



A novel assessment of an evidence-based practice course using an authentic assignment

Valerie Dory, Robert Gagnon, Therese De Foy, Corentin Duyver & Sophie Leconte

To cite this article: Valerie Dory, Robert Gagnon, Therese De Foy, Corentin Duyver & Sophie Leconte (2010) A novel assessment of an evidence-based practice course using an authentic assignment, Medical Teacher, 32:2, e65-e70, DOI: [10.3109/01421590903199718](https://doi.org/10.3109/01421590903199718)

To link to this article: <https://doi.org/10.3109/01421590903199718>



Published online: 17 Feb 2010.



Submit your article to this journal [↗](#)



Article views: 954



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

WEB PAPER

A novel assessment of an evidence-based practice course using an authentic assignment

VALERIE DORY¹, ROBERT GAGNON², THERESE DE FOY¹, CORENTIN DUYVER¹ & SOPHIE LECONTE¹

¹Universite catholique de Louvain, Belgium, ²Universite de Montreal, Canada

Abstract

Background: Evidence-based practice (EBP) is now a component of most medical curricula. Summative assessment instruments are often of debatable quality, do not cover the full spectrum of EBP or lack authenticity.

Aim: To develop and evaluate the quality of an authentic assessment instrument for use in summative assessment of general practice trainees.

Methods: An assignment was designed based on the ask, acquire, appraise and apply steps of EBP. Content validity was evaluated by external EBP experts. Concurrent validity was tested with the Fresno test. Inter-rater agreement and internal consistency were measured. Acceptability and feasibility were also assessed.

Results: EBP experts agreed that the instrument had good content validity. Concurrent validity was good (disattenuated intraclass correlation coefficient 0.75). Inter-rater agreement varied from 0.70 to 0.83. Internal consistency was high (Cronbach's alpha 0.70–0.86). The procedure was feasible but only moderately acceptable to students.

Conclusion: Our authentic assignment provided a valid, reliable and feasible procedure to assess our students. Acceptability was moderate, probably due to teething problems in instructions given and unfamiliarity with the format. Consequential validity data are lacking and would be of value. Our instrument could be an interesting alternative to other validated tests that may be less authentic.

Introduction

Teaching evidence-based practice (EBP) is an important part of any modern medical curriculum. Several reviews on the effectiveness of EBP teaching have commented on the generally poor methodological quality of studies (Audet et al. 1993; Norman & Shannon 1998; Green 1999; Taylor et al. 2000; Coomarasamy & Khan 2004; Flores-Mateo & Argimon 2007). One of the issues raised concerns the validity and reliability of outcome measurements (Green 1999; Taylor et al. 2000; Coomarasamy & Khan 2004; Flores-Mateo & Argimon 2007).

At Universite catholique de Louvain, Belgium, the master of medicine curriculum is divided into a mostly theoretical component for a duration of 2½ years followed by a year of core clerkships and a final semester of elective clerkships in the chosen vocational track. While an EBP course is currently being phased into the master's core curriculum, the general practice vocational track has included an EBP course for its final year medical students for several years now. Up until now, assessment of the course was included in the objective structured clinical examination (OSCE) assessment of the semester through two OSCE stations. Assessors have found this assessment strategy somewhat superficial, probably due to the time constraint (7.5 min), leading to a tendency to prompt students. They felt that some students were obtaining satisfactory marks despite an obviously weak grasp of concepts. This led to the development of a new assessment strategy.

Practice points

- Many instruments used for summative assessment of EBP are of debatable quality.
- Several validated instruments have been developed but either do not test all the steps of EBP or lack authenticity.
- An authentic assignment marked by at least two raters using a grid is a valid, reliable and feasible summative assessment procedure for students following an EBP course in the general practice track.
- Consequential validity and reproducibility data are lacking.

A review specifically regarding EBP assessment instruments was conducted by Shaneyfelt et al. (2006). They analysed instruments according to the context of use, i.e. individual formative/summative assessment, curriculum evaluation and behaviour evaluation. Few of the reviewed instruments evaluate the whole spectrum of EBP from asking to applying. Few instruments have documented validity and reliability. Those that validly and reliably evaluate all the EBP steps for individual formative or summative assessment are written tests, using a multiple-choice (Fritsche et al. 2002) and/or question open-ended format (Ramoset al. 2003; Weberschock et al. 2005). The limitation of these tests in our view was that they

Correspondence: Valerie Dory, Centre Academique de Medecine Generale, Universite Catholique de Louvain, Avenue Emmanuel Mounier 53 (boite 5360), 1200 Brussels, Belgium. Tel: 32 2 7645347; fax: 32 2 7645327; email: valerie.dory@uclouvain.be

impose clinical scenarios and data, thus lacking task authenticity in selection of a clinical query and actual searching. We hoped to stimulate students to use EBP in their future practice by demonstrating how EBP could help them solve clinical problems encountered during their clerkship. We, therefore, chose to use an authentic assignment as the basis for course assessment and we sought to evaluate the quality of our assessment.

Methods

Course objectives and basis for assessment

The objectives of the course are in line with a practical view of EBP as based on the integration of best available medical evidence, clinical expertise and patient preferences (Straus et al. 2005). We seek to demonstrate that EBP can be integrated into daily practice, and that it should be patient-driven, i.e. that the process should usually be triggered by a query stemming from a particular encounter with a patient. The objectives are, therefore, that students learn to 1) transform a question arising from practice into an answerable and searchable clinical query (using the PICO technique), 2) search validated websites (e.g. PubMed, Cochrane database, validated guideline sites) and/or use validated search motors (i.e. SUMSearch, TRIP Database) to find relevant information, 3) critically appraise sources (with a focus on therapeutic articles, diagnostic articles and clinical guidelines), and 4) findings critically to the patient case (ask, acquire, appraise and apply). We also hope that students will develop appropriate attitudes including a critical mindset. The course uses small group learning (i.e. three groups of around 15–20 students). Three seminars are devoted to critical appraisal using grids adapted from those developed by the Department of Family Medicine of Université Laval (available from <http://cetp.fmed.ulaval.ca/cetp>). One plenary session is devoted to Internet searching in the library computer room. The final three sessions are devoted to presentations of student assignments. The assignment was initially designed as a practical exercise where students worked in pairs, taking a clinical question arisen in their clerkship, writing a searchable question, searching the Internet and appraising one selected source. Up until now, the assignment was not marked. Our objective was to maximize educational impact with an authentic assessment which would hopefully encourage these final year medical students to use EBP in residency and in their future practice as general practitioners (GPs). We, therefore, felt that the assignment we already used for formative purposes would be an ideal basis for summative assessment.

Assessment instrument

Students were instructed to work in pairs. The assignment was to select a clinical query, search the Internet for answers, retrieve one document and critically appraise it, and draw practical conclusions. The assignment was to be three pages long and to include: a justification of their choice of topic, a detailed description of the steps they took in searching the

literature with a justification of choices made, a detailed analysis of the chosen document and a conclusion. Article critiques (e.g. from Clinical Evidence) or articles which had been appraised by Minerva (a Belgian EBM journal) were not allowed for the purposes of the assignment. Following enquiries by students, further instructions were given that consensus reports were not allowed either. Students had 7 weeks to prepare their assignment and a slide-show presentation. Assignments were marked by three raters independently, with a maximum score of 20. Following presentations, students were given the opportunity to amend their initial assignment with a 2-point bonus at stake (results not reported).

Two of the four research assistants who were the course tutors developed an initial marking grid which was modified following discussions with the two other course tutors and the course coordinator, an experienced EBP expert.

The marking grid (Table 1) was designed to mark each step described in the instructions:

- (1) choice of topic;
- (2) literature search and document selection; and
- (3) critical appraisal including practical implications.

Relative weights were given, with critical appraisal accounting for half the points, literature searching for around a third (7/20) and the rest for choice of topic. Choice of topic included 3 items scored out of a maximum of 1. Literature searching included 7 items scored out of a maximum of 1. Critical appraisal included 12 items, 9 of which were scored out of a maximum of 1, 2 out of a maximum of 0.3 and 1 out of a maximum of 0.4. Items were designed by adapting the Laval critical appraisal grids. Relevance, for instance, was not included in the critical appraisal section but rather in the selection of topic section. Items on validity were designed in parallel form for different types of documents (i.e. article on an intervention, article on a diagnostic study, guideline, systematic review). The wording of items was intended to allow raters to evaluate the students' comprehension and application ability, avoiding simple present/absent criteria. Partial credit was allowed. However detailed scoring rubrics were not provided. The assignments were marked independently by three raters (the two course tutors who had developed the initial marking grid, VD and TDF, and the course coordinator, PC).

Evaluation of the instrument

Messick has defended a more unified conception of validity of the interpretation of data from measurement instruments. Downing (2003) has illustrated the different kinds of evidence required for validity assessment which generally include elements pertaining to 1) content, 2) response process, 3) internal structure, 4) relationship to other variables, and 5) consequences. We also examined the acceptability to students and the feasibility of the assessment procedure. The course was followed by a course evaluation session during which students were asked to fill in a questionnaire and to sit the Fresno test (see below concurrent validity). The questionnaire included 16 statements with answers on a 5-point Likert scale (+2 completely agree; +1 agree; -1 disagree; -2 completely

Table 1. The marking grid.

General criteria	Detailed criteria	Score	Max
1. Choice of topic	1.1. The student starts from his/her clinical experience		1
	1.2. The student argues that the problem is common and/or serious		1
	1.3. The student argues that potential management options exist		1
	2.1. The student searches one of these validated websites: <i>Primary sources via Medline, pubmed, Embase, Biomed Central, British Medical Journal, Lancet, New England Journal of Medicine, Journal of the American Medical Association, Family Practice, Canadian Medical Association Journal, British Journal of General Practice, Family Medicine</i> <i>Tertiary sources via: Cochrane, Minerva, Evidence-Based Medicine, Clinical Evidence</i> <i>Guidelines: National Guideline Clearinghouse, Societe Scientifique de Medecine Generale, Doris Medica, Haute Autorite de Sante, Canadian Medical Association, National Health Service Evidence, National Institute for Clinical Excellence, New Zealand Guidelines Group, Arztlisches Zentrum für Qualität in der Medizin, Agency for Healthcare Research and Quality, Guidelines International Network</i> <i>Search engines: Ovid, Trip Database, SUMSearch, Catalogue et Index des Sites Medicaux de langue Francaise</i> <i>Other: Institut National d'Assurance Maladie-Invalidite, Centre Belge d'Information Pharmacotherapieutique, Universite Laval</i>		1
2. Literature search	2.2. The student searches at least 2 validated websites or uses a validated search engine		1
	2.3. The student asks a good question (PICO)		1
	2.4. The student uses MeSH or thesaurus terms		1
	2.5. The student argues his choice of article according to the type of article (he/she identifies the type of article and gives valid reasons for his/her choice): <i>Primary sources: rare problem or recent developments</i> <i>Tertiary sources: analysis, summarising of data, putting data in perspective</i> <i>Quaternary sources: analysis, synthesis, recommendations for practice with levels of evidence, applicability</i>		1
	3.1 The student checks in the abstract that the article/guideline/systematic review answers their query		1
	3.2 The student check in the abstract that the article/guideline/systematic review is recent		1
	3.3 The student correctly analyses validity (using the following criteria)		5
	3.3.1. Study type	/1	1
	3.3.2. Groups are comparable	/1	1
	3.3.3. Double blinding	/1	1
3. Critical appraisal of the article	3.3.4. Relevant endpoints (hard, appropriately measured)	/1	1
	3.3.5. All randomised subjects taken into account	/1	1
	3.3.5.1. Lost to follow up	/0.5	
	3.3.5.2. Intention to treat analysis	/0.5	
	3.4. The student correctly puts results into perspective		
	3.4.1. The student correctly analyses statistical significance : p-value<0.05 and 95% CI	3.4.1. The student correctly analyses 3.4.1. The student correctly interprets levels of 3.4.1. The student correctly analyses the evidence and/or strength of recommendations summary of results	1
	3.4.2. The student correctly analyses clinical significance (ARR or NNT and NNH)	3.4.2. The student checks whether the student correctly analyses clinical significance (interpretation of values and guideline was validated by peers potential harms)	1
	3.4.3. The student correctly analyses applicability to his/her patients		
	3.4.3.1. Similar population		0.3
	3.4.3.2. Feasibility in his/her practice		0.4
	3.4.3.3. Acceptability to patients		0.3
	3.4.4. The student states that it is important to look at consistency with other studies/guidelines/systematic reviews		1
	3.4.5. The student raises potential conflicts of interest		1
	TOTAL SCORE		19

CI: confidence interval; ARR: absolute risk reduction; NNT: number needed to treat; NNH: number needed to harm; Se: sensitivity; sp: specificity; PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio

disagree; 0 do not know). This session was programmed just after the final session of the course and attendance was voluntary, although declared as highly desirable. Students were not informed of the session's exact content other than a broad statement that we wanted to evaluate the new version of the course.

Validity. 1) Content validity was ensured by blueprinting. It was verified by submitting the assignment instructions and marking grid together with course objectives to three EBP experts. They were asked to assess the relevance of each course objective (ask, acquire, appraise, apply, attitude), and the relevance and comprehensiveness of the items in relation to each. Students were asked to assess content validity in a global way by answering the question 'The assessment adequately reflects the course objectives'.

2) Regarding the response process, several elements were taken into account. Familiarity of students to assessment format was not measured but is likely to be poor as students are rarely assessed using assignments at our medical school. The appropriateness of weighting by course objective was assessed by the three EBP experts. In order to assess the accuracy of raters' scores, we measured inter-rater agreement using an intraclass correlation between the ratings of three independent raters. Differences between raters were also assessed using Bland–Altman plots.

3) Internal consistency was assessed using Cronbach's alpha. Reproducibility was not evaluated.

4) Concurrent validity was measured by comparing the assignment mark with results on the Fresno test. Assignments were usually given in by pairs of students, whilst the Fresno test is done individually. We assessed concurrent validity by using the Pearson product-moment correlation (observed and disattenuated) between individual scores on the assignment and on the Fresno test. The Fresno test was translated into French by VD and translation was reviewed by RG. VD and CD independently scored the test. The intraclass correlation between the two raters was 0.75 (95% CI [0.58; 0.86]). The average mark was 112.87 (SD 25.92, range 55–166) out of a possible 212.

5) Finally, consequential validity was not assessed.

Acceptability. Students were asked two questions pertaining to acceptability in the questionnaire: 'The assessment was a good measure of my learning in this course' and 'The workload required by the assignment was acceptable'.

Feasibility. Feasibility was assessed using time spent by tutors collecting assignments and marking them and in the student survey by the question 'I estimate that I spent ... hours on the original assignment'.

Results

Twenty assignments were returned by 42 (1 on her own, 32 in pairs, and 9 in groups of three) out of 49. Seven students failed to return their assignment on time and were failed. Marks were poor with an average of 11.35/20 (SD 3.56). Participation in the evaluation session was high with 43 of 49 students taking part (88%).

Table 2. Inter-rater agreement for assignment marking.

Raters	ICC (95% CI)
VD–TDF	0.83 [0.61; 0.93]
VD–PC	0.70 [0.38; 0.87]
TDF–PC	0.80 [0.56; 0.92]
All three raters	0.77 [0.59; 0.89]

ICC: intraclass correlation coefficient.

Validity. The three EBP experts rated the relevance of course objectives as all 'relevant' or 'very relevant'. Relevance of items was rated as good or very good by all EBP experts for four out of rating grid items. Experts disagreed on the relevance of items pertaining to the 'ask' objective. Comprehensiveness of items was rated as good or very good by all EBP experts for four out of five objectives. Experts disagreed on the comprehensiveness of the global score as an indicator of achievement of the attitudinal objective. Students' global assessment of content validity was positive, with 70% (95% CI [56%; 83%]) of responders agreeing with the statement 'The assessment adequately reflects the course objectives'.

The three EBP experts generally felt that weighting was appropriate except for the 'apply' objective which two experts felt should be given more weight in the total score. Intraclass correlations between pairs of raters ranged from 0.70 to 0.83, with an intraclass correlation coefficient (ICC) of 0.77 (95% CI [0.59; 0.89]) for triple marking (Table 2), differences were not statistically significant. Bland–Altman plots (Figure 1) indicate a few extreme differences of 7 points (on a mark out of 20). TDF generally gave lower scores. Differences between scores attributed by VD and TDF were more marked for low scores, between VD and PC for high scores, with no clear pattern between PC and TDF.

Cronbach's alpha was 0.85 for the mean scores of the three raters and ranged from 0.70 to 0.86 for scores given by individual raters.

Scores on the assignment were moderately correlated with scores on the Fresno test (observed $r=0.57$, $p<0.001$; disattenuated $r=0.75$). There were missing data for 11 students who had failed to return their assignment and/or not participated in the Fresno test.

Acceptability. A satisfactory 65% (95% CI [51%; 79%]) of responders agreed with the statement 'The assessment was a good measure of my learning in this course'. However, only 55% (95% CI [40%; 70%]) of responders agreed with the statement 'The workload required by the assignment was acceptable'.

Feasibility. Students declared having spent an average of 13 h (SD 7) on their assignment, which is consistent with course credit (two credits or 20 h, with 10.5 h devoted to teaching sessions). It took two tutors 2 h to collect and check assignments. Two tutors kept track of their marking times which were markedly different with an average of 10 (SD 3) minutes per assignment for VD and 30 (SD 9) for TDF. Total marking time for the two assessors who kept track of time spent was 13 h.

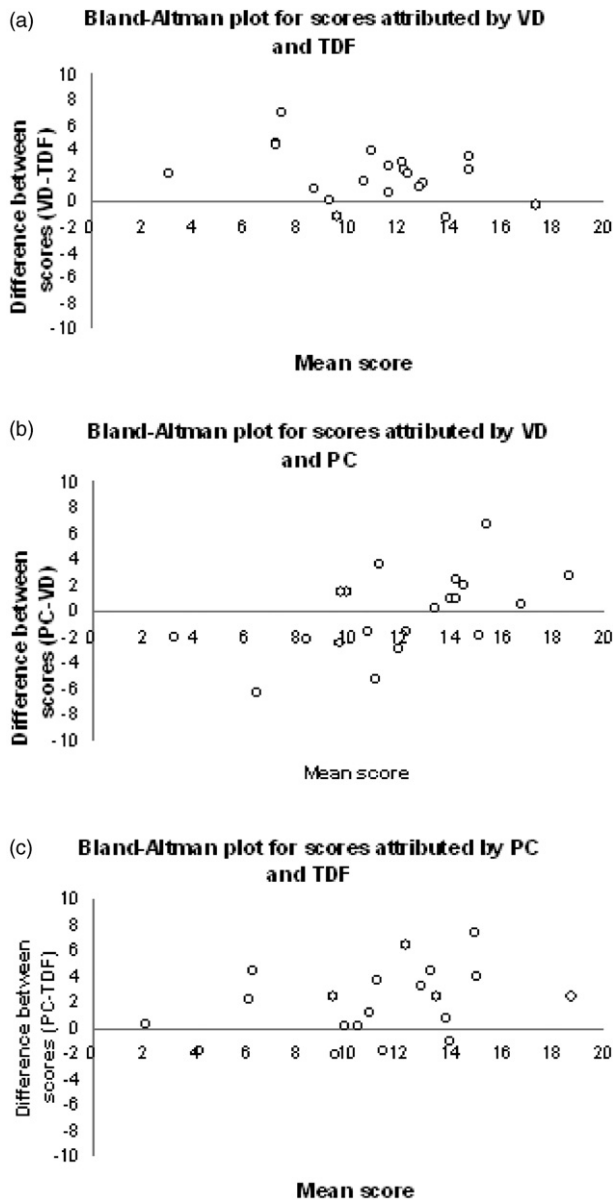


Figure 1. Bland–Altman plots for scoring by three raters of the EBP assignments. Difference of scores is plotted against the mean score for each assignment ($n=20$).

Discussion

EBP has become a familiar course within most undergraduate and postgraduate medical curricula. While most teachers have probably devised their own local assessment methods, reviews of the literature point to a paucity of validity evidence for assessments used in published studies, an ironic finding for a course which aims to train medical students to critically appraise the validity of medical data. Assessments with demonstrated validity for individual summative purposes do exist. Many do not cover the whole EBP process. Those that do usually supply students with cases and data. Only one other study has also used a written assignment along with MCQs but reported data are mainly on the MCQ part of the instrument (Weberschock et al. 2005).

Our instrument has demonstrated content validity. Its correlation (0.75) to the Fresno test which has similar content

demonstrates good concurrent validity. Internal consistency of scores given by three raters is excellent. Inter-rater agreement was above 0.80 for marking by two course tutors but was lower for other pairs and for marking by three raters. The tutors with the highest level of agreement were those who had designed the initial marking grid which may explain the close correlation of their scoring. Discussions on discrepancies occurred after independent rating. Consistent differences of interpretation emerged. Raters differed when scoring relevance of topic choice, i.e. giving credit or not if relevance is discussed when appraising the document rather than in discussing choice of topic. The vague formulation of some criteria, e.g. 'the student correctly analysed ...', revealed differences in scoring. As intended, these formulations require judgement from raters and often, the brief nature of responses led to differences in inference of level of understanding displayed by students. Providing detailed scoring rubrics such as those used in the Fresno test might help clarify expectations. Furthermore, the strict page limit may have encouraged excessive conciseness requiring the raters to infer students' level of understanding. The limit could reasonably be extended. Such modifications should improve inter-rater agreement. The instrument was feasible for further use and results suggest that two raters may be sufficient in the future, after further clarification of criteria for scoring and increased experience with the grid, thus limiting total marking time.

Requiring students to work in pairs raises the question of the validity of attributing the same score to the two members, regardless of actual participation in the assignment. This issue should be balanced with the advantages of group assignments: increased feasibility by limiting marking time and potential educational impact by encouraging teamwork, an increasingly recognized part of modern medical practice. By limiting groups to two members and allowing students to select their partner, the risk of one student contributing nothing to the assignment should be limited. This should, however, be taken into account were this instrument to be implemented as part of a high-stakes examination.

The assignment represents one measure of a unified (internally consistent) construct of EBP. Case specificity is a well-described phenomenon in medical education and requires tests to use several cases in order to attain reproducibility of the total score (Eva et al. 1998; Eva 2003; Norman 2005). Whether this also applies to EBP is unclear. The assignment covers the content of the course fully but students' ability to apply their EBP know-how may indeed vary from one clinical query to another and one source of evidence to another. Once again, this issue is one of balancing feasibility and authenticity against reproducibility. This may be a concern for high-stakes examinations.

While we hope that students will have experienced the potential benefits of EBP in doing the authentic assignment, whether students will actually use EBP in the future is still to be studied. We did not collect data on consequential validity. Behavioural instruments are few and are based either on portfolios or on record audits which may not be easily implemented in our current context of residency (Shaneyfelt et al. 2006).

Some teething problems did occur. Students complained that instructions were not sufficiently detailed. Once feedback on assignment presentations were given, students said that they understood what was expected of them better and indeed those who returned a modified assignment showed marked improvements. This experience has led us to rethink the marking of the original versus improved assignments. While scoring the original assignment drives students to do them conscientiously, we could reduce its weight to half the total score, with the other half being given to the improved version of the assignment. Students were extremely motivated to improve their assignments but somewhat disheartened by the limited impact on their final score.

Changing the assessment strategy has proved time consuming but extremely beneficial. Tutors felt that more students displayed deeper understanding and application ability at the end of the course than in previous years. As always, developing a new assessment instrument requires making course objectives explicit and sometimes uncovers differing views among the people involved. Consulting external EBP experts to validate the instrument provided us with useful feedback both on our objectives and on the actual instrument. Collecting data on the new instrument has also had secondary benefits by introducing research rigour into the teaching sphere at our department.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

VALERIE DORY is a PhD student at the General Practice Centre at Universite catholique de Louvain, Belgium. Her thesis is on self-assessment of general practice residents.

ROBERT GAGNON is a psychometrician. He is a research associate at the Health Sciences Education Centre at Universite de Montreal, Canada.

THERESE DE FOY is a general practitioner. At the time of the study, she was a research assistant at the General Practice Centre at Universite catholique de Louvain, Belgium.

CORENTIN DUYVER is a general practitioner. At the time of the study, he was a research assistant at the General Practice Centre at Universite catholique de Louvain, Belgium.

SOPHIE LECONTE is a general practitioner and a PhD student at the General Practice Centre at Universite Catholique de Louvain, Belgium. Her thesis is on cough assessment.

References

- Audet N, Gagnon R, Ladouceur R, Marcil M. 1993. How effective is the teaching of critical analysis of scientific publications? Review of studies and their methodological quality. *CMAJ* 148(6):945–952.
- Coomarasamy A, Khan KS. 2004. What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ* 329(7473):1017, DOI: 10.1136/bmj.329.7473.1017.
- Downing SM. 2003. Validity: On meaningful interpretation of assessment data. *Med Educ* 37(9):830–837.
- Eva KW. 2003. On the generality of specificity. *Med Educ* 37(9): 587–588.
- Eva KW, Neville AJ, Norman GR. 1998. Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Acad Med* 73(10 Suppl.):S1–S5.
- Flores-Mateo G, Argimon JM. 2007. Evidence based practice in postgraduate healthcare education: A systematic review. *BMC Health Serv Res* 7:119. Available from: <http://www.biomedcentral.com/1472-6963/7/119>
- Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, Kunz R. 2002. 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* 325(7376):1338–1341.
- Green ML. 1999. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: A critical review of curricula. *Acad Med* 74(6):686–694.
- Norman G. 2005. Research in clinical reasoning: Past history and current trends. *Med Educ* 39(4):418–427.
- Norman GR, Shannon SI. 1998. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: A critical appraisal. *CMAJ* 158(2):177–181.
- Ramos KD, Schafer S, Tracz SM. 2003. Validation of the Fresno test of competence in evidence based medicine. *BMJ* 326(7384):319–321.
- Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, Whelan C, Green M. 2006. Instruments for evaluating education in evidence-based practice: A systematic review. *JAMA* 296(9):1116–1127.
- Straus SE, Richardson WS, Glasziou P, Haynes RB. 2005. Evidence-based medicine: How to practice and teach EBM. 3rd ed. Edinburgh: Churchill Livingstone.
- Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. 2000. A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ* 34(2):120–125.
- Weberschock TB, Ginn TC, Reinhold J, Strametz R, Krug D, Bergold M, Schulze J. 2005. Change in knowledge and skills of year 3 undergraduates in evidence-based medicine seminars. *Med Educ* 39(7):665–671.