical practice has already been emphasized since the early 2000s [4]. Practicing EBM at the point of care was also identified as an important skill to acquire for four-fifths of internal medicine residents in the US [30].

Even in developing countries, critical appraisal of scientific publications is becoming increasingly important. More than three-quarters of graduate medicine students at a

Saudi university thought that EBM was important to their career [31]. The subject is so worthwhile that factors associated with success in critical reading exam among medical students of a school of medicine in France have been identified [32]. It was also comforting that 61% of our learners, with no significant difference between the two groups, were for the introduction of critical appraisal in examination for access to clinical residency, 85% committed themselves to writing at least one scientific article and 80% were for valuing original articles in medical teachers' promotion evaluation grids. Algeria is still one of the countries which are lagging behind considerably in the area of medical written research output [10]. To overcome the countless challenges related to the battle of production and visibility, all specialties must be involved in a huge effort to promote the critical reading of medical scientific publications.

Strengths and limitations of the study

Such a study was of interest to educators interested in approaches to the challenges of teaching critical appraisal to a largely clinically oriented faculty group in a setting largely naïve to such content. The learners, medical lecturers, were distributed randomly between both groups trained by a senior or junior lecturer. Attributing the results of performances to the degree of teaching experience became legitimate assuming the junior teacher was motivated and he was gifted from the standpoint of basic teaching skills. Calculation of the size of the two groups

of learners is based on a value of the score's standard deviation of 20 points (scale of scores from 0 to 200). This value was of the same order of magnitude, or ever less than, as that found by authors using the Fresno test in before-after studies [7, 8, 24, 29]. Mean differences well above 20 were often showed in these studies. The choice of an equivalence interval of +/- 20 points, i.e. +/- one point on a scale of scores from 0 to 10 which would not matter in practice, seemed an acceptable option for admitting equivalence. However, the monocentric nature of the study (Medical School of Blida – Algeria) and comparing two teachers of the same specialty could be a limiting factor on extrapolability of the results. The pupil-teacher effect could also lead to apprehensions regarding the teachers' choice and the occurrence of a conflict of interest.

CONCLUSION

The topic of critical appraisal and how to improve it is of broad interest. Performing equally well in providing training on critical appraisal of medical publications by either senior or junior teachers is a legitimate and relevant subject, particularly in developing countries characterized by scarcity of EBM-trained teachers. The carried out study demonstrated this equivalence between one junior lecturer and one senior lecturer. This holds out strong hope for the future to form a critical mass of young teachers, of all specialties, who can teach critical appraisal of medical scientific literature.

REFERENCES

1. Audet N, Gagnon R, Ladouceur R, Marcil M. [How effective is the teaching of critical analysis of scientific publications? Review of studies and their methodological quality]. CMAJ 1993;148:945-52.
2. Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. A systematic review of the effectiveness of critical appraisal skills training for clinicians. Med Educ 2000;34:120-5.
3. Linzer M, Brown JT, Frazier LM, DeLong ER, Siegel WC. Impact of a medical journal club on house-staff reading habits, knowledge, and critical appraisal skills. A randomized control trial. JAMA 1988;260:2537-41.
4. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. BMJ 2004;329:1017.
5. Ahmadi N, McKenzie ME, Maclean A, Brown CJ, Mastracci T, McLeod RS. Teaching evidence based medicine to surgery residents-is journal club the best format? A systematic review of the literature. J Surg Educ 2012;69:91-100.
6. Ahmadi SF, Baradaran HR, Ahmadi E. Effectiveness of teaching evidence-based medicine to undergraduate medical students: a BEME systematic review. Med Teach 2015;37:21-30.
7. McCluskey A, Lovarini M. Providing education on evidence-based practice improved knowledge but did not change behaviour: a before and after study. BMC Med Educ 2005;5:40.
8. Mohr NM, Stoltze AJ, Harland KK, Van Heukelom JN, Hogrefe CP, Ahmed A. An evidence-based medicine curriculum implemented in journal club improves resident performance on the Fresno test. J Emerg Med 2015;48:222-9 e1.
9. Maggio LA, ten Cate O, Chen HC, Irby DM, O'Brien BC. Challenges to Learning Evidence-Based Medicine and Educational Approaches to Meet These Challenges: A Qualitative Study of Selected EBM Curricula in U.S. and Canadian Medical Schools. Acad Med 2016;91:101-6.
10. Bezzaoucha A, Atif ML, Bouamra A, El Kebboub A, Benzerga M, Ben Abdelaziz A, et al. Algerian medical teachers' research output and its determinants during the 2000-2009 decade. Rev Epidemiol Sante Publique 2014;62:33-40.
11. Sheridan DJ, Julian DG. Achievements and Limitations of Evidence-Based Medicine. J Am Coll Cardiol 2016;68:204-13.

12. Straus SE, Green ML, Bell DS, Badgett R, Davis D, Gerrity M, et al. Evaluating the teaching of evidence based medicine: conceptual framework. BMJ 2004;329:1029-32.

13. Kulier R, Gulmezoglu AM, Zamora J, Plana MN, Carroli G, Cecatti JG, et al. Effectiveness of a clinically integrated e-learning course in evidence-based medicine for reproductive health training: a randomized trial. JAMA 2012;308:2218-25.

14. Bezzaoucha A. Lecture critique d'articles médicaux. Algiers - Algeria: Office des Publications Universitaires; 2016.

15. Schwartz D, Flamant R, Lellouch J. l'essai thérapeutique chez l'homme. 2ème édition ed. Paris: Flammarion médecine Sciences; 1981.

16. Salmi LR. Lecture critique et communication médicale scientifique. Comment lire, présenter, rédiger et publier une étude clinique ou épidémiologique. 3e édition. ed. Issy-les-Moulineaux (France): Elsevier Masson; 2012.

17. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. BMJ 2003;326:319-21.

18. Holloway R, Nesbit K, Bordley D, Noyes K. Teaching and evaluating first and second year medical students' practice of evidence-based medicine. Med Educ 2004;38:868-78.

19. Adeloye DO. Key challenges of evidence-based medicine in developing countries. Res J of Health Sci 2015;3:10.

20. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA 2007;298:1010-22.

21. Audet n LH. Les habiletés requises pour la lecture critique en médecine : un cadre de référence issu d'une recension des écrits. Pédagogie Médicale 2001;2:7.

22. Ilic D. Assessing competency in Evidence Based Practice: strengths and limitations of current tools in practice. BMC Med Educ 2009;9:53.

23. McCluskey A, Bishop B. The Adapted Fresno Test of competence in evidence-based practice. J Contin Educ Health Prof 2009;29:119-26.

24. Argimon-Pallas JM, Flores-Mateo G, Jimenez-Villa J, Pujol-Ribera E. Effectiveness of a short-course in improving knowledge and skills on evidence-based practice. BMC Fam Pract 2011;12:64.

25. Argimon-Pallas JM, Flores-Mateo G, Jimenez-Villa J, Pujol-Ribera E, Foz G, Bundo-Vidiella M, et al. Study protocol of psychometric properties of the Spanish translation of a competence test in evidence based practice: the Fresno test. BMC Health Serv Res 2009;9:37.

26. Lizarondo L, Grimmer K, Kumar S. The Adapted Fresno test for speech pathologists, social workers, and dieticians/nutritionists: validation and reliability testing. J Multidiscip Healthc 2014;7:129-35.

27. Bougie O, Posner G, Black AY. Critical Appraisal Skills Among Canadian Obstetrics and Gynaecology Residents: How Do They Fare? J Obstet Gynaecol Can 2015;37:639-47.

28. Raj A, Carr PL, Kaplan SE, Terrin N, Breeze JL, Freund KM. Longitudinal Analysis of Gender Differences in Academic Productivity Among Medical Faculty Across 24 Medical Schools in the United States. Acad Med 2016;91:1074-9.

29. Barghouti FF, Yassein NA, Jaber RM, Khader NJ, Al Shokhaibi S, Almohtaseb A, et al. Short course in evidence-based medicine improves knowledge and skills of undergraduate medical students: a before-and-after study. Teach Learn Med 2013;25:191-4.

30. Pereira AG, Harrell HE, Weissman A, Smith CD, Dupras D, Kane GC. Important Skills for Internship and the Fourth-Year Medical School Courses to Acquire Them: A National Survey of Internal Medicine Residents. Acad Med 2016;91:821-6.

31. Al-Faris EA, Abdulghani HM, Al-Rowais NA, Alamro N. Teaching Evidence-Based Medicine in a Saudi Medical School : A Pilot Study. J T U Med Sc 2007;2:42-9.

32. Ladner J, Tavolacci MP, Joly LM, Lequerré T, Benhamou Y, Savoye-Collet C, et al. Caractéristiques et facteurs prédictifs de la réussite des étudiants a l'épreuve de lecture critique d'article dans le deuxième cycle des études médicales. UFR de médecine de Rouen (France), 2012-2013. Presse Med 2016; 45:142-4.