Medical Teacher, Vol. 28, No. 4, 2006, pp. 318–325

Taylor & Francis
Taylor & Francis Group

Evidence-based medicine: discrepancy between perceived competence and actual performance among graduating medical students

OPHER CASPI[1,2], PATRICK MCKNIGHT[2], LILLIAN KRUSE[2], VICTORIA CUNNINGHAM[2], AURELIO JOSE FIGUEREDO[2] & LEE SECHREST[2]
[1]Rabin Medical Center, Israel; [2]The University of Arizona, USA

ABSTRACT *Since at the time of graduation from medical school physicians are expected to demonstrate adequate professional competence including mastery of critical appraisal skills, we conducted a preliminary, cross-sectional, web-based study to examine the extent to which fourth year medical students in the US are competent in core areas of evidence-based medicine (EBM). Using self-assessment instruments, subjects (n = 150) were asked to demonstrate their ability to understand the practical meaning of key methodological and data analysis constructs as they relate to patient care, to rate their perceived competence in core areas of EBM and to disclose their attitudes toward critical appraisal of the literature and EBM. The mean score in our cohort was 55% suggesting that students may have knowledge gaps that interfere with their ability to critically appraise the medical literature. There was an apparent chasm between subjects' perceived competence and their actual performance on the assessment instrument. These findings, if corroborated in larger studies, (1) suggest that better education in EBM is needed so as to avoid the possibility that patient care may inadvertently be jeopardized; and (2) cast doubt on the use of self-assessed knowledge as a proxy for actual skills with respect to EBM and medical decision-making.*

Practice points

- Graduating medical students in the US appear to have knowledge gaps that interfere with their ability to critically appraise the medical literature.
- Judicious use of the medical research literature is crucial to patient care. We must ensure that physicians-in-training develop adequate competence in EBM.
- The apparent chasm between students' perceived competence and their actual performance in EBM and medical decision-making cast doubt on the use of self-assessed knowledge as a proxy for actual skills with respect to EBM and medical decision-making.
- More research is needed in order to explore the impact that transdisciplinary training in both decision sciences and EBM may have on physicians' behavior, patient care and health outcomes.
- Both authors and journal editors should provide readers with information that is meaningful to the non-expert so as to avoid misinterpretations of the clinical research literature.

Introduction

The practice of evidence-based medicine (EBM) depends primarily on physicians' ability to critically appraise the validity of research evidence and successfully incorporate it into patient care (Guyatt & Rennie, 2002). To do so, physicians must first have good knowledge and understanding of research methodology and data analysis (Taylor *et al.*, 2000). Yet, concerns have been raised that many physicians lack methodological competence needed to make sound judgments about research quality (Schwartz *et al.*, 1982) and insufficient knowledge and inadequate competencies in research methods have been identified as barriers toward the practice of EBM (El Ansari, 2004). Improper understanding of research evidence may inadvertently result in suboptimal care and inappropriate use of resources.

Like many other professional skills, competence in critical appraisal of the literature, medical decision-making, and EBM is developmental (Epstein & Hundert, 2002). Yet, relatively, very little is known about the nature of this developmental process. Sintyce graduation from medical school represents a point in time along the training process in which physicians are expected to demonstrate adequate professional competence (Blumenthal *et al.*, 2001), we conducted a preliminary study in preparation for a large scale educational initiative in EBM. Our goals for this preliminary study were to (1) examine the extent to which graduating (fourth year) medical students in the United States are competent in these areas; (2) determine the relationship between their actual performances in these areas and their self-perceived competence; and (3) assess their attitudes toward critical appraisal of the literature and their perception of the current state of affairs of EBM. Based on a review of the literature, we hypothesized that respondents would (1) be sub-optimally prepared in these areas; (2) report greater confidence in their competence in these areas than would be warranted given their actual performance; and (3) have mixed feelings toward critical appraisal of the literature.

Correspondence: Opher Caspi, MD, PhD, Vice Chair, Recanati Center for Medicine and Research Director, Integrative Medicine, Rabin Medical Center (Beilinson Campus), Petah Tikva, Israel 49100. Tel: +972 3 9377372; fax: +972-8-9 262 144; email: ocaspi@ahsc.arizona.edu

ISSN 0142–159X print/ISSN 1466–187X online/06/040318–8 © 2006 Taylor & Francis
DOI: 10.1080/01421590600624422

Methods

Setting

The study was carried out during the 2002–2003 academic period. Given the preliminary nature of the study and privacy protection laws that prohibit access to, or information about, medical students in the US, using a random and representative sample of all fourth year graduating medical students in the US was impossible. We therefore used the Internet instead as a means of soliciting replies from as many and as diverse a range of responders as possible. This diversity seemed critical given two factors. First, US medical schools lack a standardized curriculum in research and its related fields of inquiry (Fraker *et al.*, 1996; Looney *et al.*, 1998). Second, because graduating medical students differ greatly in terms of their past experience and personal interest in these areas.

Participants

Eligible study participants had to be fourth year medical students, be fluent in English and be willing to complete the study. Potential subjects were invited to participate through e-mail messages sent to various listservs hosted by medical schools across the US. All subjects were assured confidentiality. Interested subjects were referred to an introductory web page where they found a short explanation of the purpose of the study and its format. Agreeing to participate in the study after reading the introductory page was taken to represent consent. All study protocols and procedures were approved by the University of Arizona Institutional Review Board.

Design

The online study had two parts. In Part I, we asked subjects to complete a survey summarizing their past educational and research experiences, to disclose their opinions about current topics in EBM and to assess their self-perceived competence in five related fields: biostatistics and data analysis, research methodology, critical appraisal of the literature, EBM and decision analysis and clinical reasoning. Part II constituted the assessment itself. It included ten multiple-choice questions designed to assess subjects' ability to understand the medical literature as it applies to medical decision-making. All the questions were clinically oriented, involved no calculations and resembled common problems encountered in medical practice. To avoid confusion, subjects were instructed to consider only the information provided in the questions. No prior disease-specific knowledge was required.

Instrument development

Whereas most available assessment instruments in EBM merely require subjects to define methodological constructs as a way to demonstrate knowledge acquisition (Wulff *et al.*, 1987; Young *et al.*, 2002), we were more interested in the application of that knowledge. That is, we wanted subjects to demonstrate their ability to apply the practical meaning of these constructs as they relate to patient care, rather than just provide us with dictionary denotations. We therefore had to create and pilot test our own instrument. Since a full description of the developmental and validation process of this instrument is beyond the scope of this paper, we hereby provide a short summary of this process.

To maximize the likelihood of subject participation, we developed an instrument that respondents would be able to complete within up to 30 minutes. We limited our instrument to ten questions focusing on common constructs in EBM that are directly related to the practice of medicine. The selection process of these constructs was done with the goal to maximize the instrument's construct and content validity. To do so, we reviewed at random periodicals from various medical domains so as to create a representative list of methodology and data analysis terms commonly used in the medical literature. These terms constituted the universe from which we chose the following common constructs as the subject matter of our questions: statistical significance, number needed to treat, odds ratio, confidence intervals, sensitivity and specificity, randomization and placebo control, effect size, intention-to-treat analysis, diagnostics and graphical interpretation of correlations. Two criteria guided us in choosing these constructs—they had to be used frequently in the clinical trial literature and they had to have a bearing on medical decision-making, i.e., different values of the construct were likely to result in different decisions.

To achieve maximum internal validity and maintain authenticity, for each construct we created a short question that reported on real findings from the medical literature. Students and faculty members were invited to comment on the difficulty level and face validity of those questions. Their feedback was then used in a bootstrapping process to finalize the instrument (the complete set of items is available upon request from the first author). All the questions were constructed so that responses would not depend on any prior, broad, disease-specific knowledge. For example:

> A recent randomized controlled trial of hydrocortisone treatment for chronic fatigue syndrome (CFS) showed a treatment result that neared the threshold for statistical significance, $p = 0.06$ (McKenzie *et al.* (1998), JAMA, 280, pp. 1061–1066). Considering only this study, what significance does this finding bear in the evidence-based management of future patients with chronic fatigue syndrome?
>
> (a) The p value is not significant indicating that the magnitude of hydrocortisone effects on CFS is not substantial.
>
> (b) The p value reported indicates that the findings are significant at the 95% confidence level.
>
> (c) The information provided is not sufficient to decrease my uncertainty about hydrocortisone treatment. At a minimum, I would need the effect size to better appreciate the study's contribution.
>
> (d) There exists a 6% chance that hydrocortisone is a strong treatment for CFS so I might consider prescribing it to my patients.

Data management and analysis

Respondents entered data directly into web-based forms. Internet addresses and time stamps served as the subjects'

identification variable. All analyses used all available cases (i.e., pairwise deletion) where appropriate. The web administration and measure scoring provided no prompt for missing data and made no provisions for guessing. A Rasch model approach (Wright & Stone, 1979) was used to provide information on the relative difficulty of each item and the relative performance of each respondent independently. Rasch analyses were performed using Winsteps (Linacre, 2000) and linear models analyses were performed using R (Ihaka & Gentleman, 1996). Statistical significance was determined *a priori* at $\alpha \leq 0.05$, two-tailed.

Results

Sample characteristics

A total of 202 participants responded to our online survey. One hundred and sixty-two completed the background survey (Part I) and 150 completed the assessment instrument (Part II). Forty-five percent of the respondents were female. Mean age was 27.4 (SD = 3.5). Seventy-two percent of the sample had some prior research experience. Thirty-nine percent reported planning to go into private/group practice, 21% into academics, 5% into research and 35% were not sure yet. On average, respondents reported spending 2.4 hours (SD = 2.6) reading 3.2 medical papers (SD = 7.2) and skimming 4.9 (SD = 6.6) abstracts per week.

Subjects' performance

The entire sample answered only approximately half of the questions correctly (mean raw score = 5.5 (SD = 2.0) out of a maximum of 10 points possible) despite the fact that most questions fell well within the range of ability for most respondents (Table 1). One respondent was incapable of answering any question correctly and only four respondents answered all questions correctly. As expected, subjects successfully answering the more difficult questions tended to score higher on the assessment.

Rasch model results provide information suitable for understanding the relative difficulty of each item and the relative performance of each respondent independently. Figure 1 shows the distribution of respondents' abilities along with curves representing each question. The curves model the probability a respondent will answer a question correctly given the respondent's ability (theta, in Rasch modeling terms).

Subjects' self-perceived competence

An important goal of our study was to obtain a general estimate of the extent to which subjects perceive themselves as competent in the five interrelated content domains of the study. Competence was operationally defined as the ability to understand *and* apply knowledge. Subjects were first asked to rank their current level of competence and then to compare their own competence to that of their peers. Tables 2 and 3 show the frequency in each response category for these questions.

Most subjects perceived their competence to be at least 'fair' and equal to that of their peers across all five content areas. Further indication of self-perceived competence was obtained by asking subjects to estimate the extent to which they (1) understand the terms used in the methods and results sections of clinical trial reports (9% said they typically understand all terms, 70% said they understand most terms and 20% said they typically understand only a few of the terms) and (2) believe they could design a good randomized controlled trial (60% said they believe they could do it while 40% attested to their inability to design a good RCT). Overall, 65% of the respondents were satisfied with their current competence in data analysis and research methodology.

Higher self-perceived competence ratings were correlated with higher peer comparison ratings ($r = 0.42$, $p < 0.001$),

Table 1. Results of item difficulty for each of the ten items in our instrument.

Item	Theta*	Infit**	Outfit**	Point-biserial
1. Statistical significance (*p*-values)	−0.66	0.95	0.88	0.17
2. Number needed to treat (NNT)	−1.86	0.88	0.77	0.22
3. Odds ratio (OR)	−0.34	1.04	1.1	0.03
4. Confidence intervals (CI)	0.84	1.03	1.19	0.04
5. Sensitivity and specificity	1.09	0.92	1	0.16
6. Randomization and placebo	1.09	1.02	1.02	0.05
7. Effect size interpretation	−0.2	0.91	0.89	0.21
8. Intent-to-treat (ITT)	0.72	1.05	1.07	0.01
9. Diagnostics	−0.46	0.96	0.96	0.13
10. Graphical interpretation of correlations	−0.22	1.13	1.23	−0.05
Reliability estimate: 0.95				
Separation index: 4.4				

*In the Rasch model, a low value of theta suggests an easy item whereas a high value suggests a difficult item. An analysis of the instrument at the item level reveals three clusters of questions. Item 2 was singled out as the easiest and, perhaps, the least informative question. Items 1, 3, 7, 9 and 10 tended to be moderately difficult and exemplified the present sample's ability. Finally, items 4, 5, 6 and 8 tended to be more challenging with fewer respondents correctly answering those questions.
**Item infit and outfit values are well within acceptable range (Mean Square Residual (MNSQ) < 1.5) considering the nature of the measure and the number of questions used.

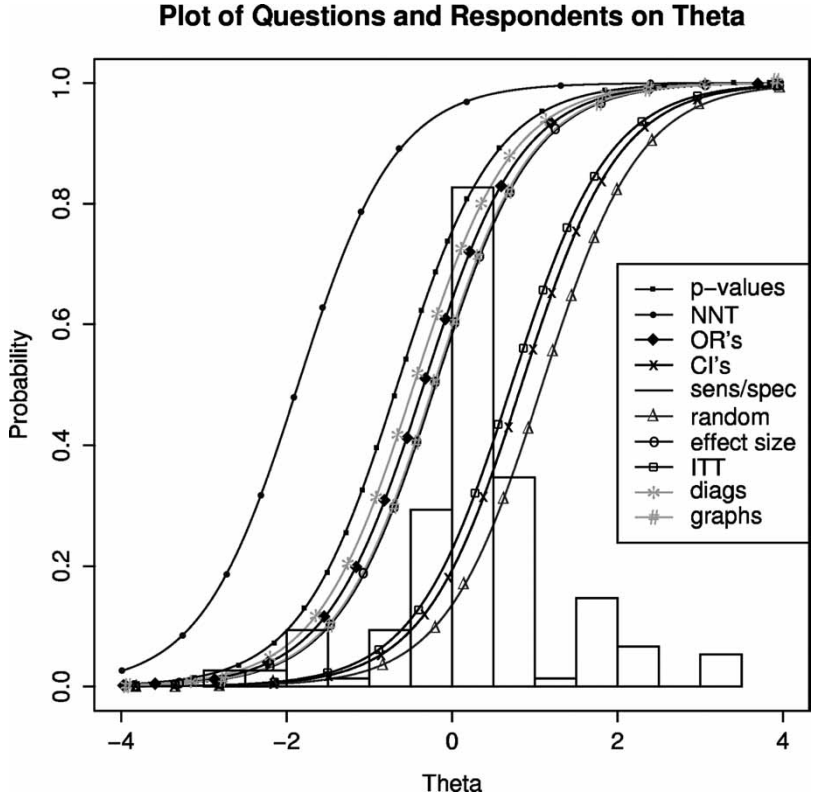

Figure 1. Distribution of respondents' abilities in relation to each of the ten questions in our instrument.
Notes: *Theta is the probability that a respondent will answer a question correctly given the respondent's ability. Respondent abilities and question difficulties are scaled on the same metric ranging from −4 to 4. Additionally, the scalar values are set by the average theta values across all questions have the mean set to 0 and the standard deviation set to 1.

Table 2. How would you rank your current competence in each of the following areas?

	Ratings		
Area	Poor	Fair	Excellent
Biostatistics/data analysis	12%	62%	26%
Research methodology	17%	62%	17%
Decision analysis/clinical reasoning	21%	73%	6%
Critical appraisal of the literature	21%	63%	16%
Evidence-based medicine	19%	65%	16%

Table 3. Compared with your peers, how would you rank your competence in each of the following areas?

	Ratings		
Area	Worse	Same	Better
Decision making	3%	75%	22%
Incorporating information into patient care	4%	72%	23%
Finding and appraising evidence	6%	59%	35%
Applying evidence to practice	5%	71%	24%

higher confidence in own ability to design an RCT ($r = 0.65$, $p < 0.02$), number of hours spent reading each week ($r = 0.18$, $p < 0.02$), satisfaction with own current competence in data analysis and research methodology ($r = 0.61$, $p < 0.001$), higher overall value ratings ($r = 0.16$, $p < 0.04$), and agreeing with the statement that clinicians must be well trained in research methods ($r = 0.16$, $p < 0.05$). Those who rated themselves as being satisfied with their current level of competence in data analysis and research methodology reported spending more time reading the literature ($r = 0.26$, $p < 0.001$), and tended to rate themselves just as good or better than their peers at clinical decision making ($r = 0.29$, $p < 0.0001$) and integration of research evidence into patient care ($r = 0.32$, $p < 0.0001$). Past research experience was correlated with higher levels of self-perceived competence in understanding research methodology ($r = 0.31$, $p < 0.001$), critical appraisal of the literature ($r = 0.19$, $p < 0.01$), and ability to design a good RCT ($r = 0.21$, $p < 0.007$). Also, subjects who reported having past research experience tended to rate themselves as better than their peers at appraising evidence from the literature ($r = 0.21$, $p < 0.007$) and were more likely to agree with the statements that clinicians must be well trained in data analysis ($r = 0.22$, $p < 0.005$), research methodology ($r = 0.28$, $p < 0.001$), and must know how to interpret research findings ($r = 0.18$, $p < 0.02$).

Subjects' attitude toward critical appraisal of the literature

In order to learn a bit about the attitudes of graduating medical students in the US toward critical appraisal of the literature and their perception of the current state of affairs

Table 4. Respondents' level of agreement/disagreement with statements concerning critical appraisal of the literature and evidence-based healthcare.

Statement	Ratings			
	Strongly disagree	Disagree	Agree	Strongly agree
'It is not important for clinicians to learn how to do research as long as they know how to interpret it.'	4%	31%	52%	13%
'Clinicians must be well trained in biostatistics/data analysis.'	3%	15%	61%	22%
'Clinicians must be well trained in research methodology.'	4%	25%	55%	15%
'Virtually all current common medical practices can be considered as evidence-based.'	25%	31%	31%	13%
'In general, publications in peer-review medical journals are free of bias and methodological flaws and the conclusions are valid and generalizable.'	31%	45%	24%	1%

of EBM, we asked them to rate the extent to which they agree/disagree with each of five statements on a four-point Likert scale (Table 4). Most respondents appreciated the importance of biostatistics and research methodology within the context of critical appraisal of the literature. Interestingly, respondents tended to have quite naïve views of the current state of affairs of EBM. Forty-four percent either agreed or strongly agreed with the statement that 'Virtually all current common medical practices can be considered as evidence-based', whereas the literature suggests the opposite (Imrie & Ramey, 2000; Booth, 2003). Also, 25% either agreed or strongly agreed that 'In general, publications in peer-review medical journals are free of bias and methodological flaws and the conclusions are valid and generalizable', whereas, in fact, current reporting is far from ideal (Moher *et al.*, 2001).

Discussion

In recent years, there has been a growing interest in critical appraisal of the medical research literature as a way to facilitate the practice of EBM. Many medical schools now teach critical appraisal skills with the intent to prepare graduates to be problem solvers and analyzers of clinical and published evidence (Radack & Valanis, 1986). However, the pertinent body of knowledge that underlies the critical appraisal skills can be difficult to teach to healthcare professionals (Campbell, 1990). The primary problem appears to be attitudinal rather than intellectual. Whereas critical appraisal skills are valued by medical students, their attitude toward statistics and methodology is often less enthusiastic (Riegelman *et al.*, 1983). Indeed, one recent study concluded that some students experience 'numerophobia'—a perceived and disproportionate fear of numbers and simple mathematical manipulation that may act as a psychological barrier for future evidence-based practitioners (Ben-Shlomo *et al.*, 2004). Moreover, the optimal timing for introducing those skills and the ideal format for such sessions remains unclear (Linzer, 1989; Holloway *et al.*, 2004). In most cases, education in these areas takes place relatively early in the medical training process at a time when their relevance and applicability to clinical practice is questionable (Looney *et al.*, 1998).

Despite the fact that different educational initiatives have been developed both at the medical school level (Landry *et al.*, 1994; Bradley *et al.*, 2005) and during residency (Rydman *et al.*, 1994; Thom *et al.*, 2004; Akl *et al.*, 2005; Thomas *et al.*, 2005), it remains largely unclear what impact such an education has on physicians' behavior. Although previous research has demonstrated that measurable increases in basic knowledge and critical appraisal skills can be seen after participation in these initiatives (Fraker *et al.*, 1996; MacAuley *et al.*, 1998; Smith *et al.*, 2000; Dorsch *et al.*, 2004; Sanchez-Mendiola, 2004) there is little evidence that knowledge and skills are taken outside the classroom and incorporated into actual patient care (Norman & Shannon, 1992; Landry *et al.*, 1994; Hyde *et al.*, 2000). Thus, the nature of the dose–response relationship between formal education in critical appraisal of the literature and physicians' competence and actual behavior still needs to be determined.

The self-assessment instrument we used in this study was developed to assess conceptual knowledge of basic methodology, data analysis, and critical appraisal skills through their application to medical decision-making. Although the question formats were modeled after previous instruments to include clinical scenarios that required a diagnostic or treatment decision, our emphasis was different. We were more interested in *procedural knowledge* (how a person does something) than in *declarative knowledge* (factual knowledge that a person has) (Anderson, 1993). This is because the ability to explain a medical term or to confront and solve medical problems with linear and sequential thinking processes may not be the kind of knowledge required for treating patients. Instead, physicians might respond to the 'feel' of the whole problem. Issues are not reduced to single elements to be dealt with by one or more forms of knowledge. Indeed, a recent test of competence in

EBM has been criticized for not asking the right questions (Mawer, 2003). We hope that our approach resulted in a better real-world simulation, although further research is needed in order to validate this proposition.

Unfortunately, the average correct score of respondents in our study, about 55%, suggests that fourth year medical students in the US may not be adequately prepared to critically evaluate the medical research literature. Wulff *et al.* (1987) concluded that the statistical knowledge of many physicians is so limited that they can neither detect inadequate statistical analyses in published papers, nor correctly interpret the results of adequate ones. Riegelman *et al.* (1983) found that fourth year medical students felt their literature-reading skills were inadequate, and Supino & Richardson (1999) reported that academic clinical faculty viewed their own knowledge as suboptimal in a number of content areas related to research and critical appraisal of the literature. In a recent study, only 22% of 50 general practitioners understood and could explain to others seven terms related to EBM (Young *et al.*, 2002). Worse yet, only four of 74 claims of understanding were confirmed in a subsequent interview and only one practitioner could provide a satisfactory explanation of any of the terms, whilst many of the explanations revealed considerable misunderstanding (Woodcock *et al.*, 2002). We believe that authors and journal editors alike should be cognizant of this situation and provide readers with information that is meaningful to the non-expert so as to avoid misinterpretations of the clinical research literature.

The most intriguing finding of our study is the apparent gap between respondents' confidence of their competence in various domains of critical appraisal of the literature and EBM, and their actual performance on the instrument. Whereas some researchers have argued that self-reported high levels of competence and preparedness are correlated with good performance (Shubert *et al.*, 1999), others have doubted whether self-assessment is a valid predictor of true competence (Gordon, 1991; Ward *et al.*, 2002). Khan *et al.* (2001), for example, found a weak association between participants' self evaluated knowledge and multiple choice test scores. Young *et al.* (2002) also found that physicians' self-ratings of their understanding of terms used in EBM differed significantly from an objective, criterion-based assessment. Since self-report preparedness has been widely used as an indicator of educational quality (Fincher *et al.*, 1993; Wickstrom *et al.*, 2000), a question arises whether self-assessed knowledge in EBM is a good proxy for actual skills.

Limitations

Our study has three important limitations. First, the non-random sampling. Due to feasibility issues related to privacy-protection laws in the US, a random sampling was (near) impossible and a choice between methods was needed. Ideal but expensive would be a cluster random sampling (randomly choose medical schools, then randomly choose students). We chose instead to do a volunteer sample via the Internet. Thus, formally, it is unknown to what extent our findings generalize to the entire population of graduating medical students. It may also be likely that our sample was biased toward subjects who are motivated to read the literature and consider themselves competent enough to appreciate findings commonly communicated in medical journals. Following this logic, and quite sadly, the bleaker picture here is probably bleaker in reality. Indeed, our findings are in agreement with previous research. Jones (2001) pointed toward the possibility that current medical educational programs can leave graduates either with important gaps in training or with an exaggerated sense of their competencies. Either discrepancy may reflect a bigger problem of education-practice training mismatch (Reuben *et al.*, 1988).

Second, our instrument covered primarily the critical appraisal element of EBM while ignoring other important components such as question formulation, searching for evidence and applying results to individual patients on a real life setting. This was done because the instrument was designed to be brief to optimize response rates. Alternatives to this type of self-assessment instrument may include formats in which subjects are asked to choose a research question that is of interest to them and develop it to a full research proposal. Whereas this type of mock project assignment is likely to better reflect subjects' abilities in research methodology, we rejected its use within the context of this preliminary study not just because we perceived it to be unfeasible, but also because we regard the skills necessary for *conducting research* to be different than the skills necessary for *using research findings* in clinical practice. Furthermore, evidence indicates that without explicit training in research, subjects' performance in this type of assignment is very poor, even among academic clinical faculty (Fraker *et al.*, 1996; Supino & Richardson, 1999). We do agree, however, that more research is needed to address gaps in physicians' competency and skills in other areas of EBM, just as a recent study showed (Green & Ruff, 2005).

Third, the small number of items, ten, in our assessment instrument may have compromised our ability to capture various degrees of subjects' performance on the measure because single questions covering broad topic areas may be inadequate to capture the breadth of a person's capability. Nonetheless, as discussed above, we think that the design approach we took in this study provided a reasonable compromise between response burden that may lead to subject attrition and missing data on the one hand, and reasonable bandwidth and fidelity of the measure, on the other hand (Cronbach & Gleser, 1965). The methodological constructs we chose overlap significantly with ones used in prior research (Weiss & Samet, 1980; Wulff *et al.*, 1987; Supino & Richardson, 1999), supporting the instrument's internal and construct validity. Furthermore, the initial Rasch analysis results indicate that the brief measure we developed has quite satisfactory psychometric properties: most questions fell well within the range of ability for most respondents; 'goodness of fit' model diagnostics show that question infit and outfit values were well within acceptable range (Mean Square Residual (MNSQ) < 1.5), and subjects successfully answering the more difficult questions tended to score higher on the test—precisely as expected. Since it often takes an ongoing effort to optimize and validate a new instrument, we are encouraged by these initial findings and are planning to continue to work on improving the self-assessment exercise we developed. Further research is needed to examine the validity of this and other EBM assessment instruments.

Conclusions and implications

Despite the fact that EBM is a required component of the Accreditation Council for Graduate Medical Education's Practice-Based Learning core competency, recent studies suggest that medical school graduates are poorly prepared in this important area (Lypson *et al.*, 2004). Since judicious use of medical research literature is critical to patient care, we must find ways to ensure that trainees acquire the fundamental critical evaluation skills and understanding necessary to assess the clinical research literature. Relying on graduating medical students' self-perceived competence is not enough. In the course of training, most physicians nowadays receive only limited education in research methodology and EBM (Looney *et al.*, 1998). The result is reduced competence that may jeopardize patient care. This may be due, in part, to failure to identify the core EBM skills that should be mandatory for the trainee to acquire in order to improve patient care (Woodcock *et al.*, 2002). We believe that the practice of good medicine is therefore likely to benefit from transdisciplinary training in both decision sciences and EBM (Schulkin, 2000). Unfortunately, there is no systematic research evaluating medical education with respect to both. More research is thus needed in order to explore the impact that the interface between these two fields may have on physicians' behavior, patient care, and health outcomes.

Goethe once said, 'Knowing is not enough; we must apply. Willing is not enough; we must do.' We call for a joint effort to resolve the apparent education–practice mismatch in the area of critical appraisal of the literature and medical decision-making.

Notes on contributors

All authors are members of the Evaluation Group for Analysis of Data (EGAD) at the University of Arizona.

Opher Caspi is an internist and methodologist by training. His research interests span from placebo effects to decision analysis and clinical reasoning to evidence-based medicine and meta-analysis.

Patrick McKnight is a research assistant professor in the department of psychology at the University of Arizona. He is an expert in measurement theory, data analysis, and computer programming.

Lillian Kruse was a student of OC at the time of the study.

Victoria Cunningham earned her PhD in program evaluation and research methodology. Her research focuses on issues related to measurement and data analysis.

Aurelio Jose Figueredo is a professor of psychology and directs the program of evolutionary psychology at the University of Arizona. He is an expert in multi-causal modeling and philosophy of science.

Lee Sechrest is a world renowned expert in program evaluation, research methodology and data analysis. He has served as mentor to all other authors at different stages of their careers.

References

Akl, E.A., Izuchukwu, I.S., El-Dika, S., Fritsche, L., Kunz, R. & Schunemann, H.J. (2004) Integrating an evidence-based medicine rotation into an internal medicine residency program, *Academic Medicine*, 79, pp. 897–904.

Anderson, J.R. (1993) *Rules of the Mind* (Hillsdale, NJ, Lawrence Erlbaum).

Ben-Shlomo, Y., Fallon, U., Sterne, J. & Brookes, S. (2004) Do medical students with A-level mathematics have a better understanding of the principles behind evidence-based medicine? *Medical Teacher*, 26, pp. 731–733.

Blumenthal, D., Gokhale, M., Campbell, E.G. & Weissman, J.S. (2001) Preparedness for clinical practice: reports of graduating residents at academic health centers, *Journal of the American Medical Association*, 286, pp. 1027–1034.

Booth, A. (2003) What proportion of healthcare is evidence based? Resource Guide. Available at: www.shef.ac.uk/~scharr/ir/percent.html

Bradley, P., Oterholt, C., Nordheim, L. & Bjorndal, A. (2005) Medical students' and tutors' experiences of directed and self-directed learning programs in evidence-based medicine: a qualitative evaluation accompanying a randomized controlled trial, *Evaluation Review*, 29, pp. 149–177.

Campbell, M.J. (1990) Response to Clayden AD. Who should teach medical statistics, when, how and where should it be taught? *Statistics in Medicine*, 9, pp. 1039–1041.

Cronbach, L.J. & Gleser, G.C. (1965) *Psychological Tests and Personnel Decisions* (Urbana, IL, Glencoe Press).

Dorsch, J.L., Aiyer, M.K. & Meyer, L.E. (2004) Impact of an evidence-based medicine curriculum on medical students' attitudes and skills, *Journal of the Medical Librarians Association*, 92, pp. 397–406.

El Ansari, W. (2004) Appraisal skills as a public health competency for evidence-based care: students' satisfaction with their research education, *Journal of Public Health Management and Practice*, 10, pp. 354–365.

Epstein, R.M. & Hundert, E.M. (2002) Defining and assessing professional competence, *Journal of the American Medical Association*, 287, pp. 226–235.

Fincher, R.M., Lewis, L.A. & Kuske, T.T. (1993) Relationships of interns' performances to their self-assessments of their preparedness for internship and to their academic performances in medical school, *Academic Medicine*, 68, pp. S47–S50.

Fraker, L.D., Orsay, E.M., Sloan, E.P., Bunney, E.B., Holden, J.A. & Hart, R.G. (1996) A novel curriculum for teaching research methodology, *Journal of Emergency Medicine*, 14, pp. 503–508.

Holloway, R., Nesbit, K., Bordley, D. & Noyes, K. (2004) *Medical Education*, 38, pp. 868–878.

Gordon, M.J. (1991) A review of the validity and accuracy of self-assessments in health professions training, *Academic Medicine*, 66, pp. 762–769.

Green, M.L. & Ruff, T.R. (2005) Why do residents fail to answer their clinical questions? A qualitative study of barriers to practicing evidence-based medicine, *Academic Medicine*, 80, pp. 176–182.

Guyatt, G. & Rennie, D. (2002) *User's Guides to the Medical Literature* (Chicago, AMA Press).

Holloway, R., Nesbit, K., Bordley, D. & Noyes, K. (2004) Teaching and evaluating first and second year medical students' practice of evidence-based medicine. *Medical Education*. 38, pp. 868–878.

Hyde, C., Parkes, J., Deeks, J. & Milne, R. (2000) Systematic review of effectiveness of teaching critical appraisal. Oxford: ICRF/NHS Centre for Statistics in Medicine, available online at: www.bham.ac.uk/arif/SysRevs/TeachCritApp.PDF.

Ihaka, R. & Gentleman, R. (1996) A language for data analysis and graphics, *Journal of Computational Graphics & Statistics*, 5, pp. 299–314.

Imrie, R. & Ramey, D.W. (2000) The evidence for evidence-based medicine, *Complementary therapies in Medicine*, 8, pp. 123–126.

Johns, M.M.E. (2001) The time has come to reform graduate medical education, *Journal of the American Medical Association*, 286, pp. 1075–1076.

Khan, K.S., Awonuga, A.O., Dwarakanath, L.S. & Taylor, R. (2001) Assessment in evidence-based medicine workshops: loose connection between perception of knowledge and its objective assessment, *Medical Teacher*, 23, p. 924.

Landry, F.J., Pangaro, L., Kroenke, K., Lucey, C. & Herbers, J. (1994) A controlled trial of a seminar to improve medical student attitudes toward, knowledge about, and use of the medical literature, *Journal of General Internal Medicine*, 9, pp. 436–439.

LINACRE, J.M. (2000) *WINSTEPS Rasch Measurement* (Chicago, MESA Press).

LINZER, M. (1989) Critical appraisal: more work to be done, *Journal of General Internal Medicine*, 4, pp. 457–459.

LOONEY, S.W., GRADY, C.S. & STEINER, R.P. (1998) An update on biostatistics requirements in U.S. medical schools, *Academic Medicine*, 73, pp. 92–94.

LYPSON, M.L., FROHNA, J.G., GRUPPEN, L.D. & WOOLLISCROFT, J.O. (2004) Assessing residents' competencies at baseline: identifying the gaps, *Academic Medicine*, 79, pp. 564–570.

MACAULEY, D., MCCRUM, E. & BROWN, C. (1998) Randomised controlled trial of the READER method of critical appraisal in general practice, *British Medical Journal*, 316, pp. 1134–1137.

MAWER, C. (2003) Does Fresno assess real-life EBM skills? *British Medical Journal*, available at: http://bmj.com/cgi/eletters/326/7384/319-29659

MOHER, D., JONES, A., LEPAGE, L. & THE CONSORT GROUP (2001) Use of the Consort statement and quality of reports of randomized trials: a comparative before-and-after evaluation, *Journal of the American Medical Association*, 285, pp. 1992–1995.

NORMAN, G. & SHANNON, S. (1992) A critical appraisal of critical appraisal skills teaching interventions, *Clinical Epidemiology Newsletter*, 12, pp. 11–12.

RADACK, K.I. & VALANIS, B. (1986) Teaching clinical appraisal and application of medical literature to clinical problem solving, *Journal of Medical Education*, 61, pp. 329–331.

REUBEN, D.B., MCCUE, J.D. & GERBERT, B. (1988) The residency-practice training mismatch. A primary care education dilemma, *Archives of Internal Medicine*, 148, pp. 914–919.

RIEGELMAN, R.K., POVAR, G.J. & OTT, J.E. (1983) Medical students' skills, attitudes, and behavior needed for literature reading, *Journal of Medical Education*, 58, pp. 411–417.

RYDMAN, R.J., ZALENSKI, R.J. & FAGAN, J.K. (1994) An evaluation of research training in a large residency program, *Academic Emergency Medicine*, 1, pp. 448–453.

SANCHEZ-MENDIOLA, M. (2004) Evidence-based medicine teaching in the Mexican Army Medical School, *Medical Teacher*, 26, pp. 661–663.

SCHULKIN, J. (2000) Decision sciences and evidence-based medicine – two intellectual movements to support clinical decision making, *Academic Medicine*, 75, pp. 816–818.

SCHWARTZ, J.S., BALL, J.R. & MOSER, R.H. (1982) Safety, efficacy, and effectiveness of clinical practices: a new initiative, *Annals of Internal Medicine*, 96, pp. 246–247.

SHUBERT, A., TETZLAFF, J.E., TAN, M., RYCKMAN, J.V. & MASCHA, E. (1999) Consistency, inter-rater reliability, and validity of 441 consecutive mock oral examinations in anesthesiology: implications for use as a tool for assessment of residents, *Anesthesiology*, 91, pp. 288–298.

SMITH, C.A., GANSCHOW, P.S., REILLY, B.M., EVANS, A.T., MCNUTT, R.A., OSEI, A., SAQUIB, M., SURABHI, S. & YADAV, S. (2000) Teaching residents evidence-based medicine skills: a controlled trial of effectiveness and assessment of durability, *Journal of General Internal Medicine*, 15, pp. 710–715.

SUPINO, P.G. & RICHARDSON, L.D. (1999) Assessing research methodology training needs in emergency medicine, *Academic Emergency Medicine*, 6, pp. 280–285.

TAYLOR, R., REEVES, B., EWINGS, P., BINNS, S., KEAST, J. & MEARS, R. (2000) A systematic review of the effectiveness of critical appraisal skills training for clinicians, *Medical Education*, 34, pp. 120–125.

THOM, D.H., HAUGEN, J., SOMMERS, P.S. & LOVETT, P. (2004) Description and evaluation of an EBM curriculum using a block rotation, *BMC Medical Education*, 4, p. 19.

THOMAS, K.G., THOMAS, M.R., YORK, E.B., DUPRAS, D.M., SCHULTZ, H.J. & KOLARS, J.C. (2005) Teaching evidence-based medicine to internal medicine residents: the efficacy of conferences versus small-group discussion, *Teaching and Learning in Medicine*, 17, pp. 130–135.

WARD, M., GRUPPEN, L. & REGEHR, G. (2002) Measuring self-assessment: current state of the art, *Advances in Health Science Education Theory and Practice*, 7, pp. 63–80.

WEISS, S.T. & SAMET, J.M. (1980) An assessment of physician knowledge of epidemiology and biostatistics, *Journal of Medical Education*, 55, pp. 692–697.

WICKSTROM, G.C., KOLAR, M.M., KEYSERLING, T.C., KELLEY, D.K., XIE, S.X., BOGNAR, B.A., *et al.* (2000) Confidence of graduating internal medicine residents to perform ambulatory procedures, *Journal of General Internal Medicine*, 15, pp. 361–365.

WOODCOCK, J.D., GREENLEY, S. & BARTON, S. (2002) Doctors' knowledge about evidence based medicine terminology, *British Medical Journal*, 324, pp. 929–930.

WRIGHT, B.D. & STONE, M.H. (1979) *Best Test Design: Rasch Measurement* (Chicago, MESA Press).

WULFF, H.R., ANDERSEN, B., BRANDENHOFF, P. & GUTTLER, F. (1987) What do doctors know about statistics? *Statistics in Medicine*, 6, pp. 3–10.

YOUNG, J.M., GLASZIOU, P. & WARD, J.E. (2002) General practitioners' self ratings of skills in evidence based medicine: validation study, *British Medical Journal*, 324, p. 9501.