Educating the examiner

How can we test clinical reasoning?

Cees P M van der Vleuten, David I Newble

Clinicians who teach medical students will acknowledge the importance of assessing the ability to think critically about diagnosis and management. In writings on medical education this process has usually been referred to as clinical problem-solving or clinical reasoning ability. Along with clinical skills it is a vital component of clinical competence,[1] and its importance is recognised in major policy documents about curriculum objectives.[2,3]

Some teachers believe they can accurately assess the clinical reasoning of their students during ward or practice attachments and during traditional viva type examinations. Such confidence is unfounded. For instance, the typical viva is conducted in an unstructured and unstandardised fashion, with subjective marking of performance in just a few clinical cases. Moreover, when only a few examiners are involved in the assessment there is much scope for bias. Thus we should not be surprised that, in terms of sensitivity and specificity, such examinations are dismally unreliable.

An important aspect of a medical school's assessment programme is the measurement of knowledge, and this can be done efficiently with methods such as the multiple-choice question (MCQ). The down-side of this strategy is that many MCQ-type examinations have measured little more than the ability to reproduce isolated facts; and a growing awareness of the detrimental effect this may have on student learning, and on acquisition of clinical reasoning skills, has encouraged the development of alternative test formats. In this article we review some of these developments, discuss the limitations and pitfalls in the assessment of clinical reasoning, and offer suggestions for future assessment practice.

Clinical simulations

Work on the assessment of clinical reasoning began in the 1960s and 1970s. At that time clinical reasoning was perceived as a generic quality—a personal attribute reflecting the ability to solve problems in the clinical domain. As such it was thought to be largely independent of factual knowledge and procedural skills; in other words, a good problem-solver would be effective in solving problems irrespective of the clinical circumstances. From this perspective clinical reasoning was akin to a psychological characteristic or personality trait.

In that era, clinical reasoning was typically measured by asking students to respond to problems that could be presented in standardised format with objective scoring of answers. Many different types of simulations were developed, ranging from oral examinations with structured patient protocols to card games. The most popular was the patient management problem (PMP),[4] a written instrument starting with a description of a patient problem. The student was required to collect data sequentially on the history, physical examination, and investigations. The data were disclosed through various technical devices such as tabs or invisible ink, and the student was invited to make diagnostic and management decisions. Some PMPs were rigid, allowing only a linear pathway; others permitted branching and alternative routes—a facility that was enhanced when they were adapted for the computer.[5] The final score was based on the data choices made and the pathway taken, and was interpreted as a measure of clinical reasoning ability.

PMPs were introduced into the assessment procedures of many medical schools and licensing bodies, particularly in North America. Another written instrument was the modified essay question.[6,7] The common denominator of all these techniques was the use of written simulated patient problems to assess ability at clinical reasoning.

These techniques were well received because of their realism rather than their proven validity and reliability. Subsequently, drawbacks emerged that tempered the original enthusiasm and forced us to explore new directions.[1,8] Indeed, the experience necessitated a fundamental reappraisal of the nature of clinical reasoning.

Difficulties with scoring

To score a PMP, it was first necessary to define the optimum data-collection strategy and pathway through the problem. This was achieved by consensus between a group of doctors experienced in management of the relevant conditions; but reaching this consensus was not easy because the doctors were found to collect different amounts of information and to use different pathways. Matters became worse with the introduction of more realistic and complex simulations. Numerous weighting schemes were developed to accommodate such difficulties and component scores were computed to reflect the proficiency, efficiency, and thoroughness of the performance. Surprisingly, these component scores correlated very highly and complex methods were no better than scores based on simple counts.

Low correlations between PMP scores

The second difficulty with PMPs cast doubts on the prevailing concept of clinical reasoning. It turned out that, however calculated, scores on one simulation were not very predictive for a score on another simulation. The correlations of scores between PMPs were consistently

Department of Educational Development and Research, University of Limburg, PO Box 616, 6200 MD Maastricht, Netherlands (C van der Vleuten PhD); **Department of Medicine, University of Adelaide, Australia** (D I Newble MD)

Correspondence to: Dr C van der Vleuten

ribozyme containing a group I intron sequence of *Tetrahymena* coupled to a 3'-portion of a wild-type RNA could repair a defective portion of RNA by targeted *trans* splicing (figure 4). The mutation of tumour suppressor genes, such as p53, leads to a loss of control of cellular proliferation. Perhaps one day we shall be able to restore normal cellular control by restoring wild-type p53 protein.

References

1 Stein CA, Cheng YC. Antisense oligonucleotides as therapeutic agents—is the bullet really magical? *Science* 1993; **261:** 1004–12.

2 Cech TR. The chemistry of self-splicing RNA and RNA enzymes. *Science* 1987; **236:** 1532–39.

3 Guerrier-Takada C, Gardiner K, Marsch T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 1983; **35:** 849–57.

4 Pyle AM. Ribozymes: a distinct class of metalloenzymes. *Science* 1993; **261:** 709–14.

5 von Ahsen V, Schroeder R. RNA as a catalyst: Natural and designed Ribozymes. *Bio Essays* 1993; **15:** 299–307.

6 Uhlenbeck OC. A small catalytic oligoribonucleotide. *Nature* 1987; **328:** 596–600.

7 Forster AC, Jeffries AC, Sheldon CC, Symons RH. Structural and ionic requirements for self-cleavage of virusoid RNAs and trans self-cleavage of viroid RNA. *Cold Spring Harb Symp Quant Biol* 1987; **52:** 249–59.

8 Pley HW, Flaherty KM, McKay DB. Three-dimensional structure of a hammerhead ribozyme. *Nature* 1994; **372:** 68–74.

9 Friedmann T. Progress toward human gene therapy. *Science* 1989; **244:** 1275–81.

10 Yuyama N, Ohkawa J, Inokuchi Y, Shirai M, Sato A, Nishikawa S, Taira K. Construction of a tRNA--embedded-ribozyme trimming plasmid. *Biochem Biophys Res Commun* 1992; **186:** 1271–79.

11 Heidenreich O, Pieken W, Eckstein F. Chemically modified RNA: approaches and applications. *FASEB J* 1993; **7:** 90–96.

12 Hendry P, McCall MJ, Santiago FS, Jennings PA. A ribozyme with DNA in the hybridizing arms displays enhanced cleavage ability. *Nucleic Acids Res* 1992; **20:** 5737–41.

13 Taylor NR, Kaplan BE, Swiderski P, Li H, Rossi JJ. Chimeric DNA-RNA hammerhead ribozymes have enhanced in vitro catalytic efficiency and increased stability in vivo. *Nucleic Acid Res* 1992; **20:** 4559–65.

14 Letsinger RL, Zhang GR, Sun DK, Ikeuchi T, Sarin PS. Cholesteryl-conjugated oligonucleotides: synthesis, properties and activity as inhibitors of replication of human immunodeficiency virus in cell culture. *Proc Natl Acad Sci USA* 1989; **86:** 6553–56.

15 Degols G, Leonetti JP, Lebleu B. Sequence-specific activity of antisense oligonucleotides conjugated to poly (L-lysine) carriers. *Ann NY Acad Sci* 1992; **660:** 331–33.

16 Snyder DS, Wu Y, Wang JL, et al. Ribozyme-mediated inhibition of bcr-abl gene expression in a Philadelphia chromosome-positive cell line. *Blood* 1993; **82:** 600–05.

17 Bertrand E, Pictet R, Grange T. Can hammerhead ribozymes be efficient tools to inactivate gene function? *Nucleic Acids Res* 1994; **22:** 293–300.

18 Dropulic B, Elkins DA, Rossi JJ, Sarver N. Ribozymes: use as anti-HIV therapeutic molecules. *Antisense Res Dev* 1993; **3:** 87–94.

19 Sarver N. Ribozymes: a new frontier in anti-HIV strategy. *Antisense Res Dev* 1991; **1:** 373–78.

20 Yu M, Ojwang J, Yamada O, et al. A hairpin ribozyme inhibits expression of diverse strains of human immunodeficiency virus type 1. *Proc Natl Acad Sci USA* 1993; **90:** 6340–44.

21 Barinaga M. Ribozymes—killing the messenger. *Science* 1993; **262:** 1512–14.

22 Cantor GH, Mcelwain TF, Birkebak TA, Palmer GH. Ribozyme cleaves rex/tax messenger RNA and inhibits bovine leukemia virus expression. *Proc Natl Acad Sci USA* 1993; **90:** 10932–36.

23 Netter HJ, Hsieh SY, Lazinski D, Taylor J. Modified HDV as a vector for the delivery of biologically-active RNAs. *Prog Clin Biol Res* 1993; **382:** 373–76.

24 Koizumi M, Hayase Y, Iwai S, Kamiya H, Inoue H, Ohtsuka E. Design of RNA enzymes distinguishing a single base mutation in RNA. *Nucleic Acids Res* 1989; **17:** 7059–71.

25 Tone T, Kashani SM, Funato T, et al. Suppression of EJ cells tumorigenicity. *In Vivo* 1993; **7:** 471–76.

26 Kashani SM, Funato T, Tone T, et al. Reversal of the malignant phenotype by an anti-ras ribozyme. *Antisense Res Dev* 1992; **2:** 3–15.

27 Kashani SM, Funato T, Florenes VA, Fodstad O, Scanlon KJ. Suppression of the neoplastic phenotype in vivo by an anti-ras ribozyme. *Cancer Res* 1994; **54:** 900–02.

28 Wright L, Wilson SB, Milliken S, Biggs J, Kearney P. Ribozyme-mediated cleavage of the bcr/abl transcript expressed in chronic myeloid leukemia. *Exp Hematol* 1993; **21:** 1714–18.

29 Shore SK, Nabissa PM, Reddy EP. Ribozyme-mediated cleavage of the bcr/abl oncogene transcript: in vitro cleavage of RNA and in vivo loss of p210 protein kinase activity. *Oncogene* 1993; **8:** 3183–88.

30 Lange W, Cantin EM, Finke J, Dolken G. In vitro and in vivo effects of synthetic ribozymes targeted against bcr/abl mRNA. *Leukemia* 1993; **7:** 1786–94.

31 Kobayashi H, Dorai T, Holland JF, Ohnuma T. Reversal of drug sensitivity in multidrug-resistant tumor cells by an MDR1 (PGY1) ribozyme. *Cancer Res* 1994; **54:** 1271–75.

32 Holm PS, Scanlon KJ, Dietel M. Reversion of multidrug resistance in the P-glycoprotein-positive human pancreatic cell line (EPP85-181RDB) by introduction of a hammerhead ribozyme. *Br J Cancer* 1994; **70:** 239–43.

33 Kiehntopf M, Brach M, Licht T, et al. Ribozyme-mediated cleavage of the MDR-1 transcript restores chemosensitivity in previously resistant cancer cells. *EMBO J* 1994; **13:** 4645–52.

34 Herschlag D, Khosla M, Tsuchihashi Z, Karpel RL. An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J* 1994; **13:** 2913–24.

35 Bertrand EL, Rossi JJ. Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. *EMBO J* 1994; **13:** 2904–12.

36 Sullenger BA, Cech TR. Tethering ribozymes to a retroviral packaging signal for destruction of viral RNA. *Science* 1993; **262:** 1566–69.

37 Sullenger BA, Cech TR. Ribozyme-mediated repair of defected mRNA by targeted trans-splicing. *Nature* 1994; **371:** 619–22.

38 Cech TR, Uhlenbeck OC. Hammerhead nailed down. *Nature* 1994; **372:** 39–40.

low. If clinical reasoning was a generic ability one would expect a high correlation. This phenomenon is not unique to tests of clinical reasoning: it applies to other aspects of clinical competence and has become known as case or content specificity.[9] In essence it is saying that the ability to deal effectively with one clinical situation does not guarantee that one can deal effectively with another—an observation that will hardly surprise the honest medical practitioner. Unfortunately it has caused major practical difficulties for those who assess students: it means that any examination based on a single clinical problem or a small number of problems will yield an unstable or irreproducible score that cannot be relied upon for pass/fail decisions. Assessments with complex simulations demand long and impractical examinations with perhaps more than a day of testing time.

High correlations between reasoning tests and knowledge tests

The third difficulty arose when a very high correlation was seen between scores of clinical-reasoning tests and those of knowledge tests. Again this casts doubt on the notion of clinical reasoning as an independent construct, reinforcing the view that clinical reasoning depends heavily on a relevant knowledge base.

Performance of experienced versus inexperienced practitioners

We might expect that, in a valid test of clinical reasoning, experienced doctors would outperform inexperienced medical students. With PMPs there was hardly any difference—indeed, students sometimes scored better.

New concepts in clinical reasoning

Much research has been done on the different mental processes of novices (eg, medical students) and experts (eg, experienced clinicians) in tackling a problem.[10,11] A few key points can be made. Firstly, expert reasoning seems to be strongly connected to knowledge. However, it is not simply the knowledge but the way the knowledge is stored, retrieved, and used that distinguishes the expert from the novice. The expert clinician accumulates knowledge in the context of concrete medical problems—a process that enhances the chance of effective retrieval. With experience, the need for a reasoning process diminishes as the mental processes become automated into patterns or "scripts". The process becomes highly efficient and effective. Professional expertise thus develops as a transition from a conceptually high and rational knowledge base (acquired from educational experience) to a non-analytical ability to recognise and handle familiar clinical situations (acquired from extensive clinical experience). Unfortunately, expertise is not easily transferable from one context to another, being clinically specific. One effect of this is evident in real life: highly specialised practitioners tend to refer problems outside their area of expertise more readily and more frequently that those with a broader practice base.

This theoretical framework offers an explanation for some of the disappointments experienced in the development of generic tests of clinical reasoning. It is now obvious why we cannot measure clinical reasoning by assessing performance, in however detailed a fashion, on a single case or problem. Relevant medical knowledge and previous experience play a central role in successful clinical problem-solving, and this naturally varies from case to case. The thinking processes of an experienced doctor are qualitatively different from those of a student. Thus, we should not be surprised that traditional simulations, which usually reward thoroughness rather than speed and quality of responses, give students an advantage over experienced clinicians.

New developments in assesment

When discussing the assessment of clinical reasoning we should not neglect the substantial developments in testing of clinical competence for certification and recertification.[12] What we have said about the content specificity of simulations applies equally to clinical examinations, and in many such examinations the traditional long and short case vivas are now being replaced or supplemented by the Objective Structured Clinical Examination (OSCE). The OSCE not only increases the sample of skills tested in a limited time but also reduces bias and improves reliability.

Clinical reasoning tests such as PMPs and the long cases were used widely because they seemed to represent what happens in real life. Now the long case has been, or should have been, abandoned on account of the small, non-standardised, and unrepresentative sample of student behaviour that is observed. The PMP has largely been abandoned because of similar drawbacks, for which the evidence has been more readily accepted. Just as the OSCE has provided a better solution to the drawbacks of traditional clinical examination, more efficient writing formats have been developed to replace the PMP.[12] An example has been the "key features" approach.[13] Such methods are based on the assertion that successful handling of a clinical case depends on a few critical elements. For instance, diagnosis in a patient with chest pain or transient loss of consciousness is more likely to depend on eliciting and evaluating key elements of the history than on physical examination findings. When testing students on their problem-solving skills in these conditions one would focus on such key pieces of information. This saves valuable testing time and allows more conditions to be assessed in a fixed period, thus lessening the hazards of content specificity.

The finding of high correlations between clinical reasoning tests and knowledge tests has challenged the notion that test format dictates what is being measured. For example, a common belief is that the writing of an essay demands higher level mental processes than does an MCQ. In reality many essay-type tests measure little more than the recall of factual knowledge, whereas a well constructed MCQ may demand a great deal of analytical thinking. In other words, the inherent measurement characteristics of test methods are less of an obstacle than they once seemed.[14] When developing an assessment procedure one should concentrate primarily on defining the task to be posed to the students (eg, the sample of relevant cases with which they should be familiar).[15] The response format is of secondary importance.

For a long time it has been obvious that examinations have enormous impact on what students learn and how they go about their learning. Very little attention has been paid to this issue in practice; for instance, over-use of MCQ-type examinations that elicit recall of isolated facts has encouraged rote learning at the expense of an understanding of the subject material and its application in the clinic. Examinations define the curriculum, often in a way that is not intended by the academic staff.[16] The

lesson is that, if we wish students to develop effective clinical reasoning skills, we must devise examinations that reflect this intent.

Implications

How then should we go about assessing clinical reasoning? Although we have no single instrument to measure this aspect of competence we have learned much about the process of clinical reasoning and how to test it. The state of the art resembles that of many aspects of medical care: as medical practitioners we do not withhold an effective treatment because it falls short of providing instant and permanent cure; as medical teachers, we should be prepared to introduce new assessment methods when they are proven superior to those currently in use.

Here are a few suggestions as to how we might begin this process. Test material used for the assessment of clinical reasoning should be as realistic as possible; for example, question-stems requiring interpretation of information from the history could be written in the words of the patient, or we could use this morning's surgery hour as a source. Since knowledge and reasoning are closely linked we should not be afraid to test knowledge, but the challenge is to test it in a way that requires application to a concrete clinical situation. A simple check would be to see whether the test question can be answered correctly without the clinical details. Why not allow students to bring their books—remembering that in real life doctors frequently consult colleagues or books to improve the quality of their diagnosis and management?

Those who devise examinations should abandon their preoccupation with test formats and methods and pay more attention to content. Essentially all methods allow assessment of clinical reasoning. Whatever the format, the test material should focus on important clinical decisions, preferably requiring the student to weigh probabilities or decide upon actions. Assessments must clearly be based on large numbers of case scenarios. For high-stake examinations such as finals, several hours of testing time are required for fair decisions. Lastly, we should monitor the influence of our examinations upon student learning, checking assiduously that the content reflects the objectives of the curriculum. In such a way we will ensure that the dog of education is no longer wagged by the tail of assessment.

References

1 Newble DI, van der Vleuten CPM, Norman GR. Assessing clinical reasoning. In: Higgs J, Jones M, eds. Clinical reasoning in the health professions. Oxford: Butterworth Heinemann (in press).

2 Physicians for the twenty-first century. Report to the Panel on the General Professional Education of the Physician and College Preparation for Medicine (GPEP). Washington, DC: American Association of American Medical Colleges, 1984.

3 General Medical Council. Tomorrow's doctors: recommendations on undergraduate medical education. London: GMC, 1993.

4 McGuire CH, Babbott D. Simulation technique in the measurement of problem solving skills. *J Educ Measurement* 1967; **4:** 1–10.

5 Norcini JJ, Meskauskas JA, Langdon LO, Webster GD. An evaluation of a computer simulation in the assessment of physician competence. *Eval Health Professions* 1986; **9:** 286–304.

6 Hodgkin K, Knox JDE. Problem centred learning: the modified essay question in medical education. Edinburgh: Churchill-Livingstone, 1975.

7 Feletti GI, Saunders NA, Smith AJ. Comprehensive assessment of final-year medical student performance based on undergraduate programme objectives. *Lancet* 1983; ii: 34–37.

8 Swanson D, Norcini J, Grosso L. Assessment of clinical competence: written and computer-based simulations. *Assessment Eval Higher Educ* 1987; **12:** 220–46.

9 Elstein A, Shulman LS, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Cambridge, Mass: Harvard University Press, 1978.

10 Schmidt H, Norman G, Boshuizen HA. Cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990; **65:** 611–21.

11 Norman G, Regeher G. Contemporary issues in psychology: implications for teaching and learning. In: Rothman A, Cohen R, Norman G, eds. Proceedings of the Sixth Ottawa Conference 1994 (in press).

12 van der Vleuten C, Newle D, eds. Methods of assessment in certification. In: Newble D, Jolly B, Wakeford R, eds. The certification and recertification of doctors: issues in the assessment of clinical competence. Cambridge: Cambridge University Press, 1994: 105–25.

13 Bordage G, Page G. An alternative approach to PMPs: the "key features" concept. In: Hart IR, Harden RM , eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1987: 59–75.

14 McGuire C. Written methods for assessing clinical competence. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Heal Publications, 1987: 46–58.

15 Newble D, Dawson B, Dauphinee D, et al, eds. Guidelines for the assessment of clinical competence. *Teaching Learning Med* 1994; **6:** 213–20.

16 Newble D, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983; **17:** 165–71.