

Validation of an Evidence-Based Medicine Critically Appraised Topic Presentation Evaluation Tool (EBM C-PET)

HANS B. KERSTEN, MD
JOHN G. FROHNA, MD, MPH
ERIN L. GIUDICE, MD

Abstract

Background Competence in evidence-based medicine (EBM) is an important clinical skill. Pediatrics residents are expected to acquire competence in EBM during their education, yet few validated tools exist to assess residents' EBM skills.

Objective We sought to develop a reliable tool to evaluate residents' EBM skills in the critical appraisal of a research article, the development of a written EBM critically appraised topic (CAT) synopsis, and a presentation of the findings to colleagues.

Methods Instrument development used a modified Delphi technique. We defined the skills to be assessed while reviewing (1) a written CAT synopsis and (2) a resident's EBM presentation. We defined skill levels for each item using the Dreyfus and Dreyfus model of skill development and created behavioral anchors using a frame-of-reference training technique to describe performance for each skill level. We evaluated the

assessment instrument's psychometric properties, including internal consistency and interrater reliability.

Results The EBM Critically Appraised Topic Presentation Evaluation Tool (EBM C-PET) is composed of 14 items that assess residents' EBM and global presentation skills. Resident presentations (N = 27) and the corresponding written CAT synopses were evaluated using the EBM C-PET. The EBM C-PET had excellent internal consistency (Cronbach $\alpha = 0.94$). Intraclass correlation coefficients were used to assess interrater reliability. Intraclass correlation coefficients for individual items ranged from 0.31 to 0.74; the average intraclass correlation coefficients for the 14 items was 0.67.

Conclusions We identified essential components of an assessment tool for an EBM CAT synopsis and presentation with excellent internal consistency and a good level of interrater reliability across 3 different institutions. The EBM C-PET is a reliable tool to document resident competence in higher-level EBM skills.

Editor's Note: The online version of this article contains the Evidence-Based Medicine Critically Appraised Topic Presentation Evaluation Tool (EBM C-PET).

Introduction

Evidence-based medicine (EBM) integrates the best research evidence with clinical expertise and patient values

Hans B. Kersten, MD, is Associate Professor of Pediatrics at Drexel University College of Medicine and St. Christopher's Hospital for Children; **John G. Frohna, MD, MPH**, is Professor of Pediatrics and Medicine at the University of Wisconsin School of Medicine and Public Health; **Erin L. Giudice, MD**, is Assistant Professor of Pediatrics at the University of Maryland School of Medicine.

Funding: This work was supported by the Association of Pediatric Program Directors Special Projects Program.

Corresponding author: Hans B. Kersten, MD, Department of Pediatrics, St. Christopher's Hospital for Children, 3601 A Street, Philadelphia, PA 19134-1095, 215.427.8150, hkersten@drexelmed.edu

Received February 24, 2012; revision received July 3, 2012; accepted September 9, 2012.

DOI: <http://dx.doi.org/10.4300/JGME-D-12-00049.1>

to optimize clinical decision-making and outcomes for patients.^{1,2} An important part of clinical training, EBM has become an essential requirement for certification in different health professions.³ For residents and fellows, EBM is part of the practice-based learning and improvement competency requirements of the Accreditation Council for Graduate Medical Education.² The teaching of EBM is more effective if it is moved into clinical practice,⁴ and many residency programs have begun to incorporate the teaching of EBM skills into ambulatory rounds,⁵ short courses,^{6,7} block rotations,⁸ and integrated curricula.^{4,5} Residency programs have moved beyond developing focused EBM goals and objectives and are interested in tools to evaluate and document proficiency in EBM.

At the same time, assessing EBM skills remains a challenge.⁶ Despite the existence of many assessment tools, most of them focus on narrow aspects, such as measurement of EBM knowledge,⁷ or basic EBM skills, such as literature searches.^{6,8} Existing tools do not provide immediate feedback on an EBM presentation, can be

time-consuming to complete, may not help programs assess learners across the broad spectrum of venues where EBM is taught,^{6,8–11} and may lack evidence about reliability and validity.³ The 2011 Sicily Statement on evidence-based practice suggested that multiple assessments of learner's higher-level skills (eg, the completeness and relevance of a learner's clinical question, not just whether the question is structured) are needed, using different evaluation tools to facilitate an assessment of learners' skills in using EBM.³ Therefore, additional tools to assess the EBM skills of different levels of learners during their EBM presentations are needed.^{6,8}

An EBM-learning activity to appraise a research article and create a synopsis of the findings is called a critically appraised topic (CAT). Learners can also present their findings verbally in a journal club format. Faculty can assess trainee skills in critical appraisal, information synthesis, synopsis creation, and oral presentation using a combination of the written CAT and verbal presentation. We reviewed the literature and found there was no validated tool for assessing the EBM skills required to create and present an EBM CAT. Such an instrument would complement tools that assess knowledge, determine whether residents are progressing in their development of higher-level EBM skills, and provide a more global assessment of learners' EBM skills.^{3,7,12,13} Our primary aim was to develop a reliable tool to evaluate residents' skills in critically appraising a research article, developing a written EBM CAT synopsis, and presenting findings to colleagues. We called this tool the *EBM CAT Presentation Evaluation Tool* (EBM C-PET).

Methods

Setting

Each of the participating academic institutions (St. Christopher's Hospital for Children, the University of Wisconsin, and the University of Michigan) had existing EBM curricula for pediatrics residents, which included the preparation of a CAT and presenting findings to peers and faculty. Pediatrics residents completing this project were required to identify a clinical scenario, ask a clinical question, perform a literature search, select an article, critically appraise the selected

What was known

Competence in evidence-based medicine (EBM) is expected of pediatrics residents, yet few validated tools exist to assess EBM skills.

What is new

The 14-item Evidence-Based Medicine Critically Appraised Topic Presentation Evaluation Tool (EBM C-PET) assesses residents' EBM and global presentation skills.

Limitations

Raters who wish to use this tool would benefit from added frame-of-reference training.

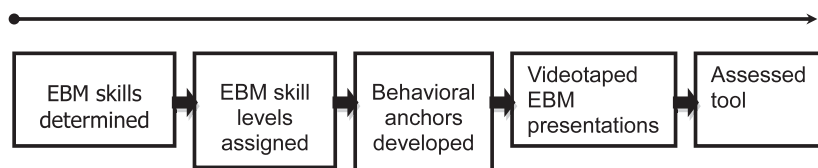
Bottom line

The EBM C-PET is a validated tool for documenting resident competence in higher-level EBM skills.

article, and apply the evidence back to their clinical scenario. The EBM learning activity included creating and presenting a CAT synopsis to their peers and faculty members in a small group setting. The existing EBM curricula at each institution were not modified. Residents developed their EBM-learning activities for presentation using a standardized format originally developed by David Sackett.¹

Evaluation Tool Development

Three investigators (the 3 authors), all with expertise in teaching and evaluating EBM for pediatrics residents, convened to develop the EBM C-PET tool (FIGURE). The investigators are a part of a group of 7 colleagues, nationally recognized within the pediatrics community for their contributions to curriculum development in EBM.^{14,15} The skills under consideration for the EBM C-PET tool were derived from a comprehensive EBM curriculum developed by the 7-member working group after a rigorous, iterative process that included searching the literature and reviewing existing EBM curricula and assessment tools. We used a modified Delphi technique to reach consensus on the optimal areas for inclusion in an EBM curriculum.¹⁶ We identified the items from the comprehensive EBM curriculum that would be incorporated into the EBM C-PET, resolving disagreements through discussion. The final tool underwent further review by additional experts to ensure adequate



FIGURE

EBM C-PET DEVELOPMENT TIMELINE

Abbreviation: EBM, evidence-based medicine.

representation of the items that would be included in the EBM C-PET tool.

Once the relevant EBM skills were determined, the investigators sought to categorize resident performance into 3 categories: *below expectation*, *developing*, or *proficient*. The 3 assessment categories were initially based on an existing EBM evaluation at one of the institutions. Behavioral anchors were created to describe performance in each of the 3 assessment categories for each of the items measured. The 3 investigators independently pilot-tested the tool on 2 sample CAT presentations/synopses, and the criteria to measure those domains were refined using a frame-of-reference training technique.^{17,18} The frame-of-reference training included evaluating the sample presentations and comparing the 3 investigators' ratings of the same presentations.¹⁸ This process informed the development of the tool and demonstrated that the initial 3 assessment categories were insufficient to reflect the full range of resident skills and to show meaningful differences in resident performances. The initial 3 categories were expanded to 5, using the Dreyfus and Dreyfus model of skill development. That model was adopted because it is based on skill acquisition, is familiar to educators, and could be readily used in the development of EBM assessment tools.^{19,20} The 5 terms in the Dreyfus and Dreyfus model (*novice*, *advanced beginner*, *competent*, *proficient*, and *expert*) were chosen to assess resident performance with the evaluation tool. For each of the 14 items measured, specific behavioral anchors were created to describe performance that would fit each of the Dreyfus and Dreyfus categories. After additional pilot testing, those behavioral anchors were further refined using an iterative process. Again, we used frame-of-reference training on our final EBM C-PET evaluation tool, discussing the 3 investigators' ratings of the same pilot presentations to help "calibrate" the investigators so they were able to discriminate between different levels of performance.¹⁸

Data Collection

Pediatrics residents' EBM CAT presentations were videotaped, and written CAT synopses were copied for distribution to the 3 investigators. Each investigator independently reviewed and scored each EBM CAT presentation and its corresponding written CAT synopsis using the EBM C-PET. Data were stored anonymously in an Excel (Microsoft Inc, Redmond, WA) database.

Each institution's Institutional Review Board granted exemption to the project before the start of data collection.

Statistical Analysis

Cronbach α coefficients²¹ were calculated to assess internal consistency using the individual item scores from each

participant, that is, items 1 to 14, for the complete EBM C-PET tool. A Cronbach α of 0.7 to 0.8 was considered acceptable, a value from 0.8 to 0.9 was considered good, and a value greater than 0.9 was considered excellent. The results of the scoring of the individual EBM presentations were analyzed with SPSS software (IBM Corporation, Armonk, NY) to determine the interrater reliability of the EBM C-PET. An intraclass correlation coefficient (ICC) was used to assess the interrater reliability. An ICC of 0.5 to 0.75 was considered moderate, and an ICC above 0.75 was considered good. A sample size of 30 was selected to provide a sufficiently narrow confidence interval around the observed ICC for the overall instrument.

Results

Participants in this study were 32 postgraduate year 2 residents from 3 different residency programs during a 1-year period. The presentation setting was a formal, 30-minute journal club conference at 2 of the institutions and a 15-minute preclinic conference at the third institution. Although a similar format for the EBM presentations and CAT synopses were used at all of the institutions, the median length of the presentations was 19.1 minutes (range, 6.2–41.2 minutes), and the length of the CAT synopses ranged from one-half of a page to 3 pages. Five EBM presentations and CAT synopses were used to pilot test the EBM C-PET and were not included in the final analysis, providing 27 subjects for analysis.

Tool Development

The final EBM C-PET assessment tool included 14 items (tool provided as online supplemental material). Eight of the items (57%) focused on EBM: develop a well-built question from a clinical scenario, understand and use key search terms, critically appraise the article, apply the critical appraisal to the clinical scenario, prepare a CAT, place the current study in context, explain EBM concepts, and provide an overall assessment of EBM skills. Five items (36%) emphasized the resident's teaching ability: organization and clarity, effective use of teaching tools, synthesis and integration, presentation conducted at level appropriate for audience, and an overall assessment of teaching skills. One global item (7%) assessed the resident's overall performance.

As an example of item development, the original scale for the item, "Places the current study in the context of other relevant research on the topic," included: (1) Does not place the current article into the larger context (*below expectation*), (2) Makes reference to 1 other source of evidence on the topic (*developing*), and (3) Integrates information from the current study into the larger literature (*proficient*). After adaptation to the Dreyfus and Dreyfus model, the scale included the following 5 items:

TABLE AGREEMENT AMONG 3 RATERS ACROSS SITES

Items	ICC Score (95% CI)
Develop a well-built (PICO) question from a clinical scenario	0.51 (0.28–0.71)
Understand key search terms and use them to identify relevant literature	0.63 (0.43–0.79)
Critically appraise an article in the style outlined in EBM text	0.51 (0.28–0.71)
Apply critical appraisal to clinical scenario	0.61 (0.40–0.78)
CAT alone addresses all salient points of the article	0.31 (0.09–0.55)
Places the current study in the context of other relevant research on the topic	0.74 (0.57–0.86)
Organization and clarity of presentation	0.42 (0.18–0.64)
Effective use of teaching tools (eg, PowerPoint, handout)	0.35 (0.11–0.59)
Synthesized and integrated relevant information	0.52 (0.29–0.72)
Presentation conducted at level appropriate for audience	0.50 (0.27–0.70)
Explanation of EBM concepts	0.63 (0.42–0.79)
Overall teaching skills, as demonstrated through this exercise	0.34 (0.11–0.58)
Overall EBM skills, as demonstrated through this exercise	0.49 (0.26–0.70)
Overall assessment	0.48 (0.26–0.69)
Overall score (all 14 items)	0.67 (0.48–0.82)

Abbreviations: 95% CI, 95% confidence interval; CAT, critically appraised topic; EBM, evidence-based medicine; PICO, Patient, population, or problem—Intervention, exposure, or prognostic factor—Comparisons—Outcomes.

- **Novice:** Does not place the current article into the larger context (ie, no background information)
- **Advanced beginner:** Makes reference to a few other sources of evidence on the topic *but does not* critically analyze their relationship to the current study
- **Competent:** Makes reference to a few other sources of evidence on the topic *and begins* to analyze their relationship to other studies
- **Proficient:** Integrates information from the current study into the larger literature using critical analysis
- **Expert:** Seamlessly integrates information from the current study into the larger literature using critical analysis *and* communicates how the article affects practice

Tool Performance

The EBM C-PET had excellent internal consistency, with a Cronbach α of 0.94 across all 3 raters. When considering the items on the tool that were used to assess the EBM presentation, Cronbach α was 0.91, and when considering the items assessing the CAT, Cronbach α was 0.84.

We used the ICC to assess interrater agreement. The ICC for each individual item on the assessment tool ranged from 0.31 to 0.74 (TABLE), and the average ICC for the 14 items was 0.67. We examined the lowest performing items to see whether the tool could be modified further, but there was no

clear pattern for excluding specific items. Thus, we maintained the original 14 items on the final EBM C-PET tool.

Discussion

The EBM C-PET, a novel tool for assessing higher-order EBM skills in residents, demonstrates good reliability and interrater agreement across 3 different residency programs. Our EBM C-PET tool was designed to be flexible and can be used in different clinical settings (clinic, inpatient, journal club, etc) within residency programs where EBM presentations may occur. The tool includes items that assess both the written CAT and the oral presentation. It is also easy to use while watching the presentation and reviewing the CAT. Perhaps most important, the tool uses behavioral anchors for each of the different assessment levels. These anchors facilitate faculty training with this tool and can be used to provide more specific feedback to the learner.²²

It has been well documented that additional tools are needed to teach and assess EBM skills in clinical practice.^{3,4,6} This tool assesses EBM skills in 4 areas identified in the Classification Rubric for Evidence-Based Practice Assessment Tools in Education (CREATE) framework in the latest Sicily statement on evidence-based practice³: ask, appraise, assess, and apply. Although not assessed in this study, the use of the Dreyfus and Dreyfus model as the assessment framework can facilitate its use

with other levels of learners (medical students, subspecialty fellows, etc) because it spans a novice-to-expert continuum.

We have documented excellent internal consistency for this tool and a good level of interrater reliability across 3 different institutions. The multi-institutional nature of this study adds to the generalizability of our findings. These strengths also form a foundation to begin to assess content validity, which involves an evaluation of the “relationship between a test’s content and the construct it is intended to measure.”²³ Along that line, we developed and selected items for the EBM C-PET based on a literature review and expert consensus of established EBM curricula. This process involved a modified Delphi technique and external review of the items before pilot testing. The use of frame-of-reference training allowed us to clarify the wording of individual items. Further implementation of the EBM C-PET by other educators would allow for further validation of the tool.

Our study has several limitations. First, the tool was validated in 3 settings and would benefit from further validation in other settings, particularly in settings with learners with varying levels of EBM skills to assess the sensitivity and specificity of the tool. Second, there was some variability in the intraclass correlation coefficients for the items on this tool. Although our primary interest was making those items practical and concise, we believe that more detail could have helped raters with scoring decisions and improved the ICC. Although a high level of correlation is always better, given that the EBM C-PET is a formative assessment, it should be combined with other assessment tools to evaluate competency in EBM. The variation in ICC also highlights the need for further frame-of-reference training for raters who may wish to use this instrument.

Conclusion

Being able to reliably assess the EBM skills of residents will help educators ensure that their trainees are meeting the goals outlined in their EBM curricula and promote lifelong learning. It will allow programs to demonstrate that their residents are meeting an essential subcompetency of the practice-based learning and improvement domain. Perhaps more important, it can also serve as a key step in fostering the use of EBM to improve the care of patients. The EBM C-PET is reliable and provides an important assessment tool for measuring and documenting residents’ critical appraisal skills and proficiency in EBM.

References

- 1 Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. 3rd ed. Edinburgh: Churchill Livingstone; 2005.
- 2 Accreditation Council for Graduate Medical Education. ACGME Program Requirements for Graduate Medical Education in Pediatrics. 2007. http://www.acgme.org/acgmeweb/Portals/o/PFAssets/ProgramRequirements/320_pediatrics_07012007.pdf. Accessed April 1, 2013.
- 3 Tilson JK, Kaplan SL, Harris JL, Hutchinson A, Ilic D, Niederman R, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Ed*. 2011;11:78–87. doi:10.1186/1472-6920-11-78.
- 4 Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? a systematic review. *BMJ*. 2004;329(7473):1017–1021.
- 5 Parkes J, Hyde C, Deeks JJ, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database of Systematic Reviews* 2001; Issue 11: Art. No.: CD001270. doi: 10.1002/14651858.
- 6 Shaneyfelt T, Baum KD, Bell D, Feldstein D, Houston TK, Kaatz S, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*. 2006;296(9):1116–1127.
- 7 Chernick L, Pusic M, Liu H, Vazquez H, Kwok M. A pediatrics-based instrument for assessing resident education in evidence-based practice. *Acad Pediatr*. 2010;10(4):260–265.
- 8 Straus SE, Green ML, Bell DS, Badgett R, Davis D, Gerrity M, et al. Society of General Internal Medicine Evidence-Based Medicine Task Force. Evaluating the teaching of evidence based medicine: conceptual framework. *BMJ*. 2004;329(7473):1029–1032.
- 9 Christakis DA, Davis R, Rivara FP. Pediatric evidence-based medicine: past, present, and future. *J Pediatr*. 2000;136(3):383–389.
- 10 Green ML, Ruff TR. Why do residents fail to answer their clinical questions? a qualitative study of barriers to practicing evidence-based medicine. *Acad Med*. 2005;80(2):176–182.
- 11 Frohna JG, Gruppen LD, Fliegel JE, Mangrulkar RS. Development of an evaluation of medical student competence in evidence-based medicine using a computer-based OSCE station. *Teach Learn Med*. 2006;18(3):267–272.
- 12 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(7376):1338–1341.
- 13 Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *BMJ*. 2003;326(7384):319–321.
- 14 Kersten HB, Thompson ED, Frohna JG. The use of evidence-based medicine in pediatrics: past, present and future. *Curr Opin Pediatr*. 2008;20(3):326–331.
- 15 Kersten H, Thompson ED, Frohna JG. Competency-based evaluation of EBM skills in pediatric residency and fellowship programs [abstract E-PAS2006:59]. In: *Pediatric Academic Societies Annual Meeting 2006*; San Francisco, CA: Pediatric Academic Societies; 2006.
- 16 Powell C. The Delphi technique: myths and realities. *J Adv Nurs*. 2003;41(4):376–382.
- 17 Holmboe ES. Direct observation by faculty. In: Holmboe ES, Hawkins RE, eds. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby Co; 2008:119–129.
- 18 Norcini JJ, Holmboe ES, Hawkins RE. Evaluation challenges in the era of outcomes-based education. In: Holmboe ES, Hawkins RE, eds. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby Co; 2008:1–9.
- 19 Dreyfus HL, Dreyfus SE. *Mind Over Machine*. New York, NY: Free Press; 1986.
- 20 Carraccio CL, Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Acad Med*. 2008;83(8):761–767.
- 21 Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. *Educ Psychol Meas*. 2004;64(3):391–418.
- 22 Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med*. 2003;15(4):270–292.
- 23 Cook DA, Beckham TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):e7–e16.