Evidence based checklists for objective structured clinical examinations

Christopher Frank

BMJ 2006;333:546-548
doi:10.1136/bmj.38943.463565.68

These include:

References
This article cites 12 articles, 3 of which can be accessed free at:
http://bmj.com/cgi/content/full/333/7567/546#BIBL

1 online articles that cite this article can be accessed at:
http://bmj.com/cgi/content/full/333/7567/546#otherarticles

Rapid responses
5 rapid responses have been posted to this article, which you can access for free at:
http://bmj.com/cgi/content/full/333/7567/546#responses

You can respond to this article at:
http://bmj.com/cgi/eletter-submit/333/7567/546

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top left of the article

Topic collections
Articles on similar topics can be found in the following collections

- Undergraduate (261 articles)
- Postgraduate: Family/General Practice (163 articles)
- Postgraduate: Residency / House staff (192 articles)
- Exams (54 articles)

Notes

To order reprints follow the "Request Permissions" link in the navigation box

To subscribe to BMJ go to:
http://resources.bmj.com/bmj/subscribers
Analysis and comment

Summary points

Medical education needs to adapt to society's changing attitudes

Work based training must be made more effective to counter reduced working hours

New methods of assessment are needed to reflect the focus on competencies

High quality, relevant research requires more interdisciplinary collaboration

Overcoming negative attitudes to assessment will involve a cultural shift

where the goal is not to be better than the others but to be better today than you were yesterday.

We have described four major challenges in medical education. One conclusion from all of these is that a close collaboration between doctors and educationalists is indispensable for good medical education and development of better education. Any monoclinical endeavour will lead to a suboptimal result.

Contributors and sources: LWTS and CPMvdV have the key responsibility for the development, implementation, and quality control of the assessment system of the medical school of the Maastricht University and are consultants for assessment both nationally and internationally: this article reflects long and frequent discussions between both authors. LWTS wrote the first draft. CPMvdV commented and made the necessary changes and additions. Both authors approved the final version. LWTS is guarantor.

Competing interests: None declared.

2001;86:930-42.
2001;86:930-42.
2001;86:930-42.

Medical education

Evidence based checklists for objective structured clinical examinations

Christopher Frank

How doctors examine a patient is often influenced more by tradition than by evidence. But trainees should be assessed on what works and not personal preferences

The objective structured clinical examination has been used to evaluate the clinical skills of medical trainees for more than 25 years. Many examinations use checklists as the main indicator of performance, although some people advocate global ratings.17 The development of these checklists is challenging. Doctors devising checklists often disagree about what should be included and the weighting given to items. This is particularly true with checklists for physical examinations, where tradition, different training sites, and specialty backgrounds influence opinion.

Reasons for concern

The physical examination represents a link with the history of medicine, and many clinicians have strong opinions on the merits of specific clinical signs. Clinicians have begun to critically review many aspects of physical examination, but this critical approach does not seem to have been applied to the development of checklists for objective structured examinations. Gorter and colleagues reviewed the literature on developing these examinations and found that only 41% of the 29 papers described the process of checklist development. None of the papers reported that checklists were based on the available literature, and only three reported use of published evidence.14

Examples

I reviewed the checklists for physical examination stations from undergraduate and postgraduate exami-
nations at a Canadian health sciences centre and from a multicentre examination to provide examples. I also reviewed the checklists in an examination preparation book. The items discussed below were chosen because there was controversy about the clinical value of a particular aspect of a physical examination or the possible weighting of marks.

I used JAMA's rational clinical examination series as a distillation of evidence for specific procedures. For each of the procedures potentially included in an examination checklist, I searched Medline for articles published after the JAMA articles. My approach was intended to simulate an evidence based approach that a committee could use and did not include steps used to answer clinically based questions, such as evaluating the effect after application of the evidence.

Respiratory examination for pneumonia

Students' skills in performing the respiratory examination are often tested by using scenarios in which patients have pneumonia or shortness of breath. The checklists commonly use items that reflect longstanding traditions for chest examination and include several controversial manoeuvres, most notably vocal and tactile fremitus and aegophony.

Studies have found the precision of respiratory examinations to be poor for various conditions, including community acquired pneumonia. Several manoeuvres that might be included on examination checklists had poor measures of precision: tactile fremitus ($k$ value of agreement $= 0.19$), bronchial breath sounds ($k$ = 0.19), and whispered pectoriloquy ($k$ = 0.11). Even the most reliable items (percussion dullness and wheezes) had $k$ values of 0.51 and 0.52 respectively, which means that clinicians agreed about half of the time beyond chance. These values are partly explained by the rarity of findings like aegophony and the fluctuating presence of clinical signs. For large scale examinations, differences of opinion between examiners about physical examination techniques can become an issue because of the need for multiple examiners.

Gennis and colleagues found that abnormal chest findings, including fremitus, decreased breath sounds, crackles, wheezes, aegophony, and percussion dullness, were present in fewer than half of patients with pneumonia. Indicated item is unclear.

Aegophony, fremitus, and whispered pectoriloquy are less clear and they are likely to lead to poor inter-rater reliability. For this reason checklist weighting should emphasise the more basic aspects of the respiratory examination and should include temperature and vital signs, especially if the examination station is intended to evaluate assessment of pneumonia rather than a generic chest examination. Assigning separate marks for anterior and posterior examination may be much less important than for comparing sides or considering examination in the lateral position.

Capillary refill for volume status or shock

The assessment of hypovolaemia can be difficult, especially with older patients. Examining the rate of capillary refill is often considered part of the assessment of volume status and shock, as well as peripheral arterial disease. Assessment of hypovolaemia in adult patients may be included in examination stations dealing with trauma or the management of shock.

The definition and range of normal values for capillary refill time is not commonly cited. The test involves compressing the distal phalanx of the middle finger (at heart level) for five seconds and then timing the return of normal colour to the finger. Measurements done by two examiners using stopwatches have been shown to be within 0.3 seconds of each other. The normal range depends on age, temperature of the room, and technique. At 21°C the upper limit of normal is 2 seconds for children and men and 3 seconds for women, although it can rise to 4 seconds in very elderly people. It is unclear if most clinical examinations provide these criteria for examiners. Low lighting has been found to make it harder to detect capillary refill. This may be relevant to examinations, where the physical environment may vary.

I found only one study that looked at the role of capillary refill in people with possible hypovolaemia (emergency departments and people who had just donated blood). The test did not have good sensitivity (6%) for identification of 450 ml blood loss (specificity of 93% and a likelihood ratio for a positive result of 1.0). The use of an arbitrary cut-off point (2 seconds) did not improve diagnostic performance. Concerns about the use of capillary refill led to its exclusion from the trauma scale and are echoed in the literature about its use in the assessment of peripheral arterial disease.

Strategies that may help

Assessment is linked closely to the content of teaching. Many undergraduate examinations are developed to assess student's progress in a clinical skills course. For
this reason, shortcomings in the checklist may reflect a lack of evidence being used in the development of the clinical skills curriculum, as well as decreased attention to teaching clinical skills, especially at the bedside. The interest in studying and critically reviewing the validity and reliability of physical examination procedures is relatively recent. It is only in the past five years that I have heard trainees citing recommendations on physical examination from journal articles or seen citations referred to in examination answers. The contents of the checklist are often controversial, especially for physical examinations. Development of checklists should be based on available evidence of effectiveness.

Use of the available evidence is particularly important for checklists in national or multicentre examinations, when it is more difficult to ensure a match between teaching and examination content than it is for examinations in individual centres. Although trainees from different centres may have been taught a wide range of manoeuvres, it is reasonable to assume that those backed by evidence will be included in the core curriculum. Using evidence based checklists may also improve the quality of teaching of clinical skills teaching if universities attempt to "teach to the exams."

Recognition of the potential pitfalls of development of checklists without formal processes is a first step to improving quality. The recommendations of Gorter and colleagues for developing checklists included the use of evidence based performance standards and use of a documented consensus approach to combine expert opinion with published evidence. By considering the evidence behind checklist items, people preparing examination checklists can feel more confident justifying their choices to trainees and examiners.

Help with assessing evidence

Reviewing evidence can be time consuming and challenging. As most practitioners become familiar with using the tools of evidence based medicine and with available resources, this task will presumably become easier. The rational clinical examination series is an excellent starting point for clinicians trying to improve their clinical skills as well as for reviewing individual examination manoeuvres. The Standards for Reporting Diagnostic Accuracy initiative (STARD) has published a checklist to assess the quality of new studies of tests of diagnostic accuracy. The Society of General Internal Medicine clinical examination research interest group and the American College of Physicians have compiled resources on the web that may help to optimise checklist items more quickly and more accurately than a review done by an individual. Evidence based textbooks on clinical skills have also been published.

Evidence is also needed about other aspects of objective structured clinical examinations. History taking is a major focus of these exams, but there is little research on the merits of the questions used in taking a history of common medical problems. Studies of the effect of an evidence based approach on the success of these exams would be important in promoting the use of this approach for examinations and for the teaching of clinical skills.

Contributors and sources: CF has been involved in undergraduate and postgraduate medical education in geriatrics for many years. He has been involved in development of the content of objective structured clinical examinations at a university and national level and has been an examiner in many occasions. This article arose at a committee meeting preparing a multicentre OSCE exam.

Competing interests: None declared.

Summary points

Objective structured examinations are an important tool for testing clinical performance.

Students' performance is usually assessed by using a checklist.

The contents of the checklist are often controversial, especially for physical examinations.