


RESEARCH ARTICLE

Open Access



Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners

Nicolas Rousselot^{1,5*} , Thomas Tombrey¹, Drissa Zongo^{2,3}, Evelyne Mouillet^{2,3}, Jean-Philippe Joseph^{1,3}, Bernard Gay¹ and Louis Rachid Salmi^{2,3,4}

Abstract

Background: There is currently an absence of valid and relevant instruments to evaluate how Evidence-based Practice (EBP) training improves, beyond knowledge, physicians' skills. Our aim was to develop and test a tool to assess physicians' EBP skills.

Methods: The tool we developed includes four parts to assess the necessary skills for applying EBP steps: clinical question formulation; literature search; critical appraisal of literature; synthesis and decision making. We evaluated content and face validity, then tested applicability of the tool and whether external observers could reliably use it to assess acquired skills. We estimated Kappa coefficients to measure concordance between raters.

Results: Twelve general practice (GP) residents and eleven GP teachers from the University of Bordeaux, France, were asked to: formulate four clinical questions (diagnostic, prognosis, treatment, and aetiology) from a proposed clinical vignette, find articles or guidelines to answer four relevant provided questions, analyse an original article answering one of these questions, synthesize knowledge from provided synopses, and decide about the four clinical questions. Concordance between two external raters was excellent for their assessment of participants' appraisal of the significance of article results ($K = 0.83$), and good for assessment of the formulation of a diagnostic question ($K = 0.76$), PubMed/Medline ($K = 0.71$) or guideline ($K = 0.67$) search, and of appraisal of methodological validity of articles ($K = 0.68$).

Conclusions: Our tool allows an in-depth analysis of EBP skills, thus could supplement existing instruments focused on knowledge or specific EBP step. The actual usefulness of such tools to improve care and population health remains to be evaluated.

Keywords: Evidence-based practice, Critical appraisal, Medical education, Kappa reliability, General practice, Skills

Background

Evidence-based Practice (EBP) is the integration of best research evidence, clinical expertise, and patient values, in a specific care context [1]. This way of practicing medicine developed in the 1980's and has subsequently been integrated worldwide within new teaching approaches, centred on problem-based learning. EBP teaching was introduced in many initial and continuing

medical education curricula to improve health care by better integrating relevant information from the scientific literature [2–14].

EBP has been described as having five steps [15, 16]: 1) Formulate a clear clinical question about a patient's problem; 2) Search the literature, with an appropriate strategy, for relevant articles [17]; 3) Critically appraise the evidence for its validity, clinical relevance and applicability; 4) Implement the useful findings back into clinical practice [18]; and 5) Evaluate the impact. This approach is particularly useful in general practice (GP) to manage primary care situations, where it has been described as the sound simultaneous use of a

* Correspondence: nicolas.rousselot@u-bordeaux.fr

¹Department of General Practice, University of Bordeaux, F-33000 Bordeaux, France

⁵Département de Médecine Générale, Université de Bordeaux, Case 148, 146 rue Léo Saignat, 33076 Bordeaux cedex, France

Full list of author information is available at the end of the article



critical research-based approach and a person-centred approach [19, 20].

Whilst many potential advantages have been suggested [16, 21], some criticisms have also been made [22]. A serious drawback is that it has not been clearly shown that EBP can improve physician skills or patient health [23–25]. Very few randomized clinical trials have documented the effect of EBP, with these trials frequently including non-comparable groups. Further, these trials were often based on subjective judgements, due to the lack of reliable and valid tools to assess EBP skills [13, 14, 25–28].

Indeed, some tools have been proposed, but are not easily accessible or validated [14, 28–32]. Most existing tools focus on assessing knowledge, rather than skills, particularly for the literature search [21, 33]; they do not assess skills for each step of EBP [34], but rather focus on article critical assessment [30, 31, 33, 35, 36], sometimes without any relation to a clinical situation [35].

Our aim was to develop a tool to assess the skills necessary for the first four steps of the EBP process, and to evaluate whether independent raters could reliably use the tool to assess acquired skills.

Methods

To assess EBP skills, we developed a comprehensive tool, including a test of skills and a scoring grid, based on literature and expert advice. We tested the applicability of the test and evaluated whether independent observers could reliably use the scoring tool to analyse answers to the test to assess acquired skills (Fig. 1). Our validity approach was based on a classical model of clinical evaluation of tool validity [37], which provides a strategy to develop and evaluate the performance of tests. This conceptualisation is similar to the “validity as a test characteristic” described in the health professions education literature [38]. This approach is shared in a large part of the French GP teachers who are also clinicians.

Tool development

Literature sources

Our tool was developed based on syntheses of the medical literature on EBP, published in the *Journal of the American Medical Association* [2, 13, 17, 18, 30, 39–42], and in the *British Medical Journal* [3, 23, 33, 43]. We also considered previous published tools’ strengths and limitations [29, 33, 34, 36].

Expert input on content and purpose of tool

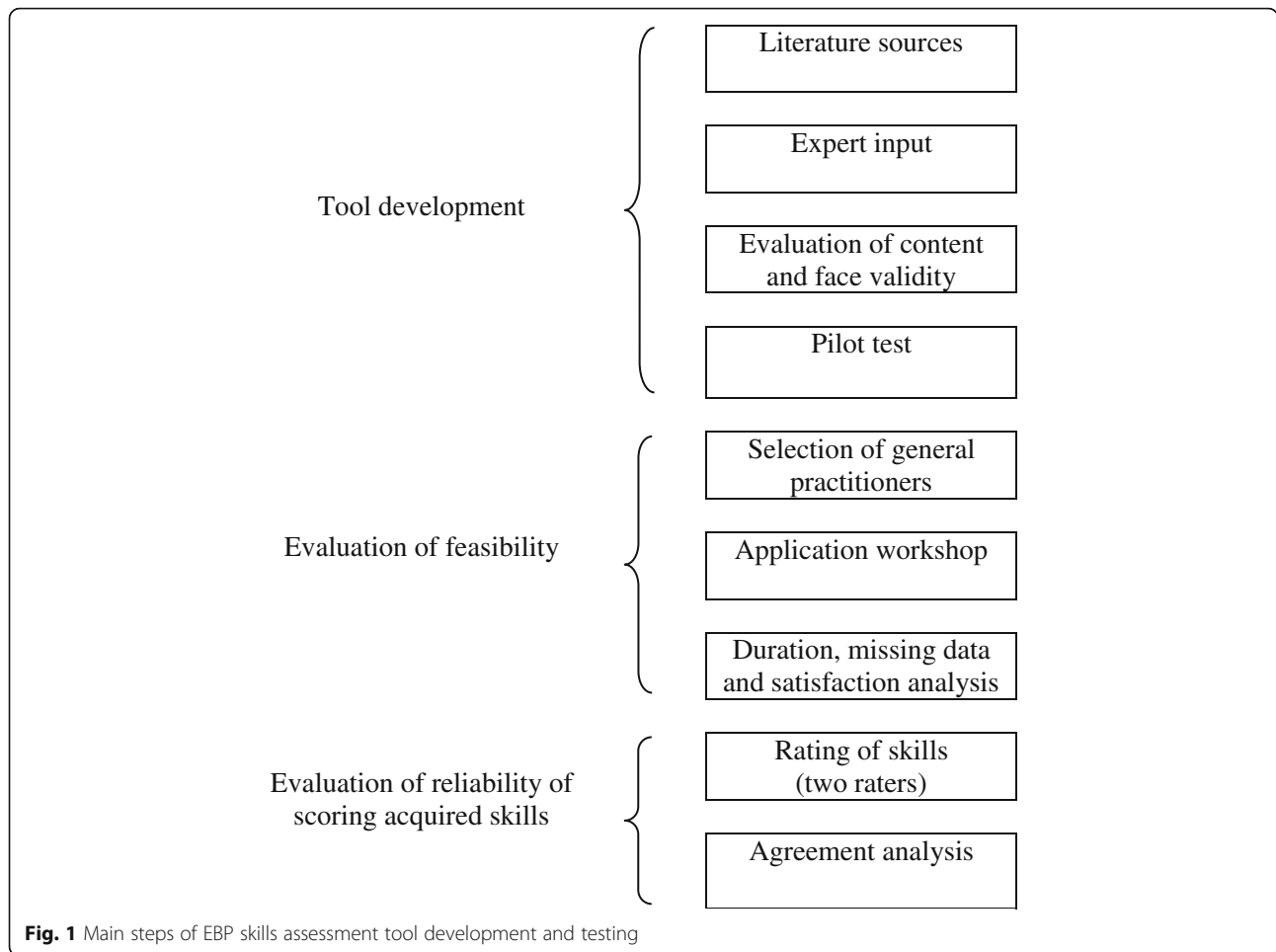
Three of the authors supervised tool development: a senior general practitioner (BG), a senior epidemiologist (LRS), both with recognised experience in EBP teaching in both initial and continuing medical education, and an

experienced senior librarian (EM) with experience in teaching literature search for health professionals.

Whereas previous tools mostly assessed knowledge [44], our aim was to assess skills, defined as the participant using knowledge by actually carrying out EBP steps about a clinical scenario [14, 28]. To assess participants’ skills, we asked them to perform tasks associated with the different EBP steps [14], with open but precise instructions, rather than only asking them how they would undertake those tasks. Then, we observed their ability to actually complete these tasks.

We assessed all first four steps of EBP independently, thus allowing participants to undertake all tasks, even if they were wrong in one of the earlier steps. This also allowed participants to receive feedback regarding their results as part of a formative assessment for each step. Our test was also built as a continuum from problems described in a clinical situation to decisions made to deal with these problems. Physician daily constraints (computer and Internet access, time... [45–47]) were also considered when designing the test.

Our tool was divided into four parts to assess necessary skills for each of the first four steps of EBP (Table 1): A clinical vignette (Table 2), on a common and complex situation likely to be seen in primary care, was used to assess the ability to formulate a clear clinical question about a patient’s problem. We asked participants to formulate four clinical questions on diagnostic, prognosis, aetiology, and treatment. The scoring grid for that part was inspired by the first question of the Fresno test [33] and assessed whether the formulated question respected the PICO (Population, Intervention, Comparison, Outcomes) criteria [48]. To assess the ability to search the literature for relevant documents related to the previous clinical questions, we asked participants to find the full text of an original article or guideline for each question. Scoring of this ability was based on recording the participants’ computers screenshots, using the Wink Screen Recording Software 2.0 (available at <http://www.debug-mode.com/wink/>), which registered one screenshot every three seconds during the test. The scoring grid was adapted from a published tool [34] to assess literature search strategies. To assess critical appraisal skills, we selected four English-language full-text original articles, covering each one of the four search questions (diagnostic, prognosis, aetiology, and treatment). Each participant was to appraise the validity of methods, relevance for care, and significance of results of only one of these articles. The scoring grid was based on previous works [1] and specific criteria to appraise the quality of articles on diagnostic [39], prognosis [40], treatment or prevention [41], and harm [42]. To assess the ability to synthesize and decide about a specific clinical situation, we developed four synopses reporting the critical appraisal of the four



articles responding to each of the initial clinical questions. The scoring grid assessed clarity of the decision, and elements used to justify the decision, including consideration of the clinical context and a question on the degree to which the participant trusted study results (Additional file 1).

Content and face validity

To improve our tool adequacy for its purpose, as part of the “content and face validity” step [37], we asked a panel of experts from the CNGE (French National College of Teachers in General Practice) for a critical review. We asked them to judge the relevance of

Table 1 Main characteristics of the EBP skills assessment tool used for each participant during the test

Test part	EBP step	Task	Support used	Skills: performance assessment ^a
First	Formulate a clinical question	Build 4 search questions to answer a clinical problem	1 case vignette	How complete and relevant are the GPs’ PICO questions?
Second	Search relevant clinical articles	Find 4 relevant articles in medical literature (with different strategies)	4 bibliographic retrieval questions	How thoroughly and efficiently do GP conduct searches?
Third	Critically appraise literature	Appraise validity, relevance and results significance of an article	1 original article	Can GP complete critical appraisals?
Fourth	Implement useful findings in clinical practice	Answer 4 clinical questions	4 synopses (of 4 original articles)	Can GP come to a reasonable interpretation of how to apply the evidence?

GP General practitioners ; ^aaccording to Tilson et al. [14]

Table 2 Summary of the case vignette

A 75-years-old man visits his general practitioner. In his medical history: an ischemic stroke 2 years before, atrial fibrillation, smoking, hypertension, and hypercholesterolemia. He was worried by a risk of epilepsy because of his stroke; asked if his use of coffee was excessive; asked to refill his prescription (with no anticoagulant but aspirin); and complained about a calf pain (without any deep vein thrombosis sign).

included items, whether any item was missing, and the format of the tool. Their comments were considered in a pre-test version of the assessment tool and the scoring grid.

Pilot test

We tested the assessment tool with a senior GP teacher of the Department of General Practice of Bordeaux and a volunteer second year GP resident, to evaluate its technical applicability and their understanding of instructions. The scoring grids were adapted and filled in once, jointly by two GP raters (TT, DZ), to formalize and homogenize the scoring procedure.

Evaluation of feasibility

We documented [28, 37]: acceptability of the tool as reflected by participation, number of undocumented items, and satisfaction of participants, time required to complete the test, time required to rate the test; for undocumented items, we tried to judge whether this was related to comprehension or technical problems, for instance failure of the Internet connection.

Selection of participants

Participants to a full test were GP residents in internship with general practitioners near Bordeaux, and GP teachers from the Department of General Medicine of Bordeaux. All had a general practice activity and were contacted by phone. Verbal informed consent was obtained from all participants.

Application workshop

The test was conducted in computer rooms of the University of Bordeaux, during a three-hour session. Each participant was provided with a computer and Internet access. Once the participants had carried out one part of the process, they sent their output by E-mail to the organizer (TT) and then received instructions for the next part. The first part was expected to last 20 min, i.e. 5 min to formulate each of the four clinical questions. The second part was one-hour long, i.e. 15 min to search one document. Each participant had to find four documents: two original articles using PubMed/MEDLINE, one document using research tools to specifically identify guidelines, and one document using a free search on the Web. The order in which participants were to find the different types of documents was

randomly allocated, so that three faculty and three residents were searching in the same order. The third part was 45-min long. Each participant had to analyse one of four articles. Here again the article was randomly allocated so that each type of article was analysed by three faculty and three residents. The last part was 40-min long, i.e. 10 min to analyse each of the four synopses and write the decision.

Duration, missing data and satisfaction analysis

The duration of tests and scoring was measured and missing or ambiguous data analysed. An anonymous satisfaction questionnaire (Additional file 2) was filled in by participants at the end of the test. After the test, participants received a synopsis of what was expected from them.

Evaluation of reliability of scoring acquired skills

Rating of acquired skills

Two of the authors (TT, DZ) independently corrected all anonymized tests, filling the scoring grids. They judged, on a four-level Likert scale the conformity of output to what was expected to reflect a given skill (for example, completely conform to expected PICO; rather conform; rather not conform; completely not conform). They separately scored: each of the four clinical questions; each of the three search strategies; appraisal of the methodological validity, relevance for care, and significance of results; each of the four decisions (Table 3, Additional file 1).

Agreement analysis

Analyses were done from data where neither the participant nor the rater was identified, with the SAS statistical software package, version 9.0 (SAS Institute Inc.). A linear weighted Kappa coefficient and its 95% confidence interval (CI) was calculated for each Likert scale to measure concordance between the two assessments [37]. Kappa was considered excellent if higher than 0.8, good if between 0.6 and 0.8, medium if between 0.4 and 0.6, and low if under 0.4 [49]. The main analysis considered missing data as completely not conform. A second analysis excluded missing data. An analysis of the sources of discrepancies between the two raters was done collegially, with the two raters and a senior epidemiologist (LRS).

Results

Feasibility

Selection of participants

Of the 28 general practice residents who were contacted, 12 agreed to participate. Of the 85 GP teachers of the Department of General Practice of Bordeaux, 46 could be contacted by phone, and 14 agreed to participate;

Table 3 Results of Likert scales for each assessed task of the EBP steps

Step	NC		Completely conform		Rather conform		Rather not conform		Completely not conform	
	n		n	%	n	%	n	%	n	%
Formulating a focused question										
Diagnostic	1		0	0.0	3	13.0	9	39.1	10	43.5
Prognosis	1		2	8.7	7	30.4	5	21.7	8	34.8
Etiologic	1		1	4.3	15	65.2	3	13.0	3	13.0
Therapeutic	1		0	0.0	2	8.7	15	65.2	5	21.7
Best information search										
PubMed/MEDLINE	8		0	0.0	0	0.0	4	17.4	11	47.8
Guidelines	17		0	0.0	3	13.0	3	13.0	0	0.0
Free search (Web)	4		4	17.4	4	17.4	7	30.4	3	13.0
Critical appraisal										
Methodological validity	3		1	4.3	2	8.7	9	39.1	8	34.8
Relevance for patient care	3		0	0.0	1	4.3	12	52.2	7	30.4
Significance of results	3		0	0.0	0	0.0	4	17.4	16	69.6
Synthesis and decision										
Diagnostic article	1		1	4.3	8	34.8	12	51.2	1	4.3
Prognostic article	1		13	56.5	4	17.4	5	21.7	0	0.0
Etiologic article	1		12	51.2	5	21.7	5	21.7	0	0.0
Therapeutic article	2		5	21.7	10	43.5	3	13.0	3	13.0

n = number of participants, NC= not completed (missing data)

three withdrew after initially agreeing, including one who cancelled three days before the workshop and could not be replaced. Eventually, 12 GP second-year residents, two men and 10 women, and 11 GP teachers, 10 men and one woman, participated. The GP teachers were one associate professor, three assistant professors and seven part-time instructors; they were aged 53 years on average.

Test and scoring duration

The workshop followed all steps as planned. The average response time was 171 min for teachers and 185 min for residents. There was a difference in the last part of the workshop (33 min for teachers and 44 min for residents), and the set time was exceeded for the third part of the test (53 min for teachers and 56 min for residents). The scoring lasted on average 44 min by test for the first rater (total: 17 h), and 30 min by test for the second rater (total: 11 h 50 min).

Missing data

Data on the test was missing in 14.6% of the Likert scales, 16.9% for teachers and 12.5% for residents (Table 3). Most missing data was for the second part of the test: four of the 23 participants' computer screenshot files were lost (3 for teachers), possibly due to handling errors by participants. Such errors were also seen once

in the first part, three times in the third part, and once in the last part. Instructions were not followed for bibliographic retrieval for 17 of the 69 Likert scales scored: 11 for residents; four were for PubMed/MEDLINE and 13 for guideline searches.

Satisfaction

Satisfaction questionnaires were filled by 22 participants. All participants were satisfied: they found the experience interesting (100%), relevant (82%), useful for clinical practice (100%), but difficult (97%). They expressed that the workshop underscored the need for training (91%) and the tool assessed well participant familiarity with EBP (91%) and could be used to assess progress with training (86%). Only 46% reported using EBP in their usual practice with the main reasons for not using it being: lack of time (94%), poor understanding of English (59%) and lack of skills to use necessary tools (71%).

Reliability of acquired skills scoring

Agreement analysis

Concordance between the two raters was excellent for their assessment of participants' appraisal of the significance of article results (Table 4). It was good for the formulation of a diagnostic question, PubMed/Medline or guideline search, and for methodological validity appraisal. It was lower for all other aspects.

Table 4 Concordance between the two raters' Likert scale for each question of the EBP steps

Step	Agreement		Weighted Kappa (K)		Weighted Kappa excluding missing data	
	n	%	K	95% CI	K	95% CI
Formulating a focused question						
Diagnostic	19	82.6	0.76	0.53–0.99	0.75	0.51–0.99
Prognosis	13	56.5	0.58	0.36–0.81	0.56	0.33–0.79
Etiologic	13	56.5	0.40	0.07–0.72	0.34	–0.00–0.68
Therapeutic	13	56.5	0.32	–0.02–0.65	0.27	–0.08–0.61
Best information search						
PubMed/MEDLINE	21	86.7	0.75	0.42–1.00	0.71	0.34–1.00
Guidelines	22	83.3	0.93	0.79–1.00	0.67	0.10–1.00
Free search (Web)	13	47.4	0.58	0.31–0.85	0.39	0.10–0.67
Critical appraisal						
Methodological validity	18	78.3	0.68	0.40–0.95	0.72	0.47–0.97
Relevance for patient care	17	73.9	0.59	0.32–0.86	0.53	0.23–0.83
Significance of results	22	95.7	0.83	0.51–1.00	0.83	0.50–1.00
Synthesis and decision						
Diagnostic article	11	47.8	0.21	–0.01–0.44	0.23	–0.04–0.50
Prognostic article	9	39.1	0.45	0.24–0.65	0.39	0.20–0.59
Etiologic article	9	39.1	0.27	0.00–0.53	0.26	0.07–0.45
Therapeutic article	12	52.2	0.44	0.19–0.70	0.37	0.11–0.63

n = number of participants with agreement between raters, *CI* = confidence interval

The main sources of discrepancy were: differences in appreciation of PICO criteria (the difference between an “incomplete” and “not conform” response depending on response precision, which was not assessed equally by the two raters); raters' entry errors and irrelevant response not scored as “not conform”; errors and omissions in filling scoring grid; discrepancies in assessment of articles and website quality for free research; differences in appreciation of decision making and synthesis, depending on rater's harshness and expectation for decisions to be explained. In case of disagreement between raters, we chose to keep the most favourable assessment for this last question only.

Discussion

We developed the first French-language tool to assess EBP skills of general practitioners. Concordance between raters was excellent for assessment of the participants' appraisal of the significance of article results. It was good for the formulation of a diagnostic question, PubMed/MEDLINE and guideline searches, and for article methodological validity appraisal. It was lower for all other aspects.

Our tool covers all relevant skills, as the main four steps of the EBP process are assessed. In that regard, it completes existing tools, such as the Fresno test [33] and the Berlin questionnaire [36], as both only include the first three steps, and focus mostly on critical

appraisal [14]. The only published validated test assessing those four steps is the ACE tool [21]. Our tool is again complementary, as the ACE tool assesses more knowledge than skills, using simple true-false questions, whereas our tool includes observation of actual searches and critical appraisals. This more focused assessment of knowledge rather than skills is also a limitation of the Fresno test, which mostly covers literature search and critical appraisal, and of the Berlin Questionnaire.

We assessed physicians' skills with open-ended questions, asking for the completion of specific tasks; for instance, our observation was innovative with the recording of screenshots, and assessed them with objective items. These features make our tool and its application closer to and more relevant for clinical practice. It has been developed using various kinds of complex questions relating to real-life situations, which, to our knowledge, has not been done before; we believe it could be transposed to many complex clinical situations.

We still have to improve parts of the tool before it can be proposed to the EBP teaching community. Concordance between raters was low, notably for the last part of the test related to synthesis and decision making. More precise scoring grids and a better application of assessment items are needed to reduce raters' subjectivity when assessing skills. This was also sometimes seen for the first part of the test, regarding formulation of a search question. This first part, based on the

Fresno test for which good inter-rater reliability has been documented [33], was composed of questions on short and simple case vignette. This part of the Fresno test had a low variability of possible responses, whereas our test was closer to practice.

Another potential limitation of our test is the time needed for its completion; three hours, much longer than the ACE tool and Berlin Questionnaire (15–20 min), and Fresno test (one-hour long) [21, 33, 36]. Simplifying our tool might shorten this completion time, but is likely to reduce its relevance for practice. Moreover, time devoted to each part (5 min to build a search question, 15 min to find an original article, 45–60 min to analyse it, and 10 min to synthesize and decide) is a realistic reflection of what can be done in practice.

Two possible reasons for the low level of reliability of some items of our tool are the low level of skills, and the variation in the harshness of raters. Another hypothesis is that the tool is not a valid reflection of the actual skills. Indeed, a tool well-perceived by users (the so-called “face validity”), of which the content has been agreed by experts (content validity) and which showed acceptable reliability, might still not adequately measure what it is supposed to measure [37, 50]. Therefore, we still need studies of the construct or criterion validity of our tool. However, the latter is difficult to assess, as there is no gold standard for all EBP skills. A gold standard could be developed through expert judgement based on formal consensus methods [51].

As our tool yields 14 independent scores, it is well suited to identify which of the skills a student or a physician should focus his future training on (formative assessment). However, we still need to develop a way to provide profiles for the four main skills and a judgment of an individual's overall EBP skills, as a way to compare participants and evaluate our tool's validity. Other perspectives to further develop our test and evaluate its performance should take into consideration limitations of our study: small number of testers, precluding the use of other analytical techniques to evaluate reliability such as log linear models.

As our work was initiated by the GP Department of the University, we selected participants with a practical experience in GP. Indeed, we wanted to assess the ability to use EBP skills to improve patients care in a GP setting. Moreover, the use of the same clinical scenario throughout the whole assessment process is an indirect way to evaluate the potential impact of acquired skills in clinical practice. We also selected GP residents and teachers to get a heterogeneous sample, as recommended to evaluate reliability [52]. Nevertheless, we believe, by looking at the responses,

that all residents were probably not EBP fledglings and all GP teachers, given their age, were not EBP experts, as already shown elsewhere [45]. This generation contrast, the small number of participants and raters [53], and the focus on a population linked with the University probably limit the generalizability of our results.

Conclusions

Our tool is relevant for practice as it allows an in-depth analysis of EBP skills. It could respond to a real need to better assess EBP skills of general practitioners. It can also be seen as usefully complementing existing tools, but further validation, including comparison with the latter, is needed. The actual usefulness of such tools to improve care and population health remains to be evaluated.

Additional files

Additional file 1: Two parts of the tool to assess EBP skills: 1) Content of the skill assessment form; and 2) Scoring grid. This file gives more information about our tool. (DOCX 61 kb)

Additional file 2: Satisfaction questionnaire. This file presents the satisfaction questionnaire filled in by participants at the end of the test. (DOCX 16 kb)

Abbreviations

EBP: Evidence-based practice; GP: General practice

Acknowledgements

We thank Mrs. Wendy R. McGovern for reviewing the manuscript.

Funding

This study was funded by the College of Aquitaine general practitioners/teachers. The sponsor had no influence on the study design, the collection, analysis or interpretation of data, on the writing of the manuscript or on the decision to submit it for publication.

Availability of data and materials

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

Authors' contributions

NR Literature review, analysis and interpretation of data; drafting of the article. TT Conception and design, acquisition of data, analysis and interpretation of data. DZ Acquisition of data. EM Conception and design. JPJ, BG and LRS Conception and design, interpretation of data. All authors revised the article and approved the version to be published.

Authors' information

NR: MD, MSc Assistant professor (Department of General Practice, Bordeaux University). TT: MD. DZ: MD. EM: Librarian (ISPED/INSERM U-1219), Instructor (Bordeaux University). JPJ: MD, Professor. Assistant director at the Department of General Practice (Bordeaux University). BG: MD, Professor. Head of the DGP (Bordeaux). LRS: MD, PhD, Professor. PU-PH. Head of ISPED.

Ethics approval and consent to participate

This study was approved by the University of Bordeaux. This study did not need formal ethics approval. This complies with French national guidelines (reference: Article L1121-1 du Code de la santé publique). Verbal informed consent was obtained from all participants. Written consent was unnecessary according to French national regulations (reference: Article L1121-1 du Code de la santé publique).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of General Practice, University of Bordeaux, F-33000 Bordeaux, France. ²ISPED/Bordeaux School of Public Health, University of Bordeaux, F-33000 Bordeaux, France. ³Centre INSERM U-1219 Bordeaux Population Health, F-33000 Bordeaux, France. ⁴CHU de Bordeaux, Pole de sante publique, Service d'information médicale, F-33000 Bordeaux, France. ⁵Département de Médecine Générale, Université de Bordeaux, Case 148. 146 rue Léo Saignat, 33076 Bordeaux cedex, France.

Received: 29 May 2018 Accepted: 31 October 2018

Published online: 09 November 2018

References

- Sackett DL, Straus S, Richardson WS, Rosenberg WM, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.
- Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420–5.
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71–2.
- Haute Autorité de Santé. Épreuves Classantes Nationales (ECN) - Sommaire et Mode d'emploi. 2018. https://www.has-sante.fr/portail/jcms/c_646948/fr/epreuves-classantes-nationales-ecn-sommaire-et-mode-d-emploi. Accessed 15 Aug 2018.
- Ministère de l'éducation nationale de l'enseignement supérieur et de la recherche. Arrêté du 21 avril 2017 relatif aux connaissances, aux compétences et aux maquettes de formation des diplômés d'études spécialisées et fixant la liste de ces diplômés et des options et formations spécialisées transversales du troisième cycle des études de médecine. *Journal Officiel de la République Française*. n°0100 du 28 avril 2017. <https://www.legifrance.gouv.fr/eli/arrete/2017/4/21/MENS1712264A/jo>. Accessed 15 Aug 2018.
- Royal College of general practitioners. GP curriculum: overview. 2016. <http://www.rcgp.org.uk/GP-training-and-exams/GP-curriculum-overview.aspx>. Accessed 15 Aug 2018.
- The Royal Australian College of General practitioners. The RACGP Curriculum for Australian General Practice 2016. 2016. <http://curriculum.racgp.org.au>. Accessed 15 Aug 2018.
- Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Family Medicine. 2018. <https://www.acgme.org/Specialties/Program-Requirements-and-FAQs-and-Applications/pfcattid/8/Family%20Medicine>. Accessed 15 Aug 2018.
- Meats E, Heneghan C, Crilly M, Glasziou P. Evidence-based medicine teaching in UK medical schools. *Med Teach*. 2009;31(4):332–7.
- The College of Family Physicians of Canada. Evaluation objectives. Defining competence for the purposes of certification by the College of Family Physicians of Canada: the evaluation objectives in family medicine. 2010. <https://www.cfpc.ca/EvaluationObjectives/>. Accessed 15 Aug 2018.
- Coppus SFPJ, Empananza JI, Hadley J, Kulier R, Weinbrenner S, Arvanitis TN, et al. A clinically integrated curriculum in evidence-based medicine for just-in-time learning through on-the-job training: the EU-EBM project. *BMC Med Educ*. 2007;7:46.
- Thangaratnam S, Barnfield G, Weinbrenner S, Meyerrose B, Arvanitis TN, Horvath AR, et al. Teaching trainers to incorporate evidence-based medicine (EBM) teaching in clinical practice: the EU-EBM project. *BMC Med Educ*. 2009;9:59.
- Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA*. 2002;288(9):1110–2.
- Tilson JK, Kaplan SL, Harris JL, Hutchinson A, Ilic D, Niederman R, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Educ*. 2011;11:78.
- Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayan K, et al. Sicily statement on evidence-based practice. *BMC Med Educ*. 2005;5(1):1.
- Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ*. 1995;310(6987):1122–6.
- Hunt DL, Jaeschke R, McKibbon KA. Users' guides to the medical literature: XXI. Using electronic health information resources in evidence-based practice. Evidence-Based Medicine Working Group. *JAMA*. 2000;283(14):1875–9.
- Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, et al. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. Evidence-Based Medicine Working Group. *JAMA*. 2000;284(10):1290–6.
- WONCA Europe. The European definition of General Practice/Family Medicine. 2011. <http://www.woncaeurope.org/gp-definitions>. Accessed 15 Aug 2018.
- Galbraith K, Ward A, Heneghan C. A real-world approach to evidence-based medicine in general practice: a competency framework derived from a systematic review and Delphi process. *BMC Med Educ*. 2017;17(1):78.
- Ilic D, Nordin RB, Glasziou P, Tilson JK, Villanueva E. Development and validation of the ACE tool: assessing medical trainees' competency in evidence-based medicine. *BMC Med Educ*. 2014;14:114.
- Straus SE, McAlister FA. Evidence-based medicine: a commentary on common criticisms. *CMAJ*. 2000;163(7):837–41.
- Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence-based medicine changes anything? A systematic review. *BMJ*. 2004;329(7473):1017.
- Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev*. 2001;(3):CD001270.
- Horsley T, Hyde C, Santesso N, Parkes J, Milne R, Stewart R. Teaching critical appraisal skills in healthcare settings. *Cochrane Database Syst Rev*. 2011;(11):CD001270.
- Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med*. 1999;74(6):686–94.
- Ilic D, Maloney S. Methods of teaching medical trainees evidence-based medicine: a systematic review. *Med Educ*. 2014;48(2):124–35.
- Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a systematic review. *BMC Health Serv Res*. 2007;7:119.
- Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*. 2006;296(9):1116–27.
- 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(17):2537–41.
- Taylor R, Reeves B, Mears R, Keast J, Binns S, Ewings P, et al. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. *Med Educ*. 2001;35(6):544–7.
- Ilic D. Assessing competency in evidence based practice: strengths and limitations of current tools in practice. *BMC Med Educ*. 2009;9:53.
- Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence-based medicine. *BMJ*. 2003;326(7384):319–21.
- Rana GK, Bradley DR, Hamstra SJ, Ross PT, Schumacher RE, Frohna JG, et al. A validated search assessment tool: assessing practice-based learning and improvement in a residency program. *J Med Libr Assoc*. 2011;99(1):77–81.
- MacRae HM, Regehr G, Brenneman F, McKenzie M, McLeod RS. Assessment of critical appraisal skills. *Am J Surg*. 2004;187(1):120–3.
- Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, Kunz R. Do short courses in evidence-based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence-based medicine. *BMJ*. 2002;325(7376):1338–41.
- Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. 5th ed. Oxford: Oxford University Press; 2015.
- St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Adv Health Sci Educ Theory Pract*. 2017;22(4):853–67.
- Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1994;271(5):389–91.

40. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. *JAMA*. 1994;272(3):234–7.
41. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1993;270(21):2598–601.
42. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA*. 1994;271(20):1615–9.
43. Guyatt GH, Meade MO, Jaeschke RZ, Cook DJ, Haynes RB. Practitioners of evidence-based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ*. 2000;320(7240):954–5.
44. Malick SM, Hadley J, Davis J, Khan KS. Is evidence-based medicine teaching and learning directed at improving practice? *J R Soc Med*. 2010;103(6):231–8.
45. Te Pas E, van Dijk N, Bartelink MEL, Wieringa-De Waard M. Factors influencing the EBM behaviour of GP trainers: a mixed method study. *Med Teach*. 2013;35(3):e990–7.
46. Zwolsman SE, van Dijk N, Te Pas E, Wieringa-de Waard M. Barriers to the use of evidence-based medicine: knowledge and skills, attitude, and external factors. *Perspect Med Educ*. 2013;2(1):4–13.
47. van Dijk N, Hooft L, Wieringa-de Waard M. What are the barriers to residents' practicing evidence-based medicine? A systematic review. *Acad Med*. 2010;85(7):1163–70.
48. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12–3.
49. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. 3rd ed. Hoboken: Wiley; 2003.
50. Downing SM. Face validity of assessments: faith-based interpretations or evidence-based science? *Med Educ*. 2006;40(1):7–8.
51. Bourrée F, Michel P, Salmi LR. Consensus methods: review of original methods and their main alternatives used in public health. *Rev Epidemiol Sante Publique*. 2008;56(6):415–23.
52. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64(1):96–106.
53. Sadatsafavi M, Najafzadeh M, Lynd L, Marra C. Reliability studies of diagnostic tests are not using enough observers for robust estimation of interobserver agreement: a simulation study. *J Clin Epidemiol*. 2008;61(7):722–7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

