

Development and Validation of an Instrument to Measure Knowledge of Evidence-Based Practice and Searching Skills.

Peter Bradley[†], Jeph Herrin[‡]

[†]Norwegian Medicines Agency, Oslo, Norway

[‡]Flying Buttress Associates, Charlottesville, VA, USA

Abstract: The aim of this study was to develop and validate three instruments which measure knowledge about searching for and critically appraising scientific articles (evidence-based practice-EBP). Twenty-three questions were collected from previous studies and modified by an expert panel. These questions were then administered to 55 delegates before and after two international conferences in EBP; the responses were assessed for discriminative ability and internal consistency. Five questions were discarded and three instruments of six questions each were developed. Finally, the instruments were re-validated in a randomized controlled trial comparing two educational interventions at the University of Oslo, Norway by 166 of 175 eligible medical students. In the re-validation, the instruments showed satisfactory level of discriminate validity ($p < 0.05$), but borderline levels of internal consistency (Cronbach's α 0.52-0.61). More research is needed to develop a suitable instrument which includes questions on searching for evidence.

Key words: Evidence-based practice, validation, medical education

In 1972, Archie Cochrane published an influential book about the effectiveness and efficiency of health services which is often considered to be the birth of evidence-based medicine (EBM) or evidence-based practice (EBP). Further articles drew attention to; the over and under use of some treatments, the potentially harmful effects of some treatments, and variation in health care standards between hospitals, diagnostic groups and social classes.^{1,2,3} A well-known example was the delay in implementing thrombolysis treatment (an effective treatment to dissolve blood clots) which prevented death after acute myocardial infarction (heart attack).⁴ In response to this, Cochrane emphasised the importance of using existing empirical research in medical decision-making and challenged the medical profession to organize critical summaries of all randomized controlled trials within each medical speciality.²

EBM (evidence-based medicine) has more recently been defined as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” and consists of the following 5 steps:

- formulating clinically important questions
- efficient gathering of clinical evidence (research)
- critical appraisal (assessment) of evidence
- applying evidence to practice
- evaluating own practice⁵

The impact of short courses in evidence-based practice (EBP) has been evaluated in many trials. Most studies have evaluated the impact of educational interventions in terms of either knowledge of EBM; skills to critically appraise scientific articles; changes in self-reported attitudes to EBM; self-reported reading behaviour or course satisfaction. However, few authors have used validated instruments,⁶ which has contributed to difficulty in interpreting trial results.⁷ More recently, attempts have been made to validate instruments, but these have not included questions about searching for evidence^{8,9} and have thus excluded a major component of EBM.

In this study, we aimed to develop and validate instruments which measured knowledge about critical appraisal *and* searching skills for use in a randomized controlled trial. The trial aimed to compare two educational programs in evidence-based practice (EBP) for medical students at the University of Oslo, Norway in 2002/3.¹⁰ Students were recruited to the trial from three separate semesters to ensure an adequate number of study participants. EBP knowledge (and skills) were assessed in each semester in an end-of-term examination. Since students' examination papers were made public soon after testing, each cohort of students in the trial required separate, but similar instruments to test EBP knowledge – three instruments in all.

Methods

Development and validation comprised three stages: development of a pool of questions, validation of the questions to develop three instruments (including before and after assessment) and re-validation/implementation of the instruments in a randomized controlled trial.

Development - Existing questions from previous studies^{11,12} were modified by a panel of four experts with experience in teaching EBP for several years (a research librarian, a professor in public health medicine, a further medical public health specialist and a research physiotherapist). Questions were then selected for comprehensiveness and relevance (*content and face validity*) to the syllabus of the educational intervention in the trial.¹⁰

A pool of 23 multiple-choice questions was created in English and translated to Norwegian, using a process of back translation for quality control. Each question consisted of a stem and three items, to which the response options were “True”, “False”, and “Don't Know”. Each stem introduced a key learning

1. To increase the sensitivity of a search you should:

 - a) only use MESH terms to obtain relevant articles
 - b) use MESH terms and a free text search to obtain a greater number of articles, even if some are irrelevant
 - c) only use a free text search to ensure few irrelevant articles are obtained

Figure 1 – Example of multiple choice question

concept within EBP and the items following each stem were intended to reflect differing degrees of familiarity with the concept introduced. An example is shown in figure 1.

Piloting and validation - In this second stage, 3 similar instruments of 6 questions each were developed by removing questions which reduced overall internal consistency. Content and face validity were maintained, but the wording of two of the questions in the third instrument was changed, as they were felt to be ambiguous.

The pool of 23 questions was given to a total of 55 delegates at two international, one-week conferences in EBP. Both conferences were based on the educational model developed by McMaster

University, Canada.¹³ Fifteen potential questions were answered by delegates before and after the first conference, the “6th Nordic workshop in evidence-based health care” (May 2000) to develop the first two instruments. The delegates were varied in terms of their educational background but few delegates had previously attended more than one course on EBP or taught EBP to others. Thirty-two percent of the delegates completing the questionnaire were doctors in clinical practice, 47% were other health qualified professionals (mainly nurses, and physiotherapists), and 20% were health care librarians. Eight additional questions were answered by delegates before and after the “Central-Eastern European Evidence-Based Health Care Workshop” (December 2002) to develop the third instrument. The delegates at this conference were generally similar in terms of their educational and professional background to those attending the first conference. Sixty-six percent of the delegates completing the questionnaire were doctors with clinical practice, 22% other qualified health professionals (mostly nurses), whereas 11% were health care librarians. This knowledge level of participants at both conferences was assumed to be similar to the medical students in the forthcoming trial.

Items were scored as +1 if the answer was correct, -1 if incorrect, and 0 if unanswered or marked as “don't know”. Scores for the three items of each question were then averaged for that stem. Thus the total possible score for the entire set of knowledge questions ranged from -15 to +15.

Using the follow-up responses to the first fifteen questions, principal component analysis was used to identify the potential number of educational themes (or factors) being tested in the instrument and which factors were being measured by each individual question.

Internal consistency was measured by calculating values for Cronbach's α and inter-item covariance from the follow-up responses from all questions. We considered a Cronbach's α greater than 0.7 to be satisfactory.¹⁴ Discriminative ability was defined as the ability to determine whether the three instruments were able to detect changes in knowledge about EBP and searching skills. It was calculated as the mean difference between baseline and follow-up scores which were compared to zero with a simple t-test. We considered a p-value of ≤ 0.05 to be significant.

Implementing items in a RCT - In this final stage, the three instruments were re-validated by 175 medical students enrolled in the randomised

Table 1
Baseline characteristics of Students in Randomized Trial per Study

Characteristic	Total (n)
N	175
Female	48.3% (84)
Age	
20-29	89.7% (156)
30-39	9.8% (17)
39+	0.6% (1)
Previously attended course in EBM	10.3% (18)
Skills in using Internet	80.6% (141)
Access to Internet at home	53.7% (94)
Performed search in biomedical database in last 6 months	79.9% (139)
Ever received training in:	
a) research methodology	21.7% (38)
b) epidemiology	2.3% (4)
c) statistics	24.0% (42)
d) literature searching	12.6% (22)

controlled trial at the University of Oslo, Norway between Spring 2002 and ending in Spring 2003. The baseline characteristics of the students are reported in Table 1. The syllabus¹⁰ covered was similar to those covered by the EBP conferences in the second validation stage.

Each instrument was re-validated by one cohort of students (approximately 60 students from each semester) using post-intervention scores alone. As the instruments had already been shown to have adequate discriminative ability, only internal consistency was considered by re-calculating Cronbach's α scores. This time scores were not averaged for each question stem, but were treated individually. Otherwise, the same scoring system and statistical assumptions were employed as in the second stage of validation. The completed instruments are available from the authors by request.

Results

Factor identification - Principal components factor analysis (results not shown) indicated that there were two distinct factors (educational themes) which were being measured – which responded to the questions on searching and critical appraisal.

Discriminative ability - Table 2 shows the mean differences in scores obtained before and after the international workshops and the subsequent p-values and confidence intervals. All values are ≤ 0.05 .

Internal consistency - Table 3 shows the Cronbach's α scores from the second stage of validation and Table 4 from the third stage. Scores are given for the three instruments overall and for the two educational themes identified. In the both stages suboptimal values for Cronbach's α were obtained as all were ≤ 0.7 .

Discussion

A pool of 23 questions were developed from previous outcome measured, using expert opinion. Each question contained three related items. The instruments were assessed for discriminatory ability and internal consistency. Five questions were subsequently removed to try to improve internal consistency.

Three instruments were developed, consisting of six questions each. Prior to the medical students' trial, the instruments showed borderline levels of internal consistency. For the third instrument, the scores were felt to be due to ambiguities in the wording of two items in the latter instrument, which were subsequently reformulated before the final stage of validation in the randomized controlled trial. In the trial all instruments again showed suboptimal levels of internal consistency. This may suggest that not all items in the instruments reliably measured the knowledge acquired by students and may have made real differences in knowledge more difficult to detect. On the other hand, the mean increase in scores pre and post teaching showed a statistically significant

Table 2
Discriminative Ability (mean difference) Measured in Second Stage of Validation

	No. of respondents	Mean difference	95% confidence interval	P value
Instrument 1	n=29	2.07	0.46-3.68	0.014
Instrument 2	n=29	2.28	0.70-3.85	0.007
Instrument 3	n=26	2.49	0.67-3.85	0.001

Table 3
Internal consistency (Cronbach's α) Measured in Second Stage of Validation

	No. of respondents	Cronbach's alpha (searching)	Cronbach's alpha (critical appraisal)	Cronbach's alpha (overall)
Instrument 1	n=29	0.56	0.66	0.64
Instrument 2	n=29	0.67	0.61	0.58
Instrument 3	n=29	0.17	0.54	0.03

Table 4
Internal Consistency (Cronbach's α) Measured in Third stage of Validation

	No. of respondents n=166	Cronbach's alpha (searching)	Cronbach's alpha (critical appraisal)	Cronbach's alpha (overall)
Instrument 1	n=42	0.53	0.52	0.52
Instrument 2	n=58	0.65	0.30	0.47
Instrument 3	n=66	0.43	0.52	0.61
All		0.32	0.36	0.57

change for all instruments. This suggests that the instruments were able to discriminate between students on the basis of overall EBP knowledge acquired.

One possible explanation for the borderline scores is that the instruments were inappropriate for the study group or specific training intervention. Nonetheless, this would be a surprising result given the long process of validation, and that the scores obtained by medical students post-intervention were very similar to those obtained by expert groups in a previous validation study⁸. An alternative explanation is that high levels of internal consistency are difficult to obtain from such a short examination.

The study illustrates the difficulties of developing valid, generalizable instruments for measuring EBP knowledge in trials. Further work is required to develop better instruments which are

suitable for use in different training contexts and include items on searching for evidence.

Acknowledgements

The work was supported by the Medical Faculty at the University of Oslo, the Norwegian Institute for Public Health and the Norwegian Directorate for Health and Social Welfare

We would like to thank Arild Bjørndal, Gro Jamtvedt, Doris Kristoffersen, Lena Nordheim and Irene Wiik at the National Centre for Health Services Research, Oslo for their advice and support

References

1. Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust, 1972

2. Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. In: Medicines for the year 2000. London: Office of Health Economics, 1979, I-II
3. Morgan M, Beech R. Variations in lengths of stay and rates of day case surgery: implications for the efficiency of surgical management. *J Epidemiol Community Health* 1990; 44:90-105
4. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC, A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;268:240-248
5. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2
6. Hyde C, Parkes J, Deeks J, Milne R, Systematic review of effectiveness of teaching critical appraisal. ICRF/NHS Centre for Statistics in Medicine, Oxford: 2000
7. Taylor R, Reeves B, Ewings P, Binns S, Khan PE, Keast J, Mears R A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ* 2000; 34:120-125
8. Taylor R, Reeves B, Mears R, Keast J, Binns S, Khan PE, Development of a questionnaire to evaluate the effectiveness of evidence practice teaching, *Med Educ*, 2001; 35(6): 544-547
9. Firsche I, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, Kunz R, Do short courses in evidence based medicine improve knowledge and skills? Validation of Berlin questionnaire and before and after study of courses in evidence based medicine, *BMJ* 2002; 325:1338-41
10. Bradley P, Oterholt C, Herrin J, Nordheim L, Bjørndal A, A comparison of directed and self-directed learning programmes in evidence-based medicine for medical students: a randomised controlled trial, submitted *BMJ*, April 2004
11. Taylor R et al, A randomised controlled trial of the effectiveness of critical appraisal skill workshops for health service decision-makers in the South and West Region. 1999:Bristol, unpublished.
12. Enoch K, An evaluation of Computer Aided Learning within Evidence-Based Health Care educational resources, 2000: Oxford, unpublished
13. Guyatt G, Rennie D. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago; JAMA and archives journals, 2002.
14. Cronbach LJ, Coefficient α and the internal structure of tests. *Psychometrika* 1951;16: 197-234

Correspondence

Peter Bradley, Acting head,
Pharmaceutical reimbursement section
Norwegian Medicines Agency,
Sven Oftedals vei 8, 0950
Oslo, Norway
Tel + (47) 22 16 84 33
Fax + (47) 22 89 77 99
E-mail: peter.bradley@NoMA.no

Address for correspondence

Brekkekroken 1, 1430 Aas, Norway