

A validated search assessment tool: assessing practice-based learning and improvement in a residency program

Gurpreet K. Rana, MLIS; Doreen R. Bradley, MILS; Stanley J. Hamstra, PhD; Paula T. Ross, MA; Robert E. Schumacher, MD; John G. Frohna, MD, MPH; Hilary M. Haftel, MD; Monica L. Lypson, MD

See end of article for authors' affiliations.

DOI: 10.3163/1536-5050.99.1.013

Objective: The objective of this study was to validate an assessment instrument for MEDLINE search strategies at an academic medical center.

Method: Two approaches were used to investigate if the search assessment tool could capture performance differences in search strategy construction. First, data from an evaluation of MEDLINE searches from a pediatric resident's longitudinal assessment were investigated. Second, a cross-section of search strategies from residents in one incoming class was compared with strategies of residents graduating a year later. MEDLINE search strategies formulated by faculty who had been identified as having search expertise were used as a gold standard comparison. Participants were presented with a clinical scenario and asked to identify the search question and conduct

a MEDLINE search. Two librarians rated the blinded search strategies.

Results: Search strategy scores were significantly higher for residents who received training than the comparison group with no training. There was no significant difference in search strategy scores between senior residents who received training and faculty experts.

Conclusion: The results provide evidence for the validity of the instrument to evaluate MEDLINE search strategies. This assessment tool can measure improvements in information-seeking skills and provide data to fulfill Accreditation Council for Graduate Medical Education competencies.

INTRODUCTION

The acquisition of strong information-searching skills is one of the many building blocks upon which medical knowledge and clinical skills are built. The Accreditation Council for Graduate Medical Education (ACGME) has listed practice-based learning and improvement (PBLI) as a core competency, defining PBLI as the ability to (1) investigate and evaluate the care of patients, (2) appraise and assimilate scientific evidence, and (3) continuously improve patient care based on constant self-evaluation and lifelong learning [1]. Of these, the ability to appraise and assimilate scientific evidence is infrequently taught in many postgraduate medical education programs, and currently no one standardized method is used to assess this competency [2].

Being able to search the medical literature effectively is an essential skill in the practice of evidence-based medicine (EBM). Moreover, developing and maintaining strong literature searching skills (use of appropriate search keywords, use of appropriate limits, development of evidence-based search strategies, and use of correct search syntax) help physicians identify appropriate literature that affects medical decision making [3]. Due to their training in the areas

Highlights

- The University of Michigan MEDLINE Search Assessment tool can be used to assess search skills in residency education.
- Five elements were identified as *critical elements* in the development of an effective MEDLINE search strategy: inclusion of all search concepts, appropriate use of Medical Subject Headings, appropriate use of search limits, successful combination of all concepts, and search efficiency.

Implications

- This validated assessment tool can serve as an effective means to measure improvements in residents' information-seeking skills and provide data to fulfill Accreditation Council for Graduate Medical Education competencies.

of MEDLINE search strategies and search retrieval, medical librarians are recognized as specialists in biomedical literature searching and search assessment [4].

Historically, a variety of methods and tools have been used to assess residents' search skills [5]. Instruments have typically been based on the development of a gold standard search strategy, as determined by an expert searcher, against which all residents' searches for a specific clinical question are compared. The comparison includes identified ele-



This article has been approved for the Medical Library Association's Independent Reading Program <<http://www.mlanet.org/education/irp/>>.



A supplemental appendix is available with the online version of this journal.

ments determined by the expert searchers and agreed upon search criteria that would be included in an effective and efficient search. Although the ACGME mandates outcome assessment in residency programs, the authors do not currently know of any validated instruments that specifically measure residents' MEDLINE searching performance [6, 7]. The goal of this study was to validate an assessment instrument for MEDLINE search strategies.

METHODS

Tool development

Interest at the University of Michigan in developing a validated tool dates back to 2001, when a residency program director and two librarians developed a blinded randomized controlled trial to measure librarians' impact on residents' searching and retention of skills while on rotation in the neonatal intensive care unit [6]. Upon determining that a viable MEDLINE search assessment tool was not available, the team began developing a new assessment tool. In evaluating the resident searches, a gold standard search on a relevant topic was established by using all identified search criteria and searching elements. Search scores were determined by the librarians, and points were deducted for common search errors identified in the literature [6]. This method of search strategy assessment was found to have limitations. It was too subjective and lacked the flexibility needed to measure the variety of search skills among searchers.

Subsequently, in 2002, Nesbit and Glover developed a search assessment instrument based on feedback gathered in a national survey of medical librarians, in which the librarians were asked to weight the importance of a series of search elements for a successful search [8]. This instrument was an improvement over the "gold standard" assessment technique, because it allowed greater objectivity and provided a more efficient means to measure search skills. The structure of the instrument allowed assessment with greater flexibility. With permission from the authors, the instrument template was adapted by adding additional search elements and a section to allow subtraction of points for search errors. The new instrument is called the University of Michigan MEDLINE Search Assessment (UMMSA). UMMSA maintains flexibility and can be used in various biomedical and health sciences curricula. It is not necessarily limited to assessing search skills in clinical medicine.

The UMMSA tool is a matrix that measures the use of identified search elements and other important criteria. UMMSA is customized to apply to a specific question. Because the first step in effectively searching the medical literature is to identify a "searchable question," the searcher is expected to identify an appropriate search question, based on an introduced clinical scenario, as well as an appropriate search strategy. There are eleven search criteria for which

points are awarded and two search elements for which points are deducted (Table 1).

Data collection

Data were collected from residents in the department of pediatrics using the UMMSA. Because the ACGME competencies had been recently introduced and required supporting documentation of competency achievement, the department of pediatrics had a specific interest in finding a meaningful method to assess information retrieval skills as a part of the PBLI competency. The pediatrics liaison librarian in collaboration with department of pediatrics designed the assessments.

Participants

Participants in this study included all 22 incoming pediatric and medicine-pediatric first-year residents in 2004 (cohort 1); 10 self-selected pediatric residents who graduated the year later, in 2005 (cohort 2); and 9 faculty volunteers who had been identified by their expertise in EBM concepts and principles. These faculty members had EBM teaching responsibilities, demonstrated professional knowledge, and published on the topic of EBM. The 10 self-selected residents in cohort 2 were recruited through an email message sent out to all third-year residents by the department of pediatrics. These 10 residents represented a 50% participation rate, out of a total 20 third-year residents. No incentives were offered for their participation. The incoming residents (cohort 1) were tested twice, once in 2004, just after receiving training in MEDLINE searching and once just before graduation. The 9 faculty volunteers and the self-selected pediatric residents were tested in July 2008. Searchers were provided with directions to email the searches directly from the Ovid MEDLINE interface to the librarians conducting the assessment.

In 2004, incoming pediatric residents received basic MEDLINE training from librarians during their residency orientation, after which their skills were assessed using the UMMSA. Prior to that, pediatric residents had not received formal MEDLINE training. Opportunities for additional MEDLINE training varied depending on emphasis placed on MEDLINE search skills in the different clinical rotations and the significance placed on EBM skills, including searching of the medical literature, by various attending physicians.

Search scoring procedure

Two librarians from the University of Michigan's Health Sciences Libraries (authors Rana and Bradley) rated each search strategy performance in a blinded fashion, using the new instrument. Study participants were presented with a case of a pediatric patient with bronchiolitis, asked to identify an appropriate search question, and conduct a MEDLINE search (Appendix, online only). After reading the clinical scenario,

Table 1
Search assessment criteria

Criterion	Points available	Significance
Points awarded		
1. Included all search concepts	5 per concept	Searcher identified all the key search terms
2. Used Medical Subject Headings (MeSH)	5 per concept	Searcher mapped search terms to and used the precision of MeSH terms whenever possible
3. Exploded MeSH concepts	5 per concept	Searcher exploded MeSH terms used, thus creating a more comprehensive search by picking up literature indexed under narrow terms to the broader exploded MeSH term
4. Used proper MeSH subheadings	5 per concept	Searcher used subheadings in conjunction with a MeSH term in the proper context
5. Used appropriate age limits	5 per concept	Searcher used age limits to narrow search results to the population under consideration (e.g., all adult, all child, infant, preschool, senior, aged)
6. Used appropriate search limits	5 per concept	Searcher limited search retrieval by rational limits (e.g., English language, human [versus animal] population)
7. Focused one or more MeSH term	5	Searcher created a more precise search by tagging one or more MeSH search terms with a "major topic" designation
8. Used appropriate Boolean term	5	Searcher correctly applied Boolean logic using "AND," "OR," and "NOT"
9. Combined all concepts	10	Searcher included all the key search concepts in the final search strategy
10. Looked for evidence (<i>Limit by publication type, use of evidence-based medicine (EBM) filter, or limit to EBM reviews</i>)	10	Searcher recognized the clinical search question appropriately (i.e., therapy, diagnosis, prognosis, or causation/harm) and used a search strategy to find literature reflecting the search category
Points subtracted		
1. Used inappropriate limits	-5 per limit	Searcher used search limits that did not make sense based on the clinical question or limited every term separately
2. Used incorrect term, including subheading	-5 per incorrect term or subheading	Searcher used a search concept that changed the nature of the search to retrieve irrelevant literature or used a search term in an incorrect context

Please contact Gurpreet K. Rana at preet@umich.edu for a copy of the actual search assessment tool.

searchers were expected to identify the clinical question: "Are bronchodilators effective in treating bronchiolitis in infants?"

Based on the clinical question derived from the prepared scenario, the two librarians derived a gold standard search strategy in Ovid MEDLINE. The identified gold standard search strategy included the elements that both librarians agreed were required in an effective MEDLINE search to address this specific clinical question. The raters used UMMSA to independently assess the search strategy effectiveness for each of the study participant's searches.

The Ovid MEDLINE search interface was used because, at the time of this study, the process of search strategy development was distinguishable and search strategies were easily captured using the Ovid interface, as opposed to the PubMed interface. The searcher's thought process was more apparent when viewing a submitted Ovid MEDLINE strategy. The searcher's choice of Medical Subject Headings (MeSH) based on the natural language term entered could be assessed. Thus, an aspect of the searcher's information-seeking behavior would be revealed by their selection of appropriate MeSH terms. In the basic PubMed search, terms are mapped to MeSH automatically without the searcher needing to select appropriate headings, so the searcher's judgment could not be assessed in this same fashion.

Study design

Two approaches were used to validate the instrument's ability to measure improvements in search strategy that would be expected from training and experience.

1. Twenty-two incoming residents in pediatrics received training in MEDLINE searching as part of their residency orientation. Their search skills were tested following instruction in 2004 and again, upon graduation, in 2007. To assess improvement in scores, fifteen of the incoming residents of cohort 1 were assessed on the same case prior to graduation (of the original twenty-two residents, seven either dropped out of the program or moved on quickly near graduation time).

2. The scores of these same twenty-two incoming residents were compared with the scores of the ten volunteer senior residents, who had not received any training, and the nine faculty volunteers.

An apparent and noted trend during the grading of searches was that some searchers ran "test searches" to measure retrieval using various strategies or tried different combinations of searchable concepts and limits (Table 2). These "test searches" were easily distinguished when the librarians viewed the submitted search strategies. Due to such experimentation with, testing of, and repetition of terms or search features (i.e., limiting every term separately), it might appear as if a searcher conducted a very lengthy search strategy. For instance, it might seem as if a very in-depth and lengthy search strategy was created when, in truth, the searcher only used two search terms in the final search strategy. The librarians only graded the final identified search strategy. In many cases, finding the search elements interwoven through "test searches" and recognizing how they were combined in the very final search strategy was a complicated process.

A level of "test searching" is acceptable, as it can be considered a natural part of generating a search strategy. However, if a searcher conducts a great deal

Table 2
 "Test searching": a submitted MEDLINE search strategy

1	exp Bronchiolitis/th [Therapy] (518)
2	limit 1 to (abstracts and english language and humans) (241)
3	exp Bronchodilator Agents/ (189078)
4	2 and 3 (25)
5	limit 4 to "all infant (birth to 23 months)" (24)
6	from 5 keep 2,6 (2)
7	kellner j.au. (35)
8	gadomski a.au. (25)
9	gadomski am.au. (20)
10	exp Bronchiolitis/dt [Drug Therapy] (907)
11	3 and 10 (213)
12	limit 11 to (abstracts and english language and "all infant (birth to 23 months)") (108)
13	from 12 keep 9,14,17,25,34-37,45 (9)
14	exp Bronchodilator Agents/tu [Therapeutic Use] (22784)
15	10 and 14 (163)
16	limit 15 to (abstracts and english language and "all infant (birth to 23 months)") (83)
17	from 16 keep 17,24,28 (3)
18	9 and 10 (7)
19	9 and 14 (7)
20	limit 19 to (abstracts and english language and "all infant (birth to 23 months)") (3)
21	from 13 keep 1,4,6 (3)
22	from 19 keep 1 (1)
23	21 or 22 (4)
24	from 23 keep 1-4 (4)

of "test search" strategizing, creating very long search strings, it is no longer considered a strategy that is both efficient and effective. Points were deducted for lack of search efficiency in such searches.

RESULTS

Reliability

To establish inter-rater reliability, both raters rated each of the 9 faculty search results. Inter-rater reliability was very high (Pearson's correlation coefficient=0.962, $P<0.0001$).

Construct validity

Significant improvements were found when comparing the 2004 and 2007 scores for pediatric and medicine-pediatric first-year residents (cohort 1) (mean improvement: 51.7 to 78.7; $t(14)=5.43$, $P<0.0001$). Search strategy scores were significantly higher for 2004 residents (who received training) upon graduation than for the 9 senior residents who had received no training (cohort 2) (median 85.0 vs. 65.0; Wilcoxon chi-square(1)=4.09, $P=0.043$). There was no significant difference in search strategy scores between the graduating residents who received training (cohort 1) and the 9 faculty experts (Wilcoxon chi-square(1)=3.82, $P=0.050$).

Item analysis

Using the average ratings of both raters, item statistics were evaluated to determine whether any of the items could be consolidated with other items. Item-total score correlations ranged from -0.46 to 0.91, with the internal consistency for the overall scale of 0.63 (Cronbach's alpha). After removing the items with

the lowest item-total correlations, the range of item-total score correlations improved to 0.67 to 0.84, with a corresponding increase in internal consistency to 0.84 (Cronbach's alpha). The items removed were "exploded MeSH concepts," "used proper subheadings," "used limits: human & English," "focused on one or more concepts," "inappropriate limits," and "incorrect terms." In addition to statistics identifying these items as yielding the lowest item-total scores, librarians, in their judgment as expert searchers, concurred that these would be low predictors of quality searches.

This process resulted in five specific search elements being identified as critical in the development of an effective MEDLINE search strategy:

1. included all search concepts
2. used MeSH
3. used appropriate search limits
4. combined all concepts
5. had an efficient search

A sixth item, "looked for evidence," had a moderate correlation, so, while not as critical, it was retained in the final assessment form.

DISCUSSION

Residency programs have identified information competencies in their programs, including the skill to conduct an effective search of the medical literature search as one of several information competencies [1]. Historically, such skills have not been assessed rigorously in a standardized manner or considered to be a foundation stone on which medical knowledge and clinical skills are built. This is despite the need to develop competency-based outcomes for PBLI as put forth by the ACGME [1]. Some might consider literature searching skills to be intangible and somewhat subjective. Successful searching of the medical literature is often described as a combination of art and science, and there is more than one way that an effective and efficient search can be conducted. Thus, development of a rubric for effective search strategies (that garner germane articles in search results and that are recognized as having identifiable characteristics and elements) is all the more important for clinical learning. This study demonstrates that it is possible to create a search assessment tool.

In this study, a formal assessment was conducted of the construct validity of the UMMSA to further develop and refine the instrument according to accepted psychometric standards. The main criterion for this analysis was determining the sensitivity of the instrument in detecting expected differences in performance. To investigate the reliability and construct validity of this assessment instrument, two groups were used that could be expected to differ in performance, interns and resident. This study also sought to determine the performance of faculty on such a tool. The rationale was that if a difference could be measured between the two groups of interns and residents, it would provide promise for further refinement of the assessment. This study has shown

that the tool can measure the difference in quality of search strategies between internship and residency. In addition, it showed no difference in skill level between faculty and senior residents. These findings demonstrated construct validity for the tool itself as a useful means to measure search strategy.

Evidence for validity included: (1) The tool was developed by expert librarians, and when used, it assessed search strategies that led to useful clinical information; (2) the structure of the tool had been piloted in other settings; and (3) there was a difference in performance expected for the level of expertise of the learner [9].

CONCLUSION

The results of this study provide evidence of the validity of an instrument to evaluate MEDLINE search strategies. This assessment tool can serve as an effective means to measure improvements in residents' information-seeking skills and provide data to fulfill ACGME competencies. Steps to further the utility of the search strategy assessment tool will include revision of the search instrument to focus on the identified *critical search elements* and use of the data collected from the UMMSA instrument to consider its utility in measuring the search elements that are vital to measuring success in information seeking.

REFERENCES

1. Accreditation Council for Graduate Medical Education. Outcomes project: common program requirements: general competencies [Internet]. Chicago, IL: The Council [cited 30 Sep 2009]. <<http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>>.
2. Morrison L, Headrick L. Teaching residents about practice-based learning and improvement. *Jt Comm J Qual Patient Saf.* 2008 Aug;34(8):453-9.
3. Doig GS, Simpson F. Efficient literature searching: a core skill for the practice of evidence-based medicine. *Intensive Care Med.* 2003;29(12):2119-27.
4. Smith CA. An evolution of experts: MEDLINE in the library school. *J Med Libr Assoc.* 2005 Jan;93(1):53-60.
5. Shaneyfelt T, Baum K, Bell D, Feldstein D, Houston T, Kaatz S, Whelan C, Green M. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA.* 2006 Sep 6;296(9):1116-27.

6. 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 Apr;90(2):194-201.
7. Brettle A. Evaluating information skills training in health libraries: a systematic review. *Health Libr Rev.* 2007;24 (suppl 1):18-37.
8. Nesbit K, Glover J. Ovid Medline search strategy scoring sheet. Rochester, NY: University of Rochester; 2002.
9. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.

AUTHORS' AFFILIATIONS

Gurpreet K. Rana, MLIS* (corresponding author), preet@umich.edu, Coordinator, Library Global Initiatives; **Doreen R. Bradley, MILS†**, dbradley@umich.edu, Director, User Education Initiatives; University Library, 1135 Catherine Street, University of Michigan, Ann Arbor, MI; **Stanley J. Hamstra, PhD**, stan.hamstra@gmail.com, Associate Director, Academy for Innovation in Medical Education, Faculty of Medicine, University of Ottawa, Ottawa, Canada; **Paula T. Ross, MA**, paulat@umich.edu, Research Area Specialist, Department of Health Behavior and Health Education, School of Public Health; **Robert E. Schumacher, MD**, neoschu@umich.edu, Professor, Department of Pediatrics, Medical School; University of Michigan, Ann Arbor, MI; **John G. Frohna, MD, MPH**, jfrohna@pediatrics.wisc.edu, Associate Professor, Departments of Pediatrics and Department of Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, WI; **Hilary M. Haftel, MD**, hils@umich.edu, Associate Professor, Departments of Pediatrics and Internal Medicine, Medical School, University of Michigan, Ann Arbor, MI; **Monica L. Lypson, MD**, mlypson@umich.edu, Assistant Dean of Graduate Medical Education and Associate Professor of Internal Medicine, Arbor VA Healthcare System and Department of Internal Medicine, Medical School, University of Michigan, Ann Arbor, MI

Received May 2010; accepted August 2010

* At the time of this study, Gurpreet Rana was clinical education librarian, Health Sciences Libraries, University of Michigan, Ann Arbor, MI.

† At the time of this study, Doreen Bradley was liaison services coordinator, Health Sciences Libraries, University of Michigan, Ann Arbor, MI.