REVIEW

Instruments for Evaluating Education in Evidence-Based Practice

A Systematic Review

Terrence Shaneyfelt, MD, MPH
Karyn D. Baum, MD, MSEd
Douglas Bell, MD, PhD
David Feldstein, MD
Thomas K. Houston, MD, MPH
Scott Kaatz, DO
Chad Whelan, MD
Michael Green, MD, MSc

Context Evidence-based practice (EBP) is the integration of the best research evidence with patients' values and clinical circumstances in clinical decision making. Teaching of EBP should be evaluated and guided by evidence of its own effectiveness.

Objective To appraise, summarize, and describe currently available EBP teaching evaluation instruments.

Data Sources and Study Selection We searched the MEDLINE, EMBASE, CINAHL, HAPI, and ERIC databases; reference lists of retrieved articles; EBP Internet sites; and 8 education journals from 1980 through April 2006. For inclusion, studies had to report an instrument evaluating EBP, contain sufficient description to permit analysis, and present quantitative results of administering the instrument.

Data Extraction Two raters independently abstracted information on the development, format, learner levels, evaluation domains, feasibility, reliability, and validity of the EBP evaluation instruments from each article. We defined 3 levels of instruments based on the type, extent, methods, and results of psychometric testing and suitability for different evaluation purposes.

Data Synthesis Of 347 articles identified, 115 were included, representing 104 unique instruments. The instruments were most commonly administered to medical students and postgraduate trainees and evaluated EBP skills. Among EBP skills, acquiring evidence and appraising evidence were most commonly evaluated, but newer instruments evaluated asking answerable questions and applying evidence to individual patients. Most behavior instruments measured the performance of EBP steps in practice but newer instruments documented the performance of evidence-based clinical maneuvers or patient-level outcomes. At least 1 type of validity evidence was demonstrated for 53% of instruments, but 3 or more types of validity evidence were established for only 10%. High-quality instruments were identified for evaluating the EBP competence of individual trainees, determining the effectiveness of EBP curricula, and assessing EBP behaviors with objective outcome measures.

Conclusions Instruments with reasonable validity are available for evaluating some domains of EBP and may be targeted to different evaluation needs. Further development and testing is required to evaluate EBP attitudes, behaviors, and more recently articulated EBP skills.

JAMA. 2006;296:1116-1127 www.jama.com

PHYSICIANS OFTEN FAIL TO IMPLEMENT clinical maneuvers that have established efficacy.[1,2] In response, professional organizations have called for increased training in evidence-based practice (EBP) for all health care professions and at all levels of education.[3-6] Evidence-based practice may be defined as the integration of the best research evidence with patients' values and clinical circumstances in clinical decision making.[7]

As educators implement EBP training, they need instruments to evaluate the programmatic impact of new curricula and to document the competence of individual trainees. Prior systematic reviews of EBP training summarized the effectiveness of educational interventions,[8-13] but only 1 that was conducted in 1999 also included a detailed analysis of evaluation instruments.[8] Although there are multiple components of EBP (BOX), as of 1998 the published instruments focused on critical appraisal to the exclusion of other EBP steps, mea-

See also Patient Page.

Author Affiliations: Department of Medicine, University of Alabama School of Medicine, and Department of Veterans Affairs Medical Center, Birmingham (Drs Shaneyfelt and Houston); Department of Medicine, University of Minnesota Medical School, Minneapolis (Dr Baum); Department of Medicine, Division of General Internal Medicine, David Geffen School of Medicine at University of California, Los Angeles (Dr Bell); Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison (Dr Feldstein); Henry Ford Hospital, Detroit, Mich (Dr Kaatz); Department of Medicine, University of Chicago, Chicago, Ill (Dr Whelan); and Department of Medicine, Yale University School of Medicine, New Haven, Conn (Dr Green).
Corresponding Author: Terrence Shaneyfelt, MD, MPH, University of Alabama School of Medicine, Veterans Affairs Medical Center, 700 S 19th St, Birmingham, AL 35233 (terry.shaneyfelt@med.va.gov).

Box. Definitions of Variables and Terminology Used in This Study

Description: Format of instrument; choices include written or Web-based test, self-report survey, OSCE with standardized patients, other OSCE, portfolio, audiotape of teaching sessions, record audit, chart-stimulated recall, direct observation (clinical evaluation exercise), rating scale, and other

Development: Free-text description of development

EBP domains

Knowledge: Knowledge about EBP

Skills: EBP skills are distinguished from knowledge by participants *applying* their knowledge by performing EBP steps in some type of clinical scenario, such as with a standardized patient, written case, computer simulation, OSCE, or direct observation.

Ask: Converting the need for information (about prevention, diagnosis, prognosis, therapy, causation, etc) into an answerable question

Acquire: Tracking down the best evidence with which to answer that question

Appraise: Critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect), and applicability (usefulness in one's own clinical practice)

Apply: Applying the evidence in clinical decision making (includes both individualizing the evidence [such as recasting number needed to treat for the patient's baseline risk] and integrating the evidence with the patient's preferences and particular clinical circumstances)

Attitude: Attitudes toward EBP

Behaviors: Actual performance of EBP in practice

Enacting EBP steps in practice: Actually enacting EBP steps (such as identifying clinical questions) in the course of patient care activities

Performing evidence-based clinical maneuvers: Performing evidence-based maneuvers in trainee's actual practice, such as *prescribing* angiotensin-converting enzyme inhibitors for congestive heart failure with depressed left ventricular function or *checking* hemoglobin A_{1c} in patients with diabetes

Affecting patient outcomes: Trainee's patients experience improved or favorable outcomes, such as lower blood pressure

Feasibility: Documentation of some measure of ease of implementation; choices include time required to administer instrument, time required to score instrument, expertise required to score instrument, cost to administer and score, administrative support required, other

Interrater reliability: Statistical test (κ or correlation coefficient) of the agreement among 2 or more raters' scoring of the responses. Applied only to instruments that required some level of judgment to score, such as free-text responses. In contrast, reliability testing was deemed not applicable for instruments that required no rater judgment to score, such as multiple-choice tests. Credited as "tested" if a quantitative assessment was done. Credited as "established" if the corresponding statistical test was significant.

Participants (number, discipline, and level): Participants in whom the instrument was tested; options include undergraduate medical students (year), residents (specialty), fellows (specialty), faculty physicians, practicing physicians, nurses in training, practicing nurses, allied health professionals, and other health care professionals

Validity: For all types except content validity, credited as "tested" if a quantitative assessment of a particular type of validity was done; credited as "established" if the corresponding statistical test was significant*

Based on content: External review of the instrument by experts in EBP

Based on internal structure

Internal consistency: Statistical test to establish the relationship between items within either the entire instrument or a prespecified section of the instrument

Dimensionality: Factor analysis to determine if the instrument measured a unified latent construct or, if specified in advance, discrete subthemes

Based on relationship to other variables

Responsive: Ability to detect the impact of an EBP educational intervention; requires statistical comparison of same participant's scores before and after an EBP educational intervention

Discriminative: Ability to discriminate between participants with different levels of EBP expertise; requires statistical comparison of instrument scores among participants of different levels of EBP ability

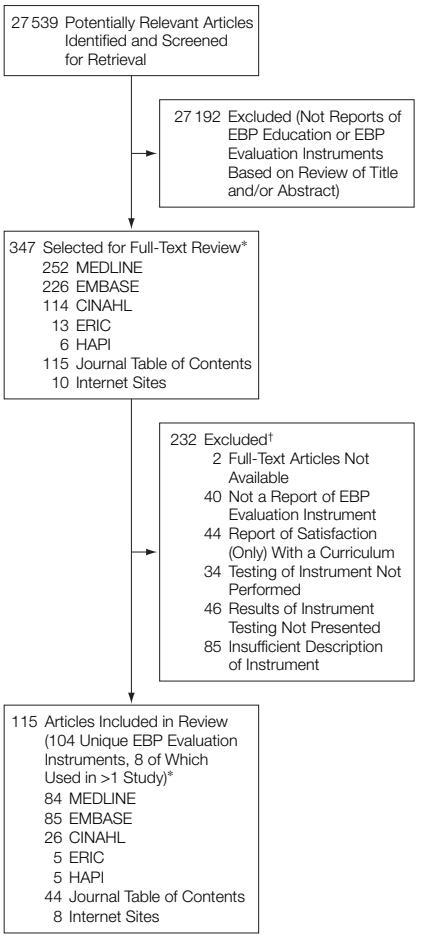
Criterion: Statistical test of the relationship between the instrument scores and participants' scores on another instrument with established psychometric properties

Abbreviations: EBP, evidence-based practice; OSCE, observed structured clinical examination.

*Classification of validity is based on the *Standards for Educational and Psychological Testing* of the Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education[14] and other recommendations.[15,16]

sured EBP knowledge and skills but did not objectively document behaviors in actual practice, and often lacked established validity and reliability.[8] In 2002, Hatala and Guyatt[17] noted that "ironically, if one were to develop guidelines for how to teach [evidence-based medicine] based on these results, they would be based on the lowest level of evidence." Since then, instruments have been developed to try to address the deficits in evaluation. In addition, EBP has become more sophisticated, requiring additional skills. For example, in identifying evidence, practi-

Figure. Search for and Selection of Articles for Review



EBP indicates evidence-based practice.
*Articles could be found in more than 1 database (see "Methods" section of text for details of search strategies, databases, and names of the 8 journals whose tables of contents were searched).
†Reasons for exclusion not mutually exclusive.

tioners must be able to appraise, select among, and search emerging electronic secondary "preappraised" information resources.[18] In applying evidence to decision making, they must explicitly integrate patient preferences and clinical context.[7]

Because of these changes, we performed a systematic review of EBP evaluation instruments and strategies, documenting their development, format, learner levels, EBP evaluation domains, psychometric properties, and feasibility. Our 2 goals were to provide guidance for EBP educators by highlighting preferred instruments based on evaluation needs and to make recommendations for EBP education research based on the current state of the EBP evaluation science.

METHODS

Identification of Studies

To identify evaluation instruments, we searched the MEDLINE, EMBASE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Health and Psychosocial Instruments (HAPI), and Educational Resources Information Center (ERIC) databases from 1980 through April 2006. Search terms included *evidence-based medicine*; *critical appraisal*; *clinical epidemiology*; *journal club*; *clinical question*; *medical informatics*; *medical informatics applications*; *information storage and retrieval*; *databases, bibliographic*; *integrated advanced information management systems*; *MEDLARS*; *education*; *clinical trials*; *controlled clinical trials*; *multicenter studies*; and *program evaluation*. We also manually searched the reference lists of retrieved articles, tables of contents of 8 major medical education journals (*Academic Medicine*, *Medical Education*, *Teaching and Learning in Medicine*, *Medical Teacher*, *Advances in Health Sciences Education*, *Medical Education OnLine*, *Journal of Continuing Education in the Health Professions*, and *BioMed Central Medical Education*), several EBP Internet sites,[4,19-23] and the authors' personal files. The Internet sites were chosen based on author experience as loci that might contain instruments not identified by other strategies.

We included studies that (1) reported an instrument or strategy that evaluated EBP knowledge, skills, attitudes, behaviors, or patient outcomes; (2) contained a sufficient description of the instrument or strategy to permit analysis; and (3) presented results of testing the performance of the instrument or strategy. We did not exclude any articles based on study design. Given the breadth of our review and the large number of articles initially captured by our search strategy, it was not feasible to translate the non–English-language articles to determine their suitability for inclusion. Thus, we limited our analysis to studies published in English. For 1 study, we contacted the authors for clarification. Studies that reported only satisfaction with a curriculum were excluded. Two authors (T.S. and M.G.) independently evaluated each article in the preliminary list for inclusion, and disagreements were resolved by consensus.

Data Extraction

We developed and piloted a standardized data form to abstract information from the included articles. A randomly assigned set of 2 raters, representing all permutations of the 6 raters, independently abstracted information from each of the included articles. In this process and in the article inclusion process, raters were not blinded to any portion of articles. After submitting their original abstraction forms to a central location, the pairs of raters resolved their differences by consensus. The abstraction variables included description and development of the EBP evaluation instrument; number, discipline, and training levels of participants; EBP domains evaluated; feasibility assessment; and type, method, and results of validity and reliability assessment[14] (see Box for definitions). We determined interrater reliability for the article inclusion process and for the data abstraction process based on data from all included articles. κ Statistics were calculated and interpreted according to the guidelines of Landis and Koch.[24]

Quality Categorization of Studies

We did not use restrictive inclusion criteria related to study quality. However, we did define 3 levels of instruments, based on (1) the type, extent, methods, and results of psychometric testing and (2) suitability for different evaluation purposes. For use in the summative evaluation of individual trainees, we identified instruments with the most robust psychometric proper-

ties generally and, in particular, the ability to distinguish between participants of different levels of EBP experience or expertise (level 1). These instruments had to be supported by established interrater reliability (if applicable), objective (non–self-reported) outcome measures, and multiple (≥3) types of established validity evidence (including evidence of discriminative validity).

For use in evaluating the programmatic effectiveness of an EBP educational intervention, we identified a second group of instruments supported by established interrater reliability (if applicable) and "strong evidence" of responsive validity, established by studies with a randomized controlled trial or pre-post controlled trial design and an objective (non–self-reported) outcome measure (level 2). These instruments generally have less robust psychometric properties than level 1 instruments, which must be supported by 3 or more different types of validity evidence. However, level 2 instruments must be supported by higher-level ("strong") evidence for responsive validity in particular. The criteria for "strong evidence" are stricter than the definition of responsive validity (Box) used for the general classifications in this review. Instruments meeting all of the criteria for level 1 may also have "strong evidence" for responsive validity (as indicated in the table footnotes) but this is not required for this designation.

Finally, considering the evaluation of EBP behaviors, we anticipated that few of the instruments would meet either of the preceding thresholds. Therefore, we used a single criterion of an objective (non–self-reported) outcome to distinguish a group of relatively high-quality measures in this domain (level 3).

In cases in which an instrument included 2 distinct pieces (with different formats) intended to evaluate 2 distinct EBP domains, we applied the quality criteria separately to each. For descriptive purposes, we included both subinstruments in the tables and indicated if one or both met the psychometric threshold.

We calculated descriptive statistics for the characteristics and psychometric properties of the evaluation instruments. Analyses were performed using Stata Special Edition version 9.0 (Stata Corp, College Station, Tex).

RESULTS

Inclusion criteria were met by 115 articles[25-140] representing 104 unique assessment strategies (8 instruments were used in >1 study, and 1 study was reported in 2 articles) (FIGURE). There was substantial interrater agreement for the article inclusion process (κ=0.68; 95% confidence interval [CI], 0.47-0.89), as well as for the assessments of validity based on content (κ=0.70; 95% CI, 0.49-0.91) and based on internal structure (κ=0.71; 95% CI, 0.47-0.95). There was moderate agreement on the assessment of validity based on relationships to other variables (κ=0.52; 95% CI, 0.35-0.70).

Characteristics of EBP Evaluation Instruments

The participants' health care professions discipline and training level and the evaluated EBP domains are shown in TABLE 1 (see Box for definitions). The majority of instruments targeted students and postgraduate trainees, while nonphysicians were rarely evaluated. The instruments most commonly evaluated EBP skills (57%), followed by knowledge and behaviors (both 38%), followed by attitudes (26%). Among the EBP skills, critical appraisal of evidence was included in the greatest proportion of instruments.

Thirty (86%) of the 35 evaluation approaches for the "acquire" step related exclusively to skills in searching MEDLINE or similar bibliographic databases for original articles. Of the 5 instruments considering alternative electronic information sources, 4 specifically evaluated awareness, preference for, or skills in searching specific secondary evidence-based medical information resources (including the Cochrane Library, Database of Abstracts of Reviews of Effectiveness, *ACP Journal Club*, and *Clinical Evidence*)[25,33,43,135] while the remaining one[42] merely referred to "Web sites." Similarly, among the instruments evaluating the "apply" step, only 5 (38%) of 13 went beyond the ability to consider research evidence to also assess the ability to integrate the evidence with the patient's particular clinical context and preferences. Evaluation approaches included standardized patient ratings of students explaining a therapeutic decision after reviewing research evidence,[38,39] scoring of residents' free-text justification of applying results of a study to a "paper case,"[28] and documenting decision making before and after access to a research abstract[41] or MEDLINE search.[49]

Most of the instruments evaluating EBP behaviors measured the use of EBP steps in practice. Of these, only 6 (18%) of 34 used objective outcome measures[31,52-54,113,137] with the remaining relying on retrospective self-

Table 1. Characteristics of EBP Evaluation Instruments*

Characteristics	Instruments, No. (%) (N = 104)
Participants' health care profession discipline and training level	
Students†	43 (41.3)
Postgraduate trainees‡	35 (33.7)
Practicing physicians	30 (28.8)
Nonphysicians§	13 (12.5)
EBP evaluation domains	
EBP knowledge	39 (37.5)
EBP skills	59 (56.7)
Ask	13 (12.5)
Acquire	35 (33.7)
Appraise	40 (38.5)
Apply	13 (12.5)
EBP attitudes	27 (26.0)
EBP behaviors	39 (37.5)
Performing EBP steps in practice	34 (32.7)
Performing evidence-based clinical maneuvers in practice	3 (2.9)
Patient outcomes	2 (1.9)

Abbreviation: EBP, evidence-based practice.
*See Box for definitions. Categories are not mutually exclusive.
†Medical students (n = 43), dental students (n = 1), and nursing students (n = 1).
‡Internal medicine (n = 19), emergency medicine (n = 1), surgery (n = 2), obstetrics/gynecology (n = 3), pediatrics (n = 1), and family medicine (n = 8) residents.
§Nurses (n = 7), physical therapists (n = 1), researchers (n = 1), and not specified (n = 4).

Table 2. Psychometric Characteristics of Evidence-Based Practice Evaluation Instruments*

Characteristics	No. (%) (N = 104)	
	Tested	Established
Content validity	17 (16.3)	17 (16.3)
Validity based on internal structure	13 (12.5)	13 (12.5)
Internal consistency	13 (12.5)	13 (12.5)
Dimensionality	6 (5.8)	6 (5.8)
Validity based on relationship to other variables	59 (56.7)	46 (44.2)
Responsive validity	51 (49.0)	41 (39.4)
Discriminative validity	10 (9.6)	9 (8.7)
Criterion validity	7 (6.7)	4 (3.9)
Instruments with ≥3 types of validity tests	11 (10.6)	10 (9.6)

*See Box for validity definitions. Categories are not mutually exclusive.

reports. Only 3 instruments measured the performance of evidence-based clinical maneuvers in practice,[57,58,140] and 2 evaluated the effect of an EBP teaching intervention on patient outcomes.[57,58]

Feasibility and Psychometric Testing

Feasibility of implementation was reported for 19 (18.3%) of the 104 instruments. Among these, 13 reported the time required to administer or score the instrument,* 4 described the expertise required for scoring,[37,47,65,72] and 4 estimated the financial costs of implementation.[28,54,100,114] Investigators performed interrater reliability testing on 21 (41.2%) of the 51 instruments for which it was appropriate, most commonly using κ statistics and correlation coefficients.

Investigators conducted at least 1 type of validity testing in 64% and established it in 53% of the 104 EBP evaluation instruments (TABLE 2). However, multiple (≥3) types of validity evidence were established for only 10% of the instruments. Investigators most commonly sought (57%) and established (44%) evidence for validity based on relationships to other variables. Among these, responsive validity was most commonly tested and established, followed by discriminative and criterion validity.

*References 40, 45, 60, 61, 64, 66, 69, 72, 88, 99, 113, 123, 126.

Eight instruments were used in subsequent studies, either for further validation or to evaluate programmatic impact of an EBP curriculum. One instrument[60] was used in 3 later studies[61-63]; 1 instrument[140] was used in 2 later studies[55,56]; and 6 instruments[33,41,59,64,133,137] were used in 1 subsequent study each.[32,65,87,134,136,139]

Quality Categorization of Instruments

Level 1 Instruments. TABLE 3 summarizes the EBP evaluation domains, format, and psychometric properties of the instruments supported by established interrater reliability (if applicable), objective (non–self-reported) outcome measures, and multiple (≥3) types of established validity evidence (including evidence for discriminative validity). These instruments are distinguished by the ability to discriminate between different levels of expertise or performance and are therefore suited to document the competence of individual trainees. Furthermore, the robust psychometric properties in general support their use in formative or summative evaluations. The Fresno Test[25] and Berlin Questionnaire[59] represent the only instruments that evaluate all 4 EBP steps. In taking the Fresno Test, trainees perform realistic EBP tasks, demonstrating applied knowledge and skills. However, more time and expertise are required to grade this instrument. The multiple-choice format of the Berlin Questionnaire restricts assessment to EBP applied knowledge but also makes it more feasible to implement. The other instruments in Table 3 evaluate a narrower range of EBP.

Level 2 Instruments. In addition to 4 of the instruments in Table 3,[26,59,60,64] 9 instruments fulfilled the criteria for strong evidence of responsive validity (TABLE 4). These are appropriate to consider for evaluating programmatic (rather than individual) impact of EBP interventions. Six evaluated EBP knowledge and skills.[27-31,37] Among these, only one[27] measured all 4 EBP steps. Residents articulated clinical questions, conducted MEDLINE searches, performed calculations, and answered free-text questions about critical appraisal and application of the evidence. In this study, gains in skills persisted on retesting at 6 months, indicating both concurrent and predictive responsive validity. The instrument described by Green and Ellis[28] required free-text responses about the appraisal of a redacted journal article and application of the results to a patient. The 3 multiple-choice tests[29-31] detected improvements in trainees' EBP knowledge. However, in 2 of the studies, this gain did not translate into improvements in critical appraisal skills as measured with a test article[29] or the incorporation of literature into admission notes.[30] Finally, in Villanueva et al,[37] librarians identified elements of the patient-intervention-comparison-outcome (PICO) format[141] in clinical question requests, awarding 1 point for each included element. In a randomized controlled trial of instruction in clinical question construction, this instrument detected improvements in this skill.

Four EBP behavior instruments met the criteria for strong evidence of responsive validity and an objective outcome measure.[31,52,57,113] Among these, 3 measured the enactment of EBP steps in practice.[31,52,113] Ross and Verdieck[31] analyzed audiotapes of resident-faculty interactions, looking for phrases related to literature searching, clinical

Table 3. Level 1 Instruments (Individual Trainee Formative or Summative EBP Evaluation)*

Source	Study Settings/Participants	EBP Domains	Description	Interrater Reliability†	Validity
Ramos et al[25] (Fresno Test)	53 "Experts" and 43 family practice residents and faculty in instrument development study	Formulate a focused question Identify appropriate research design for answering the question Show knowledge of electronic database searching (including secondary sources) Identify issues important for the relevance and validity of an article Discuss the magnitude and importance of research findings	Open-ended free-text questions, fill-in-the-blank questions, and calculations relating to 2 pediatric clinical scenarios; scored using a standardized grading rubric that includes examples of acceptable answers and specifies 4 or 5 grading categories (not evident, minimal and/or limited, strong, excellent), each of which is associated with a point value	Yes (R = 0.72-0.96)	Content Internal consistency Discriminative
Bennett et al[26]	79 Medical students in various clerkships in pre-post controlled trial	Critical appraisal skills	Set of case-based problems that require a diagnostic or treatment decision matched with an article advocating the test or treatment; students must "take a stand" and "defend" it in writing; graded on preset criteria	Yes (κ = 0.74-1.00)‡	Content Discriminative Responsive
Fritsche et al,[59] Akl et al[136] (Berlin Questionnaire)§	43 "Experts," 20 third-year students, 203 participants in EBP course in instrument development study[59] 49 Internal medicine residents in nonrandomized controlled trial of EBP curriculum[136]	Knowledge about interpreting evidence Skills to relate a clinical problem to a clinical question Best design to answer a question Use quantitative information from research to solve specific patient problems	2 Separate sets of 15 multiple-choice questions built around "typical" clinical scenarios	NA	Content Internal consistency Discriminative Responsive
Taylor et al, [60,61] Bradley and Herrin,[62] Bradley et al[63]§	152 Health care professionals in instrument development study[60] 145 General practitioners, hospital physicians, allied health professionals, and health care managers in RCT of critical appraisal training[61] Modified and "revalidated" instrument on 55 delegates at international EBP conferences[62] 175 Students in RCT of self-directed vs workshop-based EBP curricula[63]	Knowledge of critical appraisal Knowledge of MEDLINE searching	Sets of 6 multiple-choice questions with 3 potential answers, each requiring a true, false, or "don't know" response; best score on each set = 18	NA	Content Internal consistency Discriminative Responsive
MacRae et al[64,65]§	44 Surgery residents in instrument development study[64] 55 Surgeons in RCT of Internet-based EBP curriculum[65]	Critical appraisal skills	3 Journal articles, each followed by a series of short-answer questions and 7-point scales to rate the quality of elements of the study design; short-answer questions based on cards from an EBP texbook[141]	Yes (R = 0.78-0.91)	Internal consistency Discriminative Responsive
Weberschock et al[66]	132 Third-year medical students and 11 students with advanced training in "EBM working group" in development and pre-post uncontrolled study of peer-teaching EBP curriculum	EBP knowledge and skills (specific EBP steps not specified)	5 Sets of 20 multiple-choice questions (5 "easy," 10 "average," and 5 "difficult") linked to clinical scenarios and pertaining to data from published research articles	NA	Internal consistency Discriminative Responsive Criterion
Haynes et al,[87,137] McKibbon et al[138]§‖	308 Physicians and physicians in training in RCT of one-one precepting and searching feedback[87] 158 Clinicians (novice end users), 13 "expert searcher" clinicians (expert end users), and 3 librarians[137,138]	MEDLINE searching skills	Search output scored by comparison to searches (for same clinical questions) by an expert end user physician and a librarian; "relative recall" calculated as number of relevant citations from a given search divided by number of relevant citations from the 3 searches (participant, expert physician, and librarian); "precision" calculated as the number of relevant citations retrieved in a search divided by the total citations retrieved in that search; article "relevance" rated reliably on a 7-point scale	Yes (κ = 0.79)‡	Content Discriminative Responsive
		EBP behavior (enacting EBP steps—MEDLINE searching–in practice)	Library system electronically captures questions that prompt the search, search strategy, and search output	NA	

Abbreviations: EBP, evidence-based practice; NA, not applicable; RCT, randomized controlled trial.
*These instruments had to be supported by established interrater reliability (if applicable), objective (non–self-reported) outcome measures, and 3 or more types of established validity evidence (including evidence for discriminative validity).
†Reliability testing was deemed not applicable for instruments that required no rater judgment to score, such as multiple-choice tests (see Box).
‡Demonstrated both interrater and intrarater reliability.
§Instruments evaluated in more than 1 study. Results from all of the studies were used to determine number of trainees, reliability, and validity.
‖Met level 1 criteria for searching skill portion of overall instrument, not for the searching behavior portion.

epidemiology, or critical appraisal. Family practice residents' "evidence-based medicine utterances" increased from 0.21 per hour to 2.9 per hour after an educational intervention. Stevermer et al[113] questioned residents about their awareness and knowledge of findings in recent journal articles relevant to primary care practice. Residents exposed to academic detailing recalled more articles and correctly answered more questions about them. Focusing on the "acquire" step, Cabell et al[52] electronically captured trainees' searching behaviors, including number of log-ons to databases, searching volume, abstracts or articles viewed, and time spent searching. These measures were responsive to an intervention including a 1-hour didactic session, use of well-built clinical question cards, and prac-

Table 4. Level 2 Instruments (Programmatic EBP Curriculum Evaluation)*

Source	Study Settings/Participants	Knowledge and Skill Domains	Description	Interrater Reliability†	Validity
Smith et al[27]	55 Medical residents in pre-post controlled crossover trial of 7-week EBP curriculum, which included interactive sessions and computer laboratory training	Skills in formulating clinical questions Skills in MEDLINE searching Skills in critical appraisal Skills in applying evidence to individual patient decision making Knowledge of quantitative aspects of diagnosis and treatment studies	Test including sets of questions (format not specified) relating to 5 clinical cases	Yes (not reported)	Responsive‡
Green and Ellis[28]	34 Residents in controlled trial of a 7-session EBP curriculum	Skills in critical appraisal Skills in applying evidence to individual patient decision making	9-Question test (requiring free-text response) relating to a case presentation and a redacted journal article	Yes (R = 0.87)	Content Responsive
Linzer et al[29]§	44 Medical residents in RCT of journal club curriculum	Epidemiology and biostatistics knowledge	Multiple-choice test (knowledge); 15 questions chosen so that perfect score would allow access to 81% of medical literature[142]	NA	Content Responsive
		Skills in critical appraisal	Free-text critical appraisal of text article; scoring based on "gold standard" criteria developed by consensus of faculty	Yes (difference in observer variability = 1.1)	Content Discriminative
Landry et al[30]§	146 Medical students in controlled trial of 2 90-minute seminars	Research design and critical appraisal knowledge	10-Item test	NA	Content Responsive
		Skills in applying medical literature to clinical decision making	Blinded review of patient "write-ups" looking for literature citations	No	None
Ross and Verdieck[31]	48 Family practice residents in controlled trial of 10-session EBP workshop (control residents in different program)	EBP knowledge (specific steps not specified)	50-Item "open-book" multiple-choice test	NA	Content Responsive
		EBP behavior (enacting EBP steps in practice)	Analysis of audiotapes of resident-faculty interactions looking for phrases related to literature searching, clinical epidemiology, or critical appraisal	No	Content Responsive
Villanueva et al[37]	39 Health care professional participants in a library "evidence search and critical appraisal service" in an RCT of providing instructions and clinical question examples	Skills in formulating clinical questions	Librarians identified elements of the patient-intervention-comparison-outcome (PICO) format in clinical question requests[141]; 1 point awarded for each of 4 elements included	Yes (κ = 0.68)	Responsive
Cabell et al[52]	48 Internal medicine residents in RCT of EBP curriculum, which included a 1-hour didactic session, the use of well-built clinical question cards, and practical sessions in clinical question building	EBP behavior (performance of EBP steps in practice)	Library system electronically captures MEDLINE searching behavior, including number of log-ons, searching volume, abstracts viewed, full-text articles viewed, and time spent searching	NA	Responsive
Langham et al[57]	Primary care practice teams in RCT of practice-based training in EBP and/or patient information management	EBP behaviors (performing evidence-based clinical maneuvers) EBP behaviors (affecting patient outcomes)	Record audit for physician performance and patient-level quality indicators relating to cardiovascular risk reduction‖	NA	Responsive
Stevermer et al[113]	59 Family practice residents in RCT of EBP academic detailing	EBP behavior (performance of EBP steps in practice)	Test of awareness and recall of recently published articles reporting "important findings about common primary care problems" (selected by faculty physicians)	NA	Responsive

Abbreviations: EBP, evidence-based practice; NA, not applicable; RCT, randomized controlled trial.
*Instruments supported by established interrater reliability (if applicable), "strong evidence" of responsive validity, established by studies with a randomized controlled trial or pre-post controlled trial design, and an objective (non–self-reported) outcome measure. Four instruments from Table 3 also had "strong evidence" of responsive validity and are appropriate for similar uses.[26,59,60,64]
†Reliability testing was deemed not applicable for instruments that required no rater judgment to score, such as multiple-choice tests (see Box).
‡Gains in skills persisted after 6 months, indicating both concurrent and predictive (responsive) validity.
§Met level 2 criteria for the EBP knowledge portion of overall instrument, not for the EBP skill portion.
‖Quality indicators included recording of serum cholesterol and changes in serum cholesterol levels, recording of blood pressure and blood pressure control, recording of smoking status, and aspirin use.

tical sessions in clinical question building. One EBP behavior instrument in this category evaluated EBP practice performance and patient outcomes using medical record audits. Langham et al[57] evaluated the impact of an EBP curriculum, documenting improvements in practicing physicians' documentation, clinical interventions, and patient outcomes related to cardiovascular risk factors.

Although 3 controlled studies demonstrated the responsive validity of having librarians score MEDLINE search strategies[34,36] or clinical question formulations[43] according to predetermined criteria, these did not meet the criteria for interrater reliability testing.

Level 3 Instruments. In addition to the 5 EBP behavior instruments included in levels 1 and 2,[31,52,57,113,137] 4 others used objective outcome measures but did not demonstrate strong evidence of responsive validity or multiple sources of validity evidence (TABLE 5).[53,54,58,140] Two of these consisted of electronic learning portfolios that allowed trainees to document their enactment of EBP steps.[53,54]

The remaining 2 instruments measured the performance of evidence-based maneuvers or patient outcomes. Ellis et al[140] devised a reliable method for determining the primary therapeutic intervention chosen by a practitioner and classifying the quality of evidence supporting it. In this scheme, interventions are (1) supported by individual or systematic reviews of randomized controlled trials, (2) supported by "convincing nonexperimental evidence," or (3) lacking substantial evidence. This instrument was subsequently used in 2 pre-post (but uncontrolled) studies of EBP educational interventions.[55,56] Finally, Epling et al[58] performed a record audit before and after residents developed and implemented a diabetes clinical guideline.

COMMENT

We found that instruments used to evaluate EBP were most commonly administered to medical students and postgraduate trainees and evaluated skills in searching for and appraising the evidence. At least 1 type of validity evidence was demonstrated in 53% of instruments (most commonly based on relationship to other variables), but multiple types of validity evidence were established for very few.

Educators need instruments to document the competence of individual trainees and to evaluate the programmatic impact of new curricula. Given the deficits of instruments previously

Table 5. Level 3 Instruments (EBP Behavior Evalution)*

Source	Study Settings/Participants	Knowledge and Skill Domains	Description	Interrater Reliability†	Validity
Crowley et al[53]	82 Medical residents in prospective cohort study	EBP behavior (enacting EBP steps in practice)	Internet-based portfolio ("compendium") that allows residents to enter their clinical questions, information searching (via MEDLINE reference links), appraisal of articles, and impact on patient care decisions	NA	None
Fung et al[54]	41 Obstetrics/gynecology residents across 4 programs in prospective cohort study	EBP behavior (enacting EBP steps in practice)	Internet-based learning portfolio that allows residents to describe an initiating clinical scenario, enter their clinical questions, link to information resources, and document learning points and implications for their practice	NA	None
Straus et al,[55] Lucas et al,[56] Ellis et al[140]‡	35 Internal medicine faculty physicians and 12 residents in pre-post uncontrolled trial of multicomponent EBP curriculum, which included 7 one-hour sessions and provision of evidence-based resources on the hospital network[55] 33 inpatient internal medicine physicians in pre-post uncontrolled trial of providing standardized literature searches relating to primary diagnosis[56] A physician "team" on a teaching hospital general medicine service in cross-sectional study[140]	EBP behaviors (performing evidence-based clinical maneuvers)	Record audit to determine the "primary therapeutic intervention" chosen by a practitioner and rate the level of evidence supporting it: (1) supported by individual or systematic reviews of RCTs, (2) supported by "convincing nonexperimental evidence," or (3) lacking substantial evidence	Yes (R = 0.76-0.92[55]; R = 0.20-0.56[56]; not reported[140])	Responsive
Epling et al[58]	11 Family practice residents in pre-post trial of EBP curriculum that included development of a clinical guideline	EBP behaviors (performing evidence-based clinical maneuvers) EBP behaviors (affecting patient outcomes)	Record audit for physician performance and patient-level quality indicators relating to care of patients with diabetes mellitus§	NA	Responsive

Abbreviations: EBP, evidence-based practice; NA, not applicable; RCT, randomized controlled trial.
*EBP behavior instruments with objective (non–self reported) outcome measures. Five EBP behavior instruments from previous tables also have objective outcome measures and are appropriate for similar uses.[31,52,57,113,137]
†Reliability testing was deemed not applicable for instruments that required no rater judgment to score, such as multiple-choice tests (see Box).
‡Instrument evaluated in more than 1 study. Results from all of the studies were used to determine number of trainees, reliability, and validity.
§Quality-of-care indicators for diabetes included blood pressure measurement, fingerstick glucose measurement, hemoglobin A_{1c} control, documentation of foot examinations, and referral to nutritionists.

available, it is not surprising that in 2000 only a minority of North American internal medicine programs objectively evaluated the effectiveness of their EBP curricula.[143] Currently, there is a much wider selection of instruments, some of which are supported by more robust psychometric testing. While, like their predecessors, the currently available instruments most commonly evaluate critical appraisal, many more also measure the other important EBP steps. Among the instruments evaluating EBP behaviors, most continue to measure the performance of EBP steps by self-report. However, new instruments objectively document EBP steps and document the performance of evidence-based clinical maneuvers.

The choice of an EBP evaluation instrument should be guided by the purpose of the evaluation and the EBP domains of interest. The instruments in Table 3 are appropriate for evaluating the competence of individual trainees. Although they have reasonably strong psychometric properties, we believe that in the absence of well-defined passing standards for different learner levels, they should not yet be used for high-stakes evaluations, such as academic promotion or certification.

To evaluate the programmatic impact of EBP educational interventions, educators may turn to instruments with strong evidence of responsive validity (Table 4) and whose evaluation domains correspond with the objectives of their curricula. A conceptual framework for evaluating this aspect of EBP teaching has been developed by the Society of General Internal Medicine Evidence-Based Medicine Task Force.[144] It recommends considering the learners (including their level and particular needs), the intervention (including the curriculum objectives, intensity, delivery method, and targeted EBP steps), and the outcomes (including knowledge, skills, attitudes, behaviors, and patient-level outcomes). With the exception of the instruments also included in Table 3, educators should use caution in using these instruments to assess the EBP competence of individual trainees because they were developed to evaluate the effectiveness of specific curricula and lack evidence for discriminative validity.

Only 5 EBP behavior instruments met the 2 highest quality thresholds in our analysis. Notwithstanding the psychometric limitations, however, it is important to document that trainees apply their EBP skills in actual practice. Our review identified several studies that documented EBP behaviors through retrospective self-report. However, this approach may be extremely biased, as physicians tend to underestimate their information needs and overestimate the degree of their pursuit.[145] We recommend that educators restrict their selection of instruments to those with objectively measured outcomes.

Regarding the enactment of EBP steps in practice, analyzing audiotapes of teaching interactions[31] and electronically capturing searching behavior[52] showed responsive validity. However, we believe that these approaches fail to capture the pursuit and application of information in response to particular clinical questions, rendering them poor surrogates for EBP behaviors. Evidence-based practice learning portfolios,[53,54] which serve as both an evaluation strategy and an educational intervention, may represent the most promising approach to document the performance of EBP steps. However, their use in any assessment with more serious consequences than formative evaluation must await more rigorous psychometric testing.

In addition to documenting the performance of EBP steps, educators are charged with documenting behavioral outcomes of educational interventions further downstream, such as performance of evidence-based clinical maneuvers and patient-level outcomes.[146] The reliable approach of rating the level of evidence supporting clinical interventions has been widely used.[140] In 2 studies, this approach detected changes following an EBP curriculum[55] or supplying physicians with a literature search[56] but, in the absence of controlled studies, did not meet our threshold for strong evidence of responsive validity. This system appears most suited to evaluating changes in EBP performance after an educational intervention or over time. To use it to document an absolute threshold of performance would require knowing the "denominator" of evidence-based therapeutic options for each trainee's set of patients, making it impractical on a programmatic scale. The performance of evidence-based maneuvers may also be documented by auditing records for adherence to evidence-based guidelines or quality indicators. Hardly a new development, this type of audit is commonly performed as part of internal quality initiatives or external reviews. Our review found 2 examples of quality audits used to evaluate the impact of EBP training.[57,58]

Assessing EBP attitudes may uncover hidden but potentially remediable barriers to trainees' EBP skill development and performance. However, while several instruments contain a few attitude items, few instruments assess this domain in depth.[33,50,51,134] Moreover, no attitude instruments in this review met our quality criteria for establishment of validity. One instrument demonstrated responsive validity in an uncontrolled study[51] and another demonstrated criterion validity in comparison with another scale.[50]

There are limitations that should be considered in interpreting the results of this review. As in any systematic review, it is possible that we failed to identify some evaluation instruments. However, we searched multiple databases, including those containing unpublished studies, using a highly inclusive search algorithm. Because our search was limited to English-language journals, we would not capture EBP instruments described in other languages. This might introduce publication bias if such instruments differ systematically from those appearing in English-language journals. Our exclusion of insufficiently described instruments may have biased our analysis if these differed systematically from the others. Our abstraction process showed

good interrater reliability, but the characteristics of some EBP evaluation instruments could have been misclassified, particularly in determining validity evidence based on relationship to other variables. In 2 similar reviews of professionalism instruments, there was considerable inconsistency among experts in assigning types of validity evidence.[147,148]

Our findings, which identified some gaps in EBP evaluation, have implications for medical education research. First, it must be determined whether the current generation of evaluation approaches can be validly used to evaluate a wider range of clinicians, such as nurses and allied health professionals. This is supported by the Institute of Medicine's call for interdisciplinary training.[3] Second, there is a need for development and testing of evaluation approaches in 2 content areas of EBP knowledge and skills. Within the "acquire" step, approaches are needed to document trainees' ability to appraise, select, and search secondary electronic medical information resources to find syntheses and synopses of original research studies.[18] There is also a need to evaluate trainees' competence in applying evidence to individual patient decision making, considering the evidence (customized for the patient), clinical circumstances, and patient preferences.[7] Finally, the science of evaluating EBP attitudes and behaviors continues to lag behind the evaluation of knowledge and skills. Medical education researchers should continue to explore approaches that balance psychometric robustness with feasibility.

Author Contributions: Dr Shaneyfelt had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Shaneyfelt, Green.
Acquisition of data: Shaneyfelt, Baum, Bell, Feldstein, Kaatz, Whelan, Green.
Analysis and interpretation of data: Shaneyfelt, Houston, Green.
Drafting of the manuscript: Shaneyfelt, Green.
Critical revision of the manuscript for important intellectual content: Shaneyfelt, Baum, Bell, Feldstein, Houston, Kaatz, Whelan, Green.
Statistical analysis: Shaneyfelt, Houston, Green.
Administrative, technical, or material support: Shaneyfelt, Bell, Green.
Study supervision: Shaneyfelt, Green.
Financial Disclosures: None reported.
Disclaimer: Drs Baum and Green have published articles that were included as part of this review. Neither abstracted data from their own published works.
Acknowledgment: This study was conducted as a charge of the Society of General Internal Medicine Evidence-Based Medicine Task Force, of which 3 authors (Drs Shaneyfelt, Whelan, and Green) are members. We thank Heather Coley, MPH, Department of Medicine, University of Alabama School of Medicine, for her assistance in compiling data on the citation of articles in various electronic databases.

REFERENCES

1. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348:2635-2645.
2. Hayward RA, Asch SM, Hogan MM, Hofer TP, Kerr EA. Sins of omission: getting too little medical care may be the greatest threat to patient safety. *J Gen Intern Med*. 2005;20:686-691.
3. Institute of Medicine. *Health Professions Education: A Bridge to Quality*. Washington, DC: National Academies Press; 2003.
4. Accreditation Council for Graduate Medical Education Outcome Project: general competencies. http://www.acgme.org/outcome/assess/compList.asp. Accessed April 2006.
5. Association of American Medical Colleges. *Contemporary Issues in Medicine, II: Medical Informatics and Population Health*. Washington, DC: Association of American Medical Colleges; 1998.
6. American Board of Internal Medicine Self-Evaluation of Practice Performance. http://www.abim.org/moc/sempbpi.shtm. Accessed November 2005.
7. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *ACP J Club*. 2002;136:A11-A14.
8. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med*. 1999;74:686-694.
9. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ*. 1998;158:177-181.
10. 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;34:120-125.
11. Ebbert JO, Montori VM, Schultz HJ. The journal club in postgraduate medical education: a systematic review. *Med Teach*. 2001;23:455-461.
12. Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev*. 2001;(3):CD001270.
13. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? a systematic review. *BMJ*. 2004;329:1017.
14. Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.
15. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830-837.
16. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38:1006-1012.
17. Hatala R, Guyatt G. Evaluating the teaching of evidence-based medicine. *JAMA*. 2002;288:1110-1112.
18. Haynes RB. Of studies, syntheses, synopses, and systems: the "4S" evolution of services for finding current best evidence. *ACP J Club*. 2001;134:A11-A13.
19. Generalist Faculty Development Project. http://www.im.org./facdev/gimfd/index. Accessed January 2004 [no longer available].
20. Association of American Medical Colleges. http://www.aamc.org. Accessed April 2006.
21. Alliance for Academic Internal Medicine. http://www.im.org. Accessed April 2006.
22. Centre for Evidence-based Medicine at the University of Toronto. http://www.cebm.utoronto.ca/. Accessed April 2006.
23. Centre for Evidence-based Medicine at the University of Oxford. http://cebm.net. Accessed April 2006.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
25. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno Test of competence in evidence based medicine. *BMJ*. 2003;326:319-321.
26. Bennett KJ, Sackett DL, Haynes RB, Neufeld VR, Tugwell P, Roberts R. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA*. 1987;257:2451-2454.
27. Smith CA, Ganschow PS, Reilly BM, et al. Teaching residents evidence-based medicine skills: a controlled trial of effectiveness and assessment of durability. *J Gen Intern Med*. 2000;15:710-715.
28. Green ML, Ellis PJ. Impact of an evidence-based medicine curriculum based on adult learning theory. *J Gen Intern Med*. 1997;12:742-750.
29. 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;260:2537-2541.
30. Landry FJ, Pangaro L, Kroenke K, Lucey C, Herbers J. A controlled trial of a seminar to improve medical student attitudes toward, knowledge about, and use of the medical literature. *J Gen Intern Med*. 1994;9:436-439.
31. Ross R, Verdieck A. Introducing an evidence-based medicine curriculum into a family practice residency—is it effective? *Acad Med*. 2003;78:412-417.
32. Markey P, Schattner P. Promoting evidence-based medicine in general practice—the impact of academic detailing. *Fam Pract*. 2001;18:364-366.
33. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ*. 1998;316:361-365.
34. Bradley DR, Rana GK, Martin PW, Schumacher RE. Real-time, evidence-based medicine instruction: a randomized controlled trial in a neonatal intensive care unit. *J Med Libr Assoc*. 2002;90:194-201.
35. Rosenberg WM, Deeks J, Lusher A, Snowball R, Dooley G, Sackett D. Improving searching skills and evidence retrieval. *J R Coll Physicians Lond*. 1998;32:557-563.
36. Gruppen LD, Rana GK, Arndt TS. A controlled comparison study of the efficacy of training medical students in evidence-based medicine literature searching skills. *Acad Med*. 2005;80:940-944.
37. Villanueva EV, Burrows EA, Fennessy PA, Rajendran M, Anderson JN. Improving question formulation for use in evidence appraisal in a tertiary care setting: a randomised controlled trial. *BMC Med Inform Decis Mak*. 2001;1:4.
38. Bradley P, Humphris G. Assessing the ability of medical students to apply evidence in practice: the potential of the OSCE. *Med Educ*. 1999;33:815-817.
39. Davidson RA, Duerson M, Romrell L, Pauly R, Watson RT. Evaluating evidence-based medicine skills during a performance-based examination. *Acad Med*. 2004;79:272-275.
40. Fliegel JE, Frohna JG, Mangrulkar RSA. Computer-based OSCE station to measure competence in evidence-based medicine skills in medical students. *Acad Med*. 2002;77:1157-1158.
41. Schwartz A, Hupert J. Medical students' applica-

tion of published evidence: randomised trial. *BMJ*. 2003;326:536-538.
42. Berner ES, McGowan JJ, Hardin JM, Spooner SA, Raszka WV, Berkow RL. A model for assessing information retrieval and application skills of medical students. *Acad Med*. 2002;77:547-551.
43. Cheng GY. Educational workshop improved information-seeking skills, knowledge, attitudes and the search outcome of hospital clinicians: a randomised controlled trial. *Health Info Libr J*. 2003;20(suppl 1): 22-33.
44. Bergus GR, Emerson M. Family medicine residents do not ask better-formulated clinical questions as they advance in their training. *Fam Med*. 2005;37: 486-490.
45. Frasca MA, Dorsch JL, Aldag JC, Christiansen RG. A multidisciplinary approach to information management and critical appraisal instruction: a controlled study. *Bull Med Libr Assoc*. 1992;80:23-28.
46. Burrows SC, Tylman V. Evaluating medical student searches of MEDLINE for evidence-based information: process and application of results. *Bull Med Libr Assoc*. 1999;87:471-476.
47. Vogel EW, Block KR, Wallingford KT. Finding the evidence: teaching medical residents to search MEDLINE. *J Med Libr Assoc*. 2002;90:327-330.
48. Toedter LJ, Thompson LL, Rohatgi C. Training surgeons to do evidence-based surgery: a collaborative approach. *J Am Coll Surg*. 2004;199:293-299.
49. Reiter HI, Neville AJ, Norman GR. Medline for medical students? searching for the right answer. *Adv Health Sci Educ Theory Pract*. 2000;5:221-232.
50. McAlister FA, Graham I, Karr GW, Laupacis A. Evidence-based medicine and the practicing clinician. *J Gen Intern Med*. 1999;14:236-242.
51. Baum KD. The impact of an evidence-based medicine workshop on residents' attitudes towards and self-reported ability in evidence-based practice. *Med Educ Online*. 2003;8:4-10.
52. Cabell CH, Schardt C, Sanders L, Corey GR, Keitz SA. Resident utilization of information technology. *J Gen Intern Med*. 2001;16:838-844.
53. Crowley SD, Owens TA, Schardt CM, et al. A Web-based compendium of clinical questions and medical evidence to educate internal medicine residents. *Acad Med*. 2003;78:270-274.
54. Fung MF, Walker M, Fung KF, et al. An internet-based learning portfolio in resident education: the KOALA multicentre programme. *Med Educ*. 2000;34: 474-479.
55. Straus SE, Ball C, Balcombe N, Sheldon J, McAlister FA. Teaching evidence-based medicine skills can change practice in a community hospital. *J Gen Intern Med*. 2005;20:340-343.
56. Lucas BP, Evans AT, Reilly BM, et al. The impact of evidence on physicians' inpatient treatment decisions. *J Gen Intern Med*. 2004;19:402-409.
57. Langham J, Tucker H, Sloan D, Pettifer J, Thom S, Hemingway H. Secondary prevention of cardiovascular disease: a randomised trial of training in information management, evidence-based medicine, both or neither: the PIER trial. *Br J Gen Pract*. 2002;52:818-824.
58. Epling J, Smucny J, Patil A, Tudiver F. Teaching evidence-based medicine skills through a residency-developed guideline. *Fam Med*. 2002;34:646-648.
59. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer HH, 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: 1338-1341.
60. Taylor R, Reeves B, Mears R, et al. Development and validation of a questionnaire to evaluate the effectiveness of evidence-based practice teaching. *Med Educ*. 2001;35:544-547.
61. Taylor RS, Reeves BC, Ewings PE, Taylor RJ. Critical appraisal skills training for health care professionals: a randomized controlled trial. *BMC Med Educ*. 2004;4:30.
62. Bradley P, Herrin J. Development and validation of an instrument to measure knowledge of evidence-based practice and searching skills. *Med Educ Online*. 2004;9:15-19.
63. Bradley P, Oterholt C, Herrin J, Nordheim L, Bjorndal A. Comparison of directed and self-directed learning in evidence-based medicine: a randomised controlled trial. *Med Educ*. 2005;39:1027-1035.
64. MacRae HM, Regehr G, Brenneman F, McKenzie M, McLeod RS. Assessment of critical appraisal skills. *Am J Surg*. 2004;187:120-123.
65. MacRae HM, Regehr G, McKenzie M, et al. Teaching practicing surgeons critical appraisal skills with an Internet-based journal club: a randomized, controlled trial. *Surgery*. 2004;136:641-646.
66. Weberschock TB, Ginn TC, Reinhold J, et al. Change in knowledge and skills of year 3 undergraduates in evidence-based medicine seminars. *Med Educ*. 2005;39:665-671.
67. Johnston JM, Leung GM, Fielding R, Tin KY, Ho LM. The development and validation of a knowledge, attitude and behaviour questionnaire to assess undergraduate evidence-based practice teaching and learning. *Med Educ*. 2003;37:992-1000.
68. Stern DT, Linzer M, O'Sullivan PS, Weld L. Evaluating medical residents' literature-appraisal skills. *Acad Med*. 1995;70:152-154.
69. Thomas KG, Thomas MR, Dupras DM. Assessment tools for evaluating critical appraisal skills. *Med Educ*. 2004;38:569.
70. Badgett RG, Paukert JL, Levy LS. Teaching clinical informatics to third-year medical students: negative results from two controlled trials. *BMC Med Educ*. 2001;1:3.
71. Barnett SH, Kaiser S, Morgan LK, et al. An integrated program for evidence-based medicine in medical school. *Mt Sinai J Med*. 2000;67:163-168.
72. Bazarian JJ, Davis CO, Spillane LL, Blumstein H, Schneider SM. Teaching emergency medicine residents evidence-based critical appraisal skills: a controlled trial. *Ann Emerg Med*. 1999;34:148-154.
73. Beasley BW, Woolley DC. Evidence-based medicine knowledge, attitudes, and skills of community faculty. *J Gen Intern Med*. 2002;17:632-639.
74. Cartwright CA, Korsen N, Urbach LE. Teaching the teachers: helping faculty in a family practice residency improve their informatics skills. *Acad Med*. 2002; 77:385-391.
75. Cork RD, Detmer WM, Friedman CP. Development and initial validation of an instrument to measure physicians' use of, knowledge about, and attitudes toward computers. *J Am Med Inform Assoc*. 1998;5:164-176.
76. Cramer JS, Mahoney MC. Introducing evidence-based medicine to the journal club, using a structured pre and post test: a cohort study. *BMC Med Educ*. 2001;1:6.
77. Crites GE, Chrisagis X, Patel V, Little D, Drehmer T. A locally created EBM course for faculty development. *Med Teach*. 2004;26:74-78.
78. Cuddy PG, Elenbaas JK, Coit KJ. The effectiveness of a slide-tape program on literature evaluation. *J Biocommun*. 1984;11:2-4.
79. DeLisa JA, Jain SS, Kirshblum S, Christodoulou C. Evidence-based medicine in physiatry: the experience of one department's faculty and trainees. *Am J Phys Med Rehabil*. 1999;78:228-232.
80. orsch JL, Frasca MA, Wilson ML, Tomsic ML. A multidisciplinary approach to information and critical appraisal instruction. *Bull Med Libr Assoc*. 1990;78: 38-44.
81. Edwards R, White M, Gray J, Fischbacher C. Use of a journal club and letter-writing exercise to teach critical appraisal to medical undergraduates. *Med Educ*. 2001;35:691-694.
82. Erickson S, Warner ER. The impact of an individual tutorial session on MEDLINE use among obstetrics and gynaecology residents in an academic training programme: a randomized trial. *Med Educ*. 1998; 32:269-273.
83. Fox NJ, Dolman EA, Lane P, O'Rourke AJ, Roberts C. The WISDOM project: training primary care professionals in informatics in a collaborative "virtual classroom." *Med Educ*. 1999;33:365-370.
84. Gehlbach SH, Bobula JA, Dickinson JC. Teaching residents to read the medical literature. *J Med Educ*. 1980;55:362-365.
85. Ghali WA, Saitz R, Eskew AH, Gupta M, Quan H, Hershman WY. Successful teaching in evidence-based medicine. *Med Educ*. 2000;34:18-22.
86. Grad R, Macaulay AC, Warner M. Teaching evidence-based medical care: description and evaluation. *Fam Med*. 2001;33:602-606.
87. Haynes RB, Johnston ME, McKibbon KA, Walker CJ, Willan AR. A program to enhance clinical use of MEDLINE: a randomized controlled trial. *Online J Curr Clin Trials*. 1993;(No. 56).
88. Hunt DP, Haidet P, Coverdale JH, Richards B. The effect of using team learning in an evidence-based medicine course for medical students. *Teach Learn Med*. 2003;15:131-139.
89. Ibbotson T, Grimshaw J, Grant A. Evaluation of a programme of workshops for promoting the teaching of critical appraisal skills. *Med Educ*. 1998;32:486-491.
90. Jerant AF, Lloyd AJ. Applied medical informatics and computing skills of students, residents, and faculty. *Fam Med*. 2000;32:267-272.
91. Kellum JA, Rieker JP, Power M, Powner DJ. Teaching critical appraisal during critical care fellowship training: a foundation for evidence-based critical care medicine. *Crit Care Med*. 2000;28:3067-3070.
92. Khan KS, Awonuga AO, Dwarakanath LS, Taylor R. Assessments in evidence-based medicine workshops: loose connection between perception of knowledge and its objective assessment. *Med Teach*. 2001; 23:92-94.
93. Kitchens JM, Pfeifer MP. Teaching residents to read the medical literature: a controlled trial of a curriculum in critical appraisal/clinical epidemiology. *J Gen Intern Med*. 1989;4:384-387.
94. Kronick J, Blake C, Munoz E, Heilbrunn L, Dunikowski L, Milne WK. Improving on-line skills and knowledge: a randomized trial of teaching rural physicians to use on-line medical information. *Can Fam Physician*. 2003;49:312-317.
95. Lieberman SA, Trumble JM, Smith ER. The impact of structured student debates on critical thinking and informatics skills of second-year medical students. *Acad Med*. 2000;75(10)(suppl):S84-S86.
96. Linzer M, DeLong ER, Hupart KH. A comparison of two formats for teaching critical reading skills in a medical journal club. *J Med Educ*. 1987;62:690-692.
97. Lloyd FJ, Reyna VF. A web exercise in evidence-based medicine using cognitive theory. *J Gen Intern Med*. 2001;16:94-99.
98. McGlade KJ, McKeveney CJ, Crawford VL, Brannigan P. Preparing tomorrow's doctors: the impact of a special study module in medical informatics. *Med Educ*. 2001;35:62-67.
99. Neville AJ, Reiter HI, Eva KW, Norman GR. Critical appraisal turkey shoot: linking critical appraisal to clinical decision making. *Acad Med*. 2000;75(10)(suppl) :S87-S89.
100. Poyner A, Wood A, Herzberg J. A project to improve information technology skills for flexible trainees and overseas doctors. *Health Info Libr J*. 2004;21: 57-60.
101. Riegelman RK. Effects of teaching first-year medical students skills to read medical literature. *J Med Educ*. 1986;61:454-460.
102. Riegelman RK, Povar GJ, Ott JE. Medical students' skills, attitudes, and behavior needed for literature reading. *J Med Educ*. 1983;58:411-417.

103. Romm FJ, Dignan M, Herman JM. Teaching clinical epidemiology: a controlled trial of two methods. *Am J Prev Med*. 1989;5:50-51.
104. Rosenfeld P, Salazar-Riera N, Vieira D. Piloting an information literacy program for staff nurses: lessons learned. *Comput Inform Nurs*. 2002;20:236-241, 242-243.
105. Schoenfeld P, Cruess D, Peterson W. Effect of an evidence-based medicine seminar on participants' interpretations of clinical trials: a pilot study. *Acad Med*. 2000;75:1212-1214.
106. Schwartz DG, Schwartz SA. MEDLINE training for medical students integrated into the clinical curriculum. *Med Educ*. 1995;29:133-138.
107. Schwartz K, Northrup J, Israel N, Crowell K, Lauder N, Neale AV. Use of on-line evidence-based resources at the point of care. *Fam Med*. 2003;35:251-256.
108. Scott I, Heyworth R, Fairweather P. The use of evidence-based medicine in the practice of consultant physicians: results of a questionnaire survey. *Aust N Z J Med*. 2000;30:319-326.
109. Seelig CB. Affecting residents' literature reading attitudes, behaviors, and knowledge through a journal club intervention. *J Gen Intern Med*. 1991;6:330-334.
110. Srinivasan M, Weiner M, Breitfeld PP, Brahmi F, Dickerson KL, Weiner G. Early introduction of an evidence-based medicine course to preclinical medical students. *J Gen Intern Med*. 2002;17:58-65.
111. Starr SS, Renford BL. Evaluation of a program to teach health professionals to search MEDLINE. *Bull Med Libr Assoc*. 1987;75:193-201.
112. Steele G, Greenidge E. Integrating medical communication skills with library skills curricula among first year medical students at the University of the West Indies. *Health Info Libr J*. 2002;19:206-213.
113. Stevermer JJ, Chambliss ML, Hoekzema GS. Distilling the literature: a randomized, controlled trial testing an intervention to improve selection of medical articles for reading. *Acad Med*. 1999;74:70-72.
114. Thomas PA, Cofrancesco J Jr. Introduction of evidence-based medicine into an ambulatory clinical clerkship. *J Gen Intern Med*. 2001;16:244-249.
115. Wadland WC, Barry HC, Farquhar L, Holzman C, White A. Training medical students in evidence-based medicine: a community campus approach. *Fam Med*. 1999;31:703-708.
116. Wainwright JR, Sullivan FM, Morrison JM, MacNaughton RJ, McConnachie A. Audit encourages an evidence-based approach to medical practice. *Med Educ*. 1999;33:907-914.
117. Young JM, Glasziou P, Ward JE. General practitioners' self ratings of skills in evidence based medicine: validation study. *BMJ*. 2002;324:950-951.
118. Godwin M, Seguin R. Critical appraisal skills of family physicians in Ontario, Canada. *BMC Med Educ*. 2003;3:10.
119. Thom DH, Haugen J, Sommers PS, Lovett PC. Building on a block rotation to integrate evidence-based medicine into a residency program. *BMC Med Educ*. 2004;4:19.
120. Bergus G, Vogelgesang S, Tansey J, Franklin E, Feld R. Appraising and applying evidence about a diagnostic test during a performance-based examination. *BMC Med Educ*. 2004;4:20.
121. Khan KS, Dwarakanath LS, Pakkal M, Brace V, Awonuga A. Postgraduate journal club as a means of promoting evidence-based obstetrics and gynaecology. *J Obstet Gynaecol*. 1999;19:231-234.
122. Oliveri RS, Gluud C, Wille-Jorgensen PA. Hospital doctors' self-rated skills in and use of evidence-based medicine: a questionnaire survey. *J Eval Clin Pract*. 2004;10:219-226.
123. Naidr JP, Adla T, Janda A, Feberova J, Kasal P, Hladikova M. Long-term retention of knowledge after a distance course in medical informatics at Charles University Prague. *Teach Learn Med*. 2004;16:255-259.
124. Dorsch JL, Aiyer MK, Meyer LE. Impact of an evidence-based medicine curriculum on medical students' attitudes and skills. *J Med Libr Assoc*. 2004;92:397-406.
125. Cayley WE Jr. Evidence-based medicine for medical students: introducing EBM in a primary care rotation. *WMJ*. 2005;104:34-37.
126. Mukohara K, Schwartz MD. Electronic delivery of research summaries for academic generalist doctors: a randomised trial of an educational intervention. *Med Educ*. 2005;39:402-409.
127. McKenna HP, Ashton S, Keeney S. Barriers to evidence-based practice in primary care. *J Adv Nurs*. 2004;45:178-189.
128. Jette DU, Bacon K, Batty C, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. *Phys Ther*. 2003;83:786-805.
129. Connor E. Using clinical vignette assignments to teach medical informatics. *Med Ref Serv Q*. 2003;22:31-44.
130. Lawrence JC, Levy LS. Comparing the self-described searching knowledge of first-year medical and dental students before and after a MEDLINE class. *Med Ref Serv Q*. 2004;23:73-81.
131. Dee C, Stanley EE. Information-seeking behavior of nursing students and clinical nurses: implications for health sciences librarians. *J Med Libr Assoc*. 2005;93:213-221.
132. Linton AM, Wilson PH, Gomes A, Abate L, Mintz M. Evaluation of evidence-based medicine search skills in the clinical years. *Med Ref Serv Q*. 2004;23:21-31.
133. Pierce S. *Readiness for Evidence-Based Practice: Information Literacy Needs of Nursing Faculty and Students in a Southern US State* [dissertation]. Natchitoches: Northwestern State University of Louisiana; 2000.
134. Pravikoff DS, Tanner AB, Pierce ST. Readiness of US nurses for evidence-based practice: many don't understand or value research and have had little or no training to help them find evidence on which to base their practice. *Am J Nurs*. 2005;105:40-52.
135. Forsetlund L, Bradley P, Forsen L, Nordheim L, Jamtvedt G, Bjorndal A. Randomised controlled trial of a theoretically grounded tailored intervention to diffuse evidence-based public health practice. *BMC Med Educ*. 2003;3:2.
136. Akl EA, Izuchukwu IS, El-Dika S, Fritsche L, Kunz R, Schunemann HJ. Integrating an evidence-based medicine rotation into an internal medicine residency program. *Acad Med*. 2004;79:897-904.
137. Haynes RB, McKibbon KA, Walker CJ, Ryan N, Fitzgerald D, Ramsden MF. Online access to MEDLINE in clinical settings: a study of use and usefulness. *Ann Intern Med*. 1990;112:78-84.
138. McKibbon KA, Haynes RB, Dilks CJ, et al. How good are clinical MEDLINE searches? a comparative study of clinical end-user and librarian searches. *Comput Biomed Res*. 1990;23:583-593.
139. Schwartz A, Hupert J. A decision making approach to assessing critical appraisal skills. *Med Teach*. 2005;27:76-80.
140. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based: A-Team, Nuffield Department of Clinical Medicine. *Lancet*. 1995;346:407-410.
141. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123:A12-A13.
142. Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med*. 1983;309:709-713.
143. Green ML. Evidence-based medicine training in internal medicine residency programs a national survey. *J Gen Intern Med*. 2000;15:129-133.
144. Straus SE, Green ML, Bell DS, et al. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ*. 2004;329:1029-1032.
145. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? *Ann Intern Med*. 1985;103:596-599.
146. Whitcomb ME. Research in medical education: what do we know about the link between what doctors are taught and what they do? *Acad Med*. 2002;77:1067-1068.
147. Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med*. 2005;80:366-370.
148. Lynch DC, Surdyk PM, Eiser AR. Assessing professionalism: a review of the literature. *Med Teach*. 2004;26:366-373.