Medical Education

Vol. 41, Issue 12,(December 2007), Page 1113-1242

Clinical Performance

My Story

Clinical Expertise

Contents

Volume 41 Issue 1 January 2007

Volume 41 Issue 2 February 2007

Volume 41 Issue 3 March 2007

Volume 41 Issue 4 April 2007

Volume 41 Issue 5 May 2007

Volume 41 Issue 6 June 2007

Volume 41 Issue 7 July 2007

Volume 41 Issue 8 August 2007

Volume 41 Issue 9 September 2007

Volume 41 Issue 12 December 2007

in this issue

Clinical expertise

Clinical diagnosis is at the heart of medicine. Since the discovery of X-rays, technology has transformed the confirmatory investigation of initial diagnoses but the subtle evaluation of signs and symptoms remains as the crucial core skill required of doctors. The educational journey from first year medical student to competent practitioner is about much more than knowledge acquisition. It is this journey from novice to expert that provides much of the special fascination of medical education as a distinct branch of educational research.

In this issue we have chosen to focus principally on the central theme of clinical decision-making. This has been achieved through commissioning papers which were then processed through our standard blind-peer review procedure. In doing so, we hope to provide our readers with an up-to-date, convenient, and authorative source of information about the concepts and evidence underpinning our understanding of clinical reasoning, judgement, and expertise. The terms employed in the collection of papers are themselves of interest as they immediately shape our approach to the practise of medicine. The papers have direct relevance to the field of medicine but also have wider implications.

Wherever important decisions have to be made in environments that are complex and rich in ambiguity and uncertainty, the educational processes which help practitioners understand and master the different problem-solving strategies that can be applied will be of relevance.

See pages *1133–1192*

Undergraduate medical students' exposure to clinical ethics

Students regularly experienced situations in clinical teaching settings that challenged their ethical values. Medical students' reported low levels of confidence in their ability to address these challenges, and perceived a need for additional support from clinical teachers. Cordingley *et al* use a cross sectional survey using web-based and paper questionnaires to explore medical students' experience of challenges to their ethical knowledge and understanding in clinical practice, and to investigate their need for support when faced with such challenges. Clinical teachers were identified as the most relevant providers of guidance. Appropriate educational provision requires medical educators to be equipped with the knowledge and the skills to engage with students' ethical concerns.

See pages *1202–1209*

The role of portfolios in medical education

The move towards competence-based medical education has created a need for instruments to support and assess competence development. Portfolios seem suitable but mixed reports of their success are emerging. Further clarification is needed about which elements in the implementation of portfolios in medical education are crucial. Using a systematic review of empirical studies on portfolios Driessen *et al* found that important factors for success were: clearly communicated goals and procedures, integration with curriculum and assessment, flexible structuring, support through mentoring, measures to heighten feasibility and reduction of time required. Without assessment, portfolios were vulnerable to competition from other summative assessment instruments. Further studies should focus on the effectiveness and user friendliness of portfolios, the merits of holistic assessment procedures, and the competencies of effective portfolio mentors.

See pages *1224–1233*

announcement

New Editor for Medical Education

Medical Education is pleased to announce the appointment of a new Editor, Dr Kevin Eva, PhD.

Dr Eva is Associate Professor and Associate Chair of the Department of Clinical Epidemiology and Biostatistics at McMaster University and the new Editor in Chief of Medical Education. His formal academic title does not fully describe the breadth of his considerable contribution to medical education and medical education research over the past decade. His research interests span the acquisition, development, maintenance, and assessment of clinical judgment and expertise. This places his research and his expertise at the heart of medicine. His contributions to our understanding of clinical reasoning have been recognised internationally through formal awards including, in 2007, the Award for Best Paper by the Professions Education Division of the American Educational Research Association. His reputation and impact clearly extend beyond medicine and into mainstream education. Equally, his reputation within our discipline is now global in its reach. He is in constant demand as a consultant, speaker, and research collaborator across the world.

As an editorial colleague, I have had the privilege of seeing at first hand the quality of Dr Eva in his unfailingly constructive approach to the work of other researchers. The combination of intellectual rigour with generosity of spirit, I believe, characterises the community of medical educators and researchers. Kevin is the embodiment of these values. He has become increasingly involved in the ethics of medical education research, acting as the Chair of the Quality and Standards Group of the journal until 2006. We can expect that he will continue to be a staunch supporter of initiatives to uphold the integrity of the academic publishing process and the ethical foundations of educational research in his new role.

Kevin brings to the journal not only his own considerable expertise in research. As Deputy Editor of Medical Education and with his other editorial responsibilities, he has a deep understanding of the intellectual and practical issues faced by academic journals. We welcome him to his new role, confident that with his total commitment to the improvement of medical education through high quality research he will add lustre to the journal.

Dr Graham Buckley
Executive Editor

commentaries

Clinical expertise research: a history lesson from those who wrote it

Remy M J P Rikers & Peter P J L Verkoeijen

People have always been intrigued by exceptional performances demonstrated in different domains by experts. It is mind-boggling how experts can perform extremely complex tasks almost effortlessly and without mistakes. After decades of research into expertise, our understanding of expert performance has grown substantially, but there are still many unresolved questions.

> *In order to be considered expert, doctors must excel not only in diagnostic performance and therapy, but must also be highly proficient in communication skills*

In this quest for understanding expert performance, research into the development of medical proficiency plays a prominent role. Almost everybody will have to deal with a serious illness sooner or later in life and hence will benefit from competent doctors who are capable of dealing with the problem effectively. However, unlike many other areas of expertise research, medical expertise is extremely multi-faceted.[1] That is, in order to be considered an expert, doctors (i.e. practitioners) must excel not only in diagnostic performance and therapy, but must also be highly proficient in communication skills, so that they are able to explain a problem to patients, colleagues and students. Furthermore, being considered an expert by patients or students does not necessarily imply that colleagues share this opinion, and vice versa. Given this complexity of medical expertise, one might wonder whether it is possible at all for a doctor to excel in all these different facets. This intricacy of medical expertise has led many researchers to concentrate first on areas of expertise that are less multi-dimensional, but still highly complex. It is perhaps not surprising that chess has become the drosophila of expertise research. Chess has qualities that other domains, including medicine, lack.[1] For instance, chess skill can be measured (i.e. ELO rating) and experimental studies can often be carried out in laboratory settings without sacrificing ecological validity.

> *Given this complexity of medical expertise, one might wonder whether it is possible for a doctor to excel in all these different facets*

Moreover, one can quite easily distinguish between good and bad decisions (i.e. in terms of chess moves) and the player's communication skills are almost completely irrelevant. Expertise in chess boils down to the ability to play the most optimal move in any position. Although expertise in chess has many unique characteristics that set it apart from expertise in other domains, the pioneering studies of expertise in chess by De Groot[2] and later by Chase and Simon[3] have had substantial impact on (medical) expertise research. As a matter of fact, many of the studies reported in this themed section are based upon the experimental approach that originates from studies of chess expertise. For instance, the prominent role of free recall in medical expertise studies is an example of its influence. It is assumed that a measure, such as free recall, tells us something about, or is a reflection of, the organisation of knowledge in the doctor's or student's mind. However, this influence of chess on medical expertise has been subject to only limited research because its perspective is primarily drawn from cognitive science.[4] Many of the skills demonstrated by doctors in clinical settings have been ignored because they cannot be investigated in laboratory settings, where, for example, doctors

> *Major challenges for future research into medical expertise will involve identifying and describing the differences between expertise levels in terms of knowledge and skills*

Department of Psychology, Erasmus University Rotterdam, Rotterdam, The Netherlands.

Correspondence: Remy M J P Rikers, Department of Psychology, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Tel: +31 10 408 9655; Fax: +31 10 408 9009; E-mail: Rikers@fsw.eur.nl

doi: 10.1111/j.1365-2923.2007.02920.x

work with paper cases instead of real patients with whom they are able to interact. Nonetheless, the artificiality of tasks performed in laboratory settings does not prevent doctors from exhibiting their superior skills and has provided us with better understanding of the changes that occur in the structure and content of medical knowledge with increasing levels of expertise.

Major challenges for future research into medical expertise will not only involve identifying and describing differences between expertise levels in terms of knowledge and skills, but also determining how these expert characteristics can be effectively passed on to medical students. It is extremely difficult, not only in medicine, but in all areas of expertise research, to identify those characteristics or criteria that are essential to the development of expertise. Many of the (selection) criteria we use may not be adequate and may even be harmful. One of our favourite illustrations of this point uses the athletic event of the high jump. Most people will agree that, in order to become a successful high jumper, height is essential. At least, this is what Stefan Holm, a Swedish high jumper, was told in the early years of his career. Although Holm was only 181 cm tall, this did not discourage him from practising. His perseverance paid off and, during the last Olympic Games in Athens, he won a gold medal. Thus, whatever else it may take to become a good high jumper, it is not necessarily height that determines the outcome.

It is extremely difficult, not only in medicine, but in all areas of expertise research, to identify those characteristics or criteria that are essential to the development of expertise

Although many of us might assume that talent plays an essential role in achieving excellence in a particular domain, recent studies have shown that this premise is not entirely true and have demonstrated that practice is more important to the development of excellence than a particular set of genes.[5,6] However, it isn't just any type of practice that will turn somebody into an expert. Practice that aims at achieving excellence should be focused on those aspects of performance that have not yet been mastered, and should allow room for the making and correcting of errors, with adequate feedback on performance. Ericsson and colleagues have dubbed this type of intense training *deliberate practice*.[7] It forces an individual – often with the help of a trainer – to identify and to overcome their limitations. Studies on different areas of expertise have shown that it takes at least 10 years of deliberate practice for someone to excel in a domain. This is a very long period and it involves many personal sacrifices, which explains why so many people drop out on the way.

The aim of this special section is to critically revisit this and other views on the development of medical expertise by involving those who developed them and those who stand on their shoulders. Theories of medical expertise will be discussed and evaluated from both theoretical and empirical perspectives, as well as with regard to their implications for medical education. We believe that this perspective will help us to better understand what it takes to become a medical expert, and we hope that it will serve as a launching pad for future research into expertise.

Studies on different areas of expertise have shown that it takes at least 10 years of deliberate practice for someone to excel in a domain

REFERENCES

1 Ericsson K.A. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004; **79** (10 Suppl): 70–81.

2 De Groot A. *Thought and Choice in Chess* [Original work published 1946] The Hague: Mouton, 1978.

3 Simon HA, Chase WG. Skill in chess. *Am Sci* 1973;**61**:394–403.

4 Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;**39**:418–27.

5 Ericsson KA, Lehmann AC. Expert and exceptional performance: evidence on maximal adaptations on task constraints. *Annu Rev Psychol* 1996;**47**:273–305.

6 Howe MJA, Davidson JW, Sloboda JA. Innate talents: reality or myth? *Behav Brain Sci* 1998;**21**:399–442.

7 Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 1993;**100**:363–406.

Prototypes and semantic qualifiers: from past to present

Georges Bordage

Clinical reasoning has been the focus of a rich body of scholarship in medical education, for which the work of Arthur Elstein and his colleagues, on the Medical Inquiry Project at Michigan State University, has served as an early platform. A major thread in this research on clinical reasoning concerns our investigations of prototypes and semantic networks, conducted over the last several years. This essay reviews the development and main results of our work and formulates 3 main lessons gleaned from it, related to the dynamic interaction among theories, methods and practice, the importance of theory testing and theory building, and the versatile nature of knowledge organisation in memory.

> *According to structural semantics, knowledge is given meaning through networks of relationships represented by dichotomous abstract qualifiers*

One of the main conclusions of the work by Elstein *et al.* was already becoming apparent when I first met Arthur Elstein in 1975. It was later reported in their book, *Medical Problem Solving: an Analysis of Clinical Reasoning*: 'The difference between experts and weaker problem solvers are more to be found in the repertory of their experiences, organised in long-term memory, than in differences in the planning and problem-solving heuristics employed.'[1] A couple of years later, this conclusion led me to ask how medical knowledge is organised in memory and to test an accepted theory of knowledge organisation at the time, prototype theory (led by the work of Rosch and Mervis[2–4]), as it might apply to medical knowledge. According to prototype theory, medical categories would be organised in memory around representative exemplars, the prototypes that serve as anchors for other members of the category. For example, gastrointestinal disorders would be organised around duodenal ulcers, gastritis and Crohn's disease. Other more peripheral members of the category might include cirrhosis, malabsorption, Meckel's diverticulitis, ileus and colon cancer. I conducted 4 studies to gather converging evidence for the presence of prototypes in the memory of medical students and experienced family doctors. The 4 studies consisted of:

1 a free recall task of disorders belonging to 14 categories of medical knowledge (i.e. organ systems [e.g. respiratory, endocrine, musculoskeletal], pathophysiological mechanisms [e.g. inflammation] and major patient complaints [e.g. abdominal pain]);

2 a judgement task measuring the degree of representativeness of the category members from study 1;

3 a free recall task of the attributes of the same disorders, and

4 a measure of response time for prototypical (central) versus peripheral members as measured in study 2.

The results from these 4 studies provided converging evidence for the presence of prototypical disorders in the memory of medical students and experienced doctors.[5] The disorders recalled first were judged to be more representative than those recalled later, had more attributes in common (family resemblance), and were recalled faster and more accurately than the peripheral members.

> *Successful diagnosticians use semantic qualifiers more frequently and in more diversified sets in their discourses than diagnosticians who are less successful*

Subsequently, 2 education corollaries from prototype theory were tested using observational studies, namely, that prototype formation was related to medical school courses that contained fewer disorders ($r = -0.58$) and more intermediate level disorders ($r = +0.73$; e.g. angina pectoris compared with coronary insufficiency, a superordinate category, or Prinzmetal angina, a subordinate category).[6] This reflected a strategy policy of 'less is more' in teaching and learning prototypes, where the initial presentation was limited to 4 or 5 main causes of shoulder pain, for example, rather than all 32 causes in the textbook. Prototype formation was also fostered by presenting intermediate level disorders to students. Osler had understood and applied these

Department of Medical Education, College of Medicine, University of Illinois at Chicago, Chicago, Illinois, USA

Correspondence: Dr Georges Bordage, Department of Medical Education (M/C 591), College of Medicine, University of Illinois at Chicago, 808 South Wood, Chicago, Illinois 60612-7309, USA. Tel: 00 1 312 996 7349; Fax: 00 1 312 413 2048; E-mail: bordage@uic.edu

doi: 10.1111/j.1365-2923.2007.02919.x

principles a century before by dwelling on 2 disorders, pneumonia and typhoid, with the clerks working on his ward because '…if thoroughly understood by the students, [pneumonia and typhoid] give them a satisfactory foundation on which to build their later experience' (in Cushing[7]). He also encouraged his peers to teach less, for 'the student tries to learn too much, and we the teachers try to teach him too much – neither, perhaps, with great success'.[8]

Theory, practice and research methods constantly interact to provide a dynamic interplay that helps move theory and practice forward

By the mid-1980s, the questions I was asking had evolved and now focused on the nature of the relationships that link pieces of knowledge in memory. Prototype theory did not provide an adequate conceptual framework with which to address this new question, but structural semantics provided a promising theoretical basis. According to structural semantics, knowledge is given meaning through networks of relationships represented by dichotomous abstract qualifiers (axes), such as disorders associated with acute versus chronic onsets or local versus systemic manifestations. For example, 4 sets of semantic qualifiers (axes) go a long way towards providing a mental scaffolding for knowledge about low back pain, that is, acute–chronic, immediate–delayed, above–below the knee and local–systemic. Thus, when an otherwise healthy man complains of a lower back pain that began the previous day when he lifted a heavy object and that ran down his right leg, the clinician thinks: 'Here's an *acute*, *immediate* lower back pain radiating *below-the-knee* pain, with no *systemic* manifestations; he's more likely to suffer from a herniated disc with a compression than a sprain, an inflammation, associated with *delayed* and *above-the-knee* manifestations.' Furthermore, links between clinical and basic science knowledge are brought to mind through the semantic qualifiers, such as 'immediate pain', related to a nerve compression, compared with 'delayed pain', related to an inflammation.

In a series of observational studies conducted over a decade, using qualitative think-aloud protocols for neurological, gastrointestinal, intensive care and rheumatological conditions, we observed that successful diagnosticians used semantic qualifiers more frequently and in more diversified sets in their discourses than diagnosticians who were less successful. They 'organised the symptoms and signs into coherent systems of relationships of abstract qualities, [and have] broader and deeper representation and understanding of the problems'.[9] (For a summary and educational implications, see Bordage.[9]) Those with diagnostic difficulties or inaccuracies stuck with a factual, literal view of the case, and failed to see the more abstract semantic dimensions of the problem, expressed with semantic qualifiers. This was true for both medical students and experienced doctors; it represents a common characteristic that cuts across levels of experience and taps into the abstract mental scaffolding used to relate bits and pieces of knowledge.

This work on semantic qualifiers provided the opportunity not only to test structural semantic theory in medicine, but also to contribute to further development of the theory by proposing 4 categories of discourses based on 2 organisational dimensions, a semantic dimension, as measured by the number of semantic qualifiers used, and a syntactic dimension, as measured by the extent (length) of the discourses:

1 reduced discourses (limited semantic content and limited discourses);

2 dispersed (limited semantic content and extended discourses);

3 elaborated (semantic richness and extended discourses), and

4 compiled (semantic richness and limited discourses).[10]

The fourth category is similar to the notion of encapsulation described by Schmidt and Boshuizen.[11] Greater diagnostic accuracy and understanding was associated with semantically rich discourses, either elaborated or compiled.

Programmatic research provides depth of understanding over time

Based on think-aloud protocols and the hypothesis that more successful diagnosticians would use more semantic qualifiers to represent a patient's chief complaint, a case-control observational study was conducted.[12] The results showed that successful diagnosticians (case) used over twice as many semantic qualifiers in their characterisation or representation of the chief complaints (means of 3.8 versus 1.6) as unsuccessful diagnosticians and entertained twice as many diagnoses in their differential (means of 2.9 versus 1.5); for example: 'This looks like a *localised*, *acute*, *recurrent*, *large* joint *mono* arthritis that is more likely associated with *gout* than a *septic arthritis* or any small joint arthritis like *rheumatoid arthritis*.'[13] The role of problem representation, as portrayed here by the use of semantic qualifiers to construct a more abstract mental image of the problem, is an important analytical strategy for solving problems (see, for example, Brenner *et al.*[14] in mathematics education).

In the late 1990s, we tested an education corollary of the structural semantic theory, hypothesising that promoting elaborated semantic representations of chief complaints during case presentations would improve diagnostic accuracy. The results from this experimental study were mixed.[15] Although the students in the experimental group increased their use of semantic qualifiers to describe their cases, this was not associated with improved accuracy. It is likely that when the students initially learned this knowledge, earlier in their education, it was not organised according to semantic networks, and so although they increased their use of semantic qualifiers, they did not have existing relational semantic networks to tap into. The mixed results led to the design of a new, and yet to be conducted, study based on experimental procedures developed by Woods *et al.*[16] to teach basic science concepts related to clinical knowledge; the same procedures will be used here but to teach semantic networks and scaffolding.

Three main lessons can be gleaned from our work on prototypes and semantic networks over the past 3 decades:

1 theory, practice and research methods constantly interact to provide a dynamic interplay that helps move theory and practice forward;

2 programmatic research provides depth of understanding over time, and

3 the human mind and memory are flexible and versatile entities.

First, medical education is essentially a practical field where theories and evidence can inform practice and vice versa. In an American Education Research Association symposium 15 years ago (1992), the late Terry Mast advocated that: 'Theory should drive research; research should drive theory; practice should drive both research and theory… moreover, research and theory should have some influence on practice' (in Bordage and Williams[17]). In our latest test of semantic theory, mixed results have prompted us to go back and test the theory from a different perspective, thus highlighting the dynamic nature of the research enterprise. Norman[18] has argued that theories are dynamic entities set forth to be proven or disproven and to be built upon. This has been and continues to be the case over the years in our work on prototypes and semantic qualifiers. Similarly, a single method is unlikely to provide satisfactory answers to all questions. Competing hypotheses and multiple, converging methods are more productive than single hypotheses and methods. The theory being tested and the questions addressed should determine which methods to use, whether protocol analyses, observational studies or experimental studies – not the other way around.

> *The human mind and memory are flexible and versatile entities*

Second, theory testing, theory building and depth of understanding are facilitated through long-term, programmatic research as opposed to isolated, opportunistic research. A programme of research allows for concepts and variables to be measured and tested systematically. Most research in medical education today consists of isolated studies conducted without any conceptual or theoretical foundation.[19] Theories should not only frame research questions and educational innovations, although that alone would help greatly in medical education, but should also be tested and built upon. As the interactions among theories, methods and practice evolve, the more likely it is that sound educational principles can emerge to better inform fellow researchers, clinical teachers, administrators, policy makers, and licensing and certification agencies, and eventually have some impact on patient care, the ultimate goal of medical education.

Third, in the beginning I had asked how medical knowledge was organised in memory, as if there were a single fixed, crystallised system of organising knowledge in memory. Research has shown that memory is fluid and flexible. Consequently, knowledge and its organisation can be represented in numerous ways, such as prototypes, either as abstractions, as in our work, or as instances, as in Norman and Brook's work on non-analytical reasoning,[20] or as illness scripts, causal networks and semantic structures, to name a few (see Norman[21] for a review). Eva[22] has shown that learning is maximised by promoting both analytical and non-analytical (pattern recognition) structures and processes during clinical education. The most recent results by Ark *et al.*,[23] showed enhanced diagnostic accuracy when learners compared and contrasted categories, thereby fuelling the authors' speculations about better understanding the diagnosticity of the signs and symptoms, and raise yet another hypothesis about the potential dynamic role that semantic qualifiers, with their dichotomous and contrasting features, might play in facilitating diagnosis.

Any quest to unravel how knowledge is organised in memory has to include consideration of the versatile nature of the human mind in response to situation-specific demands within and across individuals, which brings us back to the concept of case-specificity introduced by Elstein *et al.*[1] in their Inquiry Project of the 1970s (see Norman *et al.*[24] for a more recent iteration). Basic science and clinical instruction will be best served by recognising both the structural nature of knowledge and the corresponding experiential instances in memory, which emerge in learners from repeated exposure to clinical cases and didactic instruction.[25,26] This is akin to Ericsson's deliberate practice with feedback[27] and the mixed practice described by Hatala *et al.*,[28] where deliberate, mixed practice provides opportunities to build a repertoire of instances, concepts and relationships, and the feedback to highlight the accuracy of the decisions and the structural integrity of the instances, concepts and relationships. As more becomes known about the multiple facets and interdigitations of clinical reasoning and knowledge organisation, more attention can focus on instructional strategies that can optimise clinical reasoning, as exemplified by the work of Eva and collaborators, and, if needed, help identify and offer remediation to medical students and house officers who need assistance in becoming expert clinicians.

REFERENCES

1 Elstein SA, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978;276.
2 Mervis CB, Catlin J, Rosch E. Relationship among goodness of example, category norms, and word frequency. *Bull Psychon Soc* 1976;**7**:283.
3 Rosch E, Mervis CB. Family resemblance: studies in the internal structure of categories. *Cognit Psychol* 1975;**7**:573.
4 Rosch E. Principles of categorization. In: *Rosch E, Lloyd BB, eds. Cognition and Categorization*. Potomac, MD: Erlbaum Press 1978;27–48.
5 Bordage G, Zacks R. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ* 1984;**18**:406–16.
6 Bordage G. The curriculum overloaded and too general. *Med Educ* 1987;**21**:183–8.
7 Cushing H. *The Life of Sir William Osler*, Vol. 1. Oxford: Clarendon Press 1925;460.
8 Osler W. After Twenty-Five Years. *Montreal Medical Journal* 1899; **28**:823.
9 Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med* 1994;**69**:883–5.
10 Bordage G, Lemieux M. Structuralism and medical education: a comparative study of the cognitive strategies of novice physicians. *Semiotic Inquiry* 1986;**6**:143–79.
11 Schmidt H, Boshuizen H. Encapsulation of biomedical knowledge. In: Evans D, Patel V, eds. *Advanced Models of Cognition for Medical Training and Practice*. New York: Springer Verlag 1992;265–82.
12 Chang R, Bordage G, Connell K. The importance of early problem representation during case presentations. *Acad Med* 1998;**73 (Suppl)**:109–11.
13 Bordage G, Connell K, Chang R, Gecht M, Sinacore J. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. *Acad Med*, 1997;**72 (Suppl)**: 37–9.
14 Brenner ME, Mayer RE, Moseley B, Barr T, Duran R, Reed BS, Webb D. Learning by understanding the role of multiple problem representations in learning algebra. *Am Educ Res J* 1997;**34**:663–9.
15 Nendaz M, Bordage G. Promoting diagnostic problem representation. *Med Educ* 2002;**36**: 761–7.
16 Woods NN, Neville AJ, Levinson AJ, Howey EHA, Oczkowski WJ, Norman GR. The value of basic science in clinical diagnosis. *Acad Med* 2006;**81 (Suppl)**:124–7.
17 Bordage G, Williams R. Relationship between theory and practice. *Prof Educ Res Q* 1992;**14**:1–4.
18 Norman GR. Editorial: theory-testing research versus theory-based research. *Adv Health Sci Educ Theory Pract* 2004;**9**:175–8.
19 Cook D, Beckman T, Bordage G. Quality of reporting experimental studies in medical education: a systematic review. *Med Educ* 2007; **41**:737–45.
20 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract* 1997;**2**:173–84.
21 Norman GR. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;**39**:418–27.
22 Eva K. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39**:98–106.
23 Ark TK, Brooks LR, Eva K. The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ* 2007;**41**:281–7.
24 Norman GR, Bordage G., Page G, Keane D. How specific is case specificity? *Med Educ* 2006;**40**: 618–23.
25 Bowen J. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 2006;**355**:2217–25.
26 Norman G. Building on experience – the development of clinical

reasoning. *N Engl J Med* 2006;**355**:2251–2.
27 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79 (Suppl)**:70–81.
28 Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ Theory Pract* 2003;**8**:17–26.

Assessment: do we need to broaden our methodological horizons?

Ayelet Kuper, Scott Reeves, Mathieu Albert & Brian David Hodges

Although medical education is a broad field of research and practice, it has come to be dominated by issues of assessment. Reasons for this emphasis range from the focus on accountability for educational outcomes[1] to the established relationship between assessment and student motivation.[2] Researchers in the domain, especially in North America, have largely focused on methodologies taken from psychometrics and have overlooked the broader social sciences literature devoted to the analysis of social behaviour and social interaction. In this commentary we provide a critique of the ubiquitous use of psychometric methodologies and perspectives and argue that the social sciences offer other rich methodological resources for the study of assessment.

Within medical education research, evaluation is almost always carried out using a set of appraisal tools that are collectively known as psychometrics. We talk about whether a test is valid (whether it measures the thing we want to measure) and whether it is reliable (whether it measures it in a reproducible fashion). Psychometrics has been very successful in evaluating the assessment of many aspects of medical training. It has, for example, allowed the medical education community to systematically evaluate different measures of medical content knowledge,[3] as well as to show that technical skills can be assessed in a reproducible, valid manner.[4]

The social sciences offer other rich methodological resources for the study of assessment

What is rarely made explicit, however, is what the use of psychometric analysis implies about that which is being assessed. More sophisticated psychometricians do stipulate that the latent traits they measure do not really 'exist in any physical or physiological sense'[5] – that they're 'statistical constructs'.[5] However, there is a longstanding implicit reification in the literature of the existence of these underlying internal traits that can be measured over time.[6,7] In either case, it is clear that psychometric tools were initially developed by cognitive psychologists to be valid and useable for phenomena that could at least be conceptualised as stable traits within a single individual. They were designed for the assessment of personality traits such as intelligence, honesty and diligence. Despite issues of test–retest reliability and other methodological hurdles related to positive and negative changes in knowledge over time, they have since been extended for use in the assessment of knowledge and performance.

There is growing understanding that some aspects of medical education are better thought of as social constructs: instead of being considered as expressions of a single individual's abilities, they are conceived of as the products of interactions between two or more individuals or groups

With this caveat, psychometric tools have proven themselves to be very useful for assessing the aspects of medical training, such as content knowledge, that are more easily conceptualised as psychological constructs, as existing individually within each trainee. There is, however, a growing understanding that some aspects of medical education are better thought of as social constructs. That is, instead of being considered as expressions of a single individual's abilities, they are conceived of as the products of interactions between 2 or more individuals or groups.

The Wilson Centre for Research in Education, University of Toronto, Toronto, Canada

Correspondence: Ayelet Kuper, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Room HG 08, Toronto, Ontario M4N 3M5, Canada. Tel: 00 1 416 480 4994; Fax: 00 1 416 480 4563; E-mail: ayelet94@post.harvard.edu

doi: 10.1111/j.1365-2923.2007.02945.x

From a social science (e.g. sociology, anthropology, social psychology) perspective, our ideas about how people should act in different situations are context-specific and culture-bound. They are constructed by the cultures and structures of the societies in which we live, as well as by the social groups to which we belong in those societies and by our ongoing interactions with others. As each person will have had a different set of experiences and positions within one or more societies, he or she will have a slightly or radically different set of perceptions and interpretations of appropriate behaviour in various contexts. Translated into medical settings, this means that our descriptions of competence in certain areas are socially constructed and may differ from those of our fellow doctors, our non-doctor colleagues, and our patients.

Our descriptions of competence in certain areas are socially constructed and may differ from those of our fellow doctors, our non-doctor colleagues, and our patients

This point of view changes the definitions of many abilities expected of trainees from being stable and internal to being socially constructed and historically transient and, as such, situational and interpersonal. Empathy, for example, can be described as the behaviour that causes another person, such as a patient, to perceive someone, such as their trainee doctor, as being empathic; that trainee doctor's empathy (or lack of) is constructed in the encounter rather than being an innate quality of that doctor as an individual. This construction comes from each person's perception of the other and of the situation, a perception that, for its part, is grounded in each of their culture(s) and personal histories. These perceptions may differ radically not only between the trainee and the patient, but between either or both of them and an examiner who is observing the encounter.

This leads to an intriguing problem. In a domain where the touchstones have been inter-rater reliability (and a numerical understanding of validity that depends on such reliability), how do we account for the shifting, context-laden, socially constructed nature of trainee competencies, such as empathy and professionalism? Rather than trying to pin down the definitions of these abilities to a single artificial norm, how can we begin to capture their inherent variability and analyse it in a systematic, meaningful way?

There is tantalising scope for whole programmes of research in this area

There is a large body of literature in the social sciences devoted to the analysis of social behaviour and social interaction (such as the work of Pierre Bourdieu, Irving Goffman, Anselm Strauss, Howard Becker and Kurt Lewin, among others). Ethnography provides such a methodological approach.[8] Ethnography has as its central focus the understanding of social processes, behaviours and perceptions that exist within groups and organisations. Participant observations and key informant interviews are commonly employed to explore and illuminate the social actions and interactions that occur within a specific context. Discourse analysis provides another possible methodological route. By focusing in depth on the verbal interactions that occur between individuals, using techniques such as interviews, explanations of how individuals construct versions of the social world can be generated. These methodologies, among others, allow the generation of rich qualitative datasets, which can be used to create qualitative assessments.

For example, an analysis of a combination of semi-structured key informant interviews and focus groups, based in the ethnographic tradition, conducted on a medical ward with patients, nurses, students and other trainees, as well as with attending doctors, could generate rich, meaningful trainee assessments for certain aspects of clinical rotations. Such assessments would be particularly useful for non-medical expert doctor competencies, such as those of collaborator, communicator and professional.[9] Unlike standard 360-degree assessments, the emphasis would be on capturing the range and perceptions of interpersonal behaviour taken in context. Although the results of such an assessment would not be psychometrically reproducible, the robustness of such a process could evaluated in a rigorous way. Driessen *et al.* have previously described an evaluation method using strategies derived from qualitative research to show the credibility and dependability of a portfolio assessment process.[10] Although their work was not grounded in a particular methodology, it should be possible to adapt the different criteria for judging the rigour of various methodological approaches in order to evaluate assessments carried out in those traditions. In this case, for example, one could look to the criteria used for evaluating ethnographic research in the social sciences. There is, indeed, tantalising scope for whole programmes of research in this area.

> *We need to think about the nature of the constructs and competencies we are trying to assess and choose our evaluation tools accordingly*

Although we are not advocating an end to the use of psychometrics in medical education, we would like to propose that we go back to basics. We need to think about the nature of the constructs and competencies we are trying to assess and choose our evaluation tools accordingly. Rather than being tied to any single methodology, we should continue to focus our concerns on our ability, as a medical community, to know what it is that our trainees know and do, and to assess and evaluate it appropriately. That's all that matters – the rest is commentary.

REFERENCES

1 Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med* 2004;**79** (10):955–60.

2 Swanson DB, Case SM. Assessment in basic science instruction: directions for practice and research. *Adv Health Sci Educ Theory Pract* 1997;**2**:71–84.

3 Wass V, van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;**357** (9260):945–9.

4 Martin JA, Regehr G, Reznick R, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;**84** (2):273–8.

5 Anastasi A, Urbina S. *Psychological Testing*, 7th edn. Upper Saddle River, NJ: Prentice-Hall 1997;73.

6 Brown FG. *Principles of Educational and Psychological Testing*, 3rd edn. New York: Holt, Rinehart & Winston 1983;3–16.

7 Murphy KR, Davidshofer CO. *Psychological Testing: Principles and Applications*, 5th edn. Upper Saddle River, NJ: Prentice-Hall 2001;39–46.

8 Hammersley M, Atkinson P. *Ethnography: Principles and Practice*, 2nd edn. London: Routledge 1995.

9 Frank JR, ed. *The CanMEDS 2005 Physician Competency Framework. Better Standards. Better Physicians. Better Care.* Ottawa: Royal College of Physicians and Surgeons of Canada 2005;15–28.

10 Driessen E, van der Vleuten C, Schuwirth L, van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ* 2005;**39** (2):214–20.

An expert-performance perspective of research on medical expertise: the study of clinical performance

K Anders Ericsson

CONTEXT Three decades ago Elstein *et al.* published their classic book on medical expertise, in which they described their failure to identify superior performance by peer-nominated diagnosticians using high- and low-fidelity simulations of the everyday practice of doctors.

OBJECTIVE This paper reviews the results of subsequent research, with a particular emphasis on the progress toward Elstein *et al.*'s goal of capturing the essence of superior clinical performance in standardised settings in order to improve clinical practice.

RESULTS Research following publication of Elstein *et al.*'s book was influenced by laboratory research in cognitive psychology, which resulted in a redirection of its original focus on capturing clinical performance in practice to studies of changes in cognitive processes as functions of extended clinical experience. There is currently renewed interest in linking laboratory research with studies of the acquisition of superior (expert) performance in the clinic.

CONCLUSIONS Research on medical expertise and simulation training in technical procedures and diagnosis provide exciting opportunities for establishing translational research on the acquisition of superior (expert) performance in the clinic by capturing it with representative tasks in the laboratory, reproducing it for experimental analysis, and developing training activities, such as deliberate practice, that can induce measurable improvements in performance in the clinic.

Department of Psychology, Florida State University, Tallahassee, Florida, USA

Correspondence: K Anders Ericsson, Department of Psychology, Florida State University, 1107 West Call Street, PO Box 3064301, Tallahassee, Florida 32306-4301, USA. Tel: 00 1 850 644 9860; Fax: 00 1 850 644 7739; E-mail: ericsson@psy.fsu.edu

KEYWORDS clinical competence/*standards; physicians/*standards; decision making; clinical medicine/*standards.

*Medical Education 2007: **41**: 1124–1130*
doi:10.1111/j.1365-2923.2007.02946.x

INTRODUCTION

It is now over 30 years ago since Simon and Chase[1] published their first proposal for a general theory of expertise and almost 3 decades since the classic book by Elstein *et al.*, *Medical Problem Solving: an Analysis of Clinical Reasoning*,[2] appeared. It is timely to review the progress. Last year, the first *Cambridge Handbook of Expertise and Expert Performance*[3] was published, demonstrating that the field of research on expertise has matured into an integrated discipline. Within the last 5 years over 15 books on expertise have been published.[4] Today, research on expertise is taking on particular relevance as federal funding agencies are promoting translational research,[4,5] where scientists have to establish a direct connection between important phenomena in the clinic, such as outcomes of treatments in different hospitals,[6] and laboratory research on the causes of the same phenomena, such as individual differences in the skills of surgeons. Furthermore, medicine is studying methods to measure and pay for performance.[7]

This paper will begin by reviewing Elstein *et al.*'s[2] research and how some of their failures to identify superior clinical performance influenced subsequent investigators and their research agendas. The expert-performance approach will be briefly described and its framework used to put research on medical expertise into perspective.

Overview

What is already known on this subject

Research on medical expertise has traditionally identified experienced clinicians as experts and then compared their diagnostic processes with those of novices, such as medical students.

What this study adds

This article sketches an alternative approach to medical expertise based on the study and analysis of superior (expert) performance in the clinic and outlines how laboratory research on medical expertise can remain closely connected to superior clinical performance.

Suggestions for further research

Research on medical expertise might establish direct links to robust performance phenomena in the clinic, enabling empirical examination of theoretical assumptions guiding current research. This should also enable measurement of longitudinal development of performance of expert clinicians and describe their associated deliberate practice activities.

THE ORIGINS OF RESEARCH INTO MEDICAL EXPERTISE

Modern research into expertise can be traced back to de Groot's[8,9] classic research on expertise in chess, where he demonstrated the possibility of studying the thought processes of world-class chess players by giving them representative tasks that captured the challenges they would regularly encounter during matches in chess tournaments.

Elstein *et al.*[2] developed a similar research programme to capture superior (expert) performance in diagnosis and treatment. First, Elstein *et al.*[2] carefully designed representative tasks with situations that 'a general internist practicing in a community hospital of moderate size could be reasonably expected to see' with a room 'designed to resemble a physician's office' and actors who 'were carefully trained to simulate the patients'. Second, they searched extensively for expert diagnosticians with superior skills and contacted practising doctors in Michigan for nominations of the 'best diagnosticians known to the respondent', and to identify doctors with the most nominations along with a group of typical doctors. Finally, Elstein *et al.*[2] collected data on doctors' thought processes by videotaping doctor–patient interactions, especially comments made by doctors as if they were 'reviewing cases with students'. In addition, the doctors also viewed their videotapes and reported their memories of thoughts they had had during the videotaped interactions.[2] (pp 47–8)

Based on their analysis of their extensive data, Elstein *et al.*[2] concluded that the objective performance of 'the best diagnosticians' did not differ significantly from that of typical doctors, and nor did cognitive processes differ systematically between the 2 groups. Further, their original conclusions about the general structure of the diagnostic process were later shown to lack rigorous statistical support.[10,11]

Some reactions to Elstein *et al.*'s failure

The failure of Elstein *et al.*'s studies[2] to uncover significant individual differences in diagnostic performance had several consequences. In their book, Elstein *et al.*[2] reviewed possible reasons for the failure, such as problems with recruitment, problems with the tasks and lack of statistical power. In a retrospective assessment 10 years later, Elstein *et al.*[10] suggested that medical doctors would not know enough about their colleagues' diagnostic performance to be able to nominate superior diagnosticians accurately. They concluded that 'years of experience, specialty board certification and (occasionally) academic rank or responsibility have become the indicators of expertise'.[10] (p 22) If medical expertise can be assumed to develop gradually as a function of years of experience, then investigators would not necessarily have to study the highest level of expertise – the specialists – but could study the development toward expertise in more accessible populations, such as medical students and interns. Similarly, the failure to capture individual differences in expert diagnostic performance with either costly high- or low-fidelity simulations made the issue of capturing expertise mute and permitted researchers to explore phenomena related to the acquisition of diagnostic reasoning more freely. In fact, several pioneering researchers adapted experimental paradigms from cognitive psychology for studying prototypes of concepts[12] and exemplar-based categorisation.[13] Patel and Groen[14] adapted the paradigm for studying chess experts' memory for briefly presented chess positions[15] to the domain of medicine. They had medical students and specialists read a case description of a patient for a couple of minutes, then recall as much as they could in writing and subsequently write out a description of the underlying pathophysiology of the case before reporting their diagnosis. Their focus was thus more on the structure of the information about the patient held in the long-term memory *after* all the relevant information had been processed and the diagnosis generated.[16,17] More generally, these changes in paradigms reduced the connections to superior diagnostic performance in the clinic.

Subsequent research on medical expertise and clinical performance

Research by Bordage[12] comes closest to continued pursuit of Elstein *et al.*'s[2] research agenda, in part because Elstein[18] changed his research focus to statistical decision theory as a more promising area for improving the quality of clinical practice. Bordage[12] continued to use 'think aloud' protocols to study individual differences in diagnostic reasoning, with a new focus on discovering differences in the structure and organisation of knowledge. Bordage developed training interventions to induce the development of elaborated semantic representations of reported symptoms, but there was no improvement in diagnostic accuracy among medical students. He concludes his paper in this issue with a proposal for closer interaction between theoretical laboratory research and training in medical school.

The research agenda of Schmidt and Rikers[19] is directly linked to work by Patel and Groen,[14] who took issue with Elstein *et al.*'s[3] model for the generation and elimination of diagnostic hypotheses in favour of Simon and Chase's[1] theory of expertise, which favoured mechanisms with pattern recognition-based retrieval from memory. Their paper provides a very thoughtful review of the issues regarding forward and backward reasoning, the non-monotonic relationship between immediate memory and expertise, and the role of biomedical knowledge in supporting expert diagnostic reasoning. With respect to relevance to clinical practice, several of the studies conducted by their group have moved closer to analyses of concurrent thought processes on tasks representative of diagnosis in the clinic. Whereas Patel and Groen's paradigm[14] focused on the immediate recall of descriptions of patients and written posthoc explanations, studies carried out by their group had their participants think aloud during their reading of patient descriptions, with the main task of generating a diagnosis rather than immediate recall. A further step in this development is provided by Monajemi *et al.*[20], who studied diagnosis in the context of the goal of treating patients, thus reinstating the task primarily studied by Elstein *et al.*[2] Pertinent to contextual effects, Schmidt and Riker[19] discuss the possibility of differences in experts' cognitive processes during diagnosis in actual clinical practice and in experimental situations and propose further research on differences between the 2 contexts.

The research group at McMaster University, led by Norman, is advocating a research programme that is most clearly opposed to Elstein *et al.*'s[2] central methodological assumptions. Norman *et al.*[13] propose that 'expert and novice problem solving is based on, to some degree, similarity to a prior specific exemplar in memory'[13] as a result of non-analytic reasoning (NAR). Given that NAR 'is presumed to be inaccessible to introspection',[13] this group does not collect verbal reports of thinking and relies primarily on laboratory experiments with experimental paradigms from cognitive psychology. The core paradigm demonstrates experimentally that characteristics of exemplars of a diagnosis (pictures of skin conditions, electrocardiograms [ECGs]) presented during an initial study and practice phase (with a typical duration of around an hour) can bias subsequent diagnoses of other cases (presented with similar stimuli). For example, Young *et al.*[21] taught undergraduate psychology students 4 different pseudo-psychiatric diagnoses and found a bias induced by the studied case examples when subjects were tested immediately after the study and practice phase. Drawing on results with this general paradigm, Norman *et al.*[13] report significant biases for psychology students, medical students and professionals below the level of specialists. Norman *et al.*[13] acknowledge that specialists may be different and may rely on sequential reasoning to diagnose challenging cases. In an interesting study Eva *et al.*[20] demonstrate that it may be possible to reduce these biases in psychology students by inducing both analytic and non-analytic reasoning. However, Eva *et al.*[22] caution against generalisation of their results, and raise some very relevant issues about the generalisability of results from short-term training studies with novices to the acquisition of diagnostic performance that normally takes years of study and internship to acquire in clinical settings.

In their review of the current state of research on medical expertise, Mylopolous and Regehr[23] argue that researchers have focused too much on the effects of extensive experience in diagnostic performance.

According to these authors, contemporary researchers assume that 'most novices eventually become experts. It is presumed that with more experience automatically comes the accrual of a greater (or better) resource base on which to rely, suggesting that expertise is an automatic and inevitable consequence of experience'.[23] By contrast, Mylopolous and Regehr[23] propose that this assumption is only true for some experts, who are skilled in executing routine procedures (routine experts), but that there are other experts (adaptive experts), who 'continue to grow only because of their intentional engagement in ''progressive problem solving''; the continual re-investment of cognitive resources into creating not merely better performance, but in fact better understanding [of] the problem of their domain'.[23] They conclude that current medical education trains routine experts and that new directions are necessary to train adaptive experts.

All the reviews in this special issue[12,13,19,23] acknowledge directly or indirectly that research into medical expertise during the past 3 decades has not identified medical professionals with reproducibly superior performance in their daily clinical practice and thus has not captured the associated phenomena in laboratories, as Elstein *et al.*[2] set out to do. Most researchers of medical expertise seem to have been influenced by early general theories of expertise,[1] with their focus on the differences between the cognitive processes of medical students (novices) and those of experienced clinicians (experts).[23]

Toward the study of superior (expert) performance in clinical practice

The distinction between experienced clinicians and clinicians with superior performance has received hardly any attention among researchers of medical expertise, except for Mylopolous and Regehr's[23] review. In the broader field of expertise research this important distinction has a long history. In the early 1990s my collaborators and I started to criticise a definition of expertise based on years of experience and discussed evidence that the amount of domain-related experience is only weakly related to objective performance of representative tasks for many types of expert, such as chess players, amateur athletes, auditors, stockbrokers and other types of professionals. More recent reviews and meta-analyses of thousands of experienced health professionals show weak or non-existent correlations between performance on representative tasks and years of professional experience after the completion of education.[24–26] In fact, for many types of performance there is a negative correlation with years of experience, which suggests a decay in previously acquired skills.[24–26]

The expert-performance approach

Ericsson and Smith[9] proposed an alternative approach to the study of expertise (now referred to as the expert-performance approach). This approach is based on identifying individuals (expert performers) who exhibit superior performance on tasks that capture the essence of expertise in the critical domain. In the case of medicine, this would correspond to surgeons and clinicians who produce better treatment outcomes, and more accurate diagnoses for adverse conditions that can be treated effectively with established therapies. Recent reviews[24,27] show that studies of expert performers with objectively superior performance in clinical medical settings exist, but are rare.

Reproducing the superior (expert) performance in the laboratory

The expert performance approach starts with an analysis of naturally occurring behaviour,[8,9] such as matches between chess masters and rapid diagnoses in the emergency room. It then identifies key situations, where a critical immediate action, such as a chess move or a corrective action during a surgical problem, is demanded, and the appropriateness of the performed actions can be determined after the fact. These situations are then presented as videos, pictures or simulated environments to other individuals who differ in skill levels and who are instructed to generate their best immediate action. By presenting a sequence of these representative situations it has been possible to capture objective performance in different domains, such as chess, music, individual sports and scrabble, which is closely related to performance in tournaments and competitions. It is possible to measure these abilities using selection of the best move. For example, with 30 different selection tasks (10–20 minutes of testing), it is possible to attain a validity that approaches measures based on the outcomes of matches that correspond to 50–200 hours of tournament play.[4,28]

Since Elstein *et al.* performed their studies,[2] tasks in medical diagnosis have, nearly always, involved typical and routine cases. By contrast, demonstrations of superior diagnostic performance are nearly always associated with challenging cases,[13,25,26,29] which reflect rare, if ever, directly experienced conditions, such as infrequent diseases or unusual combinations of disease states. These are the types of cases that require 'adaptive expertise'[23] and are associated with rich protocols of pathophysiological reasoning.[13]

Analysis of the mechanisms that mediate superior performance

The establishment of reproducible superiority for representative tasks that capture performance in the clinic permits researchers to analyse the mechanisms responsible for this performance by tracing performance with latencies, eye movements and concurrent or retrospective reports.[30] Research on expertise has illustrated how 'think aloud' protocols during superior representative performance can identify complex skilled mechanisms that can later be validated by specially designed experimental manipulations.[30]

Once the mechanisms mediating the superior performance have been identified, the third step in the expert performance approach involves studying how these mechanisms developed or were acquired. Superior performance on the most challenging tasks is linked to planning, complex reasoning, self-monitoring and evaluation, which require representations and acquired skills to support increased demands for working memory in the form of long-term working memory.[16,17,31] The challenge is to explain how these complex mechanisms were acquired by appropriate practice activities.

The acquisition of superior (expert) performance with deliberate practice

A fundamental prerequisite for improvement of performance accuracy through practice is the availability of valid immediate feedback. In most medical activities it often takes weeks or months until a patient's development of his or her medical condition confirms or rejects the validity of diagnoses and associated treatments. Many interns and residents have to learn to diagnose X-rays, such as mammograms, without receiving accurate feedback and must rely on feedback in the form of their teachers' diagnoses of the same images – which is estimated to be around 70% accurate.[25,32] In some activities, such as surgery and emergency medicine, there is often immediate feedback about mistakes and failures, but these types of everyday activities are less appropriate for training and practice.

In numerous other domains it has been possible to identify special practice activities (deliberate practice) that performers' teachers or the performers themselves design to provide opportunities to improve particular aspects of their performance in an environment that allows gradual refinement after problem solving and repeated variations with immediate feedback.[33] In medicine some of the most exciting research has concerned high-fidelity training with simulators, in particular that involving laparoscopic techniques. A recent meta-analysis[34] revealed that enhanced performance resulting from simulator training depended directly on whether the training procedure incorporated the characteristics of deliberate practice,[25,34] such as goal-directed training, with opportunities for repetition with feedback. A subsequent review was able to demonstrate that increases in the amount of practice (meeting the criteria for deliberate practice) were associated with increases in performance.[35] Furthermore, skilled surgeons with excellent performance records have been tested using simulators and their levels of simulator performance were reliably superior to those of untrained interns. The expert performers' achievement has been used to benchmark the speed and control[36] that surgeons in training need to attain before completing their simulator training. This research on simulator training has even demonstrated significant transfer of skills learned in training to actual surgical and clinical practice.[37]

Conclusions

Emerging technological innovations in imaging, biochemical analysis and advanced treatment modalities are generating some very exciting opportunities for research on medical expertise. The development of simulation centres dedicated to training will provide further opportunities for research on training interventions that enhance performance in the clinic, as well as on the measurement and analysis of superior performance on high-fidelity tasks. It should be possible to return to the original vision of Elstein *et al.*,[2] where studies of superior clinical performance lead to improvement of the quality of clinical practice, but one would have to use a different methodological approach, namely, the expert-performance approach. If current research on medical expertise, discussed earlier in this article, were to establish direct links to robust performance phenomena in the clinic, it would be possible to examine empirically many of the theoretical assumptions that guide current research. For example, are the characteristics of the diagnostic learning of psychology and medical students closely related to those of select interns, residents and specialists, who gradually improve their diagnostic accuracy throughout their careers? It should be possible to measure the longitudinal development of the representative performance of future superior clinical performers and to describe their associated deliberate practice activities. The future of research on medical expertise looks very bright; I expect that this is where the most important discoveries about the optimal development of professional expertise will be made.

Acknowledgements: the author would like to thank Len Hill for comments on an earlier version of this review.
Funding: this study was supported by Florida State College for Women and Edward Conradi Eminent Scholar Endowment from the Florida State Foundation.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Simon HA, Chase WG. Skill in chess. *Am Sci* 1973;**61**:394–403.
2 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978.
3 Ericsson KA, Charness N, Feltovich P, Hoffman RR. *Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press 2006.
4 Ericsson KA, Williams AM. Capturing naturally occurring superior performance in the laboratory: translational research on expert performance. *J Exp Psychol Appl* 2007;**13**:115–123.
5 Zerhouni EA. Translational and clinical science – time for a new vision. *N Engl J Med* 2005;**353** (15):1621–23.
6 Hollenbeck BK, Dunn RL, Miller DC, Daignault S, Taub DA, Wei JT. Volume-based referral for cancer surgery: informing the debate. *J Clin Oncol* 2007;**25** (1):91–6.
7 Rosenthal MB, Dudley DA. Pay-for-performance: will the latest payment trend improve care? *JAMA* 2007;**297** (7):740–4.
8 de Groot A. *Thought and Choice in Chess* [Original work published 1946.] The Hague: Mouton 1978.
9 Ericsson KA, Smith J. Prospects and limits in the empirical study of expertise: an introduction. In: Ericsson KA, Smith J, eds. *Toward a General Theory of Expertise: Prospects and Limits*. Cambridge: Cambridge University Press 1991;1–38.
10 Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: a 10-year retrospective. *Eval Health Prof* 1990;**13**:5–36.
11 McGaghie WC. Medical problem-solving: a re-analysis. *J Med Educ* 1980;**55**:912–21.
12 Bordage G. Prototypes and semantic qualifiers: from past to present. *Med Educ* 2007;**41**:1193–1201.
13 Norman G, Young M, Brooks L. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;**41**:1140–45.
14 Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cognit Sci* 1989;**10**:91–116.
15 Chase WG, Simon HA. Perception in chess. *Cognit Psychol* 1973;**4**:55–81.
16 Ericsson KA, Patel V, Kintsch W. How experts' adaptations to representative task demands account for the expertise effect in memory recall: comment on Vicente and Wang (1998). *Psychol Rev* 2000;**107** (3):578–92.
17 Patel VL, Arocha JF, Kaufmann DR. Diagnostic reasoning and medical expertise. *Psychol Learn Motiv* 1994;**30**:187–251.
18 Elstein AS. Clinical problem solving and decision psychology: comment on 'the epistemology of clinical reasoning'. *Acad Med* 2000;**10** (Suppl):134–6.
19 Schmidt HG, Rikers RMJP. How expertise develops in medicine: knowledge encapsulation and illness script formation. *Med Educ* 2007;**41**:1133–39.
20 Monajemi A Rikers RMJP Schmidt HG Clinical case processing: a diagnostic versus a management focus. *Med Educ* 2007;**41**:1166–72.
21 Young M, Brooks L, Norman G. Found in translation: the impact of familial symptom descriptions on diagnosis in novices. *Med Educ.*
22 Eva KW, Hatala R, LeBlanc VR, Brooks LR. Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information. *Med Educ* 2007;**41**: 1152–58.
23 Mylopolous R, Regehr G. Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Med Educ* 2007;**41**:1159–65.
24 Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med* 2005;**142** (4): 260–73.
25 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**10** (Suppl):70–81.
26 Ericsson KA, Whyte J, Ward P. Expert performance in nursing: reviewing research on expertise in nursing within the framework of the expert-performance approach. *ANS Adv Nurs Sci* 2007;**30** (1):58–71.
27 Ericsson KA, Lehmann AC. Expert and exceptional performance: evidence on maximal adaptations on task constraints. *Annu Rev Psychol* 1996;**47**:273–305.
28 van der Maas HLJ, Wagenmakers EJ. A psychometric analysis of chess expertise. *Am J Psychol* 2005;**118**:29–60.
29 Norman GR, Trott AD, Brooks LR, Smith EKM. Cognitive differences in clinical reasoning related to postgraduate training. *Teach Learn Med* 1994;**6**:114–20.
30 Ericsson KA. Protocol analysis and expert thought: concurrent verbalisations of thinking during experts' performance on representative task. In: Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press 2006;223–42.
31 Ericsson KA, Kintsch W. Longterm working memory. *Psychol Rev* 1995;**102** (2):211–45.

32 Nodine CF, Kundel HL, Mello-Thoms C, Weinstein SP, Orel SG, Sullivan DC, Conant EF. How experience and training influence mammography expertise. *Acad Radiol* 1999;**6**:575–85.
33 Ericsson KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson KA, Charness N, Feltovich P, Hoffman RR, eds. *Cambridge Handbook of Expertise and Expert Performance*. Cambridge: Cambridge University Press 2006;685–706.
34 Issenberg SB, McGaghie WC, Petrussa ER, Lee Gordon D, Scalese R. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;**27**:10–28.
35 McGaghie WC, Issenberg SB, Petrussa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ* 2006;**40**:792–97.
36 van Sickle KR, Ritter EM, McClusky DA III, Lederman A, Baghai M, Gallagher AG, Smith CD. Attempted establishment of proficiency levels for laparoscopic performance on a national scale using simulation: the results from the 2004 SAGES Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) learning centre study. *Surg Endosc* 2007;**21** (1):5–10.
37 Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolam S, Reznick RK. Randomised controlled trial of virtual reality simulator training: transfer to live patients. *Am J Surg* 2007;**194** (2):205–11.

Received 24 September 2007; accepted for publication 10 October 2007

A modest proposal: impressions of a Year one medical student

Brandon R Allen & Jose E Rodriguez

In the US health care system, uninsured people face multiple barriers to adequate care. This personal essay reflects on the effect that caring for these patients can have on a Year 1 medical student. The student reports his observations of the disparities that exist between private care and the care provided for uninsured patients. This comparison leads to suggestions for an inexpensive solution that could be provided by the medical community. The experience of 1 student with his first patient spurs a call to action to the medical community as it strives to lead the debate and shape the solution to providing care to the uninsured. As the student ponders the problem, his observations may uncover a universal challenge that all providers must resolve for themselves.

My first patient at the community health clinic was a middle-aged man with concern written all over his face. I had just completed Year 1 of medical school and had armed myself with a tool-belt of clinical pearls and examination manoeuvres. I gathered a complete history with confidence. As I reported the findings to the attending doctor, we came to the conclusion that this man needed a colonoscopy. His family history of colon and oesophageal cancer made this an obvious step. *Not so fast.* I found that a set number of gastroenterology referrals are allotted per year (around 30) and my patient was not eligible because he did not have a positive faecal occult blood screening. Little did I know that the US Preventive Services Task Force (USPSTF) Grade A recommendation does not apply to the uninsured.

We work in a special community clinic. Our clinic was set up with a mission to serve the uninsured and the uninsurable. When any of our patients get Medicaid or Medicare, they are no longer eligible for our services. In fact, if any of our patients obtain any type of health care coverage, they are referred to the private providers in this community. Our patients by definition cannot afford their medications, so most medications are purchased at a discount pharmacy that is part of our clinic. Our patients can barely afford the $5 they are asked to contribute.

Many individuals donate their time and expertise to our under-served – the uninsured. A cardiologist, gynaecologist, dermatologist and 2 orthopaedic specialists give their services at the clinic once a month. Many more specialists have agreed to see a certain number of our uninsured patients free of charge though a volunteer referral network. There are also a few full-time providers, as well as an army of volunteers from my local medical school. Our medical community would not deny anyone life-saving treatment, regardless of his or her ability to pay. We are grateful for the time that these doctors donate. This is a stride in the right direction, but unfortunately the need is much greater than that which they – on their own –can meet.

I have been to many private offices in this community and others, where the sample closets are filled to capacity with new medications for virtually every condition encountered in primary care. When I look into the sample closet at our community health clinic, I see many bare shelves. The 'drug lunches' that are commonplace at most providers' offices are not available to our employees and volunteers. The few drug representatives I have seen come by, drop off a box of samples, flash a smile and speed off. I imagine they are travelling to a private office where providers can write prescriptions for paying customers. Although we are grateful for their help, it is simply not enough – it's like putting a band-aid on a gaping wound.

These brief observations represent a microcosm of the obstacles an uninsured patient must face. Our

Florida State University College of Medicine, Tallahassee, Florida, USA

Correspondence: Brandon R Allen BS, MS1, Florida State University College of Medicine, Department of Family Medicine and Rural Health, 1115 West Call Street #3210M, Tallahassee, Florida 32306, USA. Tel: 00 1 850 645 6850; Fax: 00 1 850 645 2859; E-mail: bra06@med.fsu.edu

doi: 10.1111/j.1365-2923.2007.02918.x

medical community's concern for this patient population appears to be in short supply, but I suspect this is a universal problem. In our county there are no referrals available for neurology, rheumatology, pain management or psychiatry. When our patients need specialty services that are not available through the volunteer programme, the same patients who cannot pay $5 for medicines are asked to accept a payment plan before any specialist will see them. If a patient's chief complaint cannot be handled by the clinic provider, the patient will not see a specialist for months, if ever. There is something wrong with this scenario.

What can we, as providers, do? I may have a plausible solution. Perhaps all non-surgical specialty providers could see 2 uninsured patients per day. If each provider works 180–200 days per year, this would generate an average of 360–400 available slots per year per specialist. I suspect there are at least 2 providers in every specialty within our region so this should give ample opportunity for an uninsured patient to be re-evaluated. Even if we cut that number in half, to only 1 patient per day, this would mean almost 200 available referrals per year, per provider. Perhaps, instead of a sales or income tax to help these patients, a 'services tax' could become a part of the practice of medicine. I truly believe that most doctors want to help underserved patient populations, but they may not know how.

Patients who require surgery or invasive diagnostic procedures present another problem. If a surgeon or diagnostician saw 2 uninsured patients per week for 40 weeks per year, 80 procedures per specialist would be available. This is considerably more than the number available now – 30 colonoscopies per year – none of which are screening procedures. I realise that hospitals would have to be involved for a plan like this to succeed. Although this suggestion involves some cost and the use of resources, this is a small price to pay for preventing death and disability in these uninsured patients.

Unfortunately, adenocarcinomas or myocardial infarctions don't wait to check who your insurance provider is before they rear their ugly heads. The burden lies on us, as present and future providers of medical care, to follow through where opportunities for the uninsured are limited. We must become the champions of these patients if we want to improve their health outcomes. If not, we are at risk of doing harm by default through neglect.

I have observed that the system in place for these individuals is fragmented at best. In my small sampling of patient care at our community health clinic, I see the 'American dream' (the widespread aspiration of Americans to live a better quality of life than their parents) working in reverse. For too many of these patients, compliance with strict treatment regimens for their chronic pathologies may take a backseat to finding their next meal or fulfilling other survival needs. How can a doctor reconcile his or her work with that line of thinking? Is that proven combination therapy really going to work if our patients can't obtain it?

This experience has made me realise that the care available to uninsured people in our health care system is poor. I came to medical school naïve to the plight of the uninsured. Having no experience of being uninsured myself, I had no clue about the types of barriers that might confront someone in need of care.

As I progress in my medical education, I will primarily be instructed in the offices of private doctors, and I am certain that vast sample closets and drug lunches will continue in these locations. I chose my medical college, in part, because of its stated mission to serve under-served populations within my state. As I move forward in my career I have cautious optimism towards the future of the uninsured. I intend to seek residency and employment at a public or community hospital. I want to work in clinics located in underserved communities and contribute wherever the mission takes me. I also see the need to recruit others to my vision. If we all contribute, - uninsured patients, - medical providers and students will all be the better for it. And to think, it all started with 1 patient who just needed a screening colonoscopy…

Contributors: BRA originated this idea and initiated the project. He wrote the initial draft, to which JER contributed.
Acknowledgements: we would like to thank the providers and staff of the We Care Network and the Lincoln Neighborhood Service Center for their continuing support of uninsured patients and for inspiring us to give a little more.
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

Received 25 May 2007; accepted for publication 15 August 2007

How expertise develops in medicine: knowledge encapsulation and illness script formation

Henk G Schmidt & Remy M J P Rikers

CONTEXT For over 30 years, research has focused on the question of how knowledge is organised in the doctor's mind. The development of encapsulated knowledge, followed by the formation of illness scripts, may both be considered as important stages in the development of medical expertise.

METHODS This paper reviews research on the knowledge encapsulation and illness script hypotheses since their initial formulation. Findings in support of these views of expertise development are reported and conflicting data are discussed.

RESULTS A great deal of empirical data have been collected over the years to investigate the view that, through clinical experiences, biomedical knowledge becomes encapsulated and eventually integrated into illness scripts. The findings of most studies, which have used various techniques to probe the ways by which students and doctors mentally represent clinical cases, are in line with this view of expertise development. However, there is still debate concerning the role of biomedical knowledge in clinical case processing.

CONCLUSIONS To facilitate the development of expertise in medical school, it is important to teach the basic sciences in a clinical context, and to introduce patient problems early in the curriculum in order to support the processes of encapsulation and illness script formation. In addition, during clerkships ample time should be devoted to enabling reflection on patient problems with peers and expert doctors.

Department of Psychology, Erasmus University, Rotterdam, The Netherlands

Correspondence: Henk G Schmidt, Building M, Room M7-48, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. E-mail: schmidt@fsw.eur.nl

KEYWORDS *clinical competence; *problem solving; clinical clerkship/*methods; curriculum; group processes; peer group; teaching/*methods.

*Medical Education 2007: **41**: 1133–1139*
doi:10.1111/j.1365-2923.2007.02915.x

INTRODUCTION

Since the early 1970s, several attempts have been made to formulate a theory of medical expertise that would provide a parsimonious account for the available data on how expert doctors deal with medical problems. These attempts can be characterised as either 'processing' or 'structure' theories. Processing theories seek expertise in the particular ways in which doctors *process* patient-relevant information. The best-known theory in this category is the approach of Elstein *et al.*[1] They postulated that medical experts deal with patient information through a cognitive process called hypothetico-deduction, where an expert would generate hypotheses early in the encounter with a patient and would test these hypotheses against data gathered in context. (For a critique of this theory, see Patel and Groen;[2] for a rebuttal, see Norman *et al.*[3]) Structure theories, by contrast, focus on the *underlying knowledge structures* that produce diagnostic hypotheses. Lesgold *et al.*,[4] for instance, suggested that diagnostic accuracy is mainly determined by the extent to which the expert possesses rich causal biomedical knowledge structures in memory.

Our own attempts to contribute to the field within the structure paradigm concentrated on how expertise in medicine matures: how do medical students actually develop medical expertise in the course of medical education? To that end we proposed a theory that considers the development of expertise as progressing through a number of transitory stages,

Overview

What is already known on this subject

The development of medical expertise is a process in which different stages can be identified. The emergence of encapsulated knowledge and rich illness scripts can be considered as defining characteristics of expertise development.

What this study adds

This study provides an integrative and critical overview of more than 20 years of research on knowledge encapsulation and illness scripts. Implications for teaching are discussed.

Suggestions for further research

Recent non-experimental studies suggest that biomedical knowledge plays a more prominent role in professional practice than had been assumed earlier. There is a need for experimental studies that further investigate its role.

each characterised by knowledge structures underlying diagnostic performance that are qualitatively different from those of other stages.[5–7] The purpose of the present paper is to summarise the research conducted within this framework since its initial formulation. In view of the available space, we will concentrate on research that specifically focused on our theory. This is not to say that research by other investigators in the field did not contribute to our particular point of view. In particular, the contributions of Patel and Groen[2] and Norman *et al.*[8] have been inspirational. First, we will sketch our theory.

TRANSITORY STAGES IN EXPERTISE DEVELOPMENT

Most early cognitive theories of expertise development[9] consider it to be largely a process of the extension of causal knowledge about a domain. Through education and experience, students acquire more and more relevant concepts and develop richer and more meaningful relationships between them. In other words, expertise development is largely a matter of knowledge *expansion*. The theory outlined here deviates from these proposals in that it postulates that the development of expertise in medicine can only be properly understood by assuming certain kinds of knowledge shifts or knowledge *restructuring* in the course of growth towards expertise. In this section, therefore, the development of medical expertise involving this idea of distinct stages or 'phases' will be briefly outlined.

Our position as to how expertise in medicine develops can be summarised as follows: in the course of their early medical training, students rapidly develop mental structures that can be described as rich, elaborate *causal networks* that explain the causes and consequences of disease in terms of general underlying biological or pathophysiological processes. When confronted with a clinical case in this stage of development, students focus on isolated signs and symptoms and attempt to relate each of these to the pathophysiological concepts they have learned. This is an effortful and error-prone process. In addition, as they do not yet recognise patterns of symptoms that fit together, processing is detailed. This is why intermediate-level students remember more details of such cases than medical experts, provided they have enough time to do so; this results in an *intermediate* effect in clinical case recall.[10,11] (With hindsight, this seems pretty obvious. However, before the discovery of intermediate effects in case recall, there was general agreement among researchers – and empirical evidence from domains such as chess – that experts would recall more information than novices and intermediates because they have more appropriate knowledge available with which to interpret and, hence, retain the information.) In addition, when asked to think aloud while solving a clinical case, intermediate-level students tend to use detailed knowledge of the basic sciences (e.g. physiology, biochemistry, anatomy) in explaining for themselves the signs and symptoms of the patient. These references to the basic sciences are virtually absent in the think-aloud protocols of expert doctors.[12,13]

However, through extensive and repeated application of knowledge acquired and, particularly, through exposure to patient problems, a change in the knowledge structures of these students occurs. Their networks of detailed, causal, pathophysiological knowledge become *encapsulated* into diagnostic labels or high-level, simplified causal models that explain signs and symptoms. Knowledge encapsulation is a learning mechanism that can be defined as the subsuming or 'packaging' of lower-level, detailed concepts and their inter-relations, under a smaller number of higher-level concepts with the same explanatory power.[5] An example may clarify what is meant by encapsulation. Assume that a young man who is suspected to be an intravenous drug addict enters the emergency room. He complains of shaking chills and fever. The fever and chills are accompanied by sweating, and a feeling of prostration. He also complains of some shortness of breath when he tries to climb the 2 flights of stairs to his apartment. Physical examination reveals a toxic looking young man who is having a rigor. His temperature is 41°C. His pulse is 124/minute. His blood pressure is 110/40. Mucous membranes are clear. Examination of his limbs shows puncture wounds in his left antecubital fossa.[2] A Year 6 medical student, when asked to explain these symptoms, would say this:

> 'This man must have been using contaminated needles, which led to an infection with gram-negative bacteria. These bacteria in the bloodstream lead to the activation of antibodies, which explains the fever reaction: the high temperature, the shaking chills, the sweating, the feeling of prostration, and the shortness of breath. The bacteria also release endotoxins, leading to vasodilatation of the arteries. Vasodilatation in turn leads to the observed drop in blood pressure and possibly shock. Decreased resistance may be a reason why the immune response fails…'

An internist, by contrast, would respond this way:

> 'This drugs user has developed a *sepsis* as a result of using contaminated needles.'

The concept of sepsis is sufficient to explain all relevant signs and symptoms; it encapsulates, or stands for, the student's detailed pathophysiological explanation. This implies that in response to such a case the expert would no longer activate knowledge of the behaviour of the bacteria and the reaction of the body; activation of the single concept of sepsis is sufficient to fully explain the condition of the patient. It may be clear that having available a concept such as sepsis enables one to see patterns of symptoms as wholes, and that the availability of this concept considerably speeds up the processing of a case and contributes to accurate diagnosis. Experts have many encapsulating concepts available, describing syndromes (i.e. groups of symptoms that collectively indicate or characterise a disease) or simplified causal mechanisms. This knowledge is often called clinical knowledge (as opposed to biomedical knowledge) and experts tend to use this kind of knowledge preferentially. This is why the think-aloud protocols of experts, unlike the protocols of students, feature hardly any basic science concepts;[12,13] experts simply have more efficient instruments for understanding available.

As students begin to practise extensively with actual patients, a second shift occurs. Their encapsulated knowledge is reorganised into the type of narrative structures we referred to as illness scripts, following Feltovich and Barrows.[14] These illness scripts are cognitive entities containing relatively little knowledge about pathophysiological causes of symptoms and complaints (because of encapsulation), but a wealth of clinically relevant information about the *enabling conditions of disease*, as a product of growing experience with how disease manifests itself in daily life. Possessing knowledge about enabling conditions is supposed to characterise advanced levels of expertise because it enables the doctor to rule out whole categories of disease and to focus immediately on those that are most likely. For instance, if, in the middle of an influenza epidemic, a woman enters the consulting room complaining of fever-like symptoms, the doctor might think of influenza. However, if she professes that she has recently been in a malaria-infested region of the world, then alternative hypotheses suddenly become more relevant. Thus, 'having been to Africa recently' is knowledge that can add considerably to the accuracy of diagnoses and to the speed with which decisions can be made, despite the fact that the knowledge is enabling rather than causal. The acquisition of enabling-conditions knowledge is largely based on practical experience; education seems to play a lesser role.[15]

Illness scripts exist at various levels of generality, ranging from representations of disease categories to prototypes, to representations of individual patients seen before. Indeed, a salient feature of our theory is the assumption that doctors sometimes actually use memories of previous patients when diagnosing a new case.[16,17] Thus, while solving a problem, a doctor searches for an appropriate script and when he has selected 1 (or a few), he will tend to match its elements to the information provided by the patient. In the course of this script verification process, the script becomes *instantiated*. Instantiated scripts in turn, do not necessarily become decontextualised after use but remain available in memory as episodic traces of previously analysed patients and may be used in the diagnosis of similar problems in the future.

Table 1 contains a summary of our conjectures.

Table 1 Transitory stages in the development of medical expertise

1 Development of elaborate declarative networks explaining the causes and consequences of disease in terms of general underlying pathophysiological processes
2 'Encapsulation' of these declarative networks into a limited number of diagnostic labels, syndromes or high-level, simplified causal models, explaining signs and symptoms
3 Transition into 'illness scripts' through the acquisition of experience-based, contextual or 'enabling conditions' knowledge
4 Storage of interpreted instances of these scripts as exemplars of the particular illness

SUMMARY OF RESEARCH

The encapsulation hypothesis

Several of the predictions derived from the idea that biomedical knowledge becomes encapsulated into clinical concepts have been confirmed. Encapsulated knowledge is more readily assessed by doctors than biomedical knowledge.[18] Pathophysiological explanations by experts contain less biomedical and more encapsulating concepts than those by students of different levels.[19] The recall protocols of experts contain more encapsulations than those of subexperts (i.e. experts recalling cases outside their specific specialties).[20] Moreover, biomedical knowledge is only indirectly related to clinical competence.[21] Of particular interest is the work of Woods *et al.* on the role of causal knowledge in diagnosis.[22,23] They asked students either to learn a list of features associated with a number of diseases, or to learn causal, biomedical knowledge associated with these diseases. Although initially both groups performed similarly well on a diagnostic task, students from the causal condition did better after a delay of 1 week, suggesting that causal knowledge clarifies coherence among symptoms in a way that simple associative knowledge does not. More importantly, the authors demonstrated that students spontaneously develop encapsulations, as evidenced by better performance on a recognition test presenting *new* concepts encapsulating the causal mechanisms learned.[22,23] The availability of these concepts seemed to facilitate processing speed,[24] as predicted by the encapsulation hypothesis. Finally, intermediate effects were shown in the recall of electrocardiogram traces,[25] breast cancer cases,[26] psychodiagnostic classification,[27] and even non-medical texts.[28]

Initial critiques of the encapsulation hypothesis[29] suggested that the intermediate effect might be an artefact of the procedure used: doctors actually remember more from a case, but, for various reasons, do not demonstrate their superiority in free recall. Eva *et al.*,[30] however, showed that the effect also occurs in *recognition* of patient data: intermediates actually recognise more of the patient information in a case than experts. Another early appraisal suggested that the intermediate effect only occurs when the primary task is to diagnose the case. If experts are asked to remember the case rather than to diagnose, they do better.[8] The intermediate effect has turned out to be, however, quite robust against experimental manipulations. In a series of experiments Van de Wiel[31] demonstrated that neither experimental instructions (to diagnose a case, to recall a case), nor case characteristics (laboratory test data presented in summary form or in full) influenced the magnitude of the intermediate effect. De Bruin *et al.*[32] tricked half their experts into believing that national experts would evaluate the quality of their recall protocols. However, this manipulation did not influence the amount and nature of case recall.

Two studies have yielded expertise effects in case recall rather than intermediate effects. In 1 of them, Norman *et al.*[8] asked students and nephrologists to diagnose and recall cases solely consisting of laboratory findings. The nephrologists' recall was much better than the recall of the advanced students. A first attempt to explain this anomaly failed; Verkoeijen *et al.*[33] presented the same cases with and without a clinical context to internists and students. Both conditions produced an intermediate effect in recall. Closer scrutiny of the procedure followed by Norman *et al.* revealed that they had asked their participants while processing the case to comment on each laboratory result in the light of hypotheses entertained. Wimmers *et al.*[34] repeated this procedure, asking nephrologists to elaborate on the cases, and replicated Norman *et al.*'s findings. A control condition in which the nephrologists processed the cases without elaboration revealed an intermediate effect again. This study and Norman *et al.*'s earlier study show that only if you oblige experts to process a case elaborately and deliberately, focusing on detail, will they produce better recall than intermediates. This is, however, not their natural way of approaching the diagnostic task. Under normal conditions, processing is characterised by the use of encapsulating concepts that lead to remembering only the most critical signs and symptoms. A second study demonstrating expertise effects in recall remains unexplained.[35]

Van de Wiel *et al.*[36] tested the assumption that clinicians' biomedical knowledge does not disintegrate or decay over time. Our theory states that basic science knowledge must remain available for activation when a case turns out to be difficult. The authors asked doctors, clerks and advanced students to explain 20 clinical constructs in terms of underlying pathophysiology. The resulting protocols of the doctors displayed generally more elaborate, qualitatively better, and more fluent explanations than those of the clerks and students. Pathophysiological knowledge relating to the causes and consequences of disease does not decay with experience, but, rather, forms a coherent structure of knowledge that can be easily accessed when needed. This is in line with findings from a study by Patel *et al.*,[37] which showed that when experts were confronted with a difficult case, the number of biomedical concepts in their protocols increased. It is also in line with findings from research employing (lexical) decision tasks, which demonstrate that experts have biomedical knowledge that is easily accessible when needed.[38,39]

The illness script hypothesis

We assumed that advanced expertise would be characterised by the emergence of illness scripts, rich in terms of enabling conditions knowledge. Several studies have demonstrated that the number of enabling conditions associated with particular diseases increases with expertise.[40–42] For instance, when study subjects were asked to describe the clinical picture of a set of diseases, the number of enabling conditions mentioned increased as the level of respondent expertise rose. In addition, a positive relation was found between the number of actual patients seen with a particular disease and the number of enabling conditions mentioned.[40] Another study demonstrated that expert performance was sensitive to typicality of both enabling conditions and consequences, whereas advanced students' performance was sensitive only to typicality of consequences.[41] Van Schaik *et al.*[42] studied the influence of enabling conditions on general practitioners' referral behaviour for gastrointestinal disorders. Levels of experience interacted with the use of enabling conditions: the more experienced they were, the more the doctors would use enabling condition information in the cases. However, contrary to illness script theory, evidence was also found for moderation of consequences.

These studies seem to imply that, with growing expertise, the number and richness of enabling conditions increases and experienced doctors make increasing use of knowledge of enabling conditions.

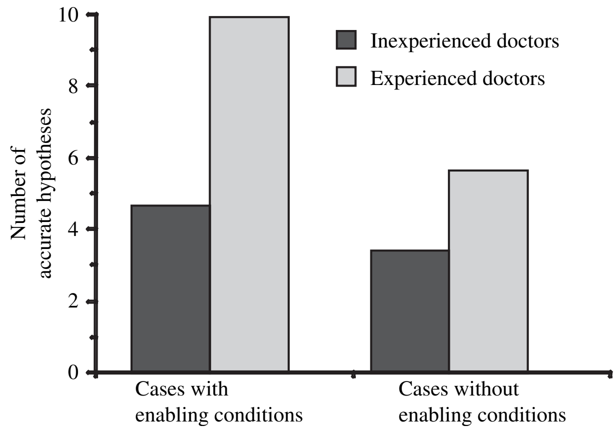

Figure 1 Influence of the availability of enabling conditions on the diagnostic performance of relatively inexperienced and experienced general practitioners. Adapted from Hobus[15]

The experts' dependence on enabling conditions for diagnosis is best exemplified by a study by Hobus.[15] He presented 16 short cases consisting solely of a number of enabling conditions and a single presenting complaint. Experienced doctors were almost twice as good as inexperienced doctors at diagnosing these cases. However, when these cases were presented without enabling conditions in a second experiment, the differences in diagnostic accuracy between experienced and inexperienced doctors disappeared entirely, demonstrating that advanced expertise in a domain seems to be solely dependent on the acquisition of enabling conditions through extended practice. Fig. 1 summarises the Hobus findings.

DISCUSSION

Many studies conducted within the encapsulation–illness script paradigm seem to support its predictions. In particular, the crucial role of knowledge of enabling conditions in advanced levels of medical expertise, and the fact that this knowledge is experiential rather than taught, seems to be well established. By contrast, the seemingly conflicting findings indicate that the extent to which experts actually use biomedical knowledge in their dealings with patients remains a thorny issue in need of further clarification. In particular, studies by Van de Wiel *et al.*[36] and Rikers *et al.*[18,39] demonstrate, contrary to predictions of the encapsulation hypothesis, that experts are actually better or faster at using biomedical knowledge than advanced students. This suggests that biomedical knowledge must be more frequently used in professional practice than assumed earlier. However, as many of these studies used non-experimental designs, alternative explanations could not always be ruled out. There is a definite need for more experimental studies here.

What are the implications of these findings for the practice of medical education? We mention 3 here.

1 Basic science should be taught only to the extent that it is – directly or indirectly – relevant to the development of encapsulating concepts. The integration of biomedical and clinical science should not be left to the students but the encapsulation process should be supported by *integrated teaching*. We firmly believe that modern curricula emphasising the organisation of disease processes around organ systems are more effective than the classic Flexnerian curriculum, which emphasises the teaching of biomedical and clinical knowledge as different phases in the medical curriculum.

2 Allow students to work with patient problems early in the curriculum, and allow them to see many and varied patients. This would certainly encourage processes of encapsulation and illness script formation. In this respect many curricula fall short. For instance, Wimmers *et al.*[43] found that, during a 12-week internal medicine clerkship, students saw fewer than 4 patients on average each week and, in fact, were confronted with little more than a single different disease per week. Thus, such clerkships can hardly be considered to contribute to the growth of expertise in these students.

3 Much time during clerkships and other postings should be spent on having students reflect and elaborate on the problems of the patients they see. Elaboration with a coach, preferably in small groups of peers, is generally considered a most effective way to integrate knowledge from different sources and to develop knowledge structures fit to the task at hand.[44,45]

Contributors: both authors discussed the ideas outlined in this paper. The first author wrote the first draft of the paper. The second author contributed parts of the final text.

Acknowledgements: none.
Funding: none.

Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978.
2 Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cogn Sci* 1986;**10**:91–116.
3 Norman GR, Brooks LR, Colle CL, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cognit Instr* 1999;**17** (4):433–48.
4 Lesgold A, Rubinson H, Feltovich PJ, Glaser R, Klopfer D, Wang Y. Expertise in a complex skill: diagnosing X-ray pictures. In: Chi MTH, Glaser R, Farr M, eds. *The Nature of Expertise*. Hillsdale, NJ: Erlbaum 1988;311–42.
5 Schmidt HG, Boshuizen HPA. Encapsulation of biomedical knowledge. In: Evans DA, Patel VL, eds. *Advanced Models of Cognition for Medical Training and Practice*. New York, NY: Springer Verlag 1992; 265–282.
6 Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;**5**:1–17.
7 Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise – theory and implications. *Acad Med* 1990;**65** (10):611–21.
8 Norman GR, Brooks LR, Allen SW. Recall by expert medical practitioners and novices as a record of processing attention. *J Exp Psychol Learn Mem Cogn* 1989;**15**:1166–74.
9 Chi MTH, Feltovich PJ, Glaser R. Categorisation and representation of physics problems by experts and novices. *Cogn Sci* 1981;**5**:121–52.
10 Rikers RMJP, Schmidt HG, Boshuizen HPA. Knowledge encapsulation and the intermediate effect. *Contemp Educ Psychol* 2000;**25** (2):150–66.
11 Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical case recall. *Mem Cognit* 1993;**21** (3):338–51.
12 Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cogn Sci* 1992;**16** (2):153–84.
13 McLaughlin KJ. *The Contribution of Analytic Information Processing to Diagnostic Performance in Medicine*. Rotterdam: Erasmus University 2007.
14 Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, eds. *Tutorials in Problem-Based Learning*. Assen/Maastricht: Van Gorcum 1984;128–142.
15 Hobus PPM. *Expertise van Huisartsen, Praktijkervaring, Kennis en Diagnostische Hypothesevorming [Expertise of Family Physicians; Practical Experience, Knowledge and the*

Formation of Diagnostic Hypotheses]. Maastricht: University of Limburg 1994.
16 Hatala R, Norman GR, Brooks LR. Influence of a single example on subsequent electrocardiogram interpretation. *Teach Learn Med* 1999;**11** (2):110–7.
17 Kulatunga-Moruzi C, Brooks LR, Norman GR. Co-ordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med* 2001;**13** (2):110–6.
18 Rikers RMJP, Loyens SMM, Schmidt HG. The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Med Educ* 2004;**38** (10):1035–43.
19 Van de Wiel MWJ, Boshuizen HPA, Schmidt HG. Knowledge restructuring in expertise development: evidence from pathophysiological representations of clinical cases by students and physicians. *Eur J Cognit Psychol* 2000;**12** (3):323–55.
20 Rikers RMJP, Schmidt HG, Boshuizen HPA. On the constraints of encapsulated knowledge: clinical case representations by medical experts and subexperts. *Cognit Instr* 2002;**20** (1):27–45.
21 De Bruin ABH, Schmidt HG, Rikers RMJP. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modelling approach. *Acad Med* 2005;**80** (8):765–73.
22 Woods NN, Brooks LR, Norman GR. The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Med Educ* 2005;**39** (1):107–12.
23 Woods NN, Neville AJ, Levinson AJ, Howey EHA, Oczkowski WJ, Norman GR. The value of basic science in clinical diagnosis. *Acad Med* 2006;**81** (10 Suppl):124–7.
24 Woods NN, Howey EHA, Brooks LR, Norman GR. Speed kills? Speed, accuracy, encapsulations and causal understanding. *Med Educ* 2006;**40** (10):973–9.
25 Gilhooly KJ, McGeorge P, Hunter J, Rawles JM, Kirby IK, Green C, Wynn V. Biomedical knowledge in diagnostic thinking: the case of electrocardiogram (ECG) interpretation. *Eur J Cognit Psychol* 1997;**9** (2):199–223.
26 Durante R, McKinlay JB, Kasten L, Potter DA. The influences of patient characteristics and physician experience on case recall. *Med Decis Making* 1997;**17** (2):199–207.
27 Witteman CLM, van den Bercken JHL. Intermediate effects in psychodiagnostic classification. *Eur J Psychol Assess* 2007;**23** (1):56–61.
28 Caillies S, Tapiero I. Text structures and prior knowledge. *Annee Psychol* 1997;**97** (4):611–39.
29 Patel VL, Groen GJ. Developmental accounts of the transition from medical student to doctor: some problems and suggestions. *Med Educ* 1991;**25**:527–35.
30 Eva KW, Norman GR, Neville AJ, Wood TJ, Brooks LR. Expert–novice differences in memory: a reformulation. *Teach Learn Med* 2002;**14** (4):257–63.
31 Van de Wiel MWJ. *Knowledge Encapsulation: Studies on the Development of Medical Expertise.* Maastricht: Maastricht University 1997.
32 De Bruin ABH, Van de Wiel MWJ, Rikers R, Schmidt HG. Examining the stability of experts' clinical case processing: an experimental manipulation. *Instr Sci* 2005;**33** (3):251–70.
33 Verkoeijen PPJL, Rikers RMJP, Schmidt HG, Van de Wiel MWJ, Kooman JP. Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 2004;**38** (6):617–27.
34 Wimmers PF, Schmidt HG, Verkoeijen PPJL, Van de Wiel MWJ. Inducing expertise effects in clinical case recall. *Med Educ* 2005;**39** (9):949–57.
35 Van de Wiel MWJ, Schmidt HG, Boshuizen HPA. A failure to reproduce the intermediate effect in clinical case recall. *Acad Med* 1998;**73** (8):894–900.
36 Van de Wiel MWJ, Boshuizen HPA, Schmidt HG, Schaper NC. The explanation of clinical concepts by expert physicians, clerks, and advanced students. *Teach Learn Med* 1999;**11** (3):153–63.
37 Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task-difficulty. *Mem Cognit* 1990;**18** (4):394–406.
38 Rikers R, Loyens S, te Winkel W, Schmidt HG, Sins PHM. The role of biomedical knowledge in clinical reasoning: a lexical decision study. *Acad Med* 2005;**80** (10):945–9.
39 Rikers R, Schmidt HG, Moulaert V. Biomedical knowledge: encapsulated or two worlds apart? *Appl Cogn Psychol* 2005;**19** (2):223–31.
40 Custers E, Boshuizen HPA, Schmidt HG. The role of illness scripts in the development of medical diagnostic expertise: results from an interview study. *Cognit Instr* 1998;**16** (4):367–98.
41 Custers E, Boshuizen HPA, Schmidt HG. The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Mem Cognit* 1996;**24** (3):384–99.
42 Van Schaik P, Flynn D, van Wersch A, Douglass A, Cann P. Influence of illness script components and medical practice on medical decision making. *J Exp Psychol Appl* 2005;**11** (3):187–99.
43 Wimmers PF, Schmidt HG, Splinter TAW. Influence of clerkship experiences on clinical competence. *Med Educ* 2006;**40** (5):450–8.
44 Chi MTH, Siler SA, Jeong H, Yamauchi T, Hausmann RG. Learning from human tutoring. *Cogn Sci* 2001;**25** (4):471–533.
45 Springer L, Stanne ME, Donovan SS. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Rev Educ Res* 1999;**69** (1):21–51.

Received 21 May 2007; accepted for publication 1 August 2007

Non-analytical models of clinical reasoning: the role of experience

Geoff Norman, Meredith Young & Lee Brooks

OBJECTIVE This paper aims to summarise the evidence supporting the role of experience-based, non-analytic reasoning (NAR) or pattern recognition as a central feature of expert medical diagnosis.

METHODS The authors examine a series of studies, primarily from their own research programme at McMaster University, that demonstrate that expert and novice diagnostic problem solving is based, to some degree, on similarity to a prior specific exemplar in the memory.

RESULTS The studies reviewed have shown NAR to be a component of diagnostic reasoning at all levels from novice to subspecialist, and in dermatology, electrocardiography and psychiatry. The retrieval process is rapid and is not available to retrospection. It may be based on visual similarity, but can also be present in verbal descriptions. Some evidence exists that the process is unlikely to be available to introspection. Further, early hypotheses based on NAR can result in the reinterpretaton of critical clinical findings.

CONCLUSIONS Non-analytic reasoning is a central component of diagnostic expertise at all levels. Clinical teaching should recognise the centrality of this process, and aim to both enhance the process through the learning of multiple examples and to supplement the process with analytical de-biasing strategies.

KEYWORDS *clinical competence; *decision making; *diagnosis; physicians/*standards/psychology.

McMaster University, Hamilton, Ontario, Canada

Correspondence: Geoff Norman, MDCL 3519, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada.
Tel: 00 1 905 525 9140 (ext 22119); Fax: 00 1 905 529 7099;
E-mail: Norman@mcmaster.ca

*Medical Education 2007: **41**: 1140–1145*
doi:10.1111/j.1365-2923.2007.02914.x

INTRODUCTION

Over 30 years ago, Elstein *et al.*[1] labelled essential characteristics of the clinical reasoning process; the so-called 'hypothetico-deductive method'. Expert clinicians, when faced with an undifferentiated diagnostic problem, reduce uncertainty by generating one or more diagnostic hypotheses and then searching for additional information to confirm or refute (mainly confirm) one or other of the hypotheses. By this strategy, the induction problem is reduced to an issue of deduction, of the form 'If the patient has X, then he must exhibit the following features'. Although this is certainly a concise description of the essential elements, a number of problems have arisen with the original formulation, and additional findings have emerged since it was established.

The first problematic finding was that this was an apt description of both experts and novices; even Year 1 medical students generated a similar number of hypotheses as experts.[2] So although the description may be accurate, it gives an incomplete account of the acquisition of expertise. Indeed, the only difference that emerged between experts and novices was that experts generated better (i.e. more accurate) hypotheses. Subsequent research has confirmed this finding. For example, in a recent study by Groves *et al.*,[3] experts made more errors in data gathering and data interpretation than novices, but were more accurate overall because they generated better hypotheses. A second feature of expert performance is that frequently there may be only 1 hypothesis, and all data gathering is directed to confirm this tentative diagnosis. Such a strategy has been called 'forward reasoning',[4] where

Overview

What is already known on this subject

Many theories of clinical reasoning have been advanced, primarily directed at the analytical knowledge structure of expert clinicians. One exception is the authors' research programme, which examines the role of experience in reasoning.

What this study adds

This study represents a synthesis of all studies examining the role of non-analytical reasoning (reasoning from prior exemplars) in medicine.

Suggestions for further research

Further research is needed to better understand how experts and novices differ in their use of non-analytic reasoning, and in the co-ordination of analytic and non-analytic strategies.

all the data apparently lead inductively to a single conclusion.

This is not to suggest that the subsequent data search and interpretation is unrelated to expertise. Bordage and Lemieux, for example, have shown that expertise is associated with coherent and compiled knowledge structures[5] and research by Schmidt and Boshuizen[6] has shown that experts do possess extensive encapsulated analytical and mechanistic knowledge that can be accessed as needed (but rarely is).

Nevertheless, left unanswered is the fundamental question of the source of these rich, accurate hypotheses. How can it be that experts, with minimal information, are able to advance tentative hypotheses about the diagnoses, seemingly effortlessly, and apparently without conscious awareness of the retrieval process? It is this question – 'Where do the hypotheses come from?' – that has been a central focus of our research programme for nearly 3 decades. One explanation arose from an unlikely source: psychological research on everyday categorisation. Categorisation is, as Murphy and Ross[7] put it: 'a central part of intelligent thought, allowing us to apply knowledge learned about a limited set of objects to a potentially infinite class of new, previously unseen, objects'. That is also precisely the role of the diagnosis in medical practice.

There are 2 dominant theories in psychology to explain the process of categorisation. The first, prototype theory, assumes that a person's experience with individual exemplars is averaged into a prototype of the category that contains most of the critical features of the category. Classification of a new object then proceeds by identifying the category in memory that contains most features in common with the new object.

An alternative view is 'exemplar theory', which posits that people are able to identify category members – a chair, a dog, a coffee mug – effortlessly and without apparent feature analysis because, in the course of their maturation, they have acquired a large number of examples or exemplars of each natural category, and are able to carry out the categorisation by making an unconscious similarity match with a particular prior example of the category.

These 2 theoretical perspectives, the first of which is based on feature-by-feature matching against a prototype, whereas the other is based on a holistic match to a prior example, imply different mental processes; however recent evidence suggests that people use a mixture of the 2 strategies. When they are novices, people tend to categorise by resemblance to a prototype, but as they gain experience, they rely more and more on similarity to individual exemplars.[8] As we shall see, a parallel developmental sequence appears to occur in clinical reasoning.

Our primary interest in this paper is the role of experience in medical diagnosis, and the application of an exemplar model of categorisation to the process of clinical reasoning. We label this process 'non-analytic reasoning' (NAR) to emphasise that its characteristics are very different from the logical analysis that characterises other theories of reasoning.

In the present paper, we will review studies that have demonstrated the presence of NAR in both experts and novices. We will then explore some empirical consequences of this formulation. First, we will examine the time taken to arrive at a diagnosis, and show that experts take less time than novices when they are correct but more when they are incorrect or unsure; this is consistent with the view that correctness is associated with rapid access to an exemplar. We will show evidence that experts are unable to anticipate the errors made by other experts, because mental exemplars are, by their nature, idiosyncratic. Finally, we will show that data interpretation is not, in general, value-free and objective. Instead, even apparently classical signs can possess considerable ambiguity and show large effects of re-interpretation consistent with a primary hypothesis.

DEMONSTRATING THE PROCESS OF NON-ANALYTICAL REASONING

A tradition in research on clinical reasoning, dating back to the first studies, is the use of introspective accounts, either from 'thinking aloud' during inquiry or from 'stimulated recall' after the process has concluded. Such methodologies are insufficient to access NAR for the simple reason that it is presumed to be inaccessible to introspection.

A different experimental strategy that does not rely on overt memory is required. This method has 2 elements: an initial learning phase where, under the guise of learning the rules and then practising them on some cases (for novices) or reviewing some cases (for relative experts), participants learn not just the rules, but also cases that exemplify the rules. A test phase then systematically manipulates similarity to particular practice examples, either in the specific feature descriptions or in some irrelevant aspect of context (e.g. the pain 'feels like a knife' versus 'the patient is a 45-year-old accountant' or 'the lesion is on a forearm'), and examines the impact of these prior cases on subsequent diagnoses.

A first demonstration of NAR was reported in 1989, by Brooks *et al.*, using dermatology materials.[9] In this study, medical students learned 6 dermatology conditions and then practised these with 4 examples per category. The experimental manipulation referred to the specific examples shown, where 1 group would, for example, see a contact dermatitis on a forearm and a lichen planus on a hand, and the other would see a lichen planus on a forearm and a contact dermatitis on a chin. They were then tested with 30 new cases, of which 9 were 'chameleons' resembling a prior example (in this case, a second contact dermatitis on a forearm). The result showed a dramatic influence of the prior example by demonstrating an accuracy of nearly 90% when the similar example came from the correct category, versus 42% when it came from an alternate category. Further unpublished studies in dermatology using similar materials showed that the influence of a prior example can be detected after delay, resulting in an increase in diagnostic accuracy for similar cases of about 26% after a 2-week delay, and can be found in experienced general practitioners (GPs).

However, although dermatology has certain affordances, as prior experience can be acquired as rapidly as it takes to change a slide, it is highly visual and holistic, and generally is viewed as a 'pattern recognition' process. Moving a step away from these visual materials, we turned to electrocardiography.[10] Individual examples are presumably much less memorable, consisting only of lines on a page. Further, the rules appear simple, inclusive and unequivocal (e.g. if the sum of the R wave on V2 and the S wave on V4 exceed 50 mm, it's congestive heart failure).

In this experiment, residents reviewed a series of cases consisting of a brief history and an electrocardiogram, and then worked through a series of test cases. These were matched or unmatched to review cases based on irrelevant demographic information (a 50-year-old banker versus an 80-year-old woman). However, test cases were always a different diagnosis (e.g. left bundle branch block versus anterior myocardial infarction), so that if subjects recalled the prior case, they would be less likely to get the diagnosis. When the case was matched to a review case on irrelevant details, diagnostic accuracy was only 23% versus 46% for unmatched cases. What is interesting about these findings is that the specific features being manipulated were objectively irrelevant to the diagnosis, so if residents were aware that they were being influenced by this information, they would not have been biased, suggesting an unconscious retrieval process.

These studies show that the effect is present in all levels of expertise and can also occur in verbal descriptions. Recent unpublished work by author MY has extended the findings. She began with 4 'pseudo-psychiatric' diseases ('pseudo-' in the sense that each was characterised by only 4 features [e.g. hallucinations]). She trained undergraduates in the 4-feature descriptions and a series of cases, in which 2 factors were experimentally manipulated: the specific feature descriptions (e.g. for hallucinations: 'She thinks that her dead mother makes her lunch every morning') and the case identifier. The research is related to the effect of familiar feature descriptions described in another paper published in this issue.

Of relevance to the present paper, MY examined the effect of a familiar identifier. She carefully created similar, but not identical, descriptions; at practice the student might meet John McIntosh, a 35-year-old public school teacher with a 5-year-old boy and a 7-year-old girl, and at test, see James McKinley, a 37-year-old elementary school teacher with twin 6-year-old daughters. Test cases were created that had 2 primary diagnoses with 2 features each, such that diagnosis A was associated with familiar features, and diagnosis B with a familiar demographic description. If familiarity were not important, the probabilities should be 50 : 50 for each. What actually emerged was that on immediate test, the probabilities were 52% for the diagnosis with familiar features and 42% for the diagnosis associated with the familiar patient identification; a day later, however, the situation had reversed and the probability of the diagnosis associated with the familiar patient identifier was 58% versus 37% for the diagnosis associated with familiar features. Presumably on immediate test, subjects recalled where they had seen a similar description and discounted it; a day later, they were less able to recall the context[11] The study suggests 3 important conclusions: firstly, familiarity effects can be induced by verbal as well as visual materials; secondly, non-analytic similarity can be induced by similar, but not identical, patient descriptions, and thirdly, the process endures, and may increase with the passage of time.

CONSEQUENCES OF NON-ANALYTIC REASONING

To the extent that expert doctors rely on NAR, we can predict certain, somewhat counter-intuitive, consequences, as mentioned in the Introduction.

Accurate diagnosis may be associated with less, not more, time

Whereas diagnostic errors are frequently attributed to haste, carelessness or 'premature closure', to the extent that experts use NAR, it can be anticipated that the process is both fast and accurate. By contrast, if no similar exemplar is accessible, the expert may have to slow down and initiate a deliberative search. Of course, in normal practice, it is difficult to isolate 'thinking time'. However, in 1 study we had dermatologists and more junior doctors diagnose a series of dermatology slides.[12] The time taken to make a correct diagnosis showed an inverse relationship to expertise: students took 13 seconds and dermatologists 8 seconds on average. The time taken to make an incorrect diagnosis was slightly positively related to level of expertise: students took 11 seconds and dermatologists 15 seconds on average. For slides that resulted in 'don't know' answers, students took 10 seconds and dermatologists 28 seconds. The interaction was highly significant ($P < 0.0001$). Thus, for experts, accuracy was inversely associated with time taken to arrive at a diagnosis.

Experts cannot predict errors of other experts

One would presume that academic clinical teachers would become expert at anticipating the kinds of errors learners make for a particular case. By contrast, if diagnosis is based on the idiosyncratic exemplars that each clinician has acquired during training and practice, it may be that one clinician cannot predict the errors another might make. To investigate this we used slides that had been diagnosed in a previous study[12] by residents, GPs and dermatologists. We had 3 dermatologists review the slides and simply indicate what errors another clinician might make. The first nomination, on average, identified only 17% of resident errors, 22% of GP errors and 27% of errors by other dermatologists. A second pass involving multiple diagnoses succeeded in identifying only 28% of resident errors, 33% of GP errors and 40% of dermatologist errors. This evidence is inconsistent with the view that errors are a consequence of specific misleading information, and consistent with a perspective that NAR based on individual exemplars is operative.

Ambiguous features are easily misinterpreted

Much of medical education is concerned with mastery of the relationships between signs and symptoms and diagnoses. It is often tacitly assumed that the difficulty of the task emanates from application of the rules, and not from the perception and labelling of features. However, if diagnosis proceeds by NAR, features may be re-interpreted to coincide with the diagnosis being entertained. LeBlanc *et al.*[13] tested this assumption by showing residents and medical students photographs of classical signs (e.g. a moon-shaped face, exopthalmus, butterfly rash, jaundice) drawn from textbooks. Each was accompanied by a brief history and a suggested diagnosis, which was either correct or represented a plausible alternate (e.g. a child with swollen parotid glands was described as showing symptoms consistent with mumps and with

Cushing's disease). Subjects at both levels were far more likely to conclude for the correct diagnosis (77% versus 9%; $P < 0.0001$) and were more likely to identify the features of the correct diagnosis (49% versus 37%; $P < 0.05$) when the correct diagnosis was suggested.

DISCUSSION

The studies reviewed in this paper provide considerable evidence that a critical component of expert and novice diagnostic reasoning is based on similarity to previously encountered examples or exemplars (NAR). The phenomenon has been demonstrated in a variety of contexts with subjects at varying levels of training and with very different materials. Nevertheless, it cannot be the case that all reasoning, all the time, is non-analytic. Certainly, there is some evidence that expert nephrologists, faced with difficult cases and minimal data, are likely to revert to a mechanism-based physiological reasoning.[14,15] Moreover, we cannot ignore the many studies suggesting that clinicians may have different representations of diagnostic knowledge, such as semantic axes,[5] causal propositions[4] or illness scripts.[16] Further, we have shown that encouraging students to use both NAR and analytical knowledge[17] can lead to improved accuracy over either alone.

There remains a further caveat. Fundamental to the experimental approach common to all these studies is the necessity to control exposure to prior exemplars in order to unambiguously show the effect of prior exemplars on reasoning. As the process is presumed to not be under conscious control, strategies used elsewhere, such as 'thinking aloud', would not be defensible. However, the consequence of this approach is that we cannot show any effect of expertise experimentally, as the acquisition of expertise cannot be adequately simulated over the short time-frames of these studies. However, in addition to the evidence of expertise effects described in the previous section, there is other suggestive evidence. Eva and Cunnington[18] have shown that with increasing age, diagnosticians rely more and more on evidence gathered early, consistent with a NAR perspective. Kulatanga-Moruzi *et al.*[19] showed in dermatology that experts reasoned better from an uninterpreted photograph of a lesion than from a verbal interpretation, whereas the opposite was true for residents, suggesting that the experts' reasoning was based on specific visual examples.

Finally, the effects of exemplars have been shown to persist over periods of a few days and up to 2 weeks.[20] Although this may appear a short time interval, it far exceeds the transition from short- to longterm memory, which occurs in seconds.

IMPLICATIONS FOR EDUCATION AND HEALTH CARE

This theory of clinical reasoning, to the extent that it reasonably characterises routine medical diagnosis, has important consequences for clinical instruction. Much clinical learning is directed at learning the 'signs and symptoms of malaria' and the 'differential diagnosis of fatigue'. Further, students spend hours mastering the 'complete history and physical'. Emphasis on this kind of knowledge and skill underestimates the importance of actual clinical experiences in the acquisition of expertise. More effort should be put into assessing the type and sequence of clinical experiences needed for some level of diagnostic mastery.

How is this related to diagnostic error? One proposed approach to reduce error – computer-based diagnostic support systems – is likely to have minimal impact,[21] for the simple reason that the doctor's search for information and interpretation of features is not and cannot be hypothesis-neutral, yet the elicited and interpreted data, translated by the doctor into signs and symptoms, represent what is input into the computer.

Instead, interventions must seek to improve and supplement, not replace, these non-analytic processes. It is clear that 1 component of expertise, in this model, is the availability of many similar exemplars to facilitate appropriate pattern recognition, and this can only result as a consequence of extensive experience, the deliberate practice of Ericsson.[22] By contrast, error-checking strategies, such as admonitions to 'think of other possibilities' or to 'think of other ways to interpret the data' that build on the initial hypothetico-deductive search, may well have some success, as demonstrated by Ark and Eva.[23]

Contributors: all authors contributed to the conception and design of the study, the collection, analysis and interpretation of data, and the write-up and revision of the manuscript.
Acknowledgements: none.
Funding: this research was supported by the Natural Science and Engineering Research Council of Canada and

the Social Sciences and Humanities Research Council, Canada.
Conflicts of interest: none.
Ethical approval: this study was approved by the McMaster University Ethics Board.

REFERENCES

1 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978;64–6.
2 Neufeld VR, Norman GR, Barrows HS, Feightner JW. Clinical problem-solving by medical students: a longitudinal and cross-sectional analysis. *Med Educ* 1981;**15**:315–22.
3 Groves M, O'Rourke P, Alexander H. The clinical reasoning characteristics of diagnostic experts. *Med Teach* 2003;**25**:308–13.
4 Patel VL, Evans DA, Groen GJ. Biomedical knowledge and clinical reasoning. In Evans DA, Patel VL, eds. *Cognitive Science in Medicine*. Cambridge, MA: MIT Press 1988;96–8.
5 Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med* 1991;**66 (Suppl)**:70–2.
6 Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical case recall. *Mem Cognit* 1993;**21**:338–51.
7 Murphy GL, Ross BH. Predictions from uncertain categorisations. *Cognit Psychol* 1994;**27**:148–93.
8 Minda JP, Smith JD. Prototypes in category learning: the effects of category size, category structure and stimulus complexity. *J Exp Psychol Learn Mem Cogn* 2001;**27**:275–92.
9 Brooks LR, Norman GR, Allen SW. The role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen* 1991;**120**:278–87.
10 Hatala RM, Norman GR, Brooks LR. Influence of a single example upon subsequent electrocardiogram interpretation. *Teach Learn Med* 1999;**11**:110–7.
11 McElree B, Dolan PO, Jacoby LL. Isolating the contributions of familiarity and source information to item recognition: a time-course analysis. *J Exp Psychol Learn Mem Cogn* 1999;**25**:563–82.
12 Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol* 1989;**125**:1063–8.
13 LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;**77 (Suppl)**:67–9.
14 Norman GR, Trott AL, Brooks LR, Smith EKM. Cognitive differences in clinical reasoning related to postgraduate training. *Teach Learn Med* 1994;**6**:114–20.
15 Verkoeijen PPJL, Rikers RMJP, Schmidt HG, van der Weil MWJ, Koomna JP. Case presentation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 2004;**38**:617–27.
16 Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ* 1987;**21**:86–91.
17 Kulatanga-Moruzi C, Brooks LR, Norman GR. Co-ordination of analytical and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med* 2001;**13**:110–6.
18 Eva KW, Cunnington JP. The difficulty with experience: does practice increase susceptibility to premature closure? *J Contin Educ Health Prof* 2006;**26**:192–8.
19 Kulatunga-Moruzi C, Brooks LR, Norman GR. The diagnostic disadvantage of having all the facts: using comprehensive feature lists to bias medical diagnosis. *J Exp Psychol Learn Mem Cogn* 2004;**30**:563–72.
20 Norman GR, Brooks LR, Allen SW, Rosenthal D. Sources of observer variation in dermatologic diagnosis. *Acad Med* 1990;**65**:S19–S20.
21 Berner ES, Maisiak RS, Cobbs CG, Taunton OD. Efforts of a decision support system on physcians' diagnostic performance. *J Am Med Inform Assoc* 1999;**6**:420–7.
22 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79 (Suppl)**:1–12.
23 Ark TK, Brooks LR, Eva KW. The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ* 2007;**41**:281–7.

Received 17 April 2007; editorial comments to authors 1 August 2007; accepted for publication 9 September 2007

Found in translation: the impact of familiar symptom descriptions on diagnosis in novices

Meredith Young,[1] Lee Brooks[1] & Geoff Norman[2]

CONTEXT The language that patients use to communicate with doctors is quite different from the language of diagnosis. Patients may describe tiredness and swelling; doctors, fatigue and oedema. This paper addresses the process by which novices, who have learned standard medical terms for symptoms, use lay descriptions of symptoms to reach a diagnosis. Data in this paper indicate that the familiarity of the language used to describe symptoms influences diagnosis in novices and diagnosis does not, therefore, involve a simple translation into standard terms that are the basis of diagnostic decision.

METHODS A total of 24 undergraduate students were trained to diagnose 4 pseudo-psychiatric disorders presented in written vignettes. Participants were tested on cases that contained 2 equally probable diagnoses, in 1 of which the symptoms were expressed using previously seen descriptions. A deviation from 50 : 50 in reported diagnostic probabilities was expected if the familiar symptom descriptions biased diagnostic decisions. Twelve participants were tested immediately after training and 12 after a 24-hour delay.

RESULTS Participants assigned greater diagnostic probability to the diagnosis supported by the familiar feature descriptions ($F[1.242] = 19.35$, $P < 0.001$, effect size = 0.40) on both immediate (52% versus 41%) and delayed (51% versus 38%) testing.

[1]Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada
[2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

Correspondence: Meredith Young, Department of Psychology, Neuroscience and Behaviour, PC 141, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada.
Tel: 00 1 905 525 9140 (ext 24824); Fax: 00 1 905 529 6225;
E-mail: youngme2@mcmaster.ca

DISCUSSION The findings indicate that diagnosis is not simply based on a process of translating patient descriptions of symptoms to standard medical labels for those symptoms, which are then used to make a diagnosis. Familiarity of symptom description has an effect on diagnosis and therefore has implications for medical education, and for electronic decision support systems.

KEYWORDS *diagnosis; decision making; education, medical, undergraduate/*methods; psychology/*education; teaching/*methods.

*Medical Education 2007: **41**: 1146–1151*
doi:10.1111/j.1365-2923.2007.02913.x

INTRODUCTION

When medical students begin to study medicine, they are taught diagnostic rules, lists of cardinal signs, and lists of textbook features for a variety of illnesses. These learned lists are usually expressed in terms of the standard medical labels for features. However, patients rarely walk into the office using medical language to describe what ails them. Patients do not use terms like retrosternal chest pain, oedema or bilateral weakness unless they are also health professionals. Instead, they talk about stabbing chest pain, swollen ankles or a loss of strength. Further, the actual presentation of features in a particular patient is rarely exactly what is shown in a textbook photograph: physical features may look somewhat different on patients of different ages, genders, races or physical conditions. Medical students have to learn to recognise these unique descriptions as *instantiations*[1] of the features that are named by the diagnostic rule. Without such a 'translational' ability, students will be unable to communicate with colleagues or relate the instantiations to the rules and disease processes discussed in their formal learning. However, the

Overview

What is already known on this subject

Medical diagnosis involves the ability to effectively 'translate' information provided by a patient into the language of medical diagnosis. However, this process has not received much attention in the literature.

What this study adds

This study investigates the biasing impact of familiar symptom descriptions on medical diagnosis in novices. We show that familiar descriptions are weighted more heavily in diagnosis than equivalent novel descriptions.

Suggestions for further research

Further research might involve investigating similar effects in more experienced doctors, and investigating how one learns to 'translate' from lay descriptions to a medical diagnosis.

question raised in this paper is whether diagnosis is strictly dependent on such a translation process. That is, do novices translate the feature instantiations present in the case into standard medical language, and then make their decision by matching them to the terms given in diagnostic rules that they have learned?

Certainly, there are alternatives to sole dependence on such a translation process. A companion paper by Norman *et al.* describes a spectrum of diagnostic strategies, whose end-points are characterised as non-analytic and analytic processing.[2] Analytic processing refers to the systematic, deliberate seeking of medically relevant features and the use of these features to make decisions on the basis of diagnostic rules. When strictly applied, this process is restricted to medically relevant features and is generally regarded as free of the cognitive biases and diagnostic errors[3] that have recently received much attention in the popular press.[4,5] It has also encouraged further development of computer-assisted diagnostic support systems.[6,7]

Non-analytic processing is often placed in opposition to this rule-based approach, proposing that decisions are heavily influenced by similarity to a prior case. Non-analytic processing is often used to refer to holistic case similarity,[3] or similarity in patient information that in itself is non-diagnostic.[8] Similarity of whole cases or patient identifiers is not conclusive evidence, but it could serve an important mnemonic function when a diagnostician is struggling with a large number of possibilities or with time pressures. Under some circumstances, the use of overall case similarity has been shown to lead to more accurate diagnosis.[9]

However, it is possible that non-analytic or similarity-based processes have an influence even when dealing with individual, medically relevant features. We might easily imagine that if a patient describes her symptoms using familiar words, this will facilitate the translation to diagnostic language. It is also possible that this familiar language would suggest a diagnosis, or be taken as more convincing evidence than would the same feature expressed in novel form. Such an effect of familiar feature instantiations has been demonstrated in cognitive psychology,[1] and the impact of familiar terms has been demonstrated in medical education when the alternative novel terms are simple synonyms. For example, Dore *et al.*[10] demonstrated that novice diagnosticians rely more heavily on familiar synonymous features (i.e. sleeplessness and insomnia) than on equally valid, novel synonyms (i.e. inability to sleep, wakefulness).

There are several reasons why familiar feature instantiations may impact diagnostic decisions. Firstly, the difficulties faced by a novice diagnostician may be strictly a matter of recognising the synonymity of different instantiations, or low confidence in recognition. A novice diagnostician might have difficulty recognising a novel presentation as an instantiation of a term in a decision-making rule. For example, students may spend a lot of time learning that Cushing's syndrome is associated with a 'moon-shaped face' and that acute myocardial infarction often presents with retrosternal chest pain, but may be unable to recognise a moon-shaped face[11] or determine that a chest pain 'that feels like my chest is put in a vice' is the same as the textbook description. Secondly, the variety between symptom presentations might trigger more direct effects of familiarity. A symptom instantiation that the student has heard before may be closely associated with the diagnostic context in which it was previously seen. For example, having previously seen a depressed patient who complained of 'tossing and turning all night' may sway the clinician to make a diagnosis of depression when a subsequent patient expresses herself in the same way. In any of these cases, the process is not strictly rule-based, but instead has some of the characteristics of non-analytic reasoning. The latter type of familiarity effects may represent an intermediate phenomenon between the traditional dichotomous poles of analytic and non-analytic processing. In either case, the novelty or familiarity of a unique presentation may affect the diagnosis that is eventually made.

In this paper, we will first investigate the impact of familiar symptom descriptions on novice diagnosticians. The specific area of study for this paper is psychiatric diagnosis, a field in which verbal descriptions from patients is a key source of information, and in which there is large variability in the description of any 1 symptom. We will then determine if our evidence allows us to select between 2 possible sources of the impact on diagnosis of familiar feature instantiations in order to establish whether familiarity of the instantiation of a feature directly impacts the diagnostic decision of novices, or whether it has only an indirect effect by allowing novices to better translate familiar instantiations into the language of rules. Any familiarity effects suggest that the novice has more than rules to learn, and thus knowing more about the source of these effects might guide the development of educational strategies to help with this initial and complex learning.

METHODS

Participants

A group of 24 Year 1 psychology undergraduate students participated in this study in return for course credit. Undergraduate students were chosen above medical students in order to ensure that participants would have limited knowledge regarding psychiatric diagnoses.

Stimuli

Four pseudo-psychiatric disorders were created for the purposes of this experiment. The psychiatric disorders with modified diagnostic rules included schizophrenia, mania, obsessive compulsive disorder and paranoid personality disorder. Each disorder was characterised by 4 unique symptoms, limiting the possibility of confusion between diagnostic categories. Subjects were told that the disorders did not reflect the real psychiatric disorders of the same name, and that they should focus on the diagnostic rules presented in the experiment. Psychiatry was chosen for this study because it involves high levels of variability in verbal symptom presentation (i.e. there are many different ways in which delusions or obsessive thoughts can be described), whereas other areas of medicine (e.g. cardiology, nephrology) are less dependent on the patient's verbal descriptions of symptoms.

Each diagnostic feature (e.g. hallucinations) was presented in a variety of instantiations within the case vignettes (for example: hearing voices when nobody else is in the room, seeing vision before going to bed, etc.). The various descriptions were constructed to differ in their 'ease of translation' into their feature label, from near synonymous terms to full behavioural descriptions. A range of difficulty of translation was included in order to expand the work of Dore *et al.*,[10] which demonstrates the impact of familiar synonymous terms. By including near-synonymous terms (e.g. needs only 2–4 hours of sleep and decreased need for sleep), and complex behavioural descriptions (e.g. because of her husband's job loss, she has started working both day and night shifts and does not feel tired), we were able to evaluate the impact of familiarity using more complex and more realistic instantiations.

Procedure

All materials were presented on a computer, programmed with RunTime Revolution Version 2.5 (RunTime Revolution Ltd, Edinburgh, UK).

Learning phase

Participants were shown the 4 features that were diagnostic of a pseudo-psychiatric disorder (e.g. hallucinations, delusions, disorganised speech and disorganised behaviour for schizophrenia). They were asked to study the list and then to identify 4 features that were characteristic of the disorder from a list of 16 features. Upon completing this task, the next disorder was presented in the same way and this sequence repeated until all 4 diagnostic categories had been presented. To complete the learning phase, participants had to complete a quiz in which they were required to identify all the features for each disorder. The pass score was set at 15/16. A pass resulted in the participant moving into the practice phase. If participants did not pass the quiz, they re-started the learning phase.

Practise phase

Participants were shown a series of 12 different cases, 3 for each of the different pseudo-psychiatric disorders, presented in random order. Each case included a patient description (including name, age, occupation and familial situation), and the 4 features characteristic of 1 of the 4 disorders. In the practise phase, features presented were instantiated in a manner consistent with the way patients might describe their symptoms (e.g. hearing voices). The practise phase served 2 purposes: it allowed participants to gain some expertise with these disorders by providing diagnostic feedback, and it exposed participants to a set of feature instantiations that would function as familiar instantiations in the test phase.

Test phase

Participants in the 'immediate test' condition moved immediately to the test phase, whereas participants in the 'delay' condition returned to the laboratory to complete the test phase after a 24-hour delay. Participants diagnosed a total of 12 test cases.

All test cases included a total of 4 features presented in unique instantiations. Each case included 2 familiar feature instantiations (drawn from the practice cases) indicative of 1 disorder, and 2 novel familiar feature instantiations indicative of another disorder. Participants were asked to assign a diagnostic probability rating to each of the 4 disorders, and to report the diagnostically relevant features. Because 2 supporting features were present for 2 different diagnoses, the 'expected' unbiased response would be a 50 : 50 split in diagnostic probabilities assigned. If familiarity of feature descriptions impacts upon the assignment of diagnostic probabilities, we would expect to see a deviation from 50 : 50 in favour of the diagnosis supported by the familiar feature descriptions.

RESULTS

The diagnostic probability assigned to each diagnosis was recorded for each participant. A repeated measures analysis of variance (ANOVA) was conducted using the 2 plausible diagnoses (from the 50 : 50 design) as the within-subjects comparison, individual cases as the repeated measure, and the testing time (either immediate or delay) as a between-subjects comparison. Across both testing conditions, participants assigned significantly more diagnostic probability to the diagnosis supported by the familiar feature descriptions ($F[1.242] = 19.352$, $P < 0.001$, effect size $[\delta] = 0.40$). There was no significant interaction between delay and the impact of familiarity ($F[1.242] = 0.193$, $P > 0.05$), indicating that the impact of familiar symptom descriptions did not change with delayed testing. Mean scores are shown in Fig. 1.

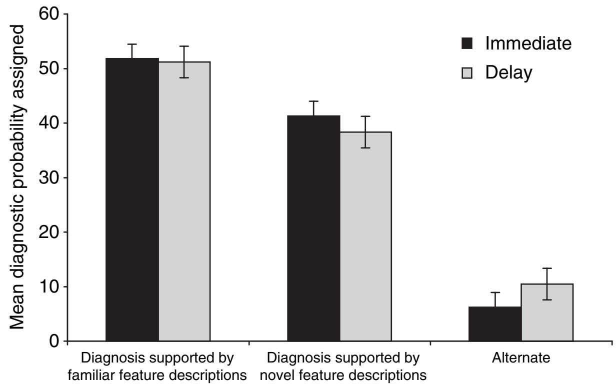

Figure 1 Mean diagnostic probability assigned to the diagnosis supported by familiar feature instantiations and the diagnosis supported by the novel feature instantiations. The other 2 diagnostic options are collapsed and shown as 'alternate'. Data for the immediate test group (***n*** = 12) and the delay test group (n = 12) are shown. Error bars indicate the standard error of the mean

In analysing the features reported by participants, we found that 10 of 24 participants reported the features entirely in the instructed 'rule' language (i.e. translated from the specific instantiation presented). These 10 participants reported nearly all features present in the test cases (99.8% of all features were reported), and reported them all in their translated form. This provides us with an indication that novices can – even when not instructed – translate correctly and effectively into diagnostic rule language. These participants showed an effect of familiarity comparable with that of the whole group (51% versus 42%), indicating that even 'perfect translators' are influenced by familiar feature descriptions.

DISCUSSION

This study demonstrates that a single, unique, prior instantiation of a medically relevant feature can affect a subsequent diagnosis. The impact of familiarity thus not only occurs at the level of holistic, non-analytic processing,[2] but also at the time of feature interpretation. This study also indicates that this effect is not transient, given that a familiar feature instantiation will continue to bias diagnostic decisions after a 24-hour delay. If a single, unique experience impacts the diagnostic decisions of novices following a 24-hour delay, then we assume this experience is stored in longterm memory,[12,13] which leads us to believe that familiarity could continue to impact diagnosis in longer delay periods, making it an influence to be taken seriously, at least for novices.

A sub-sample of subjects chose to report the diagnostically relevant features in the language of the diagnostic rule. These participants showed high levels of accuracy in their 'translations', but still showed an influence of familiarity. These results indicate that diagnosis among these novices is not just a matter of recognising instantiations of features and then matching their labels to the terms in the rules that have been learned. Future research should also examine the impact of familiar feature instantiations in expert doctors. Past research has indicated that expert doctors are influenced by familiarity of whole cases,[9] so it is possible that there may also be an effect of familiarity at the level of feature instantiation.

A rule cues us to search for particular features, or to confirm the features already found. It is important to remember that, no matter how good a rule is, it does not imply the extent to which that feature might vary in appearance or expression. This particular domain of study, psychiatry, is rich in feature variability. If we consider just the symptom of hallucinations, we understand that a patient might describe an auditory, olfactory, tactile or visual hallucination – all of which fall into the same symptom category, but a category for which there is incredible variability in terms of individual presentation.

Diagnostic error has recently received much attention in the public domain. There have been calls for more reliance on analytic, rule-based approaches to medicine, and for instructional programmes to teach doctors to be alert to potential cognitive biases that stem from pattern recognition processes (non-analytic). However, a growing literature challenges the idea that the move from a description of a symptom to a symptom label is unambiguous, and this observation has direct implications within medical education and on the development of diagnostic support systems. The current results suggest that caution should be applied in the pursuit of decision support systems, as such systems are dependent on the doctor's ability to input the correct 'translation' into them.[6,7] The present study shows that, at least in novices, this process is far from unbiased. Without this skill, diagnostic systems become ineffectual, which can be seen in the decreased accuracy of diagnoses generated by the system when medical students or doctors input the symptoms.[6] Indeed, because interpretation of features can be strongly influenced by a suggested diagnosis,[11] there is every reason to presume that the decision support system will be as prone to confirmation bias as the clinician who is supposed to be de-biased by using the system.

In conclusion, the process of medical diagnosis is a complex area, which we are slowly beginning to understand. The purpose of this paper was to investigate the impact of familiar instantiations on the diagnostic process of novices. We were successful in demonstrating such an effect for novices using materials that are plausible for psychiatric diagnosis, which lasted for at least 24 hours. This supports an interest in the much more concrete information that is provided by specific instantiations of features. From the moment a patient walks through a clinic door, she is communicating with the doctor – her gait, colouration, mannerisms and appearance all provide information that may influence a diagnosis in ways that are not captured by a strictly analytic understanding of the process of diagnosis.

Contributors: all authors contributed to the conception and design of the study, the collection, analysis and interpretation of data, and the write-up and revision of the manuscript.
Acknowledgements: the authors would like to acknowledge the continuing support of Elizabeth Howey.
Funding: this research was conducted under a larger grant provided by the Natural Science and Engineering Research Council of Canada.
Conflicts of interests: none.
Ethical approval: this study was approved by the McMaster University Ethics Board.

REFERENCES

1 Brooks LR, Hannah SD. Instantiated features and the use of 'rules'. *J Exp Psychol Gen* 2006;**132** (2):133–51.
2 Norman GR, Young ME, Brooks LR. Non-analytical models of clinical reasoning: the role of experience. *Med Educ* 2007;**41**:1140–45.
3 Croskerry P. The importance of cognitive errors in diagnosis and strategies to prevent them. *Acad Med* 2003;**87**:1–6.
4 Groopman J. What's the trouble? How doctors think. *New Yorker*, 2007; 29 January: 36–41.
5 Groopman J, Croskerry P. Is there a house in the doctor? *Globe and Mail* 2007; 24 March: F8.
6 Freidman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, Fine PL, Miller TM, Abraham V. Enhancement of clinicians' diagnostic reasoning by

computer-based consultation. *JAMA* 1999;**282** (19):1851–6.
7 Graber MA, VanScroy D. How well does decision support software perform in the emergency department? *Emerg Med J* 2003;**20**:426–8.
8 Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med* 1999;**14**:126–9.
9 Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen* 1991;**120** (3):278–87.
10 Dore K, Weaver B, Norman G. Does a rose by any other name smell as sweet? The importance of variation in case presentation to diagnostic reasoning. Abstract presented at the Research in Medical Education Conference, Washington, DC, 9–13. November 2005.
11 LeBlanc VR, Norman GR, Brooks LR. Effect of diagnostic suggestion on diagnostic accuracy and identification of clinical features. *Acad Med*, 2001;**76** (10) Suppl:18–20.
12 Squire LR, Knowlton B, Musen G. The structure and organisation of memory. *Annu Rev Psychol*, 1993;**44**:453–95.
13 Kellogg RT, Newcombe C, Kammer D, Schmitt K. Attention in direct and indirect memory tasks with short- and longterm probes. *Am J Psychol*, 1996;**109** (2):205–17.

Received 23 April 2007; editorial comments to authors 1 August 2007; accepted for publication 9 September 2007

Teaching from the clinical reasoning literature: combined reasoning strategies help novice diagnosticians overcome misleading information

Kevin W Eva,[1] Rose M Hatala,[2] Vicki R LeBlanc[3] & Lee R Brooks[4]

OBJECTIVE Previous research has revealed a pedagogical benefit of instructing novice diagnosticians to utilise a combined approach to clinical reasoning (familiarity-driven pattern recognition combined with a careful consideration of the presenting features) when diagnosing electrocardiograms (ECGs). This paper reports 2 studies demonstrating that the combined instructions are especially valuable in helping students overcome biasing influences.

METHODS Undergraduate psychology students were trained to diagnose 10 cardiac conditions via ECG presentation. Half of all participants were instructed to reason in a combined manner and half were given no explicit instruction regarding the diagnostic task. In Study 1 (n = 60), half of each group was biased towards an incorrect diagnosis through presentation of counter-indicative features. In Study 2 (n = 48), a third of the test ECGs were presented with a correct diagnostic suggestion, a third with an incorrect suggestion, and a third without a suggestion.

RESULTS Overall, the instruction to utilise a combined reasoning approach resulted in greater diagnostic accuracy relative to leaving students to their own intuitions regarding how best to approach new cases. The effect was particularly pronounced when cases were made challenging by biasing participants towards an incorrect diagnosis, either through mention of a specific feature or by making an inaccurate diagnostic suggestion.

DISCUSSION These studies advance a growing body of evidence suggesting that various diagnostic strategies identified in the literature on clinical reasoning are not mutually exclusive and that trainees can benefit from explicit guidance regarding the value of both analytic and non-analytic reasoning tendencies.

KEYWORDS teaching/*methods; *clinical competence; education, medical, undergraduate/*methods; psychology/*education; cardiology/*education; *diagnosis.

*Medical Education 2007: **41**: 1152–1158*
doi:10.1111/j.1365-2923.2007.02923.x

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2]Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada
[3]Department of Medicine, University of Toronto, Toronto, Ontario, Canada
[4]Department of Psychology, McMaster University, Hamilton, Ontario, Canada

Correspondence: Kevin W Eva PhD, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, MDCL 3522, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada. Tel: 00 1 905 525 9140 (ext 27241); Fax: 00 1 905 572 7099; E-mail: evakw@mcmaster.ca

INTRODUCTION

Since the publication of Custers *et al.*'s review[1] of the clinical reasoning literature a decade ago, there has been increasing recognition that, firstly, many of the models of clinical reasoning present in the medical education literature are not mutually exclusive and, secondly, that there is value for the diagnostician in not getting locked into any single mode of thinking about clinical cases.[1,2] To a large extent, the debate that has taken place in medical education research is reflective of many broader literatures aimed at understanding how humans make decisions.

One approach to studying and/or understanding judgement and decision making, commonly applied in economics, is based upon a model of rationality,

Overview

What is already known on this subject

When teaching electrocardiogram diagnosis there is pedagogical benefit in encouraging students to utilise analytic and non-analytic reasoning strategies rather than treating either class of strategies solely as a source of detrimental bias.

What this study adds

Maintaining an optimal balance between non-analytic and analytic reasoning strategies is fragile for students. Giving them explicit instructions to utilise a combined approach to reasoning helps them to overcome the difficulty provided by mention of a counter-indicative feature or an incorrect diagnosis, both of which are legitimate instructional techniques.

Suggestions for further research

Further research should examine these phenomena within non-visual diagnostic domains and across varying levels of training and experience.

the notion being that the utility of a decision can be derived from the product of the value gained from each particular outcome and the probability with which the outcome will occur. From this analytical perspective, it is an irrational act to buy a lottery ticket because, although the pay-out to the winner may be very large, the probability of winning is so small that the resultant utility (the product of the 2 values) is less than the cost of the ticket.[3] In medicine, severity and probability often interact in the same manner, so that low-probability diagnoses are intentionally considered because the cost of missing a serious diagnosis would be significant.

From this perspective, bias is defined as a consistent deviation from rationality. Psychologists Amos Tversky and Daniel Kahneman, among others, provided great insight into human judgement by systematically identifying and experimentally exploring biases that are prevalent in our decision making, work which resulted in Kahneman receiving the Nobel Prize in Economics in 2002. One such bias they identified is the 'availability heuristic', which refers to the tendency we have to base judgements on vivid or salient outcomes (i.e. those that are readily 'available' to us when imagining what might happen and which, thereby, seem more probable) rather than on a more rational, utility-driven approach.[4] To follow up on the above example, many continue to buy lottery tickets despite the low utility of such actions because the vividness and excitement generated through imagining what it would be like to win contributes to an over-estimation of the likelihood that one might win. Importantly, these sorts of mental short-cuts often (if not always) influence our decisions without our being aware that the heuristic is being used, thus leading many to argue that we can improve medical decision making (or decision making more generally) and reduce diagnostic errors by educating individuals about the ways in which their decision making can be biased.[5,6] Alternative, but complementary, strategies for achieving the same goal in medicine include teaching individuals to become better Bayesians by explicitly directing their attention to the need to learn (and consider) the probability of any particular diagnosis (and how that probability is conditional upon the presence or absence of variable signs and symptoms)[7] or by provision of explicit diagnostic algorithms as guidance for how to complete a diagnostic search.[8]

An alternative view of decision making derives from the argument that these biases exist with such prevalence and robustness because they are useful and, therefore, not something we should ever hope to abolish from our decision making. Gerd Gigerenzer, among others, has noted that many of these heuristics and their associated biases are adaptive; they may leave us susceptible to error in some cases, but, given our complicated world, they help us more than they hurt.[9] We might have the illusion that the use of heuristics is suboptimal, but that is because their use is noticeable only in cases where the heuristic leads to inaccurate decisions relative to what might be deemed optimal by rational standards. By analogy, in medical diagnosis, the use of pattern recognition is often noticed, and labelled as problematic, when it has led to an incorrect diagnosis based on superficial similarity. A more complete consideration, however, requires one to acknowledge that pattern recognition strategies not only often yield the same decision as analytic decision-making tendencies, but they sometimes result in better outcomes.[10] When a classic case of pneumonia presents in a doctor's office, the features that could be carefully collected and converted into a diagnosis of pneumonia will share a great deal of overlap with those of previously seen cases of pneumonia, so assigning a diagnosis on either basis will yield the same conclusion. Of real benefit, however, is that the concurrent reference to prior cases can overcome some of the limitations of the traditionally recommended careful, analytic inspection of features. That is routinely considering the features in a very careful, analytic manner may avoid the dangers of premature closure, but it can also be detrimental. Norman *et al.* have shown that prompting diagnosticians to be particularly careful in their feature calls can yield a list of features that is irreconcilable with any particular diagnosis, possibly because considering features individually rather than in context leads to 'overcalling' normal variants (i.e. leads to the diagnostician being misled by features that would not have been distracting otherwise).[11] It is arguable that, were that not the case, these sorts of non-analytic decision-making tendencies would be much more readily (and, in all likelihood, naturally) extinguishable.[12]

Taken together, these different views of our decision-making tendencies generate the impression that balance between multiple decision-making tendencies is necessary to enable a robust system that will maximise diagnosticians' accuracy. In a recent series of studies, we strived to test whether or not there is pedagogical value in instructing novice diagnosticians to avoid getting locked into any single mode of reasoning. Absolute novices (undergraduate psychology students) were given instruction in electrocardiogram (ECG) diagnosis. After a training phase, we found that groups of participants given advice to trust feelings of familiarity when making their diagnoses but to also carefully consider the features that are present on ECGs (i.e. those who were instructed to use a 'combined approach' to clinical reasoning) revealed greater diagnostic accuracy than groups of participants who were: (a) given no advice regarding how to work through new cases;[13] (b) were told to be analytic by adopting a 'feature-first' approach to diagnosis in which all features should be listed before generating a diagnosis; or (c) were told to be non-analytic by instructing them simply to diagnose based on familiarity-driven first impressions of each new ECG.[14] The combined instruction is not intended (or expected) to have prompted a particular reasoning process that could be used as a 'how-to guide' for diagnosticians, but, rather, was simply aimed at helping trainees to benefit from the suggestion in the literature on clinical reasoning that there are multiple mechanisms through which one can accurately generate a diagnosis.

In the current paper we report on the findings from a pair of additional studies aimed at testing the robustness of this phenomenon. As alluded to above, there are many ways that the diagnostic process can go wrong. If he or she relies too heavily on non-analytic processes (i.e. the feeling of having seen it all before), the diagnostician may be overly susceptible to missing key features that are indicative of alternative diagnoses, but which do not fit the overall pattern of the case. If, by contrast, the diagnostician relies too heavily on analytic processes (i.e. being particularly systematic and 'objective'), his or her perception of the cohesiveness of the case may be undermined, thus resulting in excessive weighting of unusual or distracting features. One could easily expect that novices, who have just learned a set of diagnostic rules, might be especially vulnerable to this overly analytic approach. Here we report a pair of studies in which we tried to make the case presentations particularly difficult by biasing participants' consideration of the cases away from the correct diagnosis either through mention of a feature consistent with an incorrect diagnosis (Study 1) or through mention of a plausible but incorrect diagnostic hypothesis (Study 2). These types of experimental manipulations are consistent with the uncommon but important sources of misinformation a clinician may obtain, such as a referring doctor's incorrect diagnosis or an erroneous computer-generated list of ECG features. In Study 1, the aim of the feature manipulation was to induce people towards placing greater reliance on their analytic tendencies (i.e. careful consideration of the features presented in the case), whereas the aim of the diagnostic hypothesis manipulation in Study 2 was to induce people towards greater reliance on their non-analytic tendencies (i.e. greater influence of their Gestalt impressions). Prior research by our group has shown the latter manipulation to have a strong influence on the diagnostic process.[15,16]

METHODS

General methodology

Undergraduate psychology students were recruited from McMaster University to ensure that participants had no previous experience with ECG diagnosis. They were given bonus credits in their introductory psychology class for participation.

Both studies involved 3 phases (training, practice and test) that took place in one-to-one teaching environments with a research assistant. Only the experimental manipulation during the test phase varied between the 2 studies. During the training phase, the researcher taught participants the basics regarding ECG diagnosis (i.e. the names of the 12 leads and the labels assigned to a normal waveform). They were then taught the key features of 10 diagnostic categories through presentation of a written feature list and a series of examples. The diagnostic categories were:

1 normal;
2 right ventricular hypertrophy;
3 left ventricular hypertrophy with strain;
4 left bundle branch block;
5 right bundle branch block;
6 acute anterior myocardial infarction;
7 acute inferior myocardial infarction;
8 ischaemia;
9 pericarditis, and
10 hyperkalaemia.

Within each category 4 examples were presented. During presentation of the first 2 examples, the researcher pointed to and described the features that made that particular ECG a member of the category. During the presentation of the second 2 examples, the participant was asked to do the same. This phase of the experiment required 45–60 minutes to complete per participant.

After studying all 10 categories in this way, participants entered a practice phase during which they were presented with a booklet of 10 ECGs, all of which had been seen during training, but which were randomly sorted without diagnostic labels. At the start of the practice phase, participants were randomly assigned to either a 'combined instruction' group or a 'no instruction' group. Those in the combined instruction group were encouraged to use a combination of analytic and non-analytic reasoning strategies when confronted with each new ECG through the following instructions:

> 'New ECGs often look like ECGs that have been seen before (i.e. during training). Trust this sense of familiarity, but realise that basing decisions solely on similarity can lead to diagnostic errors. So, don't "jump the gun". Consider the feature list before providing a final diagnosis.'

The other half of the participants were not given instructions that would promote any particular reasoning process. Instead, they were told:

> 'Assign a diagnosis to each ECG using whatever strategy comes naturally to you. Feel free to use the feature list, but approach each case using whatever strategy seems most appropriate to help you reach the correct diagnosis.'

All participants in both groups were allowed to view the ECG, the feature list and the generic waveform image from the training materials when assigning their diagnosis. They were given feedback after interpreting each ECG, consisting of the correct diagnosis and an indication of which features determined the correct diagnosis.

Finally, during the test phase, participants were presented with an additional booklet of ECGs. All participants were provided with the same instructions and feature list as during the practice phase, but no feedback was given regarding the accuracy of performance during the test. The biasing interventions were implemented during the test phase. No time limits were imposed, but all participants completed the entire study in 90–120 minutes.

Study-specific methodology

During the test phase of Study 1, participants were presented with 1 of 2 possible test booklets, each consisting of the same 16 ECGs. In 1 booklet the ECGs were presented along with a distracting feature. With each presentation, participants were asked to consider (before trying to diagnose the ECG) whether or not the feature named was present on the recording. Participants were unaware that the feature presented was indicative of an incorrect diagnosis in all cases. However, in all cases it was somewhat unclear whether or not the feature was actually present. The other half of the participants were shown the same set of 16 ECGs without an interfering feature included. They were simply asked to provide a diagnosis for each ECG using the instructions they had been given in the practice phase of the study. In other words, both instruction and bias were crossed, between-subject, manipulations. The number of correct diagnoses (out of 16) was calculated for each participant and a 2-way ANOVA was performed on this accuracy variable with instruction and bias as between-subject variables. Posthoc *t*-tests were used to resolve the source of a significant interaction.

The test phase of Study 2 was similar to that of Study 1 in that the biasing manipulation was implemented during the test. In Study 2, however, the bias condition was incorporated as a within-subjects manipulation such that each participant was presented with ECGs in each bias condition. Three test booklets were used, each containing the same 18 ECGs, but differing in the information presented along with the recording. Six of the ECGs were presented with no additional information, 6 were presented with the correct diagnosis, and 6 were presented with an incorrect diagnosis. To avoid the influence of prestige as much as possible, participants were told that the diagnostic suggestion had been raised by another psychology student at an equal level of training. For example, 1 ECG was presented with the statement: 'This ECG was shown to a group of psychology students and 1 of the diagnoses raised was ischaemia.' Participants were then asked to diagnose the ECG using the instructions they had been given in the practice phase of the study. Adding 2 ECGs to the set used in Study 1 enabled us to add a third bias condition while ensuring that each ECG was presented equally often in each bias condition by counter-balancing across participants. The number of correct diagnoses (out of 6) within each bias condition was calculated for each participant and a mixed design ANOVA using instruction (combined versus none) as a between-subjects factor, and bias condition as a within-subjects factor was performed on this accuracy variable. We also used *t*-tests as planned comparisons, given that we did not expect a difference across instruction within the bias-correct condition.

RESULTS

Study 1

Sixty students participated in this experiment; 30 received combined instructions and 30 received no reasoning instructions. Fifteen from each group were given ECGs with an interfering feature and 15 were not. ANOVA revealed a main effect of instruction; the group provided with the combined reasoning instruction outperformed those given no instruction on how to diagnose new ECGs ($F[1,56] = 9.06$, $P < 0.01$). The main effect of bias did not reach significance, but was in the expected direction ($F[1,56] = 1.87$, $P < 0.2$). The interaction between instruction and bias indicated that the impact of a negatively biasing feature presentation was greater in the no-instruction condition (effect size = 1.02) than in the combined reasoning condition (effect size = 0.32) ($F[1,56] = 3.7$, $P < 0.07$). *t*-tests further indicated that the accuracy rates achieved when a biasing feature was presented were lower than when no biasing feature was presented in the no-instruction condition ($P < 0.06$), but that the accuracy rates did not differ in the combined reasoning condition ($P > 0.6$). The mean scores for each condition are illustrated in Table 1. All groups performed markedly better than chance, which would result in a score of 1.6/16 given that there were 10 possible response categories.

Study 2

A total of 48 students participated in this experiment; 24 received combined instructions and 24 received no reasoning instructions. All participants were presented with 6 ECGs in each of the bias conditions (correct diagnosis, incorrect diagnosis, no bias). ANOVA revealed that none of the main effects of instruction, bias or the bias × instruction interactions were statistically significant. Planned comparisons revealed, however, that the key comparison (combined versus no-instruction within the bias-incorrect condition) revealed a statistically significant difference ($F[1,46] = 4.2$, $P < 0.05$). Similarly, although the group given combined instructions did not reveal a statistically significant effect of bias ($F[2,46] = 1.4$, $P > 0.25$), there was a significant difference between the bias-incorrect condition and the bias-correct condition in the no-instruction group ($F[1,46] = 4.3$, $P < 0.05$). The mean scores for each condition are illustrated in Table 2. Again, all groups performed markedly better than chance, which would result in a score of 0.6/6 given that there were 10 possible response categories.

Table 1 Mean diagnostic accuracy (out of 16) and standard error as a function of reasoning instruction and presence or absence of a biasing feature presentation

	Interfering feature	*No interference*	*Average*	*P*	*Effect size*
Combined reasoning instructions	8.13 (0.45)	7.80 (0.55)	7.97	> 0.6	0.17
No reasoning instructions	5.13 (0.83)	7.13 (0.54)	6.13	< 0.06	0.70
Average	6.63	7.47	7.05	< 0.2	
P	< 0.01	> 0.3	< 0.01		
Effect size	1.02	0.32			

Table 2 Mean diagnostic accuracy (out of 6) and standard error as a function of reasoning instruction and presence or absence of a biasing diagnostic suggestion

	Incorrect diagnosis presented	*No diagnosis presented*	*Correct diagnosis presented*	*Average*	*P*
Combined reasoning instructions	3.58 (0.25)	3.08 (0.28)	3.67 (0.27)	3.44	NS
No reasoning instructions	2.83 (0.27)	3.21 (0.32)	3.50 (0.24)	3.18	Incorrect < correct ($P < 0.05$)
Average	3.21	3.15	3.58	3.31	> 0.15
P	< 0.05	> 0.75	> 0.65	> 0.3	
Effect size	0.59	0.09	0.13		

NS = not significant

DISCUSSION

This pair of studies provides further empirical support for the notion that explicitly telling novice diagnosticians to utilise multiple forms of reasoning (in this case, familiarity-driven pattern recognition combined with careful consideration of the presenting features) can result in improved diagnostic accuracy. Despite the intuitive nature of statements such as 'no single tool will be right for every job', a series of studies, including these, has now been conducted that suggest students do not spontaneously adopt this approach to clinical reasoning as effectively as they might.[11,13,14] In other words, there appears to be pedagogical benefit in explicitly guiding students to reason in a way that probably sounds as if it should occur naturally. Although nobody can control (or even say with certainty) what reasoning processes are operational when a diagnostician works through a case, there is every reason to believe that participants in the no-instruction condition would have been influenced by similarity when making their diagnostic decisions[13,17] and it is inconceivable that they would not have utilised the feature list to help scan each ECG for features. However, participants in that group did not perform as successfully as participants who had been explicitly told to trust feelings of familiarity and to be careful not to overlook presenting features.

Unique to this study are findings that suggest just how frail this balance between analytic and non-analytic reasoning strategies can be and that instruction to adopt a combined reasoning approach can help maintain this balance when faced with challenging cases. Biasing participants either towards excessive reliance on feature-level information (as might occur if one were overly analytic) or towards excessive reliance on diagnosis-level information (as might occur if one were overly non-analytic) had the impact of reducing diagnostic accuracy differentially across groups of participants. In both studies, those randomly assigned to receive a combined reasoning instruction showed no decline in performance when they were intentionally misled by the researcher towards an incorrect response, by contrast with those who were not given any instruction about how to reason through these stimuli. Study 2 is the first study in 6 not to have shown an overall main effect of combined reasoning instructions compared with other reasoning instructions,[11,13,14] a fact that leads us to suspect that performance in the no-bias condition for this group was simply anomalously (and erroneously) low.

Limitations of this programme of research include the fact that it is intentionally focused narrowly on knowledge-based components of the diagnostic process in an aim to better understand how knowledge is organised in memory and how it can best be utilised in practice. We intentionally use visual materials (ECGs) in order to study the balance between analytic and non-analytic processing. There are certainly other aspects of practice that define competence for doctors, including the related, but under-studied, aspect of making treatment decisions, and broader tendencies that are usually described as being non-cognitive (e.g. communication skills or the inclination towards re-investing one's extra mental resources into continuing professional or personal development).[18] In addition, it remains to be seen whether or not the effects reported in this paper will generalise to other medical training contexts with actual medical trainees. Absolute novices were selected for this research to ensure naïveté with respect to ECG diagnosis, thus maximising the control we had over prior relevant experiences and minimising the mental database of previously seen ECGs on which participants could draw. Whether or not these strategies yield similar (and longterm) benefits for medical students at intermediate or senior levels of training or practice, the current findings should be construed as providing benefit in that they can help to set learning off on the right foot, at least in visual domains such as those provided by ECG diagnosis.

As we mentioned earlier, we stress that we do not interpret these findings to mean that we have discovered the process through which to maximise diagnostic success. On the contrary, among the principles guiding this work is the notion that diagnostic ability is somewhat amorphous and flexible.[2] Better diagnosticians are likely to be those who can overcome case-specific weaknesses through the recruitment of additional diagnostic strategies (either consciously or unconsciously) as it is unlikely that any single diagnostic pathway will be capable of resolving every patient presentation a doctor is likely to encounter. Better diagnosticians may also be those who are less likely to be satisfied (again, either consciously or unconsciously) with diagnoses that are derived from a single diagnostic pathway because these findings and the work that precedes them suggest that increased reliance on any single diagnostic strategy is likely to harm performance. Further work is required, however, to tease apart the relationships between these phenomena and ability. For now, we treat these findings as simply indicative of the value that can be derived from providing explicit guidance regarding appropriate ways to solve diagnostic problems without limiting that guidance to any single particular strategy.

Contributors: KWE and LRB conceived and designed Study 1. All 4 authors conceived and designed Study 2. LRB oversaw data collection for Study 1. KWE oversaw data collection for Study 2. KWE performed all data analyses and drafted the paper. RMH, VRL and LRB then made critical revisions. All authors approved the final manuscript.
Acknowledgements: the authors would like to thank Tavinder Ark, Lindsay Cameron, Samantha Johnson and Jessie McCready for their assistance with data collection.
Funding: none.
Conflicts of interest: none.
Ethical approval: this study was approved by the Hamilton Health Sciences Research Ethics Board.

REFERENCES

1 Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: a review. *Acad Med* 1996;**71**:24–6.
2 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39**:98–106.
3 Rosenthal J. *Struck by Lightning: the Curious World of Probabilities*. Toronto: HarperCollins 2005.
4 Tversky A, Kahneman D. Judgement under uncertainty: heuristics and biases. *Science* 1974;**185**:1124–31.
5 Klein JG. Five pitfalls in decisions about diagnosis and prescribing. *BMJ* 2005;**330**:781–4.
6 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimise them. *Acad Med* 2003;**78**:775–80.
7 Elieson SW, Papa FJ. The effects of various knowledge formats on diagnostic performance. *Acad Med* 1994;**69** (Suppl):81–3.
8 Mandin H, Jones A, Woloschuk W, Harasym P. Helping students learn to think like experts when solving clinical problems. *Acad Med* 1997;**72**:173–9.
9 Gigerenzer G, Todd P, ABC Research Group. *Simple Heuristics that make us Smart*. New York: Oxford University Press 1999.
10 Gladwell M. *Blink: the Power of Thinking without Thinking*. New York: Little Brown & Co. 2005.
11 Norman GR, Brooks LR, Colle CL, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cognit Instr* 2000;**17**:433–48.
12 Eva KW, Norman GR. Heuristics and biases – a biased perspective on clinical reasoning. *Med Educ* 2005;**39**:870–2.
13 Ark TK, Brooks LR, Eva KW. The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Med Educ* 2007;**41**:281–7.
14 Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med* 2006;**81**:405–9.
15 Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med* 1999;**14**:126–9.
16 LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002;**77** (Suppl):67–9.
17 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ* 1997;**2**:173–84.
18 Mylopolous M, Regehr G. Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Med Educ* 2007;**41**:1159–65.

Received 1 May 2007; editorial comments to authors 1 August 2007; accepted for publication 9 September 2007

Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education

Maria Mylopoulos & Glenn Regehr

CONTEXT Many approaches to the study of expertise in medical education have their roots most strongly established in the traditional cognitive psychology literature. As such, they take a common approach to the construction of expertise and frame their questions in a common way. This paper reflects on a few of the paradigmatic assumptions that have 'come along for the ride' with the traditional cognitive approach, and explores what might have been left out as a consequence.

METHODS We examine the operational definition of 'expert' as it has evolved using the traditional cognitive paradigm and we explore some alternative definitions and constructions of expert performance that have arisen in parallel education research paradigms. We address 3 inter-related aspects of expertise as manifested in the traditional cognitive approach: the construction of the expert as a (routine) diagnostician; the construction of the developmental process as the (automatic and unreflective) accrual of resources through experience, and the construction of accrued knowledge as a relatively static resource that is subsequently used and built upon with further experience.

CONCLUSIONS We hope that, by highlighting these issues, we may begin to marry the strengths of the traditional cognitive paradigm with the strengths of these other paradigms and expand the scope of cognitive research in medical expertise.

KEYWORDS *clinical competence; *cognition; *diagnosis; *education, medical; *research; knowledge.

Wilson Centre for Research in Education, Toronto, Ontario, Canada

Correspondence: Maria Mylopoulos, Wilson Centre for Research in Education, 1E565-200 Elizabeth Street, Toronto, Ontario M5G 2C4, Canada. Tel: 00 1 416 340 5086; Fax: 00 1 416 340 3792; E-mail: maria.mylopoulos@utoronto.ca

*Medical Education 2007: **41**: 1159–1165*
doi:10.1111/j.1365-2923.2007.02912.x

INTRODUCTION

The study of medical expertise has been a focus for medical education researchers for several decades. The purported value of this research enterprise appears self-evident. If we can understand the nature of expert performance and its development in individual practitioners, we will be able to structure our training programmes towards more efficient and effective development of experts and will also be able to present information to practising doctors in ways that would enable them to use their medical expertise more effectively. As the articles in this theme issue attest, this research has been informed by a variety of frameworks and approaches. However, also as evidenced by these articles, 1 of the more dominant frameworks has borrowed from the theories and methods of cognitive psychology. Against the backdrop of this traditional cognitive paradigm, efforts to understand the development and maintenance of expert doctor performance have tended to emphasise the phenomenon of clinical reasoning, and more specifically, the resources that doctors use to diagnose disease. In particular, researchers have sought to understand what it is in the expert doctor's mental processing that distinguishes expert diagnostic performance from that of a novice.

Through this lens of inquiry, a variety of interesting (often counter-intuitive) phenomena have been identified, which have shaped our understanding of expert performance in medicine. Many of these findings and phenomena, and the associated metaphors for what might therefore go on in expert mental processing, are presented in the articles contained in this issue, so we will not enumerate

Overview

What is already known on this subject

Cognitively based research in medical expertise has led to interesting and valuable inquiries regarding the nature and use of knowledge among experienced practitioners.

What this study adds

This study examines an operational definition of expertise as it has evolved using a traditional, cognitively based research approach, and explores alternative definitions and constructions of expert performance from parallel education research paradigms.

Suggestions for further research

Future research might include experimental designs incorporating alternative definitions and constructions of expert performance into our evaluation and testing of expertise. Adding this knowledge to existing understanding of ways in which experts make use of their knowledge in diagnostic reasoning tasks would enrich the cognitive medical expertise research paradigm.

them here. The overarching interpretation of these findings, however, might best be summarised with 4 important conclusions. Firstly, expertise takes years to attain and is acquired through deliberate practice in a particular domain.[1] Secondly, with this extensive experience and practice comes a set of impressively rich and well organised resources and processes with which experts are able to effectively and efficiently solve routine problems of practice. The various constructions of these resources have included prototypes,[2] scripts,[3] encapsulated concepts,[4] instances,[5] semantic networks,[6] semantic axes[7] and probability matrices.[8] Constructions of expert processes have been variously described as heuristics,[9] reasoning strategies,[10] restricted searches[11] and pattern recognition.[12] By virtue of this range of descriptions and constructions, a third general conclusion must be that the use of these resources by experts is remarkably flexible, not only for the effective diagnosis of various disease presentations, but also for the effective completion of the variety of tasks that we as researchers ask them to perform in the context of our research studies (such as generating probability matrices[8] or making relative similarity judgements about a set of concepts printed on cards[6]). As Custers *et al.* have suggested, it appears that whatever unnatural task we as researchers can think up for experts to do, they seem to be able to adapt their cognitive resources sufficiently to do it better than novices.[13] Finally, it is clearly apparent from reading this literature that the specific nature of each expert's individual set of resources and processes is remarkably idiosyncratic,[14–17] probably because of the idiosyncratic nature of each expert's personal experience and practice.

More recently, in an effort to capture the dynamic nature of medical expertise, cognitively based research has focused not on what resources are contained in an expert's mental processing, but, rather, on how these resources are used in concert during daily practice. As a manifestation of this effort, some researchers have focused on the co-ordination of analytic and non-analytic resources as a function of task.[18,19] One of the most systematic programmes of research in this area is summarised in the article by Eva *et al.* in this issue. As a second manifestation of this question of resource co-ordination, research groups have begun to explore this phenomenon more in terms of a shifting reliance on the automatic and deliberative use of resources from moment to moment in daily practice[20] in a self-regulatory process referred to variously as 'slowing down when you should'[21] and 'knowing when to look it up'.[22]

Although it is clear that the traditional cognitively based research paradigm in medical expertise has led to interesting and valuable inquiries regarding the nature and use of knowledge among experienced practitioners, it is important to recognise that with any paradigm come certain approaches and assumptions that narrow the focus of the research. Of course, this is the strength of a paradigmatic approach. However, if used unquestioningly, this is also a paradigm's weakness. For example, there is always the possibility that the nature of expertise has been presupposed by the methodologies that we use to understand it. This paper, therefore, will reflect on a few of the paradigmatic assumptions that have 'come along for the ride' with our traditional cognitive approach, and will explore what might have been left out as a consequence. In particular, we wish to examine the operational definition of 'expert' as it has evolved through this research. We will then explore some alternative definitions and constructions of expert performance that have arisen in parallel education research paradigms.[23] We will address 3 inter-related aspects of expertise as manifested in the traditional cognitive paradigm:

1 the construction of the expert as a (routine) diagnostician;
2 the construction of the developmental process as the (automatic and unreflective) accrual of resources through experience, and
3 the construction of accrued knowledge as a relatively static resource in the expert's mental processing that is subsequently used and built upon with further experience.

We hope that, by highlighting these issues, we may begin to marry the strengths of the traditional cognitive paradigm with the strengths of these other paradigms and expand the scope of cognitive research in medical expertise.

THE NATURE OF EXPERTISE IN DAILY PRACTICE

Although it is probably generally acknowledged and relatively uncontroversial that expertise, in its broadest sense, is a multi-faceted construct, the operational definition of expert performance in the context of medical education research tends to be functionally related to more routine diagnostic activities. Recognising that medical experts use their mastery of the domain to accurately diagnose patients, studies on expertise in the traditional cognitive framework have focused on the ways in which experts and novices differ in their efforts to solve diagnostic problems in which experts are usually quite accurate. Even in studies where participants are asked to engage in tasks well outside the usual activities of daily practice (such as card-sorting tasks or the generation of probability matrices), these studies tend to focus on the outputs of the task (as metaphorical windows on experts' versus novices' knowledge structures) rather than on the creative processes involved in completing them. As a result, research within this paradigm has tended to constrain experts to working well below their limits of ability and to using their knowledge in a highly constrained manner.

This approach to studying the nature of expertise contrasts interestingly with studies in other paradigms where experts are pushed to extend themselves by working 'at the edges of their competence'. In such situations, it has been found that only some 'experts' go beyond routine competencies and display flexible, innovative abilities within their domain in a process of 'extending their knowledge rather than applying it'.[24] Such findings have led researchers in these other domains to draw important distinctions between 'adaptive expertise' and 'routine expertise'[24,25] or between 'experts' and 'experienced non-experts'. [26] Routine experts (or experienced non-experts) are highly skilled technicians within their domain. They have learned complex and sophisticated sets of routines and apply them effectively and efficiently in their practice. However, when faced with a novel problem, they will tend to continue to use their existing routines, trying to adapt the problem to the solutions they are comfortable with rather than adapting their solutions to the novel problem. Both because of, and as a result of this approach to practice, additional learning tends to focus on improving efficiency by refining specific aspects of established routines. By contrast, adaptive experts will use a new problem as a 'point for departure and exploration'.[24] They consistently seek problems and challenges that stretch the boundaries of their knowledge and competency. New problems are seen as opportunities to 'explore and expand their current levels of expertise'.[24] Thus, adaptive experts are characterised by their 'flexible, innovative and creative competencies within the domain rather than in terms of speed, accuracy and automaticity of solving familiar problems'.[25] They don't 'attempt to do the same things more efficiently; they attempt to do them better'.[24]

The identification of distinct types of expertise as described in these literatures leads to a potentially troubling conclusion with regard to the cognitive paradigm. The traditional cognitive research emphasis on identifying differences in performance across levels of experience has overlooked important distinctions between types of expertise among individuals with similar levels of experience. Such distinctions, largely obscured by our cognitively based programmes of research, may have important implications for our understanding of excellence and our construction of educational programmes intended to achieve it. This issue will be further elaborated in the next section.

THE DEVELOPMENT OF EXPERTISE

As with the definition of expertise itself, there may be an important assumption implicit in how we, as researchers into the cognitive base of medical expertise, operationalise the process of developing expertise. That is, our cognitive studies, which tend to examine how novices differ from experienced practitioners in performance, are generally grounded in the operational assumption that (with the odd exception) most novices eventually become experts. It is presumed that with more experience comes the accrual of a greater (or better) resource base on which to rely, suggesting that expertise is an automatic and inevitable consequence of experience. The focus therefore, is on what has been acquired through that experience and the educational consequences are couched in questions about how to get these resources into novices faster or more efficiently.

Again, however, the literatures that draw distinctions between routine and adaptive expertise raise concerns regarding this construction of expert development. Consistent with the idea that there are different kinds of expertise (not just different levels), researchers in this paradigm have proposed that routine and adaptive expertise have distinct developmental pathways [24,27,28] that are distinguished by one's approach to daily practice. Unlike routine expertise, adaptive expertise is not merely the accrual of resources and skills through experience and practice. In fact, it has been argued that, as individuals acquire the knowledge and experience to solve typical problems, they establish routines that may work against their further growth.[29] The mark of adaptive expert development and the form of learning and practice associated with it, therefore, involves an inherent understanding of the assumptive nature of these routines. Adaptive experts continue to grow only because of their intentional engagement in 'progressive problem solving', that is, the continual reinvestment of cognitive resources into creating not merely better performance, but, in fact, better understanding of the problems of their domain.[26] The development of adaptive expertise, therefore, is not a simply a process of acquiring knowledge and skills in a domain, but, rather, it is an active process of challenging and thereby transforming one's knowledge and skills in a domain.

An important consequence of adopting the adaptive expert developmental pathway is the development of the ability not only to master the knowledge of a particular domain but also to make innovative contributions to the domain through a process of knowledge building that commences from the beginning of one's training.[30] Within the traditional cognitive paradigm, 1 approach to understanding the place of innovation in any domain has been to examine 'eminent achievement',[1] which supposes that certain individuals display a combination of knowledge (acquired through deliberate practice) and natural ability that makes them 'grand experts' capable of extraordinary performances and contributions to their field. However, such an approach rarifies these individuals and renders them less interesting to educators. Researchers outside the traditional cognitive paradigm have, instead, focused on trying to understand the types of contexts and experiences that lead to the development of adaptive experts, capable of building and creating new knowledge in their fields. This approach characterises adaptive expertise as an acquired approach to practice rather than as an innate ability and seeks to provide learning environments that foster the development of the competencies underlying adaptive expertise.[28,31,32] In this construction then, the practice of expertise is a dynamic and ongoing process. It refers to not only the development and use of a repository of knowledge in the intellect, but also the way by which the expert creates and uses knowledge in the world.

Of course, the research approach described above moves us well outside the traditional cognitive paradigm, and it is not our intention to suggest that this is the 'right' way to study adaptive expertise. Nonetheless, if we are to take this construction of adaptive expertise seriously, we cannot expect that the underlying structures and processes of routine expertise will be predictive of adaptive expertise, nor that diagnostic excellence on routine problems of practice can inform us about the processes underlying adaptive expertise. To date, as a result of the emphasis on expert–novice differences, it is likely that the traditional cognitively based studies of medical expertise have revealed more about routine expertise than adaptive expertise and have led to educational models for equipping our novices with more intellectual resources faster. Approaches to understanding the nature of adaptive expertise might be valuably incorporated into our designs and research questions.

THE CONSTRUCTION OF KNOWLEDGE RESOURCES

A third implicit assumption built into the traditional cognitive paradigm of medical expertise research involves the nature of the accrued resources themselves. That is, asking about the components of an expert's mental processing tends to lead to a treatment of the knowledge gained from previous experience as something that is acquired, stored and called upon to address future problems of practice. Consistent with a 'folk psychology' metaphor of the mind as a container, our research paradigm has come to understand expert performance as the outcome of experts applying their intellectual knowledge to their daily work. Because the knowledge is implicit in the work that is being performed, watching experts perform at particular time-points in our studies was thought to reveal the knowledge they possessed.

While this may be a reasonable assumption, it carries within it the possibility of treating knowledge solely as a stable and therefore relatively static resource that experts are able to call upon in their daily practice. This, in turn, can lead to a construction of experience as primarily adding to knowledge (or at best incrementally refining it) rather than transforming knowledge. Among the more salient realisations of this potential concern is the construction of our studies as having 'learning phases' followed by 'test phases'; this is explicitly enshrined in the title of a cognitive psychology paper[33] that has formed the basis for several subsequent studies on the expert co-ordination of analytic and non-analytic resources in the medical education literature:[18,19] 'After the learning is over: factors controlling the selective application of general and particular knowledge.'

This slippery slope to treating experience, once accrued, as a relatively static resource in memory has some interesting implications. For example, it implies that with experience, experts know more than, but do not know differently, from novices. In medical education, a manifestation of this has been a response to the recognition of context specificity in clinical reasoning. A sensible educational strategy that arises from the 'more resources' construction of expertise would be to ensure that novices get as many diverse experiences of each disease as possible in as many contexts as possible as quickly as possible in order to build an wide database of knowledge, enabling them to cope analogously with the future variance and vagaries of their anticipated clinical environment.[34] We do not have to look far to see conceptual inconsistencies with this approach: for example, Bordage[35] showed that in the early stages of learning it is often fewer, rather than more, examples that lead to better learning. Needham and Begg [36] demonstrated that it is possible to teach for broader analogical transfer if participants are encouraged to treat the problems meaningfully as problems to be solved rather than as just examples to be learned.

Perhaps more importantly from an epistemological perspective, the treatment of expert knowledge as a set of stable, previously stored resources that are simply used and added to, probably under-represents the extent to which adaptive experts treat their own knowledge and understanding as a 'conceptual artefact' that can be articulated, shared, critiqued and iteratively improved.[29] In the hands of an adaptive expert, knowledge is not a static resource, and expertise is not the culmination of possessing as much domain-specific knowledge as possible. Instead, knowledge is seen as a constantly evolving, dynamic resource, and expertise resides in the ability and willingness to not only to use and build, but also to purposefully adapt and re-engineer knowledge effectively. As a result, adaptive experts view their practice not just as a means to generate desirable outcomes, but as its own form of 'knowledge in the world'.[29]

Adopting an adaptive expert approach to the construction of knowledge does not deny the value of extensive experience as a resource for effective performance. It is undeniable that the amount of knowledge that an expert possesses is crucial to the enactment of expertise. However, narrowing the construction of expertise to the resources accrued in an expert's intellect limits our conception of what an expert can do, affects the ways in which we evaluate and teach future experts and may even impact the way that experienced practitioners themselves believe they should perform.[24] By contrast, if we find ways to incorporate the construction of adaptive expertise in our paradigm, rather than encouraging our students to solve new cases solely through recourse to past experience (admittedly an important and valuable strategy), we might additionally focus on the ways in which they can use new cases to change their understanding and construction of that past experience. We are not suggesting that expert practitioners should constantly engage in the process of building knowledge and evolving their practice, but a conception of expertise that explicitly excludes these skills has consequences for the medical profession. By definition, adaptive experts never believe the learning is over and act accordingly when appropriate.

DISCUSSION

Within the traditional cognitive paradigm, expertise has been defined broadly as the mastery of existing knowledge and techniques in a given domain.[37] In medical education research, questions about how this mastery is most effectively attained, and what is involved in the expert's intellectual resources by the time it is, have produced some interesting and important findings that have shaped the way we think about experts. However, implicit in this definition is the possibility of constructing expertise as an end state of 'complete' knowledge that evolves, at best, only as the existing knowledge and techniques of the domain evolve. Perhaps it is sensible to consider routine expertise as the accrual of resources that enable the rapid and uncomplicated solution to typical problems. However, we would argue that adaptive expertise is not a state of accomplishment, but rather is best thought of as an *approach to practice*, an ongoing process of continual reinvestment of cognitive resources in an effort to transform practice and extend the boundaries of knowledge and technique iteratively. Again, this is not to deny the crucial role of experience and knowledge in defining the clinical expert. But it does imply that adaptive expertise is not a developmental stage beyond routine expertise. Rather, it is a set of habits that must be acquired and continuously enacted from the beginning of training. As Aldous Huxley wrote: 'Experience is not what happens to a man; it is what a man does with what happens to him.'[38]

To an important degree, much of the cognitively based research on expertise in medical education has implicitly adopted the more restricted definition of expertise as routine expertise in the 'controlled' experimental designs we have created to study it and the outcome measures we have used to test it. In our decisions to compare levels of experience and equate these with levels of expertise, in our decisions to remove participants from their usual contexts (violating our own tenets of context specificity), in our decisions about the nature of the tasks we ask them to perform, and in our decision to uncover the expert's intellectual resources, we have run the risk of narrowing the definition of expertise to performance of the mundane.

Although such a narrow definition may not be inappropriate in the purely theoretical laboratories of psychology departments, in an applied domain such as medical education, where researchers and educators live side by side, it is inevitable that our constructions of expertise will have a strong impact on pedagogical models for medical education. The ways in which we operationalise expert knowledge and behaviour become the benchmarks for our training and practice. Thus, it is particularly important for us to ensure that our methodologies and constructions of expertise properly reflect the competencies we want to foster in our experts. Studies of expertise that exclusively examine the performance of routine expertise in diagnostic problems therefore run the risk of leading us to aim uncomfortably low in our curricular objectives.

What would this type of inclusion mean for our research questions and methodologies? One possible suggestion is that we might develop experimental designs that incorporate the distinction between routine and adaptive expertise into the selection of participants and into the tasks we use to evaluate and test expertise. This would move us away from diagnostic reasoning tasks that assess expert–novice differences and towards researching participants both across and within levels of experience on tasks that elicit more than routine diagnostic problem solving. Adding this knowledge to our existing understanding of the ways in which experts make use of their knowledge on diagnostic reasoning tasks would enrich the cognitive medical expertise research paradigm. In this way, a more comprehensive programme of research on medical expertise might develop and inform our medical education programmes, thereby helping to produce better and more adaptive expert practitioners rather than simply producing routine experts more efficiently and effectively.

Contributors: both authors made substantial contributions to the conception and design of this paper. MM drafted the paper and GR revised it critically for important intellectual content. MM had final approval on the manuscript to be published.
Acknowledgements: none.
Funding: MM is supported by a post-doctoral award co-sponsored by the Canadian Health Services Research Foundation (CHSRF) and the Canadian Institutes of Health Research (CIHR). GR is supported as the Richard and Elizabeth Currie Chair in Health Professions Education Research.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Ericsson KA, Krampe RT, Teschromer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 1993;**100**:363–406.
2 Bordage GRZ, Nov ME. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ* 1984;**18** (6):406–416.
3 Custers E, Boshuizen HPA, Schmidt HG. The role of illness scripts in the development of medical diagnostic

expertise: results from an interview study. *Cognit Instruc* 1998;**16**:367–398.
4 Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognit Sci* 1992;**16**:153–184.
5 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract* 1997;**2**:173–184.
6 McGaghie WC, Boerger RL, McCrimmon DR, Ravitch MM. Learning pulmonary physiology: comparison of student and faculty knowledge structures. *Acad Med* 1996;**71 (Suppl)**:13–15.
7 Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med* 1991;**66 (Suppl)**:70–72.
8 Papa FJ, Elieson B. Diagnostic accuracy as a function of case prototypicality. *Acad Med* 1993;**68 (Suppl)**:58–60.
9 Detmer DE, Fryback DG, Gassner K. Heuristics and biases in medical decision making. *J Med Educ* 1978;**53**:682–683.
10 Patel VL, Evans DA, Kaufman DR. Reasoning strategies and the use of biomedical knowledge by medical students. *Med Educ* 1990;**24**:129–136.
11 Wiley J. Expertise as a mental set: the effects of domain knowledge in creative problem solving. *Mem Cognit* 1998;**26**:716–730.
12 Norman GR. Problem-solving skills, solving problems and problem-based learning. *Med Educ* 1988;**22**:279–286.
13 Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: a review. *Acad Med* 1996;**71 (Suppl)**:55–61.
14 Norman GR, Brooks LR, Allen SW, Rosenthal D. Sources of observer variation in dermatologic diagnosis. *Acad Med* 1990;**65 (Suppl)**:19–20.
15 McGaghie WC, Boerger RL, McCrimmon DR, Ravitch MM. Agreement among medical experts about the structure of concepts in pulmonary physiology. *Acad Med* 1994;**69 (Suppl)**:78–80.
16 Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Acad Med* 1996;**71**:988–1001.
17 Grant J, Marsden P. The structure of memorised knowledge in students and clinicians: an explanation for diagnostic expertise. *Med Educ* 1987;**21**:92–98.
18 Kulatunga-Moruzi C, Brooks LR, Norman GR. Co-ordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med* 2001;**13**:110–116.
19 Regehr G, Cline J, Norman GR, Brooks L. Effect of processing strategy on diagnostic skill in dermatology. *Acad Med* 1994;**69 (Suppl)**:34–36.
20 Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;**80 (Suppl)**:46–54.
21 Moulton CE, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgement. *Acad Med* 2007 Oct;**82 (10 Suppl)**:S109–S116.
22 Eva KW, Regehr G. Knowing when to look it up: a new conception of self-assessment ability. *Acad Med* 2007 Oct;**82 (10 Suppl)**:S81–S84.
23 Alexander PA. Can we get there from here? *Educ Res* 2003;**32**:3–4.
24 Bransford JD, Brown AL, Cocking RR. *How People Learn.* Washington, DC: National Academy Press 2000;.
25 Hatano G, Inagaki K. Two courses of expertise. In: Stevenson H, Azuma H, Hakuta K, eds. *Child Development and Education in Japan.* New York: W H Freeman 1986;27–36.
26 Bereiter C, Scardamalia M. *Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise.* La Salle, IL: Open Court Publishing Company 1993.
27 Bereiter C, Scardamalia M. Learning to Work Creatively with Knowledge. In: De Corte E, Verschaffi L, Entwistle N, van Merriënboer J, eds. *Powerful Learning Environments: Unravelling Basic Components and Dimensions (Advances in Learning and Instruction Series).* Oxford: Elsevier Science 2003;55–68.
28 Hatano G, Oura Y. Commentary: reconceptualising school learning using insight from expertise research. *Educ Res* 2003;**32**:26–29.
29 Bereiter C. *Education and Mind in the Knowledge Age.* Mahwah, NJ: Lawrence Erlbaum Associates 2002.
30 Scardamalia M, Bereiter C. Knowledge Building. *Encyclopedia of Education.* New York: Macmillan Reference 2003;1370–1373.
31 Scadamalia M, Bereiter C. Computer support for knowledge-building communities. *J Learn Sci* 1994;**3**:265–283.
32 Scardamalia M, Bereiter C. Knowledge Building Environments: Extending the Limits of the Possible in Education and Knowledge Work. In: Distefano A, Rudestam KE, Silverman RJ, eds. *Encyclopedia of Distributed Learning.* Thousand Oaks, CA: Sage Publications 2003;269–272.
33 Whittlesea BWA, Brooks LR, Westcott C. After the learning is over: factors controlling the selective application of general and particular knowledge. *J Exp Psychol Learn Mem Cogn* 1994;**20**:259–274.
34 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39**:98–106.
35 Bordage G. The curriculum overloaded and too general. *Med Educ* 1987;**21**:183–188.
36 Needham DR, Begg IM. Problem-oriented training promotes spontaneous analogical transfer: memory-oriented training promotes memory for training. *Mem Cognit* 1991;**19**:543–557.
37 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79 (Suppl)**:70–81.
38 Huxley A. *Text & Pretexts: an anthology with commentaries.* Westport, CT: Greenwood Press 1976.

Received 30 April 2007; editorial comments to authors 1 August 2007; accepted for publication 9 September 2007

Clinical case processing: a diagnostic versus a management focus

Alireza Monajemi,[1] Remy M J P Rikers[2] & Henk G Schmidt[3]

CONTEXT Most studies on medical expertise research have focused on diagnostic performance, whereas patient management has been largely ignored. According to knowledge encapsulation theory, applying encapsulated knowledge is a characteristic of expert doctors' diagnostic reasoning, but it is unclear whether or not encapsulated knowledge also plays a prominent role when processing a clinical case with a management focus.

METHODS The participants were 40 medical students (20 in Year 4 and 20 in Year 6) and 20 expert doctors (internists). Participants were asked to study the cases with either a diagnostic (Dx) or a management (Mx) focus. Subsequently, participants were asked to write down what they remembered from the case.

RESULTS In both conditions, experts recalled fewer propositions and used more high-level inferences than medical students. Furthermore, they processed the cases faster and more accurately than medical students, but no significant difference between Mx and Dx conditions was found. Year 4 students also showed no significant differences in recall and processing speed between conditions. By contrast, Year 6 students recalled more in a Dx than in an Mx condition, but there was no significant difference in processing speed between conditions.

CONCLUSIONS In both conditions, findings indicate that the experts' and Year 4 students' performance was not affected by processing focus. The fact that only Year 6 students were affected by processing focus might be explained by the assumption that their diagnostic knowledge and management knowledge are not fully integrated yet, a process that has already taken place in the expert's knowledge structure.

[1]Applied Physiology Research Centre, Isfahan University of Medical Sciences, Isfahan, Iran
[2]Department of Psychology, Faculty of Social Sciences, Erasmus University, Rotterdam, The Netherlands
[3]Department of Psychology, Erasmus University, Rotterdam, The Netherlands

Correspondence: Alireza Monajemi, Applied Physiology Research Centre, Isfahan University of Medical Sciences, Hezar Jarib Avenue, Isfahan, Iran. Tel/Fax: 00 98 311 792 22 95;
E-mail: Alirezamonajemi@yahoo.com

KEYWORDS *clinical competence; education, medical, undergraduate/*methods; internship and residency/*methods; students, medical/*psychology; physicians/*psychology/standards; mental recall; *diagnosis; patient care management.

Medical Education 2007: ***41****: 1166–1172*
doi:10.1111/j.1365-2923.2007.02922.x

INTRODUCTION

In order to investigate differences in knowledge organisation between expertise levels, many studies on the development of medical expertise have used the clinical case paradigm. In these studies, participants with different levels of expertise were requested to study a clinical case description, to provide a diagnosis and, finally, to recall everything they could remember from the text. Although this paradigm has provided us with a better understanding of the differences between students' and doctors' case representations, it has almost exclusively focused on diagnostic performance, largely ignoring the important aspect of patient management.[1–3] That is, while processing case information, participants only had to work out what a patient's problem was. The recall that followed therefore reflected the information that was considered important with a diagnostic focus. However, it is not clear whether or not a change in this focus will result in a different evaluation of the findings in a case description. That is, it is unclear if a management focus (Mx) while processing

Overview

What is already known on this subject

Research on medical expertise has shown that encapsulated knowledge plays a major role in doctors' diagnostic reasoning, but its role in the processing of clinical cases with a management focus is unclear.

What this study adds

Our study investigated whether or not a management focus leads to a different case representation than a diagnostic focus. The results demonstrated that doctors processed the cases in an encapsulated mode in both conditions.

Suggestions for further research

Further work is needed to investigate the role of knowledge encapsulation in the management focus and its relationship with the diagnostic focus during the course of expertise development.

case information will lead to a different appreciation of the findings than a diagnostic focus (Dx).[4] For instance, suppose a patient with a history of peptic ulcer complains about symptoms (e.g. ankle pain) completely unrelated to her previous problem. There is no need to know about the patient's concomitant diseases and drug history in order to diagnose this new complaint. However, from a management viewpoint, this information is highly relevant, as prescribing non-steroidal anti-inflammatory drugs (e.g. Ibuprofen or Naproxen) to relieve the ankle pain might worsen the symptoms of peptic ulcer. In other words, what is important in a Dx condition does not necessarily overlap with what is important in an Mx condition and vice versa.

Most studies with a diagnostic focus have shown that experts process case information faster and more accurately than medical students.[5–9] However, these studies also showed that advanced students outperformed both novices (e.g. Year 1 students) and medical experts in recalling case information (i.e. they exhibited an intermediate effect).[5,8,10] This consistent finding was surprising because in other areas of expertise research (e.g. chess) there is a linear relationship between expertise level and performance.[11] Schmidt and Boshuizen suggested that the explanation for these intermediate effects lies in the qualitatively different knowledge that students and expert doctors bring to bear on a diagnostic task.[8,9] Medical students mainly use their extended biomedical knowledge to explain case data, leading to elaborate and detailed case processing. By contrast, biomedical knowledge only plays a minor and implicit role in experts' clinical reasoning. According to Schmidt and Boshuizen, the experts' biomedical knowledge has become fully integrated with their clinical knowledge as a result of repeated exposure to large numbers of real patients.[8,9] The integration of both types of knowledge has been dubbed 'knowledge encapsulation' and leads to a more holistic approach towards case processing, which focuses mainly on signs and symptoms. Consequently, their recall is shorter and contains more encapsulated concepts than that of students.

The present study investigates the role of encapsulated knowledge in the shift from a Dx to an Mx condition. As outlined above, the results of processing clinical case information in an Mx condition do not necessarily concur with those made in a Dx condition. Moreover, the development of diagnostic knowledge is often not completely synchronised with that of management knowledge. In the first years of their training (i.e. the pre-clinical phase), students acquire knowledge largely from textbooks and lectures without any real patient encounters. There is a strong emphasis on providing the correct diagnosis, which is often not accompanied by an equivalent emphasis on developing a patient management plan.[12] Training in management knowledge often starts considerably later (during the clinical phase) than that in diagnostic knowledge. Consequently, less advanced students, confronted with a management task, will most likely deal with it as if it were a diagnostic task (which is the only mode of processing a case they have some experience with). By contrast, more advanced students might already have acquired some clinical experience in hospitals during their clinical rotations and hence management will have become a bit more concrete. However, unlike experienced doctors, these advanced students are still in the middle of the process of linking their diagnostic knowledge with their newly acquired management knowledge.

In this study, internists, Year 6 students and Year 4 students were asked to study cases in either a Dx or an Mx condition, and subsequently to recall the provided information. According to the theory of knowledge encapsulation,[8,9,13] the following predictions were made. Internists construct their clinical case representations similarly under both conditions because their Dx and Mx knowledge has become well integrated over the years and therefore no differences are expected in terms of recall and speed. Students, by contrast, are expected in both conditions to be less accurate, slower and more elaborate in their recall than internists. Furthermore, Year 4 students and Year 6 students are expected to show similar results in the Dx condition concerning recall and processing speed. However, Year 6 students, as a result of more clinical experience, might show a better diagnostic performance than Year 4 students. Finally, in the Mx condition, Year 6 students will be more elaborate in their recall and will process the case information more slowly than Year 4 students. That is, as a result of a lack of relevant knowledge, Year 4 students will treat a case with an Mx focus similarly to a case with a Dx focus. Year 6 students who have already acquired some Mx knowledge will be able to differentiate between both focuses, but are not yet proficient enough to deal with a case in an Mx focus efficiently, and hence will have to go through the case information very thoroughly, leading to more recall and slower processing times.

METHODS

Participants

Forty medical students (20 in Year 4 and 20 in Year 6) from Isfahan School of Medicine and 20 internists from 5 hospitals in Isfahan (Iran) participated. It takes about 7 years (4 pre-clinical years and 3 clinical years) to complete the curriculum at Isfahan School of Medicine. Year 4 (pre-clinical) medical students had no or very limited experience in hospital, and their clinical knowledge was sourced from textbooks and lectures, whereas Year 6 students were in the process of hospital training as interns under the supervision of senior residents and attending doctors. The internists were practitioners with an MD degree and at least 2 years of experience.

Materials

The materials consisted of a booklet containing an instruction about the procedure, 4 written descriptions of clinical cases and 2 blank response sheets after each case for recall.

The cases were identical to cases used by Patel and Groen,[14] Rikers *et al.*[6,7] and Verkoeijen *et al.*[13] The cases involved acute bacterial endocarditis (1 case),[14] heart failure (2 cases)[6,7] and Addison's disease with tuberculosis (1 case).[13] The 4 case descriptions were each about 1 page in length and consisted of 76, 82, 105 and 107 propositions, respectively. The order of cases was randomised for each participant and the same sets of cases were used for all expertise levels.

Procedure

Participants were randomly assigned to the Mx condition or the Dx condition. In order to familiarise the participants with the procedure, they were first given the opportunity to read an unrelated text of about same length as the case texts. In the Dx condition, participants were told that they had maximally 5 minutes to diagnose the case. Subsequently, they were instructed to write down everything they could recall from the case and to write down their diagnosis. In the Mx condition, participants also read the case, but this time they were instructed to devise a management plan for the patient in a maximum of 5 minutes. After they had studied the case in an Mx condition, they were also asked to write down what they remembered from the case. However, in order to maximise the contrast between both conditions, participants were not asked to provide a diagnosis after the case. The time spent reading a case was registered using a chronometer by the researcher. Processing time was recorded from the time a participant started to read the case to the point at which he or she finished reading it. Participants were informed that they could proceed to the next task whenever they were ready. However, if they did not finish within 5 minutes, they were instructed to go to the next task. During the test, the cases were presented sequentially and were studied individually. After each case had been diagnosed or managed, the next case was handed out to the participant.

Analysis

The correct diagnoses associated with the 4 cases were divided into different diagnostic elements. In line with previous studies, the accuracy of diagnosis was determined by weighting each element according to its relative importance.[6–10] The resulting diagnostic accuracy score ranged from 0 (completely inaccurate) to 3 (completely accurate). For example, if in the case of acute bacterial endocarditis the diagnosis contained the key concept 'endocarditis', 1.5 points were given. The presence of 'sepsis/septicaemia', I.V.

drug abuse', or 'infection' each contributed 0.5 points and the maximum score was 3.

The free-recall protocols were scored according to a propositional analysis method introduced by Patel and Groen.[14] A clinical case can be segmented into several meaningful information units or propositions. Each proposition consists of 2 concepts connected by a qualifier, such as *causation*, *negation*, *specification* or *temporal information*. For instance, the text fragment 'A 45-year-old man complains about nausea and vomiting for 3 weeks' consists of 4 propositions:

1 patient *specification* (man, 45 years old);
2 complaints *specification* (nausea);
3 complaints *specification* (vomiting), and
4 complaints *temporal information* (3 weeks).

Evidence for encapsulation of case data was explored by counting the number of high-level inferences in the recall protocols. High-level inferences were considered as encapsulated concepts if they could be matched to more than 1 proposition in the protocols. For example, if a case contains, among other things, the following information about a patient: *fatigue*, *abdominal pain* and *pigmentation*, a potential high-level inference might be *adrenal insufficiency*.

For each participant the data from each focus were collapsed to obtain a mean reading time, diagnostic accuracy, free recall and high-level inferences. The data were analysed using a 3 (expertise level) × 2 (case focus) analysis of variance with expertise level and case focus as between-subject factors. The least significant difference test was used to make posthoc comparisons between the different expertise groups. Significance was set at $P < 0.05$ for all tests.

RESULTS

Diagnostic accuracy

Table 1 depicts the mean diagnostic accuracy as a function of expertise level in the Dx condition (note that participants in the Mx condition did not have to provide a diagnosis). The accuracy of the diagnoses is associated with level of expertise ($F[2,27] = 9.08$, standard error of the mean [SEM] = 3.26, $P < 0.001$, $\eta^2 = 0.40$). Pairwise comparison showed that experts provided significantly more accurate diagnoses than Year 4 and Year 6 students. There was no significant difference in accuracy between Year 6 and Year 4 students.

Table 1 Mean diagnostic accuracy and standard errors (SE) as a function of levels of expertise in diagnostic (Dx) focus

Levels of expertise	*Dx*
Year 4 students	1.50 (0.17)
Year 6 students	1.62 (0.14)
Internists	2.30 (0.16)

Processing time

Table 2 depicts the mean processing time as a function of expertise level and case focus. Analysis of variance indicated that the time each participant spent reading the cases was associated with expertise level ($F[2,54] = 20.53$, SEM = 2222.70, $P < 0.05$, $\eta^2 = 0.43$). There was no main effect of focus ($F[1,54] = 0.01$, SEM = 2222.70, $P = 0.91$, $\eta^2 = 0$) and no interaction ($F[2,54] = 0.05$, SEM = 2222.70, $P = 0.94$, $\eta^2 = 0.02$). Pairwise comparisons indicated that experts were significantly faster than both Year 4 and Year 6 students. Moreover, mean processing time did not differ between Year 4 and Year 6 students. No significant difference in processing time was observed between the Dx and Mx conditions in medical students or in experts.

Table 2 Mean processing time in seconds and standard errors (SE) as a function of levels of expertise and case focus

Levels of expertise	*Dx*	*Mx*
Year 4 students	266.80 (9.78)	270.97 (8.19)
Year 6 students	254.02 (14.95)	249.35 (14.40)
Internists	180.70 (18.58)	177.07 (19.85)

Dx = diagnostic focus; Mx = management focus

Free recall

Table 3 depicts the mean number of propositions recalled as a function of expertise level and case focus. Analysis showed a main effect of expertise level ($F[2,54] = 9.716$, SEM = 149.59, $P < 0.05$, $\eta^2 = 0.27$), a marginally significant effect of focus ($F[1,54] = 3.57$, SEM = 149.59, $P = 0.06$, $\eta^2 = 0.06$), and a significant interaction ($F[2,54] = 3.30$, SEM = 149.59, $P < 0.05$, $\eta^2 = 0.11$).

Pairwise comparisons within each focus showed that in the Mx condition, internists and Year 6 students differed significantly from Year 4 students, but not from each other. In the Dx condition, Year 4

Table 3 Mean propositions recalled and high-level inferences and their standard errors (SE) as a function of levels of expertise and case focus

Levels of expertise	*Dx*		*Mx*	
	Recall	*High-level inferences*	*Recall*	*High-level inferences*
Year 4 students	44.67 (3.60)	2.5 (0.70)	49.12 (2.50)	3.6 (1.33)
Year 6 students	51.67 (3.89)	2.8 (0.82)	36.32 (3.06)	4.10 (0.9)
Internists	34.40 (5.34)	5.8 (1.38)	27.40 (4.16)	7.40 (2.1)

Dx = diagnostic focus; Mx = management focus

students and Year 6 students did not show significant differences, but both groups did produce significantly more propositions than the experts. Finally, the Year 6 students' recall was significantly higher in the Dx condition than in the Mx condition.

Further, the presence of high-level inferences in the recall protocols was also investigated because they are considered to represent evidence for encapsulated knowledge. Table 3 also depicts the mean number of high-level inferences as a function of expertise level and focus. There was a main effect of expertise level ($F[2,54] = 4.37$, SEM = 17.32, $P < 0.05$, $\eta^2 = 0.14$), but there was no main effect of focus ($F[1,54] = 1.55$, SEM = 17.32, $P > 0.05$, $\eta^2 = 0.03$), and no interaction ($F[2,54] = 0.02$, SEM = 17.32, $P > 0.05$, $\eta^2 = 0.03$). In addition, pairwise comparisons between expertise levels revealed that experts produced significantly more high-level inferences than Year 6 and Year 4 students. There was no significant difference between Year 6 and Year 4 students. Furthermore, there was no significant difference between Dx and Mx conditions within expertise levels.

In order to correct recall for differences in time spent studying the cases, we calculated the mean number of propositions recalled per second (i.e. the total number of propositions recalled divided by the processing time) as a function of expertise level and focus (Table 4). This measure gives us a purer indication of what is remembered from the case description. Analysis showed no main effect of expertise level ($F[2,54] = 0.53$, SEM = 0.00, $P > 0.05$, $\eta^2 = 0.02$), focus ($F[1,54] = 0.1.19$, SEM = 0.00, $P > 0.05$, $\eta^2 = 0.02$), nor a significant interaction ($F[2,54] = 1.15$, SEM = 0.00, $P > 0.05$, $\eta^2 = 0.04$). Pairwise comparison showed a significant difference between Dx and Mx conditions only for Year 6 students.

Table 4 Mean number of propositions recalled per second and standard errors (SE) as a function of levels of expertise and case focus

Levels of expertise	*Dx*	*Mx*
Year 4 students	0.17 (0.01)	0.18 (0.01)
Year 6 students	0.21 (0.02)	0.15 (0.01)
Internists	0.21 (0.03)	0.19 (0.04)

Dx = diagnostic focus; Mx = management focus

DISCUSSION

By contrast with previous studies, the present study was concerned with the role of encapsulated knowledge in the management of clinical problems. Based on the view of knowledge encapsulation,[8,9] it was predicted that expert doctors would process the clinical problems in an encapsulated mode regardless of their focus (i.e. Mx or Dx). Furthermore, it was expected that Year 6 students would shift from a less elaborated processing mode when diagnosing a case to a more elaborate mode when managing a case. Year 4 students were expected to be insensitive to the manipulation and to process the cases similarly in both conditions.

The results were largely in line with our assumptions, in that medical experts processed cases faster and provided more accurate diagnoses than students in both the Dx and Mx conditions. Furthermore, the experts' recall was much lower than that of medical students, whereas their mean number of high-level inferences was significantly higher. There was, as predicted, no difference between experts in Mx and Dx. These data show that experts engaged in an encapsulated processing approach independently of their processing focus. The fact that there was no significant difference in the number of propositions recalled per second between experts and medical students (Table 4), shows that differences in recall between expertise levels are not the result of a slower processing speed. These findings are in line with those of previous studies that demonstrated that experts' performance is relatively insensitive to the nature and circumstances of the task.[5–7,13,15,16]

Year 4 students showed, as expected, no significant differences in recall in Dx and Mx. Interestingly, in both Dx and Mx there were no significant differences in speed or diagnostic accuracy between Year 4 and Year 6 medical students. These findings indicate that Year 6 students did not engage in a more elaborate processing mode than Year 4 students in the Mx condition. However, there was a significant difference between both student groups in free recall. That is, Year 6 students remembered *less* than Year 4 students in the Mx condition, which is not in line with our predictions. Moreover, Year 6 students remembered less in the Mx condition than in the Dx condition, which is also not in line with our predictions. As a matter of fact, their recall performance was similar to that of the experts in the Mx condition. However, although Year 6 students provided recall protocols that were quantitatively similar to those of experts in Mx focus, this finding does not imply that their recall was also qualitatively similar. This is substantiated by the finding that Year 6 students generated significantly fewer high-level inferences than experts: Year 6 students' protocols contained 11% high-level inferences and experts contained 27%. The observed decrease in Year 6 students' recall in Mx might be explained by the findings of previous studies on knowledge encapsulation. For instance, Schmidt and Boshuizen[8] asked participants of different levels of expertise to study a case for 30 seconds and then to write down what they remembered from the case. Their results also showed no significant difference in recall between Year 6 students and experts, but, in line with our study, experts generated significantly more high-level inferences. This finding was explained by the assumption that the reduction in processing time meant that advanced students did not have sufficient time to process the case deeply and hence lost their advantage in terms of case recall. Similarly, as a result of insufficient patient management knowledge, Year 6 students in our study were also confronted with a task that interfered with their usual (diagnostic) way of dealing with clinical cases and hence experienced difficulty in building an appropriate and coherent representation.[12,17–22] Consequently, they might have experienced difficulty in reproducing the case information, which resulted in a poor recall performance. So, although Year 6 students have acquired more expertise in management than Year 4 students, this expertise hinders them in this phase because their management knowledge is still not developed enough to be used efficiently.

In sum, our study seems to indicate that, especially for Year 6 students, there is a clear distinction between a Dx and an Mx condition when processing clinical case information because Year 6 students' more recently acquired management knowledge is not yet fully developed and integrated with their diagnostic knowledge. In most medical schools, management knowledge does not seem to play an important role during the pre-clinical years, and the integration of diagnostic and management knowledge therefore mainly starts during the student's clinical years. As a result, the development of management knowledge will lag behind the student's diagnostic competence and will only become fully integrated with diagnostic knowledge after many years of clinical experience.

Contributors: all authors contributed to the conception and design of the study. AM collected the data, contributed to data analysis and interpretation, and drafted the manuscript. RMJPR participated in the data analysis and interpretation, and was involved in drafting the manuscript and its final revision. HGS supervised the whole study and contributed to the analysis and interpretation of data.
Acknowledgements: the authors thank the internists and medical students who volunteered their time for this study.
Funding: none.
Conflicts of interest: none.
Ethical approval: this study was approved by the Isfahan University of Medical Sciences Ethics Committee.

REFERENCES

1 Gruppen LD, Frohna AZ. Clinical reasoning. In: Norman GR, van der Vleuten CPM, Newble DI, eds. *International Handbook of Research in Medical Education.* Dordrecht: Kluwer Academic Publishers 2002; 205–30.
2 McGuire CH. Medical problem solving: a critique of the literature. *J Med Educ* 1985;**60**:587–95.
3 Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;**39**:418–27.
4 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79** (10 Suppl):70–81.
5 de Bruin ABH, Schmidt HG, Rikers RMJP. The role of basic science knowledge and clinical knowledge in diagnostic reasoning: a structural equation modelling approach. *Acad Med* 2005;**80** (8):765–73.

6 Rikers RMJP, Schmidt HG, Boshuizen HPA. Knowledge encapsulation and the intermediate effect. *Contemp Educ Psychol* 2000;**25** (2):150–66.
7 Rikers RMJP, Schmidt HG, Boshuizen HPA. On the constraints of encapsulated knowledge: clinical case representation by medical experts and subexperts. *Cognit Instr* 2002;**20** (1):27–45.
8 Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical case recall. *Mem Cognit* 1993;**21** (3):338–51.
9 Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;**5**:205–21.
10 van de Wiel MWJ, Schmidt HG, Boshuizen HPA. A failure to reproduce the intermediate effect in clinical case recall. *Acad Med* 1998;**73**:894–900.
11 Vicente KJ, Wang JH. An ecological theory of expertise effects in memory recall. *Psychol Rev* 1998;**105** (1):33–57.
12 Brater DC, Nierenberg DW. Medical students' education in clinical pharmacology and therapeutics. *Ann Intern Med* 1988;**108**:136–7.
13 Verkoeijen PP, Rikers RMJP, Schmidt HG, van de Wiel MW, Kooman JP. Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 2004;**38** (6):617–27.
14 Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cognit Sci* 1986;**10**:91–116.
15 Rikers RMJP, Schmidt HG, Boshuizen HPA, Linssen GC, Wesseling G, Paas FG. The robustness of medical expertise: clinical case processing by medical experts and subexperts. *Am J Psychol* 2002;**115** (4):609–29.
16 Rikers RMJP, Loyens SM, Schmidt HG. The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Med Educ* 2004;**38** (10):1035–43.
17 Kwan CY. Learning of medical pharmacology via innovation: a personal experience at McMaster and in Asia. *Acta Pharmacol Sin* 2004;**25** (9):1186–94.
18 Maxwell S. Teaching safe and effective prescribing in the medical curriculum. *Med Educ* 2003;**37** (9):839–40.
19 Prince KJAH, van de Wiel MW, Scherpbier AJJA, van der Vleuten CPM, Boshuizen HPA. A qualitative analysis of the transition from theory to practice in undergraduate training in a PBL-medical school. *Adv Health Sci Educ* 2000;**5**:105–16.
20 Pearson SA, Rolfe I, Smith T. Factors influencing prescribing: an intern's perspective. *Med Educ* 2002;**36** (8):781–7.
21 Pearson S, Smith AJ, Rolfe IE, Moulds RF, Shenfield GM. Intern prescribing for common clinical conditions. *Adv Health Sci Educ Theory Pract* 2000;**5** (2):141–50.
22 Custers EJFM, Boshuizen HPA. The psychology of learning. In: Norman GR, van der Vleuten CPM, Newble DI, eds. *International Handbook of Research in Medical Education*. Dordrecht: Kluwer Academic Publishers 2001;163–203.

Received 3 May 2007; editorial comments to authors 1 August 2007; accepted for publication 7 September 2007

Science is fundamental: the role of biomedical knowledge in clinical reasoning

Nicole N Woods

CONTEXT Although training in basic science is generally considered a critical aspect of medical education, there is little consensus regarding its precise role in clinical reasoning. Whereas some reports suggest that biomedical knowledge is rarely used in routine diagnosis, other research has found that biomedical knowledge can become an integral part of the expert knowledge base.

OBJECTIVE The purpose of the current paper is to present evidence in support of different views regarding the role of biomedical knowledge, including the two-world hypothesis, encapsulation theory and recent work on the role of biomedical knowledge in novice diagnosticians. The implications of these models for clinical teaching will be examined.

DISCUSSION Recent work suggests that biomedical knowledge can help novices develop a coherent and stable mental representation of disease categories. As a result, learners are able to retain clinical knowledge over time and maintain diagnostic accuracy when faced with clinical challenges. This suggests that clinical teachers should attempt to make explicit connections between biomedical knowledge and clinical facts during training.

KEYWORDS *clinical competence; *decision making; *diagnosis; biological sciences/*education; education, medical, undergraduate/*methods; teaching/*methods.

*Medical Education 2007: **41**: 1173–1177*
doi:10.1111/j.1365-2923.2007.02911.x

Wilson Centre and Department of Surgery, University of Toronto, Toronto, Ontario, Canada

Correspondence: Nicole N Woods, Wilson Centre, 200 Elizabeth Street, ES 1-565, Toronto, Ontario M5G 2C4, Canada. Tel: 00 1 416 340 4803; Fax: 00 1 416 340 3792; E-mail: nikki.woods@utoronto.ca

INTRODUCTION

Clinical reasoning depends on a careful balance of several different types of knowledge, including knowledge of the clinical features of disease, of case-based exemplars and of the biomedical mechanisms that govern the functioning of the human body. Although few would argue the importance of clinical knowledge or that novices should be exposed to a variety of cases, clinical teachers continue to debate the role of biomedical knowledge in routine clinical reasoning. Whereas basic scientists may tout the importance of their discipline, some clinicians see the biomedical sciences as only peripherally relevant to daily practice.

At first glance it is easy to understand both sides. Even a layperson would agree that a trained health professional should have some understanding of anatomy, physiology, genetics, biochemistry and biology. Patients imagine that their doctors routinely consider these fundamental principles. Most people would find something unsettling about the notion that these disciplines might have little impact on everyday medical decisions. By contrast, we are all familiar with situations in which information that is important in theory becomes less so in practice. Much like the content of a high school physics class, the basic science training received in medical school might be quickly forgotten with time and practical experience. According value to basic science in routine clinical reasoning may represent a naïve perception that is not reflective of daily experience. Proponents of such an argument can find ample support in the clinical reasoning literature, citing numerous examples of expert reasoning that is seemingly independent of basic science knowledge.[1–3] Pattern recognition[1] and other forms of non-analytic reasoning[2] can lead to accurate clinical decisions with little-to-no biomedical knowledge. A clinician

Overview

What is already known on this subject

There are conflicting views on the role of biomedical knowledge. It may play a limited role in clinical reasoning or be embedded into the knowledge structure of the expert clinician and activated during diagnosis.

What this study adds

This paper presents an indirect role for biomedical knowledge in diagnosis by novices. It outlines the value of biomedical knowledge in memory and diagnosis under challenging conditions.

Suggestions for further research

Future research examining these effects in the classroom and other aspects of clinical reasoning is needed.

simply does not need to recall the specific mechanism of a disease in order to recognise the similarities between 2 patients, suggesting that there is little need for basic science knowledge to play a substantial role in any model of the nature of medical expertise.

Patel and colleagues once championed such a model in which basic science concepts and clinical knowledge form 2 entirely separate mental representations, with clinical knowledge providing the basis for most expert reasoning and biomedical knowledge serving predominantly as a communication tool.[3] Patel *et al.* sought to support this model through the most direct means possible: a series of studies that simply asked clinicians to think aloud while working through a clinical case. A qualitative analysis of the verbal reports revealed little mention of biomedical concepts.[3,4] Instead, most doctors focused on the analysis and interpretation of clinical features. Only when confronted with a diagnostic challenge did experienced clinicians begin to explicitly rely on biomedical principles.[5]

Combine the empirical findings of Patel *et al.* with the anecdotal reports of practising clinicians and it is tempting to conclude that basic science is of little value to the experienced clinician. However, the doctors in these think-aloud studies and the practising clinician who believes he does not use his basic science knowledge may simply be expressing a type of meta-cognitive bias that we all display. They simply do not recognise (and therefore cannot verbalise) how their knowledge of physiology, biochemistry and the other sciences shapes the way they view, organise and interpret clinical information. This is neither a novel concept nor unique to doctors. As human beings, we are often unaware of cues, assumptions and background knowledge that impact our decisions.[6] However, lack of awareness of the impact of basic science knowledge does not diminish its actual significance. Although the clinical sciences may have the most obvious impact on expert reasoning, basic science may still play a subtle, yet important, role.

It is precisely this type of indirect role for basic science knowledge that forms the basis of Schmidt's encapsulation theory.[7] According to Schmidt, biomedical knowledge and clinical facts becoming increasingly integrated as the clinician gains experience. For the medical expert, basic science concepts become encapsulated under clinical facts in the mental representation of a disease. With time, clinicians can seamlessly recognise a group of clinical facts linked by biomedical knowledge, without needing to overtly describe the underlying pathophysiology. This encapsulation of biomedical knowledge explains why there is little mention of basic science principles or mechanism in explicit recall or reasoning measures. Instead, the impact of biomedical knowledge will be revealed only through indirect means. Using a 'priming' paradigm, a recent series of studies by Rikers found that, unlike medical students, practising clinicians were able to quickly identify biomedical words and phrases related to the correct diagnosis of a previously presented clinical case.[7–9] This suggests that as the clinician worked through the clinical case, both clinical and biomedical knowledge were activated, enabling quick identification of biomedical target words shortly afterward. Critically, this occurred regardless of the doctor's explicit awareness.

It is important to note that, considered in isolation, encapsulation theory tells us that expert mental representation of any disease can include biomedical knowledge embedded in clinical knowledge. Encapsulation theory does not explain whether encapsulated biomedical knowledge is a causal factor in expert performance or simply the by-product of years of clinical experience. Just because biomedical knowledge *can* be encapsulated under clinical knowledge does not mean it *must* be in order for expert clinical performance to be attained. Perhaps clinicians would perform equally well without these types of knowledge structure.

This is a critical question for clinical teachers. It has been assumed since Flexner that basic science education is a critical part of medical training. Yet there is little basis for this assumption. Perhaps the instructional and mental resources devoted to the basic sciences could be better spent by having students gain additional clinical exposure or learn alternate reasoning strategies. For example, there have been several demonstrations that diagnostic accuracy can be improved by helping students consider the alternatives and think about the features of disease in terms of statistical probabilities.[10–12] In fact, students who learn using quantitative aids have been found to show greater accuracy than those who learn the same material with the benefit of verbal descriptors.[13] This would seem to suggest that mathematical probabilities may in fact be a valuable learning tool for the novice diagnostician.

However, recent research has tested the learning efficacy of a type of quantitative aid against that of a biomedical description.[14] Although the 2 learning methods led to equal levels of success immediately after learning, the basic science framework led to greater diagnostic accuracy after a 1-week delay. This suggests that the novice diagnostician can benefit from understanding biomedical knowledge even without a wealth of additional case-based knowledge and that biomedical knowledge can be the basis of a useful learning tool for even rank novices, although the added value may not manifest until it is used long after the original learning.

One possible explanation for this finding comes from cognitive psychology and research into categorisation. In order to categorise something, we must have intimate knowledge of the features of the category as well as some theory regarding the relationship between those features.[15] Critically, the features of any category are rarely random. Instead, they go together for a reason – perhaps an underlying biological or mechanical process. In medicine, each diagnostic category includes a set of key clinical features. For the participants in this study, the basic science text explained the relationship between those features, allowing the students to understand that features of each disease go together for a reason. Once the diagnostic category becomes more than a random assortment of signs and symptoms, students can develop a more coherent mental representation that is easier to retain in the longer term. Hence the biomedical knowledge served as a mnemonic device for learners.

Additional research suggests that training students with the underlying mechanisms does more than just help them remember the material. In another study, undergraduate psychology students were asked to learn a series of artificial diseases.[16] For 1 group of participants, the learning materials included the clinical features as well as simple explanations for how the features were connected. The other group learned the clinical features without the explanations. In a later speeded decision-making task, participants were asked to diagnose a set of cases as quickly as possible. They were then asked to diagnose another set of cases, taking as much time as they needed. The results showed that although students who only learned the features showed a typical, standard speed–accuracy trade-off (they were more error prone on the speeded version of the diagnostic task), students who learned the causal mechanisms did not. In fact, the students trained with causal mechanisms were *more* accurate when asked to move quickly than when told to take their time. This counter-intuitive pattern of performance is similar to a pattern demonstrated in experts performing a well learned skill.[17] This suggests that the causal mechanisms allowed the novices in this study to function more like experts. In another study, participants trained with either isolated features or features with causal explanations were presented with difficult cases that included novel terminology. Despite having the same amount of practice, students with knowledge of underlying mechanisms were better able to make the translation from their learning material to the novel terminology and arrive at the correct response, compared with those who had only studied the clinical features.[18] Like the experts in the encapsulation studies, students with the causal knowledge structure were also able to quickly recognise 'encapsulated' terms presented after studying a related clinical case.[19] Although causal knowledge was not explicitly required to complete the tasks assigned in any of these studies, learners were able to use their additional understanding to their advantage.

Taken together, these findings suggest that understanding the underlying mechanism of disease can create valuable coherence among the clinical features. Immediately after learning, or with very simple cases, students can use their knowledge of clinical features to arrive at the correct diagnosis. However, such a strategy will become less effective as memory decays over time or cases increase in difficulty. As a result, students trained using just clinical features have difficulty diagnosing cases after a time delay or when presented with novel terminology. Providing students with the appropriate theoretical knowledge gives them the means to create a coherent picture of a case when the clinical features become disorganised. This ability to rely on what makes sense rather than a step-by-step analysis of clinical information also seems to move a novice further along the road to expertise.

In order to translate the findings of these laboratory studies into effective educational practice, we must first answer a question that could have begun this entire discussion: what is basic science? The term 'basic science' probably evokes many responses in different people. It may be that a single answer to this question is not possible or even desirable. Clearly, there are many different types of information that can be used to link and explain clinical features. In psychiatry, for example, a complete biological description of a disease may simply not be possible. However, under some circumstances, the nature of the links, the depth of the scientific explanation and even the accuracy of the information seems to have little impact on the value of biomedical knowledge in helping novices retain and use clinical information to diagnose a case. Studies using artificial materials, simple causal chains and incomplete explanations suggest that what matters most is that that the links provided are clear, plausible and stable. A study using undergraduate medical students, for example, demonstrated that providing simple biochemical and pathophysiological explanations for clinical findings was sufficient to provide a stable performance in diagnosis of neurological and rheumatologic diseases.[20] The basic science explanations did not help students understand the science in great depth, but they did give them some understanding of why a particular sign or symptom occurred in a specific disease. It is this type of simple relational knowledge that seems to enhance memory, to improve the ability to diagnose challenging cases and, potentially, to act as a precursor to the encapsulated representation that is considered a hallmark of expertise.

Given these findings, perhaps the most important aspect of the issue at hand concerns how we make sure that students do in fact learn the links and mechanisms that will be of greatest value. The structure of the traditional medical curriculum in which basic sciences courses are taught first, and are followed by clinical training, may simply not be conducive to this type of learning. Unlike the laboratory studies in which participants learned the biomedical and clinical information in an integrated package, the traditional 2-stage model of undergraduate education requires that students first learn the basic science and then spontaneously recognise its relationship to the clinical information they learn 2 years later. However, it has been demonstrated that this type of accurate transfer of biomedical concepts to clinical problems is unlikely to occur, even when the biomedical information is provided only minutes prior to the clinical information.[21] If experienced clinicians do not explicitly and overtly express basic science concepts in their work-up of clinical cases, it seems highly unlikely that clinical clerks or even junior residents will be able to spontaneously see the connections and apply their knowledge correctly. Thus, medical training must be structured so that the relationship between biomedical concepts and clinical facts is made explicit, concise and clear. To this end, a key goal throughout the early stages of medical training should be to integrate clinical information and the supporting biomedical concepts into a coherent package.

Although the studies presented focused on text-based cases in a few medical domains, clinical teachers from many disciplines may be able to use these findings as a simple model for effective educational practice. The success of laboratory studies using very simple materials suggests that causal connections and explanatory links infused into clinical instruction can enhance student learning even when the basic science concepts are not covered in great detail. Rather than requiring elaborate explanations, it seems that students can benefit from a basic understanding of the links between clinical features and the pathways that lead to specific presentations. This could potentially be accomplished by clinical teachers choosing to infuse basic science concepts into traditional lectures or by having basic scientists and clinicians work together to create clinical curricula.

Further study is needed in order to determine how to best achieve this balance without the careful controls of laboratory studies. In the classroom and on the ward, it is likely to be very difficult to ensure that students pay attention to the biomedical knowledge that will support their success. With further research we will probably find that an understanding of biomedical knowledge can drastically change the way a student perceives and interprets clinical data. This may even occur in visual domains, such as dermatology or radiology. The correct application of these findings to the curriculum will also require addressing the role of biomedical knowledge outside of diagnosis. Thus far, we have focused on the impact of basic science during the processing of a clinical case. However, it is likely that biomedical knowledge impacts on other decisions regarding the treatment and management of a case.

Acknowledgements: the author would like to thank Remy Rikers and Geoff Norman for providing the motivation to write this article, and Glenn Regehr for his comments on the manuscript.
Funding: none.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Elstein AS, Schwartz A, Higgs J, Jones M. *Clinical Reasoning in Medicine. Clinical Reasoning in the Health Professions.* Oxford, UK: Butterworth Heinemann 2000:95–106.
2 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract* 1997;**2** (2):173–84.
3 Patel VL, Evans DA, Groen GJ. Reconciling basic science and clinical reasoning. *Teach Learn Med* 1989;**1** (3):116–21.
4 Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cognit Sci* 1986;**10** (1):91–116.
5 Joseph GM, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. *Med Decis Making* 1990;**10** (1):31–46.
6 Nisbett RE, Wilson TD. Telling more than we can know – verbal reports on mental processes. *Psychol Rev* 1977;**84** (3):231–59.
7 Rikers RMJP, Schmidt HG, Moulaert VA. Biomedical knowledge: encapsulated or two worlds apart? *Appl Cogn Psychol* 2005;**19** (2):223–31.
8 Rikers RM, Loyens S, te Winkel WW, Schmidt HG, Sins PH. The role of biomedical knowledge in clinical reasoning: a lexical decision study. *Acad Med* 2005;**80** (10):945–9.
9 Rikers RM, Loyens SM, Schmidt HG. The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Med Educ* 2004;**38** (10):1035–43.
10 Noguchi Y, Matsui K, Imura H, Kiyota M, Fukui T. Quantitative evaluation of the diagnostic thinking process in medical students. *J Gen Intern Med* 2002;**17** (11):839–44.
11 Kurzenhauser S, Hoffrage U. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Med Teach* 2002;**24** (5):516–21.
12 Wolf FM, Gruppen LD, Billi JE. Differential diagnosis and the competing-hypotheses heuristic. A practical approach to judgement under uncertainty and Bayesian probability. *JAMA* 1985;**253** (19):2858–62.
13 Elieson SW, Papa FJ. The effects of various knowledge formats on diagnostic performance. *Acad Med* 1994;**69** (Suppl 10):81–3.
14 Woods NN, Brooks LR, Norman GR. The value of basic science in clinical diagnosis: creating coherence among signs and symptoms. *Med Educ* 2005;**39** (1):107–12.
15 Wisniewski EJ, Medin DL. On the interaction of theory and data in concept-learning. *Cognit Sci* 1994;**18** (2):221–81.
16 Woods NN, Brooks L, Norman GR. It all makes sense: biomedical knowledge, causal connections and memory in the novice diagnostician. *Adv Health Sci Educ Theory Pract.* 2007;**12**:405–15.
17 Beilock SL, Bertenthal BI, McCoy AM, Carr TH. Haste does not always make waste: expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychon Bull Rev* 2004;**11** (2):373–9.
18 Woods NN, Brooks LR, Norman GR. The role of biomedical knowledge in diagnosis of difficult clinical cases. *Adv Health Sci Educ Theory Pract* 2007;**12**:417–26.
19 Woods NN, Howey EH, Brooks LR, Norman GR. Speed kills? Speed, accuracy, encapsulations and causal understanding. *Med Educ* 2006;**40** (10):973–9.
20 Woods NN, Neville AJ, Levinson AJ, Howey EHA, Oczkowski WJ, Norman GR. The value of basic science in clinical diagnosis. *Acad Med.* 2006;**81** (10 Suppl): S124–7.
21 Patel VL, Groen GJ, Scott HM. Biomedical knowledge in explanations of clinical problems by medical students. *Med Educ* 1988;**22** (5):398–406.

Received 11 May 2007; editorial comments to authors 1 August 2007; accepted for publication 7 September 2007

Scripts and clinical reasoning

Bernard Charlin,[1] Henny P A Boshuizen,[2] Eugene J Custers[3] & Paul J Feltovich[4]

CONTEXT Each clinical encounter represents an amazing series of psychological events: perceiving the features of the situation; quickly accessing relevant hypotheses; checking for signs and symptoms that confirm or rule out competing hypotheses, and using related knowledge to guide appropriate investigations and treatment.

OBJECTIVE Script theory, issued from cognitive psychology, provides explanations of how these events are mentally processed. This essay is aimed at clinical teachers who are interested in basic sciences of education. It describes the script concept and how it applies in medicine via the concept of the 'illness script'.

METHODS Script theory asserts that, to give meaning to a new situation in our environment, we use goal-directed knowledge structures adapted to perform tasks efficiently. These integrated networks of prior knowledge lead to expectations, as well as to inferences and actions. Expectations and actions embedded in scripts allow subjects to make predictions about features that may or may not be encountered in a situation, to check these features in order to adequately interpret (classify) the situation, and to act appropriately.

CONCLUSIONS Theory raises questions about how illness scripts develop and are refined with clinical experience. It also provides a framework to assist their acquisition.

KEYWORDS knowledge; education, medical, undergraduate/*methods; teaching/*methods; *clinical competence; review [publication type]; *decision making; *diagnosis.

*Medical Education 2007: **41**: 1178–1184*
doi:10.1111/j.1365-2923.2007.02924.x

INTRODUCTION

An adult patient comes into the outpatient office of a doctor complaining of facial pain and nasal obstruction for 2 days' duration. Instantly, from these 2 signs, knowledge about acute facial pain pops into the clinician's mind, with sinusitis being especially salient because of its frequency of occurrence in this age group. This specific knowledge then orients the questions asked and physical examinations administered. A few minutes later, a new patient comes in with vertigo signs. Instantly, knowledge about sinusitis and facial pain is dismissed from active memory, and knowledge of vertigo takes over. Each clinical encounter engages an amazing assortment of psychological events: perceiving the important features of the situation; quickly accessing relevant hypotheses; checking for signs and symptoms that confirm or rule out these competing hypotheses, and using related knowledge to guide appropriate investigations and treatment. All clinicians experience this mobilisation of organised knowledge relevant to the situations they encounter. Feltovich and Barrows[1] have termed these knowledge structures 'illness scripts', adapting the script concept from cognitive psychology for application in medicine.[2]

This essay is aimed at clinical teachers who are interested in the basic sciences of education. It will provide an overview of scripts, a concept that well explains research data accumulated on clinical

[1]Department of Surgery, University of Montreal, Montreal, Québec, Canada
[2]Open University of the Netherlands, Educational Technology Expertise Centre (OTEC), Heerlen, The Netherlands
[3]Center for Research and Development of Education, Utrecht University, Utrecht, The Netherlands
[4]Florida Institute for Human and Machine Cognition, Pensacola, Florida, USA

Correspondence: Bernard Charlin, CPASS: Direction de la Recherche, Faculté de Médecine-Direction, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada. Tel: 00 1 514 343 6111 (ext 14140); Fax: 00 1 514 343 7650; E-mail: bernard.charlin@umontreal.ca

Overview

What is already known on this subject

'Script' is a cognitive sciences concept that aims at explaining how humans understand real-world events and why this understanding in most cases occurs almost effortlessly.

What this study adds

Scripts-based clinical reasoning is very efficient because script activation is automatic and almost unconscious, activated scripts are used in a conscious and strategic way to confirm or refute corresponding hypotheses, and activated scripts serve to guide information selection, memorisation and interpretation.

Suggestions for further research

The script theory raises questions about how scripts are acquired, and how their development and refinement can be fostered.

reasoning. For clarity, theories will be summarised and few details will be given on the experimental studies that support them. Research data that support script theory will be cited during the theory description.

SCRIPT THEORY

Script theory aims at explaining how humans understand real-world events and why this understanding in most cases occurs almost effortlessly.[3] It assumes that memory functioning involves the use of abstract cognitive structures. Scripts (schemas) arise from repeated experiences with real-world events, as a result of which certain types of information come to be organised in specific ways.[3] For example, activities such as travelling by plane or visiting a restaurant consist of a sequence of events. Having experienced such sequences a number of times, people create knowledge structures that capture the activities within such sequences.

The notion of scripts emerged from attempts to enable computers to understand real events. In addition, psychologists and linguists[2] applied it to the reading process, proposing that understanding implies knowledge structures that represent what a text is about in a general way and provide expectations to enable the reader to quickly interpret complex events within the text and to make predictions about how these situations will develop. Scripts are also applicable to understanding and acting in the real world. A script is about what is normal and what acceptable variations are, and how these variations hang together. It captures what one can expect in a frequently encountered setting, such as having a meal at a restaurant. Once established, the script then allows one to make sense of different restaurant visits and differences among them, ranging from a fast food snack to a banquet in a select restaurant. Such a structured framework allows the 'understander' to deal expeditiously with a variety of otherwise difficult-to-understand situations.[3]

Thus, scripts are goal-directed knowledge structures, adapted to perform tasks efficiently.[4] They contain attributes, each of which corresponds to some aspect of the domain modelled by the script. Each attribute can be symbolically conceived as a slot that can have different values. The restaurant example helps to illustrate the role scripts play in situation interpretation. When an individual enters a restaurant, a 'restaurant script' will be activated which loads into working memory and arouses a number of expectations about what will happen next. Upon perceiving another person approach, holding documents, some slots in the activated restaurant script will be filled: presumably, the person approaching is a waiter or waitress and the 'documents' are menus. Slots include fixed slots representing what is common in such a setting and pertaining to things that are always true (like restaurants always offering food). Other slots correspond to what has a more incidental nature and pertain to looser expectations that might be filled with various particulars (such as any of a number of different types of service or types of food that might be found across diverse restaurants). It is assumed that individuals possess many hundreds of organising and interpretive scripts, and that combinations of these scripts will be invoked in any reasonably complex situation.[3]

SCRIPTS AND CLINICAL TASKS

Diseases have an underlying time-based structure, from onset to subsequent stages of development in a host. When confronted with illnesses, doctors take actions that are related to these sequences (e.g. they look for signs, order tests or prescribe). Scripts are knowledge structures associated with time sequences, that is, with developments, events or actions as they transpire. This makes them particularly well suited to describing clinical disease knowledge.[1,5]

When doctors see patients, they perceive features – symptoms, signs and context from the patient's environment. Perceptions activate illness scripts that interpret information about the characteristics and features of the situation and that include knowledge about the relationships that link those characteristics and features. Those illness scripts will lead them to make inferences, some of which are used to rule hypotheses in or out in the diagnostic process, whereas others are used for patient management. [6] In the diagnostic process, every hypothesis is an activated illness script. If only 1 illness script pops up, this is the – most likely – diagnostic hypothesis. On the basis of the activated illness script, the diagnostician will immediately infer – without having to reason deliberately – which symptoms to expect. Theory predicts that deeper reasoning only occurs if 2 or more illness scripts are simultaneously activated for a single patient, or if there are findings that do not fully fit any particular illness script. Empirical findings suggest that in the latter case doctors sometimes ignore these misfits, which may lead to gross diagnostic errors.[7,8]

Cognitive psychology has provided insights into how scripts function in the diagnostic process.[9] Illness scripts have 'slots' that correspond to attributes associated with the specific disease they describe, with expectations about values that can or cannot be found for each attribute. For each slot, the attribute value that has the greatest probability of occurrence is the default value. The illness script a doctor might have about bacterial maxillary sinusitis would contain slots (e.g. 'predisposing conditions', 'pain location', 'pain duration', 'nasal obstruction') for which different values are possible (for the slot 'predisposing conditions': viral infection, allergic rhinitis or nasal polyposis; for the slot 'pain location': dull sensation of pressure over the maxilla or infra-orbital pain).

Four other characteristics of illness scripts are important. Firstly, the information belonging to a script is not exclusive. Symptoms and signs (unless pathognomonic) can belong to several scripts. The particular script for an illness is characterised by the set of slots regarding the signs and symptoms expected in the course of the illness and by the relationships that link them, along with its predisposing conditions and actions to take in treatment. Secondly, the activation of one script can automatically lead to the activation of another. This can be the effect of shared slots, but alarm links are possible as well, such as between diseases that can be easily confused, or a possible disease that must be treated as another until the latter has been positively disconfirmed.[10] Thirdly, scripts are generic structures that can interpret any instance of an illness. Each medical encounter implies a process of finding the actual values of the attributes observed in the patient (script instantiation).[5,6] This instantiation process also tests if the script that has been invoked is in fact the right script. Fourthly, memory of previous patients is stored in the form of instantiated scripts.[11]

Typical and non-typical patients

An important characteristic of the script concept is default values.[9] Among the acceptable values for each attribute, the most common is assumed to be present until an actual value has been verified (the actual value can of course be identical to the default value). In the sinusitis example it would be viral infection for the slot 'predisposing conditions' and infra-orbital pain for the slot 'pain location'. These default values explain why clinicians do not always look for all signs and symptoms.[6] When they have enough evidence to establish their diagnosis, they often assume that other values are present and do not specifically check them (in the sinusitis example, if a patient has an acute nasal obstruction and pus emanating from the middle meatus, the doctor may not perform percussion over the infra-orbital area, or ask for sinus X-rays). The specific instance of a case in which all slots are occupied by default values represents the prototypical version of the illness. Typical instances are more easily recognised than those that are atypical.[12]

Script activation

How scripts are activated is a key issue. Early hypothesis generation (i.e. quickly finding hypotheses relevant to the situation, with their related networks of knowledge) is an important feature of expert behaviour in medicine: if experts take into consideration the correct diagnosis during the first 5 minutes of a consultation, this hypothesis becomes definite in 95% of cases; if the proper diagnosis has not yet been considered by this timepoint, there is a 95% probability that it will be missed.[13] However, the information available in the earlier part of a clinical encounter is a rather amorphous mix of clinically relevant and irrelevant information, which is available rather than actively collected, and picked up through diverse perceptual pathways. Hobus and colleagues[14] have shown clear differences between experts and novices in this regard. Experienced doctors appeared more able than inexperienced doctors to extract relevant information from such data and to generate fruitful hypotheses. Much of this information pertains to conditions that may contribute to or protect against acquisition of a specific disease, rather than its resulting signs and symptoms. Feltovich and Barrows called this class of script slots 'enabling conditions'.[1]

In many situations, script activation occurs automatically, without conscious awareness. This activation, called non-analytic,[15,16] is based on recognition of either an instance or a pattern. The former mechanism rests on the vast repertoire of previous cases stored as instantiated illness scripts, which experienced clinicians possess; the identification of possible diagnoses occurs by recognition of similar prior examples.[17] In other situations, the configuration of data elements is so familiar that the solution leaps into mind almost instantly. This mechanism is named 'pattern recognition'.[18] Here, instead of prior examples or images, it is a configuration of salient clinical features that activates an illness script and fills the relevant slots.[19]

Non-analytic reasoning probably represents the main method of script activation.[20] For non-routine situations, deliberate script induction occurs.[21] Depending on the situation, involved mental mechanisms may be inductive reasoning, explanation-based reasoning, case-based reasoning, causal biomedical reasoning, analogy, or access to external resources (consultation, electronic databases, textbooks).

Illness script processing and assessment of fit

The set of hypotheses considered by a doctor in a given clinical situation guides the doctor's interview and examination of the patient.[22,23] It represents the initial possibilities that he or she feels need to be pursued. Whether doctors are aware or not, observation of their reasoning shows that the questions they ask and the items of physical examination they perform are, for the most part, specifically chosen to rule in or rule out, or at least strengthen or weaken, the likelihood of the hypotheses they are considering.[23]

The activation of a script provides access to a set of attributes and expectations and allows an active search to find appropriate values for slots.[24] There is no fixed order for checking script attributes. Individual clinicians proceed in different orders. This accounts for the variability in data collection observed among clinicians. Different clinicians rarely use the same set of questions to solve any single clinical problem.[22,25] Experienced doctors ask questions and carry out physical examinations that are most efficient according to their own activated scripts. This processing phase of scripts, the search for evidence, to confirm hypotheses or to rule them out, is controlled and deliberate.[6,24]

Stopping the process of diagnosis at the first hypothesis or script activated, without testing it (further), would be considered risky practice. Doctors are systematically educated to test their hypotheses by an assessment of the fit with collected data. In routine cases, on the basis of the available cues, a single relevant illness script is activated; in non-routine cases, there is a set of competing illness scripts. In both cases, the doctor tries to find if the activated script, or any of the activated scripts, adequately fits the clinical findings.

According to theory, this verification requires that values be assigned to the different attributes. For each attribute slot,[9] there are acceptable and unacceptable values. If unacceptable values are found, the script is rejected (e.g. the maxillary sinusitis script would be rejected if a history of bloody rhinorrhoea were discovered), and other scripts that accept that value are activated or reinforced (e.g. maxillary sinus cancer). Among acceptable values for an attribute, some bring more weight to a hypothesis than others. The diagnostic process aims at decreasing the likelihood of all activated illness scripts except 1. This then becomes the working diagnosis. If the doctor cannot adequately fit an activated script to the findings, he rejects it and begins to verify another.

The assessment of each value in the activated scripts explains the fluid status of the set of hypotheses in clinical encounters. Hypotheses can be reinforced, or be attenuated, or disappear, whereas others are activated.[23] The accumulation of acceptable values within a script raises the level of activation of that script, and at a particular moment the clinician decides that there is enough evidence to bring closure to the diagnostic process. He or she then settles on a definitive or working diagnosis, depending on the situation. Research[26] suggests that referral rate is also affected by enabling conditions and consequences interpretation, not only as an independent effect but also as mediated by age, gender and practice characteristics. Age and experience also affect the process of weighing evidence pro and con a certain script. Less experienced doctors take counter-evidence more seriously than older doctors.[27]

The place of basic science knowledge, the acquisition of scripts

Medical diagnosis, in its contemporary conception, is an explanation of a pattern of symptoms made on the basis of an underpinning biomedical knowledge.[5] Medical curricula devote a great amount of time to the acquisition of biomedical knowledge. Yet research data[28,29] have shown that experts use less basic sciences in their explanations than novices. Schmidt and Boshuizen[11] postulate that their knowledge is encapsulated (i.e. accessible but remaining quiescent until needed) for reasoning, teaching, patient communication, etc.

According to theory, illness scripts develop as students are exposed to real patients. In their first encounters, they apply both biomedical and clinical knowledge.[30] They consciously relate symptoms to concepts in the relevant knowledge networks they possess. However, explicit reasoning and thinking causally to carry out a diagnosis is tenuous, error-prone, elaborate and time-consuming.[5] It is more efficient to use known associations between clinical features and illnesses (scripts), and each encounter with a patient with a specific disease will add bits and pieces to the related illness script. Biomedical knowledge remains, nevertheless, present and accessible. In its encapsulated form,[1] it constitutes the anatomy of the illness script. It places constraints on the acceptable values for the different attributes of scripts and on their relationships.[6,10] It also alerts clinicians when they find abnormal findings or events that violate physiological expectations that are normally found in that specific type of disease, serving as a coherence criterion for hypotheses about the patient.[10] Biomedical knowledge can also be used in situations where no available scripts are adequate. In such cases, clinicians use their biomedical knowledge to understand the situation and to find pertinent hypotheses through a chain of causal reasoning.[7,10] If this process also fails, the clinician may revert to more general procedures, such as further referral, doing nothing (recommending the patient to return if complaints do not diminish or get worse), or treatment of individual symptoms without establishing a diagnosis.

Script acquisition is of utmost importance at the beginning of a medical career. Illness scripts require continuous updating as a result of changes in the diseases themselves and the population a doctor deals with. Both explicit and implicit learning processes can contribute to these changes. This process will start immediately, as soon as a student or doctor experiences contact with patients. The process of incorporating new theoretical knowledge into illness scripts demands active study. Analysis of health care indicators[31] suggests that for most doctors the knowledge they gathered in medical school remains the basis for medication and treatment.

CONCLUSIONS

Script theory raises educational issues concerning the instructional methods that foster their construction and refinement and their implications for the assessment of clinical competence. Because illness scripts develop from the application of biomedical and clinical sciences knowledge to real cases, and are themselves the key to further development, scripts should be a focus of attention in education. As typical cases represent the default values of an illness script, script building should start from there, taking care of the deliberate application of biomedical and clinical knowledge to the case at hand. Having formed a well established image of the typical representation of the disease,[32] attention should also be drawn to natural variations and atypical representations. In this way the slots in the illness script will develop a realistic range of values.

The implications of illness script theory and empirical findings indicate that both problem-based and experience-based learning[33] facilitate the learning and adapting of enabling conditions and consequences knowledge in early career training. However, the acquisition of theoretical knowledge deserves renewed attention, not only because it places constraints on acceptable values within scripts but also because it is this type of knowledge that should enable present students to learn new scientific knowledge and incorporate it into their scripts over 20 years from now. Although the integrating of new knowledge in old has been well investigated in medical students,[7,34] this learning process at the later stages of someone's career needs more attention. We also need to rethink continuous medical education, collaborative work forms, and support and feedback structures that can help experienced doctors to stay sharp.

Contributors: all authors contributed to the conception and design of this essay, the collection, analysis and interpretation of data, and the drafting and revision of the paper. All authors approved the final version of the manuscript.

Acknowledgements: none.
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, eds. *Tutorials in Problem-based Learning: a New Direction in Teaching the Health Professions*. Assen: Van Gorcum 1984;128–42.
2 Schank RC, Abelson R. *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum Associates 1977.
3 Gardner H. *The Mind*'s New Science: a History of Cognitive Revolution. New York: Basic Books Inc. 1987;165–9.
4 Nelson K. Event knowledge and cognitive development. In: Nelson K, Gruendel J, eds. *Event Knowledge: Structure and Function in Development*. Hillsdale, NJ: Lawrence Erlbaum Associates 1986;1–19.
5 Custers E. *The development and function of illness scripts: studies on the structure of medical diagnostic knowledge*. PhD Thesis. Maastricht: University of Maastricht, 1995;1–33.
6 Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000;**75**:182–90.
7 Feltovich P, Spiro R, Coulson R. Issues of expert flexibility in context characterised by complexity and change. In: Feltovich P, Ford K, Hoffman R, eds. *Expertise in Context*. Menlo Park, CA: American Association for Artificial Intelligence/MIT Press 1997;125–46.
8 Rikers RMJP, Schmidt HG, Boshuizen HPA, Linssen GCM, Wesseling G, Paas FGWC. The robustness of medical expertise: clinical case processing by medical experts and subexperts. *Am J Psychol* 2002;**115** (4):609–29.
9 Abbott V, Black JB, Smith EE. The representation of scripts in memory. *J Mem Lang* 1985;**24**:179–99.
10 Kassirer JP. Diagnostic reasoning. *Ann Intern Med* 1989;**110**:893–900.
11 Schmidt HG, Boshuizen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the 'intermediate effect' in clinical case presentations. In: *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*. Montreal, Canada. Hillsdale, NJ: Lawrence Erlbaum Associates 1998;139–45.
12 Bordage G, Zacks R. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ* 1984;**18** (6):406–16.
13 Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem solving by medical students: a cross-sectional and longitudinal analysis. *Med Educ* 1981;**15** (5):315–22.
14 Hobus PPM, Hofstra M, Boshuizen HPA, Schmidt HG. De context van de klacht als diagnosticum [The context of the complaint as a diagnostic tool]. *Huisarts en Wetenschap* 1998;**31**:261–3.
15 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract* 1997;**2** (2):173–84.
16 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39** (1):98–106.
17 Regehr G, Cline J, Norman GR, Brooks L. Effect of processing strategy on diagnostic skill in dermatology. *Acad Med* 1994;**6** (10 Suppl):34–6.
18 Papa F, Shores J, Meyer S. Effects of pattern matching, pattern discrimination, and experience in the development of diagnostic expertise. *Acad Med* 1990;**65** (9 Suppl):21–2.
19 Norman GR, Eva K, Brooks L, Hamstra S. Expertise in medicine and surgery. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR, eds. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, MA: Cambridge University Press 2006;339–53.
20 Norman GR. Problem-solving skills, solving problems and problem-based learning. *Med Educ* 1988;**22** (4):279–86.
21 Coderre S, Mandin H., Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ* 2003;**37** (8):695–703.
22 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978;64–121.
23 Barrows HS, Feltovich PJ. The clinical reasoning process. *Med Educ* 1987;**21** (2):86–91.
24 Lesgold A. Context-specific requirements for models of expertise. In: Evans DA, Patel VL, eds. *Cognitive Science in Medicine*. Cambridge, MA: MIT Press 1989;373–400.
25 Grant J, Marsden P. Primary knowledge, medical education and consultant expertise. *Med Educ* 1988;**22**:173–9.
26 Van Schaik P, Flynn D, Van Wersch A, Douglass A, Cann P. Influence of illness script components and medical practice on medical decision making. *J Appl Psychol* 2005;**11** (3):187–99.
27 Eva KW, Cunnington JP. The difficulty with experience: does practice increase susceptibility to premature closure? *J Contin Educ Health Prof* 2006;**26** (3):192–8.
28 Boshuizen HPA, Schmidt HG. Biomedical knowledge and clinical expertise. *Cognit Sci J* 1992;**16**:153–84.
29 Patel VL, Evans DA, Groen GJ. Biomedical knowledge and clinical reasoning. In: Evans DA, Patel VL, eds. *Cognitive Science in Medicine*. Cambridge, MA: MIT Press 1988;53–112.
30 Schmidt HG, Norman GR, Boshuizen HP. A cognitive perspective on medical expertise: theory and implication. *Acad Med* 1990;**65** (10):611–21.

31 Choudry NK, Fletcher R, Soumerai S. Systematic review: the relationship between clinical experience and quality of health care. *Ann Int Med* 2005;**142** (4):260–73.
32 Bordage G, Lemieux M. Which medical textbook to read? Emphasising semantic structures. *Acad Med* 1990;**65** (9 Suppl):23–4.
33 Dornan T, Bundy C. What can practical experience add to early medical education? A consensus survey. *BMJ* 2004;**329**:834–40.
34 Boshuizen HPA, van de Wiel MWJ. Multiple representations in medicine; how students struggle with it. In: van Someren MW, Reimann P, Boshuizen HPA, De Jong T, eds. *Learning with Multiple Representations.* Amsterdam: Elsevier 1998;237–62.

Received 2 April 2007; editorial comments to authors 1 August 2007; accepted for publication 25 September 2007

Breaking down automaticity: case ambiguity and the shift to reflective approaches in clinical reasoning

Sílvia Mamede,[1] Henk G Schmidt,[2] Remy M J P Rikers,[2] Júlio C Penaforte[1] & João M Coelho-Filho[3]

CONTEXT Two modes of case processing have been shown to underlie diagnostic judgements: analytical and non-analytical reasoning. An optimal form of clinical reasoning is suggested to combine both modes. Conditions leading doctors to shift from the usual mode of non-analytical reasoning to reflective reasoning have not been identified. This paper reports a study aimed at exploring these conditions by investigating the effects of ambiguity of clinical cases on clinical reasoning.

METHODS Participants were 16 internal medicine residents in the Brazilian state of Ceará. They were asked to diagnose 20 clinical cases and recall case information. The independent variable was the degree of ambiguity of clinical cases, with 2 levels: straightforward (i.e. non-ambiguous) and ambiguous. Dependent variables were processing time, diagnostic accuracy and proposition per category recalled. Data were analysed using a repeated measures design.

RESULTS Participants processed straightforward cases faster and more accurately than ambiguous ones. The proportion of text propositions recalled was significantly lower ($t[15] = 2.29$, $P = 0.037$) in ambiguous cases, and an interaction effect between case version and proposition category was also found ($F[5, 75] = 4.52$, $P = 0.001$, $d = 0.232$, observed power = 0.962). Furthermore, participants recalled significantly more literal propositions from the ambiguous cases than from the straightforward cases ($t[15] = 2.28$, $P = 0.037$).

CONCLUSIONS Ambiguity of clinical cases was shown to lead residents to switch from automatic to reflective reasoning, as indicated by longer processing time, and more literal propositions recalled in ambiguous cases.

KEYWORDS *clinical competence; *decision making; *diagnosis; Brazil; *internship and residency; judgement.

Medical Education 2007: **41**: 1185–1192
doi:10.1111/j.1365-2923.2007.02921.x

[1]Center for Research Development, School of Public Health of the State of Ceará, Fortaleza, Brazil
[2]Department of Psychology, Erasmus University, Rotterdam, The Netherlands
[3]Faculty of Medicine, Federal University of Ceará, Fortaleza, Brazil

Correspondence: Sílvia Mamede, Avenida Alvaro Correia, 455, 802/A, Fortaleza, Ceará, CEP 60165-230, Brazil. Tel: 00 55 85 3267 8512; Fax: 0055 85 3264 9655; E-mail: silviamamede@gmail.com

INTRODUCTION

Judicious judgements and effective decision making define successful clinical problem solving. Two different approaches for processing clinical cases, non-analytical and analytical, have been shown to underlie diagnostic decisions.[1,2] Experienced doctors diagnose routine problems essentially by recognising similarities between the actual case and examples of previous patients.[3] This pattern-recognition, non-analytical, form of clinical reasoning is largely automatic and unconscious.[3,4] In the second, analytical form of case processing, clinicians arrive at a diagnosis by analysing signs and symptoms, relying on biomedical knowledge when necessary.[2,5,6]

It has been suggested that these 2 types of reasoning result from different kinds of knowledge used for diagnosing cases. According to Schmidt and Boshuizen,[3] medical expertise development entails a process of knowledge restructuring. Biomedical knowledge is gradually 'encapsulated' under clinical knowledge and, with clinical experience, scripts of diseases and exemplars of patients, is stored in the longterm memory. Empirical studies have highlighted the role of scripts, examples and encapsulated knowledge in non-analytical processing.[3,7,8]

Overview

What is already known on this subject

Two different approaches for processing clinical cases, non-analytical and analytical, have been shown to underlie diagnostic decisions. It was not previously known, however, what makes doctors shift from the usual automatic reasoning to reflective reasoning.

What this study adds

This study shows that the ambiguity of clinical cases is among the conditions that lead doctors to adopt reflective reasoning approaches for diagnosing clinical problems. It also contributes to understanding of how the 2 modes of reasoning act in case processing.

Suggestions for further research

Further investigation is required to identify other conditions underlying the switch from non-analytical to analytical diagnostic reasoning, and whether and how reflective reasoning might be learned.

These studies have often used the clinical case paradigm, in which participants are requested to:

1 read a clinical case description;
2 provide a diagnosis;
3 recall all case information, and
4 explain case findings.

Expert doctors were shown to recall less information from the case description in a literal format than did advanced students. By contrast, experts generated more high-level inferences (i.e. inferences based on more than 1 finding in the case).[8,9] Experts apparently make shortcuts in their lines of reasoning while processing cases. They easily recognise a set of signs and symptoms as characteristic of a disease, almost automatically infer relevant encapsulated concepts, and generate diagnostic hypotheses, without needing to analyse individual findings and pathophysiological mechanisms.[8,9]

Research on medical expertise has particularly investigated non-analytical case processing. Recently, interest in analytical diagnostic reasoning has risen, stimulated by concerns about medical errors. Despite its effectiveness in routine situations, non-analytical reasoning may lead doctors to fail when they encounter complex or unusual problems.[10,11] Heuristics and experienced doctors' difficulties in reformulating initial hypotheses have been pointed out as potential causes of errors.[12,13] Reflective practice, conceptualised as doctors' ability to critically reflect on their own reasoning and decisions, has been considered crucial for optimal clinical performance.[14,15] A recent study suggested a multi-dimensional structure of reflective practice in medicine.[14] It would imply an elaborate, careful consideration of case findings and critical scrutiny of one's own reasoning. This thoughtful approach is not expected to be used in routine problems, but would be triggered by troublesome situations.[16] The conditions that trigger reflection, however, are not known.

The purpose of this study was to examine factors that lead doctors to switch from their usual non-analytical reasoning to a reflective diagnostic approach.* More specifically, we aimed to investigate whether ambiguity in clinical cases would lead to a breakdown of automaticity. By ambiguous cases, we mean a patient presentation that corresponds to the typical pattern of a disease but also includes features consistent with alternative diagnoses. The present study required residents to read a clinical case, provide a diagnosis and, subsequently, recall the case information. Based on previous studies on reflective practice,[14] we hypothesised that ambiguity in clinical cases would lead doctors to shift from non-analytical to reflective reasoning. According to the knowledge encapsulation view, doctors will largely use encapsulated knowledge when dealing with routine cases,[8,17] when they do not need to evaluate the findings extensively. Reflective reasoning, however, would require more systematic consideration of individual signs and symptoms. Based on these assumptions, we made a set of predictions. Firstly, processing time will be higher for ambiguous cases than for straightforward cases because ambiguous cases lead to reflection. Secondly, recall protocols of ambiguous cases will be more elaborate than protocols of straightforward cases, not only in terms of literal propositions but

*The 2 main forms of case processing have traditionally been designated by the terms 'non-analytical' and 'analytical', the latter conceived and operationalised largely in terms of mental, cognitive processes. Reflective practice in medicine has been conceptualised as a set of behaviours and reasoning processes involving affective dimensions, which would also be expected to play a role in the reflective processing of clinical cases. As this paper is aimed at calling attention to the affective skills required for reflective diagnostic approaches, we have adopted the term 'reflective reasoning', which will be used throughout the text.

also in terms of high-level inferences. That is, an ambiguous case will trigger additional inferences resulting from the doctor's effort to understand this complex case. Finally, diagnostic accuracy will be higher for straightforward cases than for ambiguous cases, which would validate our manipulation.

METHODS

Design

In this experimental study all participants performed under both experimental conditions (i.e. repeated measures). The independent variable was case type (straightforward versus ambiguous). Dependent variables were processing time, diagnostic accuracy and propositions (per category) recalled.

Participants

The participants were 16 second-year internal medicine residents from teaching hospitals in the Brazilian State of Ceará (mean age = 27.06 years; standard deviation [SD] = 1.06 years). A total of 22 eligible residents were invited to participate and informed consent was obtained.

Materials

The materials consisted of 20 written clinical cases, covering common conditions within the domain of internal medicine (Table 1). Case descriptions reported contextual information, complaints, findings from history taking and physical examination, and test results. There were 2 versions of each case. In the straightforward version, the case corresponded to a 'text-book case', exhibiting the set of features encountered in the typical presentation of the disease. The case description strongly suggested only 1 diagnosis. A few features were added to the straightforward version to generate the clinically ambiguous version. These features consisted of information about the patient's context, medical history and/or complaints, which raised the plausibility of an alternative diagnosis. Table 2 presents an example of a case. Items in italics were included in the clinically ambiguous version only. Cases

Table 1 Diagnostics of the cases used in the experiment

1	Community-acquired pneumonia
2	Stomach cancer
3	Acute bacterial endocarditis
4	Rheumatoid arthritis
5	Inflammatory bowel disease
6	Acute pyelonephritis
7	Acute viral hepatitis
8	Acute bacterial meningitis
9	Pseudomembranous colitis
10	Hyperthyroidism
11	Deficiency of vitamin B_{12}
12	Addison's disease
13	Acute alcoholic pancreatitis
14	Nephrotic syndrome
15	Pulmonary thromboembolism
16	Liver cirrhosis
17	Coeliac disease
18	Acute viral pericarditis
19	Acute myeloid leukaemia
20	Acute appendicitis

Table 2 Example of a case used in the experiment. Items in italics were given only in the ambiguous version of the case

The patient is a 27-year-old woman, who complains of pain in the right side of the chest that started suddenly 24 hours ago. The pain becomes worse with inspiring and is associated with dyspnoea. The patient denies cough, expectoration, haemoptysis or wheezing. She describes having felt warm, but did not take her temperature. She denies oedema in the inferior limbs and says she has never had respiratory problems. She has used no medications other than oral contraceptive pills. *She raises various types of birds at home and smokes 20 cigarettes per day.* She does not consume alcohol and has no risk factors for HIV. Her family history is negative for asthma. *Her father had pulmonary emphysema and died from coronary heart disease when he was 62 years old*

Physical examination

The patient is slightly obese. She appears uncomfortable and is in mild respiratory distress

The patient's temperature is 38 °C; pulse is 115 beats/min; blood pressure is 140/80 mmHg; respiration count is 30/min. There is no jugular turgidity. Cardiac examination is normal. Lung examination does not show rhonchus, crepitations or wheeze. Abdomen examination does not show abnormalities. The extremities are normal, without oedema or cyanosis

Tests

Haematocrit: 42%; haemoglobin: 14.5g dl^{-1}

White cell count: 6000/mm^3, with 74% neutrophils and 26% lymphocytes

Chest X-ray: normal cardiac area, small infiltrate in the right lower lobe

Electrocardiogram: sinus tachycardia

Arterial blood gas values: pH: 7.49; pCO_2: 32 mmHg; pO_2: 60 mmHg contained an average of 39.67 (SD = 11.13) propositions (i.e. discrete idea units in the text).[18]

Two expert doctors with a specialty board certification in internal medicine and over 15 years of clinical practice prepared the cases. They were presented in the format of a booklet containing, for each case, a page with the case description and a space to write the diagnostic hypothesis, followed by a blank page for free recall. Each booklet contained 20 cases to be diagnosed, 10 under each experimental condition (i.e. straightforward and ambiguous). Ambiguous and straightforward cases were presented alternately. Half the booklets started with an ambiguous case and half with a straightforward case. The presentation sequence of the 20 cases was also counterbalanced in each booklet to control for order effects.

Procedure

The experiment consisted of a training phase and a test phase. In the training phase, 2 sample cases were presented to familiarise participants with the procedures. There were no time constraints but the experimenter asked participants to strive to provide the most likely diagnoses for the cases as fast as possible. This was in order to prevent participants, as much as possible, from evaluating cases in a more elaborate manner than their usual processing mode. In the test phase, each participant received a booklet containing cases to be solved by the same procedure used in the training phase. The participant was asked to first read the case and provide the most likely diagnosis, and then to turn the page and recall the case information. For each case, the experimenter recorded processing time from the moment the participant started to read the case to the moment he or she wrote down the diagnosis. Participants were tested individually in sessions lasting approximately 60 minutes.

Analysis

The accuracy of the diagnosis provided by participants was independently rated by 2 experts with specialty board certification in internal medicine, and over 18 years of professional practice in teaching hospitals. Diagnoses were rated on a 5-point scale ranging from 0 (completely incorrect diagnosis) to 4 (completely correct diagnosis). For example, 1 point was attributed to a diagnosis that was not the correct main diagnosis for the case but contained at least 1 of its constituents (e.g. *upper gastrointestinal bleeding* in a case of *stomach cancer*). The inter-rater agreement between the 2 experts was 86%. Disagreements between raters were resolved by discussion.

The free-recall protocols were scored by means of a propositional analysis method introduced by Patel and Groen.[18] Each protocol was segmented into propositions. A proposition consists of 2 concepts linked by a qualifier, such as *causation*, *specification*, *temporal information* or *location*. For instance, the protocol fragment in Table 2, 'The patient is a 27-year-old woman, with complaints of pain in the right side of the chest that started suddenly 24 hours ago' consists of 6 propositions:

1 patient *specification* (woman);
2 patient *specification* (27-year-old);
3 complaints *specification* (pain);
4 pain *location* (in the right side of the chest);
5 starting *specification* (suddenly), and
6 complaints *temporal information* (24 hours ago).

Each proposition in the recall protocol was matched against the propositions in the text of the case description. Based on their relationships with the propositions in the text, the recalled propositions were classified into 6 categories: literal (or paraphrased) propositions; low-level inferred propositions; high-level inferred propositions; non-significant mistakes; significant mistakes, and non-existing propositions. Low-level inferred propositions are inferences based on only 1 proposition in the text, whereas high-level inferences are propositions that can be matched with a number of propositions in the case description. In Table 2, *Fever* or *tachycardia* are examples of low-level inferences. *Pleuritic pain* and *respiratory alkalosis* are possible high-level inferences.

High-level inferences have been considered to be evidence of the use of encapsulated knowledge.[17,19] The total number of propositions in the recall protocol was obtained by summing the number of literal propositions, low-level inferences and high-level inferences. When 2 researchers independently scored a subset of protocols, an inter-rater agreement of 0.92 ($P < 0.05$) was found. Differences were resolved by discussion and, as the procedure turned out to be reliable, scoring of the remaining protocols proceeded with 1 judge.

Data from cases solved in each condition were collapsed for each participant. Descriptive statistics were obtained for each experimental condition (straightforward versus ambiguous), and paired sample *t*-tests were performed for comparing processing time, diagnostic accuracy and propositions recalled in both conditions. We controlled for case length by calculating processing time per proposition and proportion of propositions recalled. Repeated measures analysis of variance, with case type and proposition category as within-subject factors, was used for comparison of the number of propositions recalled in the 6 categories of propositions in the 2 experimental conditions. Posthoc paired *t*-tests were performed for comparison across the levels of the proposition category. Effects size (partial η^2) and observed power (OP) were calculated when indicated.

RESULTS

Processing time

Table 3 shows the mean total processing time and the mean time per proposition for both conditions. Paired *t*-tests revealed a significant difference between straightforward cases and ambiguous cases, both for total time ($t[15] = 5.03$, $P < 0.001$, partial $\eta^2 = 0.628$, OP = 0.997) and time per proposition ($t[15] = 4.19$, $P = 0.001$, partial $\eta^2 = 0.539$, OP = 0.974).

Free recall

Table 4 presents proportions of propositions recalled from the text in both conditions. They were lower in the ambiguous cases than in the straightforward ones and this difference was significant ($t[15] = 2.29$, $P = 0.037$, partial $\eta^2 = 0.259$, OP = 0.573). The proportion of straightforward propositions (i.e. propositions that constituted the description of straightforward cases) recalled was significantly lower in ambiguous cases than in straightforward cases ($t[15] = 6.14$, $P < 0.001$, partial $\eta^2 = 0.715$, OP = 1.0).

Table 5 presents the mean number of propositions recalled in the 6 categories. Analysis of variance showed a large significant effect of category of proposition on the number of propositions recalled ($F[5, 75] = 46.54$, $P < 0.001$, partial $\eta^2 = 0.756$, OP = 1.00). There was no significant main effect of case type ($F[1, 15] = 1.53$, $P = 0.23$, partial $\eta^2 = 0.093$, OP = 0.212), but a large interaction between case type and proposition category was found ($F[5, 75] = 4.52$, $P = 0.001$, partial $\eta^2 = 0.232$, OP = 0.962). Posthoc *t*-tests showed a significant difference between the mean number of literal propositions recalled in straightforward cases versus ambiguous cases ($t[15] = 2.28$, $P = 0.037$). Comparisons of the mean numbers of propositions recalled for the other categories did not show significant differences.

Table 3 Mean processing time and mean time per proposition (in seconds) as a function of case version

	Straightforward cases			***Clinically ambiguous cases***		
	n	***Mean***	***SD***	***n***	***Mean***	***SD***
Processing time	16	548.50	166.74	16	687.81	218.06
Processing time per proposition	16	1.47	0.43	16	1.64	0.51

Table 4 Means for propositions recalled as a function of case version

	Straightforward cases			***Clinically ambiguous cases***		
	n	***Mean***	***SD***	***n***	***Mean***	***SD***
Number of propositions recalled	16	9.94	4.46	16	10.36	5.19
Proportion of propositions recalled	16	0.27	0.12	16	0.25	0.12
Proportion of straightforward propositions recalled	16	0.27	0.12	16	0.21	0.11

Table 5 Mean number of propositions recalled in each category as a function of case version

	Straightforward cases			***Clinically ambiguous cases***		
	n	***Mean***	***SD***	***n***	***Mean***	***SD***
Number of literal propositions recalled	16	77.00	40.84	16	84.43	49.93
Number of low-level inferences recalled	16	12.31	7.15	16	10.75	5.92
Number of high-level inferences recalled	16	10.06	7.37	16	8.37	4.27
Number of non-significant mistakes	16	2.75	2.49	16	2.62	3.07
Number of significant mistakes	16	0.75	1.61	16	0.81	1.05
Number of non-existing propositions	16	1.44	1.63	16	2.56	3.40

SD = standard deviation

Diagnostic accuracy

Mean diagnostic accuracy was higher for straightforward cases (mean = 3.09, SD = 0.45) than for ambiguous cases (mean = 2.61, SD = 0.51). A paired *t*-test showed that this difference was significant ($t[15] = 2.41$, $P = 0.029$, partial $\eta^2 = 0.279$, OP = 0.616), which indicates that our ambiguous cases were more complex (i.e. our manipulation was valid).

DISCUSSION

This study was concerned with conditions that lead doctors to shift from non-analytical to reflective reasoning when solving clinical cases. Based on previous studies on reflective practice in medicine,[14] it was hypothesised that 1 of these conditions was ambiguity of a clinical problem. In this experiment, features in case presentation were manipulated to create either a straightforward or an ambiguous case. It was hypothesised that ambiguity would lead participants to engage in elaborate, reflective case processing, and, therefore, spend more time on diagnosing ambiguous cases than straightforward ones. Furthermore, based on the notion of reflective practice in medicine,[14,15] it was predicted that reflective reasoning would be expressed by more elaborate recall protocols (i.e. more literal propositions and inferences) of ambiguous cases. Results showed that participants processed straightforward cases faster than ambiguous ones. Surprisingly, the proportion of total propositions recalled was *lower* and qualitatively different in the ambiguous condition than in the straightforward condition. By contrast, participants also recalled more literal propositions from the ambiguous cases than from the straightforward cases. Differences in the number of inferences generated in the 2 conditions were not significant.

Findings are largely consistent with the hypothesis put forward in this study. Doctors spent more time processing ambiguous cases than straightforward cases. Furthermore, differences in the literal propositions recalled suggest that these propositions were more highly activated in their case representation of the ambiguous cases. Apparently, doctors realised that the ambiguous cases required more elaborate exploration of the findings, which is characteristic of reflective practice in medicine.[14] A recent study exploring question format, task difficulty and reasoning strategies also suggested that case difficulty triggers reflection.[20]

Reflective practice also entails promptness to explore features in a case that do not fit with initial hypotheses.[14] Participants indeed seem to have engaged in such exploration when faced with contradictory clinical findings. The lower proportion of straightforward propositions recalled in ambiguous cases indicates that doctors' attention was directed to the atypical features that were added in this experimental condition. This could also explain the lower total proportion of propositions recalled in ambiguous cases.

Our findings call attention to the role of clinical features in diagnostic processes. Several studies demonstrated the significance of encapsulated knowledge in experts' reasoning, but also suggested the potential importance of individual features.[8,17,19] These studies usually explored differences in case processing associated with expertise level. The stimuli, therefore, were frequently the same, whereas participants' levels of expertise varied. The present study investigated whether differences in case characteristics would affect the diagnostic reasoning of participants with similar levels of expertise. Results suggest that individual features in fact play an important role in case representation, and that doctors' reliance on analysis of signs and symptoms increases with the ambiguity of clinical problems. Recognition of ambiguity may only happen if doctors analyse to some extent individual features in their process of identifying a pattern in a set of signs and symptoms. In the course of this process, ambiguous findings may then break down the usual automatic reasoning and lead doctors to engage in reflection.

Contrary to our predictions, there were no significant differences between conditions in high-level inferences. It was expected that reflection would also manifest itself by generating more inferences to account for ambiguous data. A possible explanation for this finding may be that the straightforward cases were more complex to resolve than expected, resulting in similar performances on both types of case. This explanation is substantiated by the participants' diagnostic performance on the straightforward cases, which resulted in about 75% of the maximum possible score (i.e. rather low for the straightforward cases). Alternatively, we might argue that the ambiguous cases were not complex enough to generate extra inferences to deal with the problem. This interpretation is also, to some extent, substantiated by the participants' diagnostic performance on this type of case. They scored 65% of the maximum score, which is not much lower than their score on the straightforward cases. It is important to note that previous studies argued that, unlike diagnostic performance and processing speed, the measure of free recall, and, in particular, high-level inferences, is often not sensitive enough to reveal small differences between doctors in different experimental conditions.[17,21]

It has been suggested that clinical teaching should aim to provide students with multiple reasoning strategies that could enable them to work through problems in different situations.[1,2,22] This requires recognising when more reflection is required. By indicating that ambiguity in a clinical case apparently acts as a cue for reflection, our findings might facilitate teaching of a combined, integrated clinical reasoning. However, how this might be done is still to be explored. Findings also highlighted the role of individual features in doctors' reasoning, which reaffirms the value attributed to teaching the importance of systematic analysis of clinical cases.

Some questions remain for future investigation. Firstly, ambiguity in clinical presentation was shown to break down automaticity, but other conditions that generate a sense of complexity and are still to be identified might have a similar effect. Secondly, light was shed on the role of individual features in diagnostic reasoning, but there is much more to be discovered about how reflection manifests itself in case processing. Finally, some doctors seem to recognise, more than others, when a problem requires an elaborate mode of processing. As far as conditions leading to reflective reasoning become known, the possibility of teaching doctors to recognise and adopt reflective approaches increases. How these reflective practices might be learned remains a question yet to be answered.

The present study has some limitations. As we have outlined above, the clinical case paradigm assumes that case representation can be probed by case recall. However, we cannot exclude the notion that concepts activated during case processing in both conditions did not appear in the protocols or that inferences in the protocols were not generated during case processing, but during the recall task.[19,21] Another limitation is inherent to the laboratory environment. Participants solved written clinical cases under experimental conditions, which restricts the generalisability of findings to performance in clinical settings.

Contributors: all authors participated in the conception and design of the study. SM provided contributions to acquisition, analysis and interpretation of data, drafted the manuscript and is the guarantor. HGS supervised the whole study and contributed to the analysis and interpretation of data, and to the critical revision of the manuscript. RMJPR participated in the data analysis and interpretation, and contributed to the critical revision of the manuscript. JCP and JMC-F participated in the analysis and interpretation of data, particularly with regard to judgement of diagnostic accuracy, and contributed to the revision of the manuscript.
Acknowledgements: the authors acknowledge Dr José Gerardo Paiva for his collaboration in contacting participants during data collection. The authors are also grateful to the medical residents who dedicated their limited free time to participating in the experiment.
Funding: this study was funded by a grant provided by the Erasmus University Rotterdam to SM.
Conflicts of interest: none.
Ethical approval: this study was approved by the Committee for Ethics in Research of the School of Public Health of the State of Ceará, accredited by the Brazilian National System of Ethics in Research.

REFERENCES

1 Norman G. Research in clinical reasoning: past history and current trends. *Med Educ* 2005;**39**:418–27.
2 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;**39**:98–106.
3 Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educ Psychol Rev* 1993;**5**:1–17.
4 Norman GR, Brooks LR. The non-analytical basis of clinical reasoning. *Adv Health Sci Educ Theory Pract* 1997;**2**:173–84.
5 Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. *Mem Cognit* 1990;**18**:394–406.
6 Kulatanga-Moruzi C, Brooks LR, Norman GR. Co-ordination of analytical and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med* 2001;**13**:110–6.
7 Brooks LR, Norman GR, Allen SW. Role of specific similarity in a medical diagnostic task. *J Exp Psychol Gen* 1991;**120**:278–87.
8 Schmidt HG, Boshuizen HPA. On the origin of intermediate effects in clinical case recall. *Mem Cognit* 1993;**21**:338–51.
9 Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognit Sci* 1992;**16**:153–84.
10 Croskerry P. The theory and practice of clinical decision making. *Can J Anaesth* 2005;**52** (6):1–8.
11 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79** (10 Suppl):70–81.
12 Eva KW. The ageing physician: changes in cognitive processing and their impact on medical practice. *Acad Med* 2002;**77** (10 Suppl):1–6.
13 Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimise them. *Acad Med* 2003;**78**:775–80.

14 Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Med Educ* 2004;**38**:1302–8.
15 Epstein RM. Mindful practice. *JAMA* 1999;**282**:833–9.
16 Schön DA. *The Reflective Practitioner: How Professionals Think in Action.* New York, NY: Basic Books 1983.
17 Rikers RMJP, Schmidt HG, Boshuizen HPA. On the constraints of encapsulated knowledge: clinical case representations by medical experts and subexperts. *Cognition and Instruction* 2002;**20** (1):27–45.
18 Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. *Cognit Sci* 1986;**10**:91–116.
19 Verkoeijen PPJL, Rikers RMJP, Schmidt HG, van de Wiel MWJ, Kooman JP. Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Med Educ* 2004;**38**:617–27.
20 Heemskerk L, Norman G, Chou S, Mintz M, Mandin H, McLaughlin K. The effect of question format and task difficulty on reasoning strategies and diagnostic performance in internal medicine residents. *Adv Health Sci Educ Theory Pract* 2007 [Epub ahead of print].
21 Rikers RMJP, Loyens SMM, Schmidt HG. The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Med Educ* 2004;**38**:1035–43.
22 Ark TK, Brooks LR, Eva KW. Giving learners the best of both worlds: do clinical teachers need to guard against teaching pattern recognition to novices? *Acad Med* 2006;**81**:405–9.

Received 19 April 2007; editorial comments to authors 1 August 2007; accepted for publication 12 September 2007

A healing curriculum

J Donald Boudreau[1], Eric J Cassell[1,2] & Abraham Fuks[1]

CONTEXT The banner of patient-centredness flies over many academic institutions; however, the practice and teaching of medicine remain oriented to disease. This incongruence is the result of an original Flexnerian dichotomy between the basic and clinical sciences and is maintained by a more recent distinction between disease and illness. One mind-set emphasises basic science and pathology pegadogically, whilst clinical medicine becomes a search for disease. The second introduces the patient as the focal point, underlining the personal and social contexts of illness.

RESPONSE AT A CONCEPTUAL LEVEL We must orient ourselves to a single central theme, namely, the well-being of the individual patient. Doing so does not deny the importance of the scientific understanding of biological function. Indeed, recent advances in genetics may permit a richer view of the individual as a unique product of genetic, developmental and experiential forces. The foregoing provide a coherent framework for a scientifically guided and humanistic medicine, which replaces the false dichotomies that have plagued medical school curricula with a congruent and stereoscopic view of medical education.

RESPONSE AT A CURRICULAR LEVEL We describe an undergraduate programme, entitled 'Physicianship', based on the fundamental premise that healing is the doctor's primary obligation. Explicit training in a specific clinical method, whose cardinal features include observation, attentive listening and clinical reasoning, emphasises the knowledge and skills necessary to effect this theoretical framework. The understanding of illnesses emphasises loss of homeostasis, whereas the physical examination highlights impairments of function. The educational experience is enriched with numerous opportunities for self-reflection.

[1]Department of Medicine, McGill University, Montreal, Quebec, Canada
[2]Department of Public Health, Weill Medical College of Cornell University, New York City, New York, USA

Correspondence: Dr Donald Boudreau, Room 633, McIntyre Medical Sciences Building, 3655 Promenade Sir William Osler, Montreal, Quebec H3G 1Y6, Canada. Tel: 00 1 514 398 5613; Fax: 00 1 514 398 3595; E-mail: donald.boudreau@mcgill.ca

KEYWORDS education, medical, undergraduate/ *methods; curriculum; *professional practice; *physician's role; patient-centred care; physician–patient relations.

Medical Education 2007: ***41****: 1193–1201*
doi:10.1111/j.1365-2923.2007.02905.x

> 'Although a clinician can be both a healer and a scientist, he cannot be an effective therapist if he merely joins these two roles in tandem by oscillating between them, adding laboratory science to bedside art. A clinician's objective in therapy is not just a conjunction, but a true synthesis of art and science, fusing the parts into a whole that unifies his work and makes his two roles one: a scientific healer… As a healer, the clinician's purpose is to treat the sick person, not merely the manifestation of disease.' Alvan Feinstein[1]

INTRODUCTION

This essay describes a new undergraduate medical curriculum that redirects the medical student's gaze away from disease and towards the sick person. We present the conceptual framework, entitled 'Physicianship', and place it in historical context by contrasting it with other patient-centred approaches. We also describe the teaching of a clinical method intended to equip students with the tools required to assess, understand and heal sick persons.

HISTORICAL CONTEXT

A seminal event in medical education in North America was the publication of Abraham Flexner's

Overview

What is already known on this subject

Traditional curricula, all generally derived from the original Flexner model, are inspired by and nurture classic dichotomies such as disease/illness, and basic science/clinical medicine. These are false and counter-productive to a coherent approach to medical education.

What this study adds

This essay recognises these various dualities but argues that medicine has but a single goal: to improve the well-being of a sick person, particularly in terms of function. It describes medicine's primary mandate as healing and recommends specific modifications to traditional curricular models.

Suggestions for further research

Evidence to support the effectiveness of the conceptual framework and the clinical method associated with it is currently being gathered.

report '*Medical Education in the USA and Canada*'.[2] Flexner perceived medical education to be bereft of a solid scientific foundation. His response was to propose a 2-phase curriculum, with scientific theory preceding clinical practice. An unintended consequence has been the nurturing of the sentiment, now prevalent, that there are 2 distinct sets of medical knowledge, the first rooted in science and scientific methodology, and the second, at times more difficult to define and delineate, linked to the delivery of clinical care. The recognition that this dichotomy is fundamentally counter-productive has generated the recurring desire to integrate these dual and parallel curricular strands and has guided curricular evolution over the past century. Flexner himself alluded to 2 categories:

> 'So far we have spoken explicitly of the fundamental sciences only. They furnish indeed, the essential instrumental basis of medical education. But the instrumental minimum can hardly serve as the permanent professional minimum. It is even instrumentally inadequate. **The practitioner deals with facts of two categories.** Chemistry, physics, biology enable him to apprehend one set; he needs a different apperceptive and appreciative apparatus to deal with other, more subtle elements. Specific preparation is in this direction much more difficult; one must rely for the requisite insight and sympathy on a varied and enlarging cultural experience.'[2] [Emphasis added]

Flexner seems to have acknowledged an inherent limitation of scientific knowledge in fully equipping doctors to understand and take care of sick persons in their social worlds. Despite his cautionary note, the scientific thrust of his report was so dominant that, by mid-century, clinical aspects of the encounter between doctor and patient had virtually completely ceded centre stage to the teaching of the scientific foundations of medicine. The emphasis on science had come to occupy, not simply the foundation of the curricular edifice, but, increasingly, the upper stories as well.

A powerful alternative to the disease model found expression in the biopsychosocial approach proposed by Engel.[3] The 'bio-psycho-social' model can be viewed as a systems-based hierarchy where the person (with unique characteristics, experiences and behaviours) is placed at the centre of a social organisation that begins with the individual's internal biochemical milieu and extends outward to encompass the family and community.[4] This model, coupled with the experience of client-centred counselling,[5] inspired a reform in the approach to the doctor–patient encounter, described as the patient-centred method.[6] This has eclipsed the purely biomedical approach, although perhaps more so in teaching than in medical practice.

Descriptions of patient-centred programmes routinely identify a need to integrate the science of medicine with a focus on the patient. Many models have been adduced. One description shows a weaving, back and forth, between 2 strands: that of science and its pathophysiologic perspective of disease, and that of patients in all their human complexities[7] (Fig. 1a). This approach is grounded in the distinction between illness and disease, as introduced by Cassell[8] and illustrated by Reading[9] (Fig. 1b). We believe, however, that, as important as the biopsychosocial approach has been in the evolution of our thinking, a medical pedagogy that considers its main task to be the integration of science with concern for the patient is left with a fundamental error. The

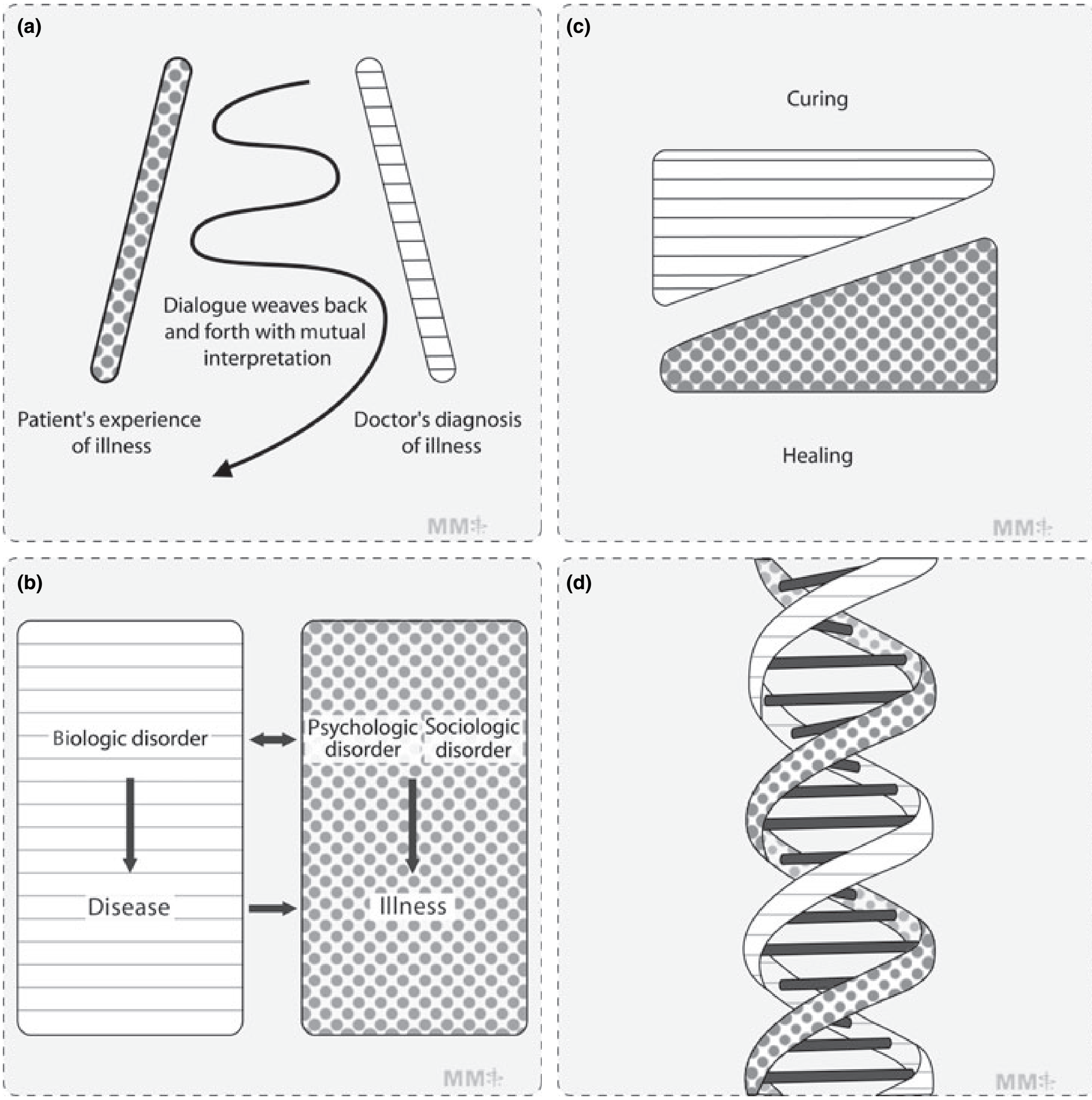

Figure 1 (a) Model showing dialogue between science and its pathophysiologic perspective of disease and patients (after McWhinney[7]). (b) Model showing the distinction between illness and disease (after Reading[9]). (c) Model showing the tension between curing and healing (after Milstein[10]). (d) Model using the structure of DNA to illustrate 2 strands of a curriculum, representing basic science and clinical medicine (after the University of Rochester School of Medicine and Dentistry[11]).

problem does not lie in the fact that this construct recognises 2 kinds of knowledge – medical science and knowledge of patients – but, rather, in the fact that it suggests 2 separate goals. Indeed, it makes explicit the duality it seeks to eliminate. The first goal, the prevalent perspective of the last century, is focused on the scientific problem of disease and its pathophysiology. The second goal is focused on the human problem(s) of the patient. Many clinicians, particularly in the context of palliative care, have grappled with this tension by making a distinction between curing and healing – between abnormalities of the body and those difficulties that arise from the patient's experience of those abnormalities (Fig. 1c). In this latter model, the 2 strands are attributed varying degrees of priority depending on the clinical situation – for example, healing is seen to predominate in end-of-life care.[10] A final example of a patient-centred curriculum, again with split fields, has been described using a contemporary image, the structure of DNA (Fig. 1d). In the 'Double Helix Curriculum',[11] the 2 strands represent basic science and clinical medicine, although it is not clear what the 'rungs' represent.

Graphic representations and statements concerning the goals and values of medicine, as outlined above, are completely understandable from a historical viewpoint, but they fail to go beyond outdated dichotomies, many of which are rooted in Cartesian mind–body duality. These various dualisms and schisms have resulted in 2 sets of goals for thoughtful clinicians, namely, treating the body by eliminating or controlling the disease and moderating the patient's experience so that it is less overwhelming or intrusive on all aspects of life. The resultant gap is a source of confusion for medical students, who are likely to witness 2 apparently non-coherent perspectives.

We believe that there is only 1 goal – the wellbeing of the patient and, more specifically, improvement in the patient's functions to allow the patient to pursue his purposes. The integration of both the scientific and humane perspectives into a single stereoscopic image takes place within the doctor. The process is analogous to that which occurs in architects as they integrate engineering and aesthetic considerations within themselves into a single coherent goal.

Except as a taxonomic abstraction, there is no disease or illness separate from the patient; there is only the patient who is ill. The patient 'lives' the pathophysiology; it is manifest in a specific manner (phenotype) in that unique patient. This may not be as readily evident in acute diseases but it is obvious in the much more common chronic diseases. Thus, the onset, presentation, diagnosis, treatment and course of any single disease differ in different patients and are influenced by domains spanning the range from genetics, development and experience, to meaning-related psychodynamics.

SETTING AND CONTEXT OF CHANGE

Our own reflections on the curriculum at McGill University made it clear that we, like many other schools, engaged in curricular change every decade or so and in each instance, a major case was made for the integration of basic science and clinical teaching. The fact that this was a recurrent feature over many decades suggests that our past successes had been limited. An additional impetus for change was provided by the Association of American Medical Colleges (AAMC) when it issued a challenge for medical schools to update their clinical teaching.[12] Lastly, it became clear that tinkering at the margins would not have the intended impact. Hence, a curriculum review process, launched in 2004, culminated in a proposal for realignment of the curriculum around the concept of the doctor as healer and professional, combining these facets of the intended 'product' under the rubric of 'physicianship'.[13]

EDUCATIONAL BLUEPRINT OF THE PHYSICIANSHIP CURRICULUM

Conceptual framework

'Physicianship' is not a word in common usage. The Oxford and Webster dictionaries define it in relation to the function(s) and role(s) of the physician. Cassell employed it approximately 2 decades ago[14] and, more recently, a 'physicianship' evaluation form has been described in the context of assessing professional behaviours.[15] Physicianship, as we understand it, is based on the following premise: the primary goal of medicine is healing. Healing encompasses the entire range of doctor–patient interactions, including treatments aimed at aberrant pathophysiologic mechanisms. Professionalism describes the requisite moral and behavioural attributes of doctors in all their guises, namely, as bedside clinicians, members of the profession, and members of the wider society.[16]

Healing has an honourable tradition. The word acquired the taint of quackery during the 20th century because it was used in a pejorative sense to discredit any medical effort not based completely on pathophysiological science. It is noteworthy that the *Oxford English Dictionary* lists the following definition for the word 'physician': 'A healer; a person who cures moral, spiritual, or political ills.'[17] This is linguistic evidence that 'healing' has been connected to non-physical ailments and again reflects the duality described previously. Decades ago, when the distinction between disease and illness was proposed, the word 'curing' was used in reference to disease and 'healing' in reference to illness.[8] We are of the opinion, however, that this dichotomy is artificial and counterproductive. In certain languages (e.g. French and Spanish), no distinction is made between curing and healing; these concepts are integrated within a single word (i.e. 'guérir' and 'curar', respectively), derived, like 'cure', from the word 'care'.

Some may be sceptical of the place of healing within medicine. Most would acknowledge, however, that the antithesis of healing may occur in any encounter and carries the potential to do considerable damage.

It has also been argued that healing may not be attributable to a 'healer'; that it depends on an innate potential within the patient.[18] Should this interpretation be accurate, it would not detract from the fact that becoming 'whole' in serious or crippling sickness virtually always requires the help of others.

Healing is fundamentally individual and irreducibly personal. Clinicians know that, even when patients are ostensibly 'cured of disease', (e.g. when successful coronary artery revascularisation or successful treatment of a malignancy has taken place), significant impairments may continue: patients may not return to work, resume their place in the family, or function psychologically or socially as they did before becoming sick. Until patients are able to function, by meeting their goals within the boundaries of their capacities and impairments, sickness is not yet over. Even when a pathophysiological source is remedied, as in, for example, the relief of pain, suffering may continue unabated in the absence of additional healing interventions. Healing depends on the knowledge of both the manifestations of sickness and the nature of the particular patient. For example, a person with diabetes shares a specific molecular basis for disease with all other patients with diabetes. However, individual experiences of the disease will vary simply because the patient, like all persons, is different from all others in every aspect of his existence. Beyond that, moreover, not only the experience but also the expression of diabetes will be unique and particular in each individual.

Organisational framework

A programme that includes elements in each year of the medical curriculum is necessary if we are to shift the frame of reference to a new coherent vision of medical training, and by extension, of medical practice. In addition, certain elements must continue in transverse fashion throughout the entire medical undergraduate experience in order to offer an integrated model of teaching and learning. The pedagogic goals and content of these elements are described in a later section of the paper. It may be useful, however, to first delineate the curricular units and settings within which these elements are delivered. The undergraduate medical curriculum at McGill University is 4 years in duration. The physicianship component consists of 5 physicianship courses and a mentorship programme called the 'Physician Apprenticeship'. A summary is presented in Table 1.

Educational objectives

The desired learning outcomes of the physicianship programme and the teaching modalities we have deployed to accomplish these are outlined in Table 2.

The first objective is to explicate to students the various capacities of the physician and to indicate that respect for and understanding of the healing function is a basic prerequisite for doctors who wish to place the patient's wellbeing at the centre of their work. A profound appreciation of the nature of suffering is required. The universal characteristics of serious sickness (e.g. disconnection from the surrounding world, vulnerability, failures of reasoning) as well as the more personal nature of suffering (e.g. loneliness, self-conflict, loss of purpose) are open to discovery through an adequate clinical method. They also represent specific opportunities for intervention. The doctor's basic tasks are to build a relationship, gather information and use that information to arrive at an understanding of the illness and its story, decipher the patient's understanding of his sickness, identify the patient's goals, plan and initiate treatment, estimate a prognosis, and

Table 1 Curricular units with primary responsibility for teaching Physicianship

Course	*Location in the 4-year programme*	*Pedagogic focus*
Physicianship 1	Year 1	Conceptual framework for physicianship; clinical observation; attentive listening; clinical thinking (and reasoning); bioethics; the professional role
Physicianship 2	Year 2	Communication skills
Physicianship 3	Year 2	Physical examination; critical appraisal and informed medical practice (i.e. evidence-based medicine)
Physicianship 4	Year 3	The healer role
Physicianship 5	Year 4	Advanced communication skills; medicine and society; professionalism and the social contract
Physician Apprenticeship	Years 1–4	Conceptual framework for physicianship; self-reflection

Table 2 List of specific objectives and teaching strategies

Elements	***Learning outcomes***	***Teaching modalities***
Theoretical framework		
Fundamental concepts	Define what is 'a person' Define health and healing Explain the nature of suffering[24] Discuss the goals of healing[25] Recite the cognitive basis and historical roots of professionalism	Assigned readings; didactic sessions followed by small-group interactions; written assignments; contact with a patient and family over a long period; discussions in the apprenticeship groups; required portfolio entries
Clinical method		
Clinical observation	Observe effectively and reliably, separating observation from interpretation, using a framework based on a modification of Berger's hierarchy of observation[26]	Student-led small groups using photographs and videos of patients
Attentive listening	Discuss the various roles of listening in the doctor–patient interaction Identify fundamental elements of language (spoken and non-verbal)[27] Explain how language works to reveal a patient's emotions and relationships to self, illness, the doctor and others; demonstrate how this skill is used as a therapeutic tool	Student-led small groups using recorded conversations of actual doctor–patient encounters
Communication skills	Discuss the role of communication in healing Demonstrate the technique of interviewing using the Calgary–Cambridge approach[28] Discover the trajectory from 'healthy' status to 'patient' status in a medical history and identify changes in function and its meaning	Introductory sessions are in didactic formats; students are then observed interviewing, progressing from role-plays to SPs and then actual patients; de-briefing and feedback is given by faculty, SPs and peers
Physical examination	Perform a complete physical examination assessing structural and physiologic abnormalities as well as a patient's capabilities in key aspects of personal function (physical, cognitive, emotional)	Students will learn basic skills by practising on themselves and will then progress to examining SPs and actual patients
Clinical thinking and reasoning	Explain the process doctors apply in formulating clinical problems underlining that clinical reasoning includes, but is not confined, to making a diagnosis Apply ethical principles Contrast different modes of inference (i.e. deduction, induction and abduction) and calculate conditional probabilities using a natural frequency approach[29] Apply a simplified version of Bayesian theory using odds and likelihood ratios	Didactic sessions followed by small groups that are led by senior medical students; exercises use recorded doctor–patient encounters and SP encounters to practise the integration of observation, listening and thinking in simulated clinical environments
Written description (and documentation)	Write an accurate and valid description of the physical appearance, speech and behaviour of patients Document a case history using the revised template based on Donnelly[30]	Throughout the programme, students document their descriptions (e.g. of visual images, spoken language, patient interviews); particular attention is paid to the learner's ability to use evidence to draw inferences from appropriate observations
Personal transformation		
Narrative competence	Formulate a narrative perspective of the patient's illness highlighting the sources of meaning Demonstrate basic textual skills (e.g. identify tense, voice and common archetypes)	Students record entries in the portfolio; the written case report template includes a section on the patient perspective
Self-reflection	Participate in activities intended to foster insights into the impact of the transition to physicianhood on personal emotions, meanings and relationships Recognise personal values, biases, strengths and liabilities Acknowledge the importance of symbols and respect celebratory acts	Guided discussions in the apprenticeship meetings; use of the portfolio; participation in events underlining key transition points (e.g. the Donning the Healer's Habit [White Coat] Ceremony)

SP = standardised patient

report the data. It is through this relationship, developed with these objectives in mind, that the healer's role is effected. The specific skills necessary to these ends are grouped under the rubric, the 'clinical method'.

The clinical method

It is only through an attempt to know the patient that one can engender the interpersonal respect necessary for the role of healer. Thus, bedside methods are not brought to bear simply in the search for a disease, but, rather, in order to know the patient and answer the cardinal question: 'Why did this particular individual (with his or her unique genetic, developmental, experiential and spiritual identities) come to visit me, the doctor, at this particular time?' Answering this question immediately accomplishes the 2 aims previously seen as disparate: that is, what is traditionally termed 'making a diagnosis' and being 'patient-centred'. These 2 goals are of a piece.

The toolbox of doctoring skills required is referred to as the 'clinical method'. Classic approaches to the clinical method, developed in the 19th century and little changed since, (with the notable exception of communication skills), and focused primarily on the search for disease, are inadequate for the teaching of physicianship. Consequently, we have introduced significant modifications in the clinical method taught under the rubric of physicianship. Thus, in Year 1, students are taught clinical observation, skilful listening, communication skills, and clinical reasoning. Physical examination is taught in Year 2 of the curriculum. Although these are described as separate, teachable entities, it is important to note that they are importantly inter-related.

Each element of the clinical method aims to equip the doctor to know his patient. Diagnostic efforts must be attuned to changes and impairments in the patient's functional capacity, as well as to morphological changes. Students must come away knowing the patient's goals, needs, concerns and preferences so that medical acts are ethically appropriate and reflect patient choices. Treatment is conceived as whatever is necessary to return the patient to as much function as possible, within the constraints of impairment and fate, and in relationship to the patient's perspective.

Description and narrative competence

The usual narrative record that is traditional in North American medicine hardly warrants the term 'history'. It is often a terse, epigrammatic and acronymic summary that concludes with a diagnosis or differential followed by suggestions for next steps and treatment. The mode of recording is clearly oriented to identifying a disease and is generally identified as belonging to a particular patient by the name-stamp on the chart. Certainly, little within the file reflects the unique individual to whom it purportedly refers. In order to counter this trend, we have introduced teaching sessions and required assignments on written description and narrative competence. The new template for the case history requires the student to detail the trajectory from 'healthy' to 'patient', integrating the functional losses. The narrative perspective is different: it entails apprehending the patient's story of illness so as to provide insight into the patient's understanding of his situation, highlighting, in particular, the sources of meaning for that person.[19] We teach the fundamental tasks in narrative competence – attention, representation, affiliation – and highlight basic textual skills such as understanding of tense, voice, common narrative archetypes and metaphors.

Transformation of the student

We consider that for effective healing, it is not only what the healer 'does' that is important, but also who the healer 'is'. Although it is often claimed that a disembodied science or technology makes the diagnosis and treats, it is clear that the doctor remains a requisite agent in clinical care. Many doctors actively avoid close knowledge of the sick – it is difficult, it appears to carry painful responsibility, and may be emotionally burdensome. The all too common admonishment to students to maintain a 'professional distance' is not acceptable. On the contrary, we expect that students will be actively immersed in such issues.

In order to provide the necessary perspective, permit time and opportunity for self-reflection and allow students to share the understandable angst and emotional turmoil that clinical care can entail, we have designed 2 strategies to provide appropriate emotional support while simultaneously promoting the values inherent to physicianship. A 4-year longitudinal course, the Physician Apprenticeship, is based on mentorship groups, each composed of 6 students, 1 senior student and 1 faculty member. Groups meet 6 times per year and provide a setting in which to explore the moral dimensions of medicine and reflect on the socialisation that occurs in medical school. Furthermore, these groups respond to an oft-expressed desire of students for a safe, non-judgemental environment in which to share their concerns, doubts and reactions about their early encounters with patients.

The second adjunct is the use of a personal (Physicianship) portfolio. Portfolios are collections of materials that can be used for assessing learner progress and documenting personal development and insight.[20] Medical students are required to make entries to this portfolio that focus on their progress in acquiring the skills of the clinical method and on their transformation from layman to doctor.

CONCLUSIONS

The model of medical education described here rests on the fundamental belief that it is the uniqueness of the person that is central to the clinical enterprise. Sir William Osler argued consistently for the centrality of the patient in medicine:

> 'One element must always be taken into account in prognosis and that is the personal equation of the patient. No two cases of the same disease are ever exactly alike. The constitution of the person, his individuality, stamps each case with certain peculiarities.'[21]

Recognition of individuality confers epistemological and ethical demands. Science and technology are among the tools employed by a contemporary doctor and these must be taught. Indeed, patients have the right to assume the presence of technical expertise. The practice of medicine without scientific methodology would be false. There is, however, a need to teach new knowledge and skills. We have briefly outlined some of these; narrative competence may, for example, provide an entry to highly personalised care.[22] The rapidly expanding field of genetics may provide a theoretical framework for the scientific basis of individuality.[23] The extent to which it will act in concert with or advance medicine's moral imperatives remains to be seen.

The subjectivity of the medical student is a key aspect of physicianship. That elusive entity, the 'self', is the central axis upon which the drama of clinical care revolves. The personhood of the patient and the doctor, and the relationship between them, are essential ingredients. Clinicians are not medical scientists – they are scientific healers. We believe this distinction to be important. It is our hope that this programme, as it continues to develop within the conceptual framework of physicianship, will represent the true synthesis of art and science expressed in Feinstein's epigram at the beginning of this essay.

Contributors: this article is primarily a description of the conceptual framework for a new curriculum. All 3 authors have been involved with the design and implementation of this curriculum from March 2004 (when the faculty's task force report on curricular renewal was finalised) to the present. All authors remain involved in ongoing curricular implementation and all are teachers on the programme. All 3 authors contributed significantly to the literature review, and the writing and editing of this article.

Acknowledgements: we are grateful to numerous colleagues who provided feedback on earlier versions of the manuscript and to the Max Bell Foundation for financial support.
Funding: this Curricular Renewal Project was funded by the Max Bell Foundation.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Feinstein AR. *Clinical Judgment*. Baltimore: Williams & Wilkins 1967;245.
2 Flexner A. *Medical Education in the United States and Canada. A Report to the Carnegie Foundation for the Advancement of Teaching*. New York: The Carnegie Foundation for the Advancement of Teaching.
3 Engel GL. The need for a new medical model: a challenge for biomedicine. *Science* 1977;**196**:129–36.
4 Reiser DE, Rosen DH. *Medicine as a Human Experience*. Baltimore: University Park Press 1984;43–60.
5 Rogers C. *Client-centered Therapy, its Current Practice, Implications and Theory*. Boston: Houghton Mifflin 1951;3–18.
6 Stewart M, Belle JB, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. *Patient-centred Medicine. Transforming the Clinical Method 2003*. Abingdon, UK: Radcliffe Medical Press 2003;1–30.
7 McWhinney IR. Focusing on lived experience: the evolution of clinical method in western medicine. In: Kay Toombs S, ed. *Handbook of Phenomenology and Medicine*. Dordrecht: Kluwer Academic Publishers 2001;331–50
8 Cassell EJ. *The Healer*'s Art. Philadelphia: Lippincott 1976;47–83.
9 Reading A. Illness and disease. *Med Clin North Am* 1977;**61**:703–10.
10 Milstein J. A paradigm of integrative care: healing with curing throughout life, 'being with' and 'doing to'. *J Perinatol* 2005;**25**:563–8.
11 University of Rochester School of Medicine and Dentistry. *The 'Double Helix' Curriculum*. http://www.urmc.rochester.edu/smd/education/medical/curriculum.cfm. [Accessed 17 April 2007.]
12 Nutter D, Whitcomb M *The AAMC Project on the Clinical Education of Medical Students*. Washington: Association of American Medical Colleges 2001;1–4.
13 Boudreau D. *Report of the Task Force on MD, CM Program Renewal*. Quebec: McGill University, Faculty of Medicine. 2004. http://www.medicine.mcgill.ca/physicianship/reports.htm. [Accessed 17 April 2007.]

14 Cassell EJ. Change concept of the ideal physician. *Deadalus* 1986;**115**:185–208.
15 Papadakis M, Osborn E, Cooke M, Healy K. A strategy for the detection and evaluation of unprofessional behaviour in medical students. Report of the University of California, San Francisco, School of Medicine Clinical Clerkships Operation Committee. *Acad Med* 1999;**74**:980–90.
16 Cruess RL, Cruess SR. Teaching medicine as a profession in the service of healing. *Acad Med* 1997;**72**:941–51.
17 Oxford English Dictionary. Electronic format. http://www.oed.com/. [Accessed 17 April 2007.]
18 Mount B. Existential suffering and the determinants of healing. *Eur J Palliat Care* 2003;**10** (Suppl 10):42–5.
19 Charon R. Narrative medicine: a model for empathy, reflection, profession and trust. *JAMA* 2001;**286**:1897–902.
20 Cole G. The definition of 'portfolio'. *Med Educ* 2005;**39**:1141.
21 Osler W. *Osler*'s Introductory Address to the 120th course of lectures in the Medical Department and 8th course in the Dental Department, in 1885, at the University of Pennsylvania, Unpublished Manuscript, BO 7664, Vol. II. 1885; 19.
22 Charon R. *Narrative Medicine. Honoring the Stories of Illness.* New York: Oxford University Press 2006;39–62.
23 Childs B, Wiener C, Valle D. A science of the individual: implications for a medical school curriculum. *Annu Rev Genomics Hum Genet* 2005;**6**:313–30.
24 Cassell EJ. The nature of suffering and the goals of medicine. *New Engl J Med* 1982;**306**:639–45.
25 Barritt P. *Humanity in Healthcare: the Heart and Soul of Medicine.* Oxon, UK: Radcliff Publishing Ltd 2005;167–87.
26 Berger L. A hierarchy of clinical observations. *Pediatrics* 1980;**65**:357–8.
27 Cassell EJ. *Talking with Patients. Volume 2: Clinical Technique.* Cambridge, MA: MIT Press 1985.
28 Kurtz S, Silverman J, Benson J. Marrying content and process in clinical method teaching: enhancing the Calgary–Cambridge guides. *Acad Med* 2003;**78**:802–9.
29 Gigerenzer G, Edwards A. Simple tools for understanding risk: from innumeracy to insight. *BMJ* 2003;**327**:741–4.
30 Donnelly WJ. Patient-centred medical care requires a patient-centred medical record. *Acad Med* 2005;**80**:33–8.

Received 29 November 2006; editorial comments to authors 4 April 2007; accepted for publication 9 July 2007

Undergraduate medical students' exposure to clinical ethics: a challenge to the development of professional behaviours?

Lis Cordingley,[1] Catherine Hyde,[2] Sarah Peters,[3] Bryan Vernon[4] & Christine Bundy[1]

CONTEXT The objective of this research was to explore medical students' experience of challenges to their ethical knowledge and understanding in clinical practice, and to investigate their need and preference for support when faced with such challenges.

METHODS We carried out a cross-sectional survey using web-based and paper questionnaires. Questions were designed using examples of ethical challenges identified in the previous literature. The study involved 3 UK university medical schools. All incorporate ethics teaching programmes in problem-based learning curricula. Participants were 732 (30% of total) senior undergraduate medical students learning within the clinical environment.

RESULTS Students regularly experienced situations in clinical teaching settings that challenged their ethical values. Despite self-reports of good levels of confidence in their knowledge of ethical principles, medical students reported low levels of confidence in their ability to address these challenges, and perceived a need for additional support from clinical teachers.

CONCLUSIONS Complex and ethically challenging situations occur commonly in medical education. Many students feel that they do not currently access sufficient support from staff to address these. Clinical teachers were identified as the most relevant providers of guidance. The nature of medicine and its delivery makes it highly likely that medical students will come into contact with ethically challenging situations. Appropriate educational provision therefore requires medical educators to be equipped with the knowledge and the skills to engage with students' ethical concerns.

[1]School of Medicine, University of Manchester, Manchester, UK
[2]North West Deanery, Manchester, UK
[3]School of Psychological Sciences, University of Manchester, Manchester, UK
[4]School of Medicine, Newcastle University, Newcastle-upon-Tyne, UK

Correspondence: Lis Cordingley PhD, School of Medicine, University of Manchester, Room 2.522 Stopford Building, Oxford Road, Manchester M13 9PT, UK. Tel: 00 44 161 275 7799;
E-mail: Lis.Cordingley@manchester.ac.uk

KEYWORDS cross-sectional studies; questionnaires; *professional practice; education, medical, undergraduate/*methods/ethics; curriculum; multicentre study[publication type]; problem-based learning; ethics, medical; Great Britain.

*Medical Education 2007: **41**: 1202–1209*
doi:10.1111/j.1365-2923.2007.02943.x

INTRODUCTION

Ethics in undergraduate education

Increasing awareness of public and professional expectations of medical practitioners' behaviours has led to greater scrutiny of how medical students are taught. In order to prepare future doctors for their professional roles, bodies such as the American Medical Association have highlighted the importance of teaching ethics[1] and the General Medical Council (GMC) has identified ethical behaviour as a 'core component' of undergraduate education.[2] Explicit in the GMC's aims is an expectation that students will increase their knowledge of ethical ideas, as well as improve their ability to understand and analyse ethical problems.[3] A core curriculum of ethics was set down by the Consensus Committee of Medical Ethicists in 1998 to produce 'doctors who will engage in good ethically and legally informed practice'.[4] (p 147) Despite this acknowledgement, provision and assessment of ethics education varies greatly

Overview

What is already known on this subject

Ethics teaching is a formal curriculum requirement for medical education in many countries. North American researchers have identified a decline in the moral reasoning and ethical sensitivity of students as they progress through medical school.

What this study adds

UK medical students frequently face complex situations in clinical learning settings which challenge their ethical values and understanding. Students reported low confidence in their ability to address ethical challenges and expressed a need for additional input, particularly from qualified doctors.

Suggestions for further research

Longitudinal studies are needed to identify whether students' beliefs as well as their behaviour change during the course of undergraduate medical education.

between undergraduate medical programmes, even within the same country.[5,6]

Ethics learning and the hidden curriculum

Although increased attention has been paid to the teaching of ethics, the impact upon student learning is less clear. Studies have demonstrated that formal ethical *knowledge* is largely retained by medical students over the course of their studies,[7] but there is evidence indicating that this does not necessarily translate into increased levels of moral reasoning. Expected improvements in this area tend not to occur in medical students, unlike others engaged in professional education, a phenomenon that has been termed 'moral levelling'.[8] A study carried out in one US medical school found that although 70% of students felt their personal code of ethics had not and would not change, the number of students who thought that derogatory comments made by doctors about patients were appropriate rose from 24% in Year 1 to 55% in Year 4.[9] Moral levelling has also been reported in Canadian universities, despite the fact that bioethics is taught in every Canadian medical school.[10,11] One study found that senior students raised fewer and different concerns than junior students around issues such as communication and respect toward patients and the authors concluded that students demonstrate a decrease in moral sensitivity with time.[12]

Ethically challenging situations

Formal ethics education is only a small part of the overall process of moral education and development[13] and students frequently witness events that raise ethical concerns. These are sometimes called 'ethical dilemmas', a term that implies a dichotomy of choice or evaluation. In fact, situations that challenge students' values are often more complex than this term indicates and may require the application of different principles, significant knowledge or considerable prior experience. A more fitting phrase may be 'ethically challenging situations'. A study of the experiences of 108 Canadian medical students identified 3 categories of ethically challenging situations: conflict between the priorities of medical education and patient care; responsibility beyond a student's capacities, and involvement in patient care perceived to be substandard.[14] Further studies have identified poor communication, objectification of patients, lack of accountability and physical harm, lapses in level of care (under- or over-treatment), lack of respect demonstrated toward patients, and failure to maintain professional norms.[12,15]

These ethical challenges have an impact on students. A study of 665 clinical students found over two-thirds felt regretful or guilty about something they had done as a student and 62% believed their ethical principles had been eroded over the course of their studies. Students were more likely to report ethical erosion if they had behaved unethically to fit in with the team or for fear of a poor evaluation.[16]

Several researchers have argued that students need support in ethically challenging experiences during their clinical years.[17,18] However, it has been suggested that these situations are seldom discussed or resolved with clinical teachers[14] and young doctors fail to report ethical disagreements to senior doctors.[19] There is no clear picture about what sources of support are available to medical students in these situations, but research suggests that only a minority of students make use of the support available.[20]

Aims

The aim of this study was to explore medical students' experiences of ethically challenging situations during the latter stages of undergraduate medical education. In addition, we investigated how confident students felt in their knowledge of key ethical and legal issues and who they would view as appropriate sources of support in this aspect of their education.

METHODS

Design

A cross-sectional survey of medical students from 3 UK medical schools was undertaken using a web-based questionnaire designed for the study.[21] Additional data were collected using hard copies of the same questionnaire. Questions addressed whether students had experienced a number of different types of ethically challenging situations identified in previous studies of medical students.[9,12,14–16,20,22–25] These covered 4 main areas: experiences that occur as part of the clinical environment; experiences specific to the role of a student; students' direct experiences, and bullying behaviours (Table 1). The questionnaire consisted mainly of closed questions regarding how frequently students experienced such challenges, their confidence in addressing them, and their perceived knowledge of core ethical and legal terms. In addition, students were given the opportunity to describe their experiences of ethically challenging situations, useful educational experiences, sources of support used in ethical challenges and who they thought appropriate to turn to for support in ethical challenges. The questions were piloted at one of the participating institutions and necessary amendments were made. Table 2 shows examples of questions used.

Sample and context

All students in Years 3–5 studying on 5-year undergraduate medicine programmes at 3 UK medical schools were invited to respond during a 2-month period in 2005. Years 3–5 were selected because this is the period when students experience the

Table 2 Examples of questions included in the survey

1 How often did you experience a clinical situation that challenged your ethical values in the *most recent week* of your clinical training?
2 How confident do you feel in challenging clinical situations which you think are ethically questionable?
3 Compared with what you receive now, how much support do you feel you need to understand and cope with ethical issues that arise in clinical situations?
4 Who do you feel would be most appropriate to provide support?
5 How confident are you that you can act ethically if you suspect that a colleague is abusing prescription drugs?
6 How confident are you that you would be able to take *valid* written informed consent from a patient?

Table 1 Types of ethically challenging situations experienced or witnessed by 732 medical students in Years 3–5 of 3 UK medical schools

	% (n)
Experiences occurring within the clinical environment	
A patient being ignored or treated as if he or she were not in the room	64.9 (475)
A procedure performed on a patient who you think cannot truly consent (e.g. too young, sedated, has a mental disorder)	37.7 (276)
A request to write, or actually writing, in the notes retrospectively	27.9 (204)
A patient being misled	21.9 (160)
A professional breaking confidentiality (e.g. giving information to a patient's family without consent)	19.4 (142)
Students' direct experiences	
A situation which challenged your ethical values	74.2 (537)
Hearing derogatory comments about patients	75.0 (549)
A situation in which you acted unethically	12.7 (93)
Experiences specific to the role of a student	
A student doctor present in an interview without the patient's consent	77.5 (565)
A patient in pain examined for educational benefit	56.6 (414)
A student doctor performing an examination without the patient's true consent	47.0 (344)
Requests for student doctors to take on responsibility greater than their skill	26.9 (197)
Bullying behaviours	
A professional using destructive innuendo and sarcasm	40.6 (297)
Persistent unjustified criticism and monitoring of your work	20.2 (148)
Physical violence by a professional towards you, another professional or a patient	4.8 (35)

majority of clinical teaching. In all 3 medical schools, ethics teaching is delivered across the whole of the 5 years by specialist ethicists (mainly non-clinical), medical educators (e.g. in communication skills settings) and clinical teachers, with the latter group concentrated in the later years of each programme. Ethics topics are integrated into problem-based learning (PBL) cases and all programmes are supported by additional lectures. (One of the participating schools incorporates significantly more lectures into its 5-year programme.) Responses were received from 773 of a possible 2458 students; 732 of these responses were useable (30%) (Table S1).

RESULTS

Types of challenges and frequency of experience

On average students reported having experienced 6 of the 15 different situations listed in the questionnaire (n = 732; range 0–15) (Table 1). There were no significant differences in reported frequency between male and female students. There was a statistically significant effect of year of study on the number of reported experiences ($F[2,729] = 36.4$). Planned comparisons indicated that students experienced more of the situations in later years of study (Table 3), possibly reflecting the increased time senior students spent in the clinical environment. No differences in the number of experiences were found between the 3 participating schools apart from among Year 5 students. Students at one school experienced significantly more of the situations (mean of 8.1 compared with 6.8 and 6.6; $F[2,178] = 6.18$, $P = 0.003$).

Table 3 Mean number of reported ethically challenging situations experienced (out of possible 15 listed) by year of study and sex of student

Sex (n)		
Male (250)	6.3	$P = 0.055$
Female (479)	5.9	
Year of study		
Year 3 (288)	5.1	$P \leq 0.01$ for all contrasts
Year 4 (263)	6.4	
Year 5 (181)	7.1	

Half the sample had experienced an ethically challenging situation in their most recent week of clinical training. A total of 50% of respondents experienced these at least once per month and 35% reported that such experiences occurred at least weekly. A total of 15% of students said they had never experienced an ethically challenging situation in a clinical setting; however, when presented with the situations listed in the questionnaire (Table 1), 95% (104) of these same students said they had experienced or witnessed one or more of these situations, which perhaps suggests that some students did not view some of these experiences as representing ethical challenges.

A total of 475 students (65% of the sample) had witnessed a patient being ignored or treated as if he or she were not in the room. Their free text comments expand on this (Table 4).

Potentially more serious was the deceit witnessed by one student:

> 'Registrar telling a patient that they would attempt to resuscitate her and then writing DNR in notes. Went unchallenged by the SHO [senior house officer] as well.' (Respondent 412; male, aged 22 years)

Knowledge and confidence

Perceived knowledge of key ethical and legal principles was reported as good: 95% of students felt they had a clear understanding of the concepts of *consent* and *confidentiality* (Table S2). By contrast, 52% of respondents (n = 373) rated their confidence in

Table 4 Examples of students' free text comments about ethically challenging situations they experienced during clinical teaching

'A consultant who persisted in having a female patient remove her bra (she had previously had her breast removed) despite her request to keep it on, for the purposes of a ''teaching'' session on chest examination in front of 4 medical students' (Respondent 754; female, aged 24 years)

'An elderly female patient being unnecessarily exposed for teaching purposes and ignored by the teaching professional when she said she was tired' (Respondent 706; female, aged 21 years)

'Consultant breaking bad news to a female patient on her own, on a ward round in front of 8 male student doctors, and a male SHO' (Respondent 58; male, aged 22 years)

'A doctor shouting at a patient who wanted more information. Consultants telling women to remove their tops in front of a group of students when they are clearly uncomfortable' (Respondent 301; female, age unknown)

'A doctor communicating almost entirely with the relative of an elderly patient who would have been quite capable of participating in her own health care decisions if given the chance to do so' (Respondent 315; female, aged 21 years) dealing with ethically challenging situations as ≤ 2 on a 5-point Likert scale.

Sources of support

Students used a variety of sources of support to understand and cope with the ethical issues that arise in clinical situations (Table 5). Most (90%) discussed ethically challenging situations with fellow students and 80% asked clinical teachers for support or guidance. Around 60% consulted written materials, but less than 10% consulted ethicists or legal specialists. A total of 61% (n = 434) felt they needed more support in addressing ethical challenges than they currently received and over half (53%) felt that the most appropriate people to provide support and guidance would be clinicians, with 20% identifying expert ethicists as the most suitable sources (see Table 6 for examples of students' reasoning). This seemed to reflect the ways in which students differentiated between the support they assumed they would receive from different professionals, for example:

> 'People who can give you a practical (rather than hypothetical) approach to ethics, and … who can guide you on how to deal with ethically difficult situations.' (Respondent 413; male, aged 21 years)

We should not be surprised that students want and expect support and guidance to come from the professional grouping with which they identify and which they expect to join. However, one respondent qualified this, saying that:

> '[The] problem often lies with the clinicians we are shadowing! Educate them about what a student should and should not be asked to do.' (Respondent 614; female, aged 24 years)

Others were more concerned with the focus of support rather than the background of the professionals and mentioned issues such as how to respond to pressure from professionals to behave unethically:

> '[It]… is not a lack of support in trying to understand what, ethically speaking, is the best thing to do, but what to do when one is expected to do something one believes to be ethically wrong.' (Respondent 594; male, aged 24 years)

Table 5 Sources of support used by students in ethical challenges

Source of support	*Number of students using this source of support*	*Percentage of students using this source of support*
Peers		
Medical student	635	87.2
Non-medical friend	378	52.5
Professionals active in clinical environment		
Professional involved	588	80.8
Clinician on firm	583	80.9
Private study		
Literature	427	59.3
Study notes	305	42.4
Electronic sources	26	3.6
Professional independent from clinical environment		
Tutor	409	56.9
Mentor	204	28.6
Expert in ethics		
Ethicist/bioethicist	69	9.6
Consultant in clinical ethics	48	6.7
Other	65	8.9

DISCUSSION

Ethical challenges in the clinical context are common. On average, respondents had experienced 6 of the examples of ethical challenges identified in previous

Table 6 Free text responses to the question of who students consider to be the most appropriate sources of support

Involvement with students

'Doesn't matter who. The most important thing is to encourage students not to be afraid of speaking up if they believe something is unethical. Anyone who could provide students with the confidence to do this would be useful' (Respondent 185; female, aged 21 years)

'A good mixture would be best. I don't expect anyone to have all the answers to everything!' (Respondent 104; male, aged 30 years)

'Someone involved in the field who is willing to involve the students rather than just give their opinion' (Respondent 285; female, aged 23 years)

Independence

'Someone with experience of the problems you face, so probably a senior doctor, but who is not linked/attached to the firm/team involved in the problem' (Respondent 309; male, aged 30 years)

'Need a no-threat environment to discuss situations which have arisen. Should be someone external to the trust and medical school' (Respondent 412; male, aged 22 years)

Practical over theoretical

'Clinicians – not ethicists as they don't appreciate what these situations are really like and I find [them] to be very unhelpful and somewhat patronising. We need practical advice and support, not wishywashy sentiment' (Respondent 414; male, aged 22 years)

'Interesting professionals [sic] with real situations instead of theoretical rubbish' (Respondent 437; male, aged 22 years) studies, such as patients in pain being examined purely for educational benefit or procedures being performed on patients who did not consent,[12,14,16] and over half the students reported that one or more such events had occurred in their most recent week of clinical training. Furthermore, the number and type of challenges increased with each year of study, which probably reflects the greater amount of time spent in clinical environments.

We recognise that students' views of incidents identified as ethically challenging in this study may not be shared by the clinical teachers present. Furthermore, students may have been unaware of mitigating factors that would enable them to reattribute the actions of the clinicians. However, it is students' *perceptions* of events that are retained and recorded, and, without the opportunity for discussion or clarification, their perceptions will influence their understanding of the values embodied in the clinical environment. What this survey makes clear is that large numbers of medical students are regularly exposed to clinical teaching situations that challenge their own ethical frameworks but which could provide valuable learning opportunities.

Although the gap between students' values and what they see in practice has been described as 'disastrous',[26] the question of whether witnessing ethically challenging behaviour in clinical settings contributes directly to moral levelling in medical students is still unanswered. For the majority of medical students, entry into clinical settings creates uncertainty about roles and behaviours, a prerequisite for conformity. Furthermore, the psychological rewards of conforming to the behaviours of senior staff are strong. Students may think that failing to act like colleagues or superiors may be socially or professionally hazardous, and that the costs of challenging established practices could leave them open to risks of social exclusion or professional marginalisation.

Our finding that medical students' confidence in their knowledge of key ethical areas was high, but that their confidence in their ability to deal with ethical challenges was low, indicates that medical schools need to go beyond their obligations to deliver a formal ethics curriculum and to ensure staff model appropriate attitudes and behaviours.[2] The analysis of students' experiences uncovered here emphasises the need to provide students with opportunities to discuss and reflect upon critical incidents with experts and with those individuals that students identify as credible role models. This aspect of personal and professional development (PPD) has been highlighted as an important way of counteracting the most damaging consequences of exposure to unacceptable behaviour:

> 'If medical schools rely on the kind of PPD that simply happens along the way, they may risk allowing the hidden curriculum to prevail. Reliance on role models as the sole means of teaching runs the risk of perpetuating current problems, if students observe that some doctors have dubious ethical values…'[27]

Knowing what to teach is not the same as knowing how to teach it. Research into the most effective methods with which to optimise the ethical development of the next generation of doctors should be a research priority.[28,29]

Knowledge and understanding of ethical and legal issues is just one part of the ethical toolkit that students require to become ethical and competent practitioners. Students also need to develop the ability to judge how and when to apply ethical principles, who they can rely on for professional guidance, and how to negotiate decisions where professional colleagues disagree. Our findings reinforce the point that clinical behaviours are subject to a number of significant influences, such as the group norms of peers and senior colleagues, social and cultural expectations such as the right to question the practice of others, and the degree of perceived support from colleagues. Students may benefit from more explicit teaching about the social influences on their professional behaviour and from opportunities to analyse and reflect upon these in clinical teaching settings.

The findings of this survey indicate that medical students view clinicians as their most appropriate sources of support, which reflects what we often hear from students directly. Although they were aware of the difficulties of engaging with inappropriate role models, students tended to view the embedded nature of clinical experience as more useful than knowledge provided by ethicists alone. This has important implications for curriculum design and delivery as it requires clinical teachers to be able to analyse medical practices from an ethical as well as a professional viewpoint. The focus of academic ethicists may need to shift from direct student contact to developing and supporting the ethical expertise of clinical teachers. Perhaps more importantly, findings such as these put the onus upon medical schools to monitor students' clinical experiences closely and create explicit learning opportunities that provide a clear counter-weight to unethical practices witnessed by students via the hidden curriculum.

Limitations and directions for future research

Although the sample size of 732 (615 via web response; 117 via hard copy response) in this study was high, the potential number of respondents was 2458, which leaves a response rate of 30%. This partially reflects the limited time period in which students were accessible to our researcher. Although this may limit the generalisability of the reported proportions of students who experienced ethically challenging situations, the actual numbers of experiences reported still presents a major challenge to medical educators. One advantage of the current study over previous research was that students were asked about their experiences during the previous week of clinical teaching. Using this approach is likely to deliver a more reliable indicator of the frequency of exposure to ethically challenging situations than some earlier studies, which relied upon student recall of events that occurred up to a whole year prior to the study.[14,23,24]

The cross-sectional nature of the current study limited our ability to tease out developmental changes suggested by previous research. Although we found some differences between year groups, more longitudinal studies of medical students are needed to identify whether their beliefs as well as their behaviours change during the course of training.

Research from social psychology informs us that when all other members of a group make a judgement that differs from that of an individual, the individual can respond in one of two ways to the discrepancy: she can question her own judgement and begin to doubt her previously held views ('What *is happening* to me?'), or she can worry about the social consequences of not conforming ('What *will happen* to me?').[30] Thus, more research is needed to identify whether exposure to perceived unethical practice changes students' internal beliefs, or whether their internal beliefs remain constant and students take active decisions to behave in ways that imply conformity to the group norm.

CONCLUSIONS

Over 10 years ago, Hafferty and Franks[13] posed the question of whether medical ethics is best viewed as a body of knowledge or, alternatively, as a part of professional identity. The special nature of medical education means that it is probably both. Evidence from this and other studies shows that medical students will witness events that challenge their own ethical values and it is unrealistic to assume this will end. Therefore, the challenge that medical schools face is to provide clear opportunities for students to scrutinise their own professional development in the light of such experience, and to make explicit an expectation that all individuals involved in clinical teaching should question their practice from ethical as well as technical standpoints.

Contributors: the study was designed by CB, LC, CH and SP. All authors contributed to the design of the questionnaire. The first draft of the paper was written by LC and CH. CH collected the data. CB, SP and BV contributed to subsequent drafts.
Acknowledgements: the authors thank Urara Hiroeh for statistical advice, and Simona Giordano and Demian Whiting for advice on the pilot version of the questionnaire.
Funding: none.
Conflicts of interest: none.
Ethical approval: at the time the research was conducted there was no institutional requirement for formal ethics approval. As there was no patient involvement, local ethics committees were not approached. However, the authors were keen to ensure the ethical treatment of all participants, and therefore senior staff with responsibility for students at each of the participating institutions were informed of the project in order that they could be approached should the survey raise issues that participants wanted to discuss.

REFERENCES

1 American Medical Association. *Code of Medical Ethics: Current Opinions with Annotations, 2006–2007.* Chicago, USA: AMA 2006.
2 General Medical Council. *Tomorrow*'s Doctors: Recommendations on Undergraduate Medical Education. London: GMC 2003.
3 Doyal L, Gillon R. Medical ethics and law as a core subject in medical education. *BMJ* 1998;**316** (7145):1623–4.
4 Hope T. Ethics and law for medical students: the core curriculum. *J Med Ethics* 1998;**24** (3):147–8.
5 Fulford K, Yates A, Hope T. Ethics and the GMC core curriculum: a survey of resources in UK medical schools. *J Med Ethics* 1997;**23** (2):82–7.
6 Goldie J. Review of ethics curricula in undergraduate medical education. *Med Educ* 2000;**34**:108–19.
7 Townsend AH, Mcllvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relation-

ship to final medical school examination performance. *Med Educ* 2001;**35** (9):841–6.
8 Self D, Schrader D, Baldwin D, Wolinsky F. The moral development of medical students – a pilot study of the possible influence of medical education. *Med Educ* 1993;**27**:26–34.
9 Satterwhite RC, Satterwhite WM III, Enarson C. An ethical paradox: the effect of unethical conduct on medical students' values. *J Med Ethics* 2000;**26** (6):462–5.
10 Herbert P, Meslin E, Dunn E. Measuring the ethical sensitivity of medical students – a study at the University of Toronto. *J Med Ethics* 1992;**18**:142–7.
11 Patenaude J, Niyonsenga T, Fafard D. Changes in students' moral development during medical school: a cohort study. *CMAJ* 2003;**168**:840–4.
12 Bissonette R, O'Shea R, Horwitz M, Route C. A data-generated basis for medical ethics education: categorising issues experienced by students during clinical training. *Acad Med* 1995;**70** (11):1035–7.
13 Hafferty F, Franks R. The hidden curriculum, ethics teaching, and the structure of medical education. *Acad Med* 1994;**69** (11):861–71.
14 Hicks LK, Lin Y, Robertson DW, Robinson DL, Woodrow SI. Understanding the clinical dilemmas that shape medical students' ethical development: questionnaire survey and focus group study. *BMJ* 2001;**322** (7288):709–10.
15 Ginsburg S, Regehr G, Stern D, Lingard L. The anatomy of the professional lapse: bridging the gap between traditional frameworks and students' perceptions. *Acad Med* 2002;**77** (6):516–22.
16 Feudtner C, Christakis D, Christakis N. Do clinical clerks suffer ethical erosion? Students' perceptions of their ethical environment and personal development. *Acad Med* 1994;**69**:670–9.
17 Kenny NP, Mann KV, MacLeod H. Role modelling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Acad Med* 2003;**78** (12):1203–10.
18 Paice E, Heard S, Moss F. How important are role models for making good doctors? *BMJ* 2002;**325**:707–10.
19 Shreves J, Moss A. Residents' ethical disagreements with attending physicians: an unrecognised problem. *Acad Med* 1996;**71** (10):1103–5.
20 Huijer M, van Leeuwen E, Boenink A, Kimsma G. Medical students' cases as an empirical basis for teaching clinical ethics. *Acad Med* 2000;**75** (8):834–9.
21 Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ* 2004;**328** (7451):1312–5.
22 Coldicott Y, Pope C, Roberts C, Nesheim B-I, MacDougall J. The ethics of intimate examinations – teaching tomorrow's doctors. *BMJ* 2003;**326** (7380):97–101.
23 Koh Y. Residents' preparation for and ability to manage ethical conflicts in Korean residency programmes. *Acad Med* 2001;**76** (3):297–300.
24 Quine L. Workplace bullying in junior doctors: questionnaire survey. *BMJ* 2002;**324** (7342):878–9.
25 Silver-Isenstadt A, Ubel PA. Erosion in medical students' attitudes about telling patients they are students. *J Gen Intern Med* 1999;**14** (8):481–7.
26 Doyal L. Closing the gap between professional teaching and practice. *BMJ* 2001;**322**:685–6.
27 Gordon J. Fostering students' personal and professional development in medicine: a new framework for PPD. *Med Educ* 2003;**37**:341–9.
28 Eckles R, Mesline E, Gaffney M, Helft P. Medical ethics education: where are we? Where should we be going? *Acad Med* 2005;**80** (12):1143–52.
29 Goldie J, Schwartz L, Morrison J. A process evaluation of medical ethics education in the first year of a new medical curriculum. *Med Educ* 2000;**34** (6):468–73.
30 Haslam S. *Psychology in Organisations: the Social Identity Approach*, 2nd edn. London: Sage Publications 2004.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Table S1. Description of sample.

Table S2. Knowledge of ethical ideas.

This material is available as part of the online article from:

http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2923.2007.02943.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 27 March 2007; editorial comments to authors 15 June 2007; accepted for publication 27 July 2007

Postmortem intubation training: patient and family opinion

Georgene W Hergenroeder,[1] Bettina C Prator,[1,2] Alice F Chow[3] & David J Powner[1]

OBJECTIVE Endotracheal intubation is a life-saving skill which requires training to master. Learning opportunities for endotracheal intubation must be balanced with patient rights and intentions. This study was conducted to explore patient and family opinions about postmortem endotracheal intubation training.

METHODS We carried out an observational, cross-sectional survey study in an urban, teaching hospital, on the day of hospital discharge. Subjects were neurologically unimpaired neurosurgical patients discharged from hospital in 2004–2005, and their relatives. We carried out interviews using a standardised script to determine whether subjects would permit postmortem intubation training on themselves or relatives, and whether permission should be granted by relatives before training.

RESULTS A total of 85% of patient and family respondents would allow intubation training after death on themselves, 76% would allow endotracheal intubation to be practised on a relative, and 81% felt the deceased's next-of-kin should be asked for permission prior to endotracheal intubation training. Subjects responded consistently as to what they would allow on self and family. Knowledge that the deceased person would have agreed to his or her body being used in endotracheal intubation training increased their likelihood of granting permission for training ($P = 0.008$). White subjects were 4.6 times more likely than non-Whites to allow intubation training on themselves ($P = 0.01$).

CONCLUSIONS Patients and families are agreeable to postmortem intubation training; however, most expect to be asked for permission. Utilising existing mechanisms which communicate desired treatment, such as advance directives, hospital admissions documents, donor registries or community health fairs may facilitate training opportunities and altruistic patient intentions.

KEYWORDS cadavers; intubation/*standards; teaching/*methods; *teaching materials; patient satisfaction; education, medical/*methods; cross-sectional studies; urban health; hospitals, teaching; family health; humans.

*Medical Education 2007: **41**: 1210–1216*
doi:10.1111/j.1365-2923.2007.02936.x

INTRODUCTION

Endotracheal intubation is a life-saving skill that requires training to master. Traditional teaching techniques include using manikins, anaesthetised animals, patients during induction of general anaesthesia, and specially prepared cadavers. Each of these methods imposes limitations in availability, anatomic accuracy, patient safety, or cost. Two studies found no difference in the ability of paramedics to intubate when trained on a manikin with or without added experience on patients or cadavers.[1,2] However, the availability of multiple training techniques enhances learning.

Supervised intubation practice on newly deceased patients is another training option to provide safe, anatomically accurate instruction and has been used in 39–54% of residency programmes.[3,4] Its utilisation

[1]Department of Neurosurgery, University of Texas Health Science Center at Houston, Houston, Texas, USA
[2]Division of General Surgery, University of South Florida College of Medicine, Tampa, Florida, USA
[3]Richard E. Smalley Institute for Nanoscale Science and Technology, Rice University, Houston, Texas, USA

Correspondence: Georgene W Hergenroeder, UTHSC-H, Department of Neurosurgery, 6431 Fannin Street, MSB 7.156, Houston, Texas 77030, USA. Tel: 00 1 713 500 6130; Fax: 00 1 713 500 7787;
E-mail: georgene.w.hergenroeder@uth.tmc.edu

Overview

What is already known on this subject

Postmortem intubation training provides safe, anatomically accurate instruction. Criticisms refer to issues of respect, privacy and consent.

What this study adds

Patients and families are agreeable to postmortem training; however, most expect to be asked for permission. Knowledge that the deceased person had agreed to his or her body being used in training increased the likelihood of granting permission.

Mechanisms should be instituted to respect patient decisions and obtain permission. Utilising existing instruments which communicate the patient's desired treatment may facilitate training and altruistic intentions.

Suggestions for further research

Research should determine which interventions succeed in obtaining consent in order to increase opportunities for training.

has been criticised because of ethical concerns about maintaining respect for the dead, assuring the privacy rights of the deceased patient, and assessing the need for acquiring informed consent from the patient through advance directives or from the family.[4–11] The legal ramifications of not seeking consent have also been explored, and include the possible violation of privacy or property rights, the infliction of emotional distress, and mistreatment of a corpse.[5,6,12]

Several published studies have evaluated public attitudes toward performing procedures on the newly deceased and current teaching hospital practices. In a general population survey, 58% of respondents were in favour of postmortem training procedures.[13] A total of 32% of participants attending a community social event supported procedures being carried out without consent, but 74% of participants wanted prior family approval.[14] Permission was granted by 59% of families of newly deceased adults[15] and 73% of families of infants in a neonatal intensive care unit.[16] Our hospital policy does not permit postmortem intubation training without permission from the family. In 2 emergency department surveys, half of those surveyed reported having had training on newly deceased patients, but hospital policy had been present in only 4–7% of cases, and family permission had been required in 1–3%.[3,17] Another study reported that consent was regarded as unnecessary by 40% of emergency department patients and families surveyed.[18] A total of 39% of families granted permission to perform a similar, albeit more invasive, procedure, cricothyroidotomy, on newly deceased patients.[19]

There is minimal literature regarding patients' opinions of the practice of endotracheal intubation on themselves. Therefore, we elicited opinions from individuals about the theoretical potential for training after their own death. The opinion of family members was obtained because advance directives rarely address this issue. It has been recognised that individuals may provide one opinion when questioned about medical procedures for themselves, yet a different opinion when asked to decide on behalf of a family member.[20] Substituted consent should reflect the patient's wishes. The specific objectives of this study were to determine if:

1. patients would permit the performance of supervised endotracheal intubation training if they had died during the current hospitalisation;
2. patients would permit the same practice for a family member following an in-hospital death;
3. family members would permit supervised intubation training if the interviewed patient had died during this hospitalisation;
4. family members would permit supervised intubation training on themselves if they were to die in a hospital, and
5. patients or family members believe that consent from next-of-kin should be obtained by health care personnel prior to postmortem intubation training.

METHODS

Institutional review board approval, permission from attending neurosurgeons and written informed consent was obtained prior to enrolling subjects in this study. A convenience sample of 66 neurologically unimpaired neurosurgical patients and their 42 family members were interviewed independently on the day of patient discharge from hospital using a standardised script (see Fig. S1). This patient group was selected because of its recent experience in a hospital environment. Likewise, family members had recent experience with hospitalised relatives. Subjects were interviewed during days when research nurses were available until data from 66 patients and accompanying family members had been accrued. The research nurses had no prior relationship with the patients.

Nursing assessments of the patients' orientation to person, place and time, and emotional status served as the mechanism to establish patient neurological status and the patients' ability to answer questions. Standardised interviews and the informed consent process were conducted by 4 trained nurse interviewers. Family members were contacted by the interviewer only after the patient granted permission, and were not interviewed if the patient was excluded, or declined to participate. Patient data were included even if no family member was available to participate. A family member was interviewed only if he or she was related by lineage or marriage to the patient and might, under appropriate circumstances, be a surrogate decision-maker for the patient. In 3 cases 2 family members were interviewed. Patients were excluded if they were discharged in a terminal medical condition, confused and unable to answer questions, < 18 years old, or non-English speaking. After the standardised interview was complete, subjects were given the opportunity to make additional comments; these comments were qualitatively coded according to theme.

Chi-square analysis (Minitab 14 Statistical Software; State College, PA, USA) was performed on independent predictors of allowing endotracheal intubation training. This was an opinion survey where the opinions of the 2 respondent groups, patients and family, were elicited. The focus was a comparison of responses of un-paired patients and non-patients. Three dichotomous dependent variables were considered:

1. if respondents would permit endotracheal intubation training on self;
2. if respondents would permit endotracheal intubation training on family members, and
3. if respondents felt permission should be obtained prior to performing endotracheal intubation training.

Factors associated with allowing postmortem endotracheal intubation training were considered statistically significant at $P < 0.05$. Binary logistic regression analysis was performed to provide an adjusted odds ratio with a 95% confidence interval for predicting outcome variables.

RESULTS

Of the patients approached for consent, 84% (66/79) agreed to participate. Demographic data are listed in Table 1. Endotracheal intubation training would be permitted by 85% of both patients and family members, and 76% were willing to allow their relatives to be subjects of endotracheal intubation training (Table 2). White subjects were 4.6 times more likely than non-Whites to allow intubation training on themselves (71/77 White versus 21/31 non-White subjects agreed; $P = 0.01$).

Subjects were 26 times more likely to agree to training on both self and a relative than to choose differently for the relative ($P < 0.01$). Specifically, if a subject agreed to allow training on himself, he tended to agree for the practice to be carried out on his relative as well. Subjects who declined to allow endotracheal intubation training to occur on their relative were 2.8 times more likely to allow training when the question was qualified with the statement, 'If your relative had made his or her wishes known to you that he or she had no objections to this procedure, would you then permit this to occur?' ($P = 0.008$). Table 3 illustrates additional odds for how family members compared with patients would answer.

Family members were available at hospital discharge for 59% (39/66) of patients. Of patients who would agree to allow training, 48% (27/56) had available family members who would also agree to allow training.

Of all responders, 81% felt the deceased's next of kin should be asked for permission prior to endotracheal intubation training. Subjects aged 19–40 years were more likely than those aged > 40 years ($P = 0.02$) to think the deceased's kin should be asked for permission prior to endotracheal intubation training. Of the 40 documented comments, 25% (10/40) related to consent or request for permission and 30% (12/40) concerned issues of respect. Additional comments related to timing and sensitivity of requests (3), organ donation (3), self-determination (3), defacing the body (3), prior precedents (2), research (2), culture (2), age of the deceased (1), value of training (1), and expectations of students at a training institution (1).

Table 1 Demographics of the study group

Description (*n* = 108)	***Combined*** *n*	%	***Patients*** *n*	%	***Family*** *n*	%
Gender						
Female	56	52	27	41	29	69
Race/ethnicity						
White	77	71	48	73	29	69
African-American	19	18	12	18	7	17
Hispanic	9	8	4	6	5	12
Asian	2	2	1	1.5	1	2
Other	1	1	1	1.5	0	0
Age						
≤ 40 years	27	25	18	27	9	21
> 40 years	81	75	48	73	33	79
Average ± standard deviation			48 ± 14.8		49 ± 11.8	
Primary discharge diagnosis (*n* = 66)						
Laminectomy/laminotomy			19	28.8		
Cervical discectomy and fusion			13	19.7		
Other			7	10.6		
Subarachnoid haemorrhage			7	10.6		
Shunt revision or placement			4	6.1		
Traumatic brain injury			4	6.1		
Craniotomy			3	4.5		
Spinal cord injury			3	4.5		
Bifrontal contusion			2	3.0		
Arteriovenous malformation			2	3.0		
Seizures			2	3.0		

Table 2 Responses by family and patient group

Questions	***Combined response***	%	***Patient response***	%	***Family response***	%
Willingness to allow endotracheal intubation training on self	92/108	85	56/66	85	36/42	86
Willingness to allow endotracheal intubation training on family member	82/108	76	51/66	77	31/42	74
Willingness to allow endotracheal intubation training with knowledge that family member would permit	97/108	90	60/66	91	37/42	88
Require permission prior to endotracheal intubation training	87/107	81	51/66	77	36/41	88

Table 3 Odds ratios: family response versus patient response

Comparison groups	***Odds ratio***	***95% CI***
Willingness to allow training on relation with knowledge that he or she would permit training versus willingness to allow training on relation	2.80	1.30–6.00*
Willingness to allow training on self: family response versus patient response	1.07	0.36–3.20
Willingness to allow training on relation: family response versus patient response	0.83	0.34–2.03
Permission required: family response versus patient response	2.12	0.71–6.35
Training on patient: family allowing training on patient versus patient allowing training on self	0.50	0.19–1.32
Training on family: family allowing training on self versus patient allowing training on family	1.76	0.62–4.99

* $P < 0.01$
95% CI = 95% confidence interval

DISCUSSION

The most significant finding is that the majority of patient and family subjects would have permitted supervised postmortem intubation practice. McNamara *et al.* reported 59% of families consented to wire-guided retrograde tracheal intubation on newly deceased relatives [15] and Benfield *et al.*

reported the consent of 70% of families approached for permission to carry out endotracheal intubation on newly deceased infants.[16] Morag *et al.* studied a group in Oslo, Norway, which, although receptive to allowing training, was more restrictive of what they would allow on relatives compared with training they would permit on themselves.[20] The respondents in both the Morag *et al.* study and our own expected to be asked for permission prior to training being performed using the newly deceased. Age was significant when requiring permission prior to postmortem endotracheal intubation training. This may be a cohort effect and will need to be replicated.

White race was predictive of willingness to be intubated, which is consistent with findings in a study on neonates.[16] Morag *et al.* noted that non-Whites were half as likely to consent to the teaching of invasive techniques compared with those from Oslo, Norway (96% White).[20] The difference between Whites and non-Whites suggests a lack of trust between families in the latter category and investigators. Lack of trust in medical researchers in the USA is well recognised.[21–23] Building trusting relationships will be a necessary step to increase participation in such teaching programmes. Obtaining consent for training purposes may help maintain or build trust.[24]

We did not experience anger directed at us for our survey questions, as others have reported.[20] However, we did note cultural influences in some respondents' comments. When asked about allowing endotracheal intubation, 1 subject stated, 'It is Chinese culture; we do not do that kind of thing in our culture.' Another subject felt the practice should not occur and commented on the need to respect the dead body, whereas others expressed the opinion that the deceased no longer have any need for their bodies.

A theme of respect flowed through the subjects' comments, which included desire for respectfully asking for permission, respect for the grieving family, respect for the body, and respect for the deceased's wishes. Patients and family members generally had little difficulty in discussing their preferences regarding postmortem intubation training. Some discussions evolved to comments on organ donation. The sensitivity and timing of approach and preparation for discussion appear to be paramount in influencing participants' decisions.

Lack of family availability to answer our survey may be comparable with a lack of family availability to consent for intubation training; 1 study reported that 16% of families were too distraught to be approached and on follow-up an additional 14% of families could not be reached.[25] In our study, even when patients agreed to allow training, families in half the cases were not available or did not agree to consent to training. If these results are replicated, requiring family consent at the time of death to perform training on a cadaver will continue to be a limiting condition.

The American Medical Association's (AMA) Guidelines for Performing Procedures on the Newly Deceased for Training Purposes state: 'When reasonable efforts to discover previously expressed preferences of the deceased or to find someone with authority to grant permission for the procedure have failed, physicians must not perform procedures for training purposes on the newly deceased patient.'[25] If the AMA position is adopted universally (eliminating training on subjects when unable to obtain consent or determine patient's wishes) a significant number of patients who would have been willing to participate would be excluded, substantial training opportunities would be lost and many training programmes would be required to modify their practices.[3,17] The intent of the AMA position is to protect and respect patient rights; in response, institutions and individuals should institute mechanisms to safeguard patients, respect their decisions and obtain permission early to use the bodies of deceased patients for training purposes.

Advance medical directives, hospital admissions documents and organ donor registries are formalised mechanisms for documenting preferences. A community-based approach to obtain consent for postmortem training is also a possibility. Advance directives have been disregarded by doctors in consideration of prognosis, quality of life and family wishes;[26] however, our results suggest that family members would be more likely to allow training if they had prior knowledge of the deceased's wishes.

If hospital admissions documents are to be used as a mechanism to relay patient preferences, they need to be worded so that they are not perceived as coercive. Hospital admission can be a stressful time; routine visits would serve as an alternative time at which to document preferences. Patient willingness to be used for training could be added to organ donor registration lists. Registration systems for organ donation have been found to be effective in educating and communicating potential donor wishes to families, although authorising donation through registries does not guarantee family agreement and eventual organ donation.[27] Likewise, authorising postmortem training may not guarantee family consent. However, our results indicate that family members would be more likely to agree if they knew the deceased had stated there was no objection to the training. A community-based approach to obtaining consent for postmortem training could be implemented at health fairs or community health education sessions, leading to the establishment of a database.

There are several limitations to the study. Participants were not surveyed about their use of advance directives nor did we survey opinions on organ donation or donor registration status, so we cannot conclude what impact this would have. However, most hospitals maintain admissions documents completed at or around the time of admission, which allow an opportunity to raise the subject for discussion. We included only English-speaking subjects and a convenience sampling of neurosurgery patients, which may not be representative of the general population. Future research should determine which interventions succeed in obtaining consent from patients and families, and increase the availability of newly deceased patients for intubation training.

CONCLUSIONS

Medical community attitudes are shifting in the direction of obtaining patient or family permission prior to postmortem training. The lay community expects to be asked for permission prior to being used for training. Effective, non-coercive mechanisms to ensure training opportunities for life-saving procedures need to be established. Advanced directives, hospital admissions documents, organ donation registries, health fairs and end-of-life discussions are all avenues to promote the fulfilling of patient preferences and to facilitate the gaining of consent for training.

Contributors: GWH contributed to the study conception and design, and acquisition, analysis and interpretation of data. DJP contributed to the study conception and design, and analysis and interpretation of data. BCP contributed to acquisition of data. AFC contributed to analysis and interpretation of data. All authors contributed to the drafting and revision of the paper and approved the final manuscript.
Acknowledgements: the authors acknowledge and thank Pam Drever MS RN and Marian Skewes FNP for their assistance in gaining respondent consent to participate and in conducting interviews.
Funding: this study was funded by the Department of Neurosurgery, University of Texas Health Science Center, Houston.
Conflicts of interests: none.
Ethical approval: institutional review board approval was received from the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston.

REFERENCES

1 Stratton SJ, Kane G, Gunter CS, Wheeler NC, Abelson-Ward C, Reich E, Pratt FD, Ogata G, Gallagher C. Prospective study of manikin-only versus manikin and human subject endotracheal intubation training of paramedics. *Ann Emerg Med* 1991;**20**:1314–8.
2 Stewart RD, Paris PM, Pelton GH, Garretson D. Effect of varied training techniques on field endotracheal intubation success rates. *Ann Emerg Med* 1984;**13**: 1032–6.
3 Fourre MW. The performance of procedures on the recently deceased. *Acad Emerg Med* 2002;**9**:595–8.
4 Schmidt TA, Abbott JT, Geiderman JM *et al.* SAEM Board of Directors. Ethics seminars: the ethical debate on practising procedures on the newly dead. *Acad Emerg Med* 2004;**11**:962–6.
5 Orlowski JP, Kanoti GA, Mehlman MJ. The ethics of using newly dead patients for teaching and practising intubation techniques. *N Engl J Med* 1988;**319**:439–41.
6 Moore GP. Ethics seminars. The practise of medical procedures on newly dead patients – is consent warranted?. *Acad Emerg Med* 2001;**4**:389–92.
7 Hudson TS. Is it ethical to practise intubation on the deceased? *JONAS Healthc Law Ethics Regul* 2000;**2**:22–8.
8 Orlowski JP. Politically correct ethical thinking and intubation practice on cadavers. *J Clin Ethics* 1994;**5**:256–60.
9 Iserson KV. Law versus life: the ethical imperative to practise and teach using the newly dead emergency department patient. *Ann Emerg Med* 1995;**25**:91–4.
10 Goldblatt AD. Don't ask, don't tell: practising minimally invasive resuscitation techniques on the newly dead. *Ann Emerg Med* 1995;**25**:86–90.
11 Ardagh M. May we practise endotracheal intubation on the newly dead? *J Med Ethics* 1997;**23**:289–94.
12 Kerns AF. Better to lay it out on the table rather than do it behind the curtain: hospitals need to obtain consent before using newly deceased patients to teach resuscitation procedures. *J Contemp Health Law Policy* 1997;**13**:581–612.
13 Brattebo G, Wisborg T, Oyen N. Using newly deceased patients in teaching procedures. *N Engl J Med* 1995;**332**:1445–7.

14 Tachakra S, Ho S, Lynch M, Newson R. Should doctors practise resuscitation skills on newly deceased patients? A survey of public opinion *J R Soc Med* 1998;**91**:576–8.

15 McNamara RM, Monti S, Kelly JJ. Requesting consent for an invasive procedure in newly deceased adults. *JAMA* 1995;**273**:310–2.

16 Benfield DG, Flaksman RJ, Lin TH, Kantak AD, Kokomoor FW, Vollman JH. Teaching intubation skill using newly deceased infants. *JAMA* 1991;**265**:2360–3.

17 Morhaim DK, Heller MB. The practice of teaching endotracheal intubation on recently deceased patients. *J Emerg Med* 1991;**9**:515–8.

18 Manifold CA, Storrow A, Rodgers K. Patient and family attitudes regarding the practice of procedures on the newly deceased. *Acad Emerg Med* 1999;**6**:110–5.

19 Olsen J, Spilger S, Windisch T. Feasibility of obtaining family consent for teaching cricothyrotomy on the newly dead in the emergency department. *Ann Emerg Med* 1995;**25**:660–5.

20 Morag RM, DeSouza S, Steen PA, Salem A, Harris M, Ohnstad O, Fosen JT, Brenner BE. Performing procedures on the newly deceased for teaching purposes: what if we were to ask? *Arch Intern Med* 2005;**165**:92–6.

21 Halbert CH, Armstrong K, Gandy OH, Shaker L. Racial differences in trust in health care providers. *Arch Intern Med* 2006;**166**:896–901.

22 Doescher MP, Saver BG, Fiscella K, Franks P. Racial/ethnic inequities in continuity and site of care: location, location, location. *Health Serv Res* 2001;**36**:78–89.

23 Boulware LE, Cooper LA, Ratner LE, LaVeist TA, Powe NR. Race and trust in the health care system. *Public Health Rep* 2003;**118**:358–65.

24 Berger JT, Rosner F, Cassell EJ. Ethics of practising medical procedures on newly dead and nearly dead patients. *J Gen Intern Med* 2002;**17**:774–8.

25 Rakatansky H, Riddick FA Jr, Morse LJ *et al* Performing procedures on the newly deceased. Council on Ethical and Judicial Affairs of the American Medical Association. <http://www.ama-assn.org/apps/pf_new/pf_online?f_n=resultLink&doc=policyfiles/HnE/E-8.181.HTM&s_t=deceased&catg=AMA/HnE&catg=AMA/BnGnC&catg=AMA/DIR&&nth=1&&st_p=0&nth=7. [Accessed 24 July 2007.] *Acad Med* 2002;**77**:1212–6.

26 Hardin SB, Yusufaly YA. Difficult end-of-life treatment decisions: do other factors trump advance directives? *Arch Intern Med* 2004; **164**:1531–3. [Erratum: 2124.]

27 Wendler D, Dickert N. The consent process for cadaveric organ procurement: how does it work? How can it be improved? *JAMA* 2001;**285**:329–33.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Figure S1. Standardised postmortem intubation training interview questions.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365.2923.2007.02936.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 12 February 2007; editorial comments to authors 23 July 2007; accepted for publication 15 August 2007

Origin bias of test items compromises the validity and fairness of curriculum comparisons

Arno M M Muijtjens,[1] Lambert W T Schuwirth,[1] Janke Cohen-Schotanus[2] & Cees P M van der Vleuten[1]

OBJECTIVE To determine whether items of progress tests used for inter-curriculum comparison favour students from the medical school where the items were produced (i.e. whether the origin bias of test items is a potential confounder in comparisons between curricula).

METHODS We investigated scores of students from different schools on subtests consisting of progress test items constructed by authors from the different schools. In a cross-institutional collaboration between 3 medical schools, progress tests are jointly constructed and simultaneously administered to all students at the 3 schools. Test score data for 6 consecutive progress tests were investigated. Participants consisted of approximately 5000 undergraduate medical students from 3 medical schools. The main outcome measure was the difference between the scores on subtests of items constructed by authors from 2 of the collaborating schools (subtest difference score).

RESULTS The subtest difference scores showed that students obtained better results on items produced at their own schools. This effect was more pronounced in Years 2–5 of the curriculum than in Year 1, and diminished in Year 6.

CONCLUSIONS Progress test items were subject to origin bias. As a consequence, all participating schools should contribute equal numbers of test items if tests are to be used for valid and fair inter-curriculum comparisons.

KEYWORDS multicentre study [publication type]; comparative study [publication type]; validation studies [publication type]; *education, medical, undergraduate; clinical competence/*standards; curriculum/*standards; selection bias; Netherlands.

*Medical Education 2007: **41**: 1217–1223*
doi:10.1111/j.1365-2923.2007.02934.x

INTRODUCTION

Comparative curriculum studies have always featured prominently on the medical education research agenda. For a long time such studies have primarily been aimed at drawing comparisons between problem-based learning (PBL) curricula and non-PBL curricula. Today, however, the concept of benchmarking is gaining ground rapidly. Whatever their goal, curriculum comparisons must be valid and fair and achieving this is fraught with pitfalls. Researchers have pointed to a myriad of potential confounders and sources of error that may thwart inter-curricular comparisons.[1] Schmidt has discussed some of these in detail.[2] Researchers who value fair inter-curricular comparisons should be aware of the problem caused by the possible differential exposure of student cohorts to the measurement instrument used. The likelihood of this type of bias is a moot question. For example, it seems theoretically legitimate for several schools which share the same curricular end objectives to use the same test to assess whether students have achieved those objectives, notwithstanding the possibility that their curricula may differ.[3] It seems reasonable to use the outcomes of a test targeted at the same end objectives to make comparisons between schools. The crucial assumption that justifies the use of a test as a measurement instrument for

[1]Department of Educational Development and Research, Faculty of Medicine, Maastricht University, Maastricht, The Netherlands
[2]Institute for Medical Education, Centre for Innovation and Research of Medical Education, University Medical Centre Groningen, Groningen, The Netherlands

Correspondence: Arno M M Muijtjens, Department of Educational Development & Research, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands. Tel: 00 31 43 388 5745; Fax: 00 31 43 388 5779; E-mail: a.muijtjens@educ.unimaas.nl

Overview

What is already known on this subject

Progress tests used by different schools and pitched at identical shared curricular end objectives are assumed to be curriculum-independent and therefore suitable for drawing comparisons between curricula.

What this study adds

Although progress tests appear to be curriculum-independent, they are subject to origin bias. Items of tests produced jointly by 3 medical schools yielded better results if students and item writers were from the same school. Validity and fairness of comparisons between schools on the basis of progress test results may be enhanced by including equal numbers of items from different schools.

Suggestions for further research

Further study of the origin of this effect would be beneficial.

comparisons is that the test is 'curriculum-independent' and thus ensures unbiased assessment of all students, irrespective of any effects of potential confounders. If this condition is met, test results can be used to compare the achievements of groups of students from different schools and to make inferences about the relative success of different curricula.[4,5] Several recent initiatives have resulted in joint test item banking by different institutions.[6,7] From the perspective of such collaborative enterprises, it is highly relevant to ensure that tests are impartial and are not biased in favour of any of the participating institutions. For instance, if test items were to favour the students at the institution that produced the item, the validity and fairness of comparisons would be significantly compromised.

Progress testing is among the assessment procedures that are assumed to be curriculum-independent.[3,8,9] It involves assessing students' achievement of curricular end objectives, based on set standards, using the same tests for all the students in every year of the curriculum. Because all students sit identical tests several times a year at regular time intervals, their consecutive test scores reflect their progress through the curriculum toward attainment of the end objectives. Seven years ago, 3 Dutch medical schools embarked on a progress test collaboration which entailed joint test production and regular, simultaneous testing of the entire student population of the collaborating schools. Although the curricula of these schools differ, they must all meet the legal requirements laid down in the national end objectives of undergraduate medical education in the Netherlands.

The current study used a best-case analysis to investigate whether tests produced jointly by the different schools are truly unbiased. Specifically, we wanted to examine whether or not the origin of items introduces bias in favour of students from the same school as the item producers. For instance, if Nijmegen students score consistently higher than students from other schools on items produced by Nijmegen item writers, the test cannot be said to be unbiased.

METHODS

Context

Since September 1999, 3 Dutch medical schools (those at the Universities of Maastricht, Nijmegen and Groningen) have jointly constructed progress tests, and every academic year all students at all 3 schools have simultaneously sat the same 4 written progress tests.[3] Each test consists of 250 true/false questions that comprise a sample from the domain of relevant and functional medical knowledge a newly graduated doctor is expected to possess. A blueprint is used to stratify tests by discipline, and by disease or complaint categories. The 4 annual tests (which combine to 1000 items) provide longitudinal information, charting the growth of students' medical knowledge over the curriculum. The 3 schools use identical test dates and times, scoring methods, and standards to determine pass/fail cut score, and the test has comparable status in the regulations of the 3 schools. The scores of approximately 5000 students are obtained for each test, and stored and analysed in 1 database.

Instrumentation

Teaching staff at the 3 schools contribute to the production of test items. In order to assure test quality, review committees consisting of 6 members of staff have been set up in each school. As all items must reflect the national end objectives of undergraduate medical education,[10] test content is assumed to be curriculum-independent. Because the test is pitched at end-of-curriculum knowledge levels and all student cohorts sit the same tests, junior students are not expected to know all the answers. For this reason, a 'Don't know' option is provided, which yields no marks but enables a negative marking procedure that penalises incorrect answers by subtracting marks. This so-called *formula scoring* is used to discourage and correct for random guessing.[11,12] Test scores are expressed as the percentage of correct minus incorrect answers.

Subjects and data

We investigated the impartiality of progress test items by comparing the item scores of the students of the 3 collaborating schools (1600, 1600 and 2100 students, respectively) on the 6 consecutive progress tests administered between September 2002 and December 2003.

Analysis

Maastricht University has some 25 years of experience with progress testing. In the initial phase of the collaboration, when the schools were working towards effective collaboration and consensus on quality control, Maastricht contributed more test items to the jointly used tests than the other 2 schools. Over the years, the relative contribution of the 3 schools has equalised, but during the period of this study the main contributors were Maastricht and Nijmegen (Table 1), which contributed mean percentages of 62% and 24% of items, respectively. Because the Groningen subset of items was relatively small and the items pertained to a restricted domain, we decided to focus on the Maastricht and Nijmegen item subsets only, but to include the Groningen students in the analysis as a comparison group. We created a situation with 2 subtests of equal size by taking a random sample from the Maastricht subset of items that was equal in size to the Nijmegen subset (referred to as the Maastricht subtest and Nijmegen subtest, respectively). Thus we obtained a balanced design, which is more robust for both analysis and interpretation of the results.

Table 1 Numbers of items contributed by each of the collaborating medical schools to each of the 6 cross-institutional progress tests analysed

	Number of items			
Progress test	*Maastricht*	*Nijmegen*	*Groningen*	*Total*
Sept 2002	176	39	27	242
Dec 2002	174	53	21	248
Mar 2003	157	53	40	250
June 2003	134	84	32	250
Sept 2003	137	67	46	250
Dec 2003	143	63	43	249
Percentage	62	24	14	100

We calculated each student's percentage correct-minus-incorrect score for each of the 2 subtests (Maastricht subtest score and Nijmegen subtest score) for each of the 6 progress tests. In order to establish a possible association between test results and shared origin of test items and students, we first calculated a difference score:

$$\text{Dscore} = \text{Maastricht subtest score} - \text{Nijmegen subtest score} \qquad (1)$$

This Dscore can be interpreted as reflecting the difference in difficulty of items for students between the Maastricht and Nijmegen subtests. A positive value indicates that the Nijmegen subtest is more difficult than the Maastricht subtest. Bias resulting from item origin should be suspected when Maastricht students have higher scores on the Maastricht subtest and Nijmegen students do better on the Nijmegen subtest (i.e. high Dscores for Maastricht students and low Dscores for Nijmegen students).

Because students' test scores increase gradually over the 6-year curriculum, the Dscore was calculated for each year group of students separately.

The variables included in the analysis are Dscore, Year group and University. The effect of interest is the Dscore for the levels of the independent variable University, because it indicates whether items favour students from the same university as the item authors. However, part of the variation in Dscores may be explained by the independent variable, Year, or by the interaction of Year and University. Therefore, we performed a 2-way ANOVA using the model:

$$\text{Dscore}_{ijk} = m + a_i + b_j + c_{ij} + e_{ijk} \qquad (2)$$

where $i = 1, 2, \ldots, 6$ refers to the i-th Year; $j = 1, 2, 3$ to the j-th University (Maastricht, Nijmegen, Groningen); $k = 1, 2, \ldots, N_{ij}$ to the k-th student in the group (of size N_{ij}) of students in the i-th Year at the j-th University; m is the general mean; a_i is the main effect of Year i; b_j the main effect of University j; c_{ij} the interaction of Year i and University j; e_{ijk} is the error term, and for the effect parameters it holds: $\sum_i a_i = \sum_j b_j = \sum_i c_{ij} = \sum_j c_{ij} = 0$.

To interpret the model in equation 2, m indicates the general difference in difficulty between the Maastricht and Nijmegen subtests; a_i is the average variation of this difference for the different levels of Year; b_j represents the variation for the 3 levels of University, and, finally, the interaction parameter c_{ij} represents the extra variation beyond $a_i + b_j$ for Year i and University j.

We were interested in the association between changes in Dscore and University. This variation is primarily represented by b_j, which indicates the average pattern of Dscore across the 3 levels ($j = 1, 2, 3$) of University. However, as this pattern may not be the same for each Year, we investigated $b_j + c_{ij}$, which indicates, for each Year ($i = 1, 2, ..., 6$) the pattern of Dscore across the 3 levels of University.

Estimates of the parameters a_i, b_j and c_{ij}, and the corresponding statistical significance were obtained in a 2-way ANOVA, using SPSS Version 12.

The above analysis was performed for each of the 6 progress tests and the results were summarised by calculating the average across 6 tests for the parameters a_i, b_j and c_{ij}.

According to Cohen's definition,[13] the effect size (ES) associated with potential origin bias of test items can be expressed as

$$\mathrm{ES} = |b_1 - b_2|/\mathrm{SD}_S, \quad (3)$$

where the numerator represents the absolute value of the average difference between the scores obtained by the Maastricht and Nijmegen students if the Maastricht subtest were replaced by the Nijmegen subtest, and $\mathrm{SD_S}$ is the standard deviation of the students' scores on a test of equal size as the subtests analysed. Similarly, the effect size can be calculated per Year by substituting $|(b_1 + c_{i1}) - (b_2 + c_{i2})|$ in the numerator of equation 3.

Using the relations $\sigma^2_{Tot} = \sigma^2_t + \sigma^2_\varepsilon$, and $\sigma^2_\varepsilon = \sigma^2_t(1-R)/R$ for the test score's total variance σ^2_{Tot}, true variance σ^2_t, error variance σ^2_ε, and reliability R, and noting that the error variance changes with (1/number of items), $\mathrm{SD_S}$ can be estimated according to

$$\mathrm{SD}^2_S = \sigma^2_{\mathrm{Tot}}[R + (1-R)\cdot(N_{\mathrm{Tot}}/N_{\mathrm{S}})], \quad (4)$$

where N_{Tot} and N_{S} are the numbers of items on a regular progress test and the analysed subtests, respectively.

RESULTS

The numbers of items contributed by the 3 schools to the 6 progress tests under study are shown in Table 1. The last row shows the overall distribution of items over the 3 universities. The item sample sizes used in the analysis correspond to the number of items in the Nijmegen column.

For each of the 6 progress tests, an ANOVA was performed according to the model specified in equation 2. The majority of the analyses reveal highly significant (F-test, $P < 0.0005$) main effects and interactions of the independent variables Year and University on the Dscore (i.e. the within-student difference between the scores on the Maastricht and Nijmegen subtests). The only exception was the non-significant main effect of Year for the June 2003 test.

Table 2 shows the averages of the resulting parameters for the 6 ANOVAs. Rows 1–6 and columns 1–3 represent the interaction parameters c_{ij}, where $i = 1, 2, \ldots, 6$ refers to the i-th Year, and $j = 1, 2, 3$ to the j-th University (Maastricht, Nijmegen, Groningen). The column 'Main' presents parameters a_i (i.e. the main effect of the factor Year) and the row 'Main' shows parameters b_j (i.e. the main effect of the factor University). Finally, the overall mean is presented in the lower right of the table. The value of − 4.72 indicates that on average the percentage correct-minus-incorrect score on the Maastricht subtest was lower than that on the Nijmegen subtest. Thus the average degree of difficulty of the items originating

Table 2 Main effects and interactions of the factors Year and University on Dscore, the within-student difference in scores on the Maastricht and Nijmegen subtests (averages of the corresponding ANOVA parameters in 6 progress tests)

Year	*University* *Maastricht*	*Nijmegen*	*Groningen*	*Main*
1	− 1.19	1.60	− 0.41	1.81
2	0.71*	− 1.53	0.82	− 1.38
3	0.44	− 1.06	0.62	− 0.43
4	0.82	− 0.34	− 0.48	− 0.15†
5	0.00	0.13	− 0.13	− 0.68
6	− 0.78	1.19	− 0.41	0.82
Main	1.77	− 2.27	0.50‡	− 4.72§

* Interaction for Year 2 students from Maastricht University (parameter c_{21})
† Main effect for Year 4 students (parameter a_4)
‡ Main effect for students from Groningen University (parameter b_3)
§ Overall mean difference score (parameter m)

from Nijmegen appears to be lower than that of the items originating from Maastricht for all students. The size of the difference in difficulty is, however, not the same for all years. The absolute difference is smaller for students in Year 1 (− 4.72 + 1.81 = − 2.91) and larger for students in Year 2 (− 4.72 − 1.38 = − 6.1).

Moreover, and this is the focus of our interest, the difference between degrees of difficulty varied between students from the different universities. In general, Maastricht students scored 1.77 higher on the Maastricht subtest than on the Nijmegen subtest, whereas for the Nijmegen students the difference between the scores on the Maastricht and Nijmegen subsets was − 2.27. For the Groningen students the difference was 0.50. However, the pattern of difference in item difficulty for the 3 levels of University was not the same in all year groups because of the Year–University interaction. For example, for Year 1 students in Maastricht, Nijmegen and Groningen, the main University effect of 1.77, − 2.27 and 0.50, respectively (Table 2, last row) was modulated by the interaction (− 1.19, 1.60, − 0.41; Table 2, first row) to result in Dscores of 0.58, − 0.67, and 0.09, respectively. So, for the Year 1 group, the pattern of the effect of item and student origin is less pronounced than for the other groups. The same applies for the Year 6 group, almost at the end of the curriculum, when the effects are also less pronounced.

For the Year 2, 3 and 4 groups, adding the interaction to the University main effect results in a more pronounced pattern across University (e.g. for the Year 2 group the resulting pattern was equal to 2.48, - 3.8 and 1.32).

Figure 1 gives an overview of the Dscore patterns across the 3 levels of University for all year groups (lines), in addition to the average pattern (shaded), which is equal to the University main effect. The graph shows that the University effect is modest in the Year 1 and 6 groups, and relatively large in the Year 2, 3 and 4 groups.

When we calculated the effect sizes using equations 3 and 4, we substituted a value of 6 for σ_{Tot}, the SD of the score in a regular progress test, because in most progress tests this is the mean value of the SD of the percentage correct-minus-incorrect score, which increases gradually from 3 in Year 1 to 9 in Year 6.[14] In accordance with the average reliability of the correct-minus-incorrect scores of progress tests, R was set to a value of 70%. Finally, we set N_{Tot} and N_S to 248 and 60, respectively, in accordance with the data in Table 1. The resulting overall effect size was ES = 0.45, which is a medium effect according to Cohen's classification.[13] The effect sizes calculated per Year (Table 3) show that large to medium effects were found in Years 2–5, and small effects in Years 1 and 6.

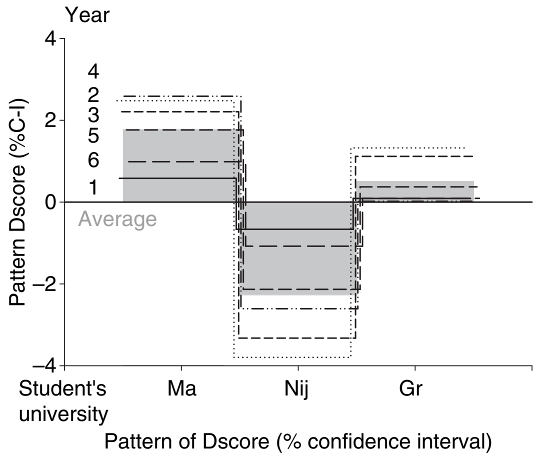

Figure 1 Effect of the student's university on between-student difference scores on the Maastricht and Nijmegen subtests (Dscores), showing the pattern of Dscores per year group of students across University (corresponding to parameters $b_j + c_{ij}$), where the corresponding Year is indicated at the left. The average effect of University is represented by the shaded area (corresponding to parameter b_j). Ma = Maastricht; Nij = Nijmegen; Gr = Groningen

Table 3 Effect sizes associated with the mean differences between Maastricht and Nijmegen students in Dscore (i.e. the within-student difference between scores on the Maastricht and Nijmegen subtests)

Year	*Mean difference*	*Effect size*	*Cohen's classification*
1	1.2	0.14	Small
2	6.3	0.70	Large
3	5.5	0.61	Medium
4	5.2	0.58	Medium
5	3.9	0.43	Medium
6	2.1	0.23	Small
All	4.0	0.45	Medium

DISCUSSION

The results of this study can be interpreted as cautioning against naive use of results on inter-institutional tests for pass/fail decisions or for making comparisons between the effects of different curricula, even when schools use identical tests aimed at the same shared curricular end objectives. Although shared objectives and identical tests may go some way towards ensuring validity and fairness of comparisons, our findings suggest that they do not automatically guarantee curriculum-independence of test items.[4,5] The analysis of the results of students from 3 medical schools on 6 progress tests aimed at identical curricular end goals points to an effect that we would like to qualify as 'origin bias' of test items. The origin of test items (i.e. the item author's school) had a marked differential impact on students' test scores in favour of students from the same school as the item writer. The differences between the scores on the Maastricht and Nijmegen subtests between Maastricht, Nijmegen and Groningen students, of 1.8, − 2.3 and 0.5, respectively, suggest that students did better on the items produced by staff from their own schools. Interestingly, the results were (relatively) indifferent for students from Groningen Medical School, whose staff did not contribute items to the subtests we studied. We found an ES of 0.45, which indicates a medium effect, of item origin on students' scores for 2 tests that would consist entirely of either Maastricht-produced or Nijmegen-produced items.

The finding that the effects of item origin were strong in Years 2–5 and moderate in Years 1 and 6 show that effects are small at the end and beginning of the curriculum and larger in the middle of it. This seems plausible because curriculum effects require some time to develop before they become manifest in test results. The small effect at the end of the programme also seems plausible because the 3 curricula are aimed at the same set of end objectives, to be attained by the end of Year 6. The emerging pattern is indicative of differences between the curricula in the pathways they take to reach the same goal, which are reflected in differences between students' results on test items originating from different schools and which diminish as students approach the end of the curriculum.

These results indicate that it may not be appropriate to draw conclusions on curricular effectiveness based on comparisons of the results on identical tests obtained by students of different schools when tests are composed by staff of 1 of the schools only. Our findings indicate that origin bias of test items tends to impair the validity and fairness of such comparisons. These problems may be avoided by ensuring that all the schools in the comparison contribute equal numbers of test items or by using stringent test review procedures involving mixed review panels from the participating schools. Nevertheless, the discriminating impact of item origin on test performance across institutions seems likely to persist, despite counter measures, particularly in the years when the curriculum is ongoing and less so at the start or end of training. Therefore, this effect should be taken into account when conclusions are drawn about the quality of curricula on the basis of comparisons between student performances on inter-institutional tests.

Contributors: AMMM conceived the study, developed the outline and method, and analysed and interpreted the data. All authors participated in acquisition of data and in drafting and revising the paper, and read and approved the final manuscript.
Acknowledgements: the authors would like to thank Ron Hoogenboom for his contributions to data collection and analysis, and Mereke Gorsira for her help in improving the English language of the manuscript.
Funding: none.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. *Med Educ* 2003;**37**:582–4.
2 Schmidt HG. Innovative and conventional curricula compared: what can be said about their effects? In: Nooman ZM, Schmidt HG, Ezzat ES, eds. *Innovation in Medical Education: An Evaluation of Its Present Status.* New York: Springer Publishing Co. 1990;1–7.
3 van der Vleuten CPM, Schuwirth LWT, Muijtjens AMM, Thoben A, Cohen-Schotanus J, van Boven CPA. Cross-institutional collaboration in assessment: a case on progress testing. *Med Teach* 2004;**26**:719–25.
4 Verhoeven BH, Verwijnen GM, Scherpbier AJJA, Holdrinet RSG, Oeseburg B, Bulte JA, van der Vleuten CPM. An analysis of progress test results of PBL and non-PBL students. *Med Teach* 1998;**20** (4):310–6.
5 Albano MG, Cavallo F, Hoogenboom R *et al.* An international comparison of knowledge levels of medical students: the Maastricht progress test. *Med Educ* 1996;**30**:239–45.
6 Prideaux D, Gordon J. Can global co-operation enhance quality in medical education? Some lessons from an international assessment consortium. *Med Educ* 2002;**36** (5):404–5.
7 Byrne GJ, Owen A, Newble D, Barton R, Garden A, Roberts T, O'Neill PA. Sharing resources for UK undergraduate written assessments – 1 year of UMAP. In: Maldonado MB, ed. *Eleventh International Ottawa Conference on Medical Education*, Barcelona, 6–8 July 2004; *Educacion Medica* 2004;**7**:192.
8 Mahadev GK, O'Neill PA, Owen AC, McCardle P, Benbow E, Byrne GJ. Seven years experience of progress testing in Manchester, UK. In: Maldonado

MB, ed. *Eleventh International Ottawa Conference on Medical Education*, Barcelona, 6–8 July 2004. *Educacion Medica* 2004;**7**:192–3.
9 Föller T, Brauns K, Fuhrmann S *et al.* Five years of progress testing at Charité Universitätsmedizin Berlin, Germany. In: Maldonado MB, ed. *Eleventh International Ottawa Conference on Medical Education.* Barcelona, 6–8 July 2004. *Educacion Medica* 2004;**7**:192.
10 Metz JCM, Verbeek-Weel AMM, Huisjes HJ. *Blueprint 2001: Adjusted Objectives of Undergraduate Medical Education in The Netherlands.* Nijmegen, The Netherlands: Driemediagroep 2001;1–62.
11 Diamond J, Evans W. The correction for guessing. *Rev Educ Res* 1973;**43**:181–91.
12 Lord FM. Formula scoring and number-right scoring. *J Educ Meas* 1975;**12** (1):7–11.
13 Hojat M, Xu G. A visitor's guide to effect sizes. *Adv Health Sci Educ* 2004;**9**:241–9.
14 Muijtjens AMM, Hoogenboom RJI, Verwijnen GM, van der Vleuten CPM. Relative or absolute standards in assessing medical knowledge using progress tests. *Adv Health Sci Educ* 1998;**3**:81–7.

Received 22 December 2006; editorial comments to authors 4 April 2007; accepted for publication 15 August 2007

Portfolios in medical education: why do they meet with mixed success? A systematic review

Erik Driessen,[1] Jan van Tartwijk,[2] Cees van der Vleuten[1] & Val Wass[3]

CONTEXT The move towards competence-based medical education has created a need for instruments that support and assess competence development. Portfolios seem suitable but mixed reports of their success are emerging.

METHODS To examine the effectiveness of portfolios, we searched PubMed and EMBASE using the keyword 'portfol*', PsychInfo and ERIC using the keywords 'portfol*' and 'medical education' and references of retrieved papers for empirical studies on portfolios in all phases of medical education. Thirty of 1939 retrieved papers met the inclusion criteria and were analysed. Data were collated against the research question, number of subjects, design, setting, findings and limitations, purpose and content, mentoring and assessment. We analysed impact using a modified version of Kirkpatrick's hierarchy.

RESULTS Because differences across studies precluded statistical meta-analysis, the data were analysed by context, goals and procedure. Positive effects were strongest in undergraduate education. Important factors for success were: clearly communicated goals and procedures; integration with curriculum and assessment; flexible structure; support through mentoring, and measures to heighten feasibility and reduce required time. Moderately good inter-rater reliability was reported and global criteria and discussions among raters were beneficial. Formative and summative assessment could be combined. Without assessment, portfolios were vulnerable to competition from other summative assessment instruments.

CONCLUSIONS For portfolios to be effective in supporting and assessing competence development, robust integration into the curriculum and tutor support are essential. Further studies should focus on the effectiveness and user-friendliness of portfolios, the merits of holistic assessment procedures, and the competences of an effective portfolio mentor.

KEYWORDS review [publication type]; education, medical/*methods; educational measurement/*methods; teaching/*methods; *learning; documentation/*methods; observer variation; clinical competence/*standards; self-assessment (psychology).

*Medical Education 2007: **41**: 1224–1233*
doi:10.1111/j.1365-2923.2007.02944.x

[1]Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
[2]Leiden University Graduate School of Teaching, Leiden University, Leiden, The Netherlands
[3]Community-based Medical Education, University of Manchester, Manchester, UK

Correspondence: Erik Driessen, Department of Educational Development and Research, Maastricht University, 6200 MD Maastricht, The Netherlands. Tel: 00 31 43 388 5774; Fax: 00 31 43 388 5779; E-mail: e.driessen@educ.unimaas.nl

INTRODUCTION

Over the last 2 decades, a significant change has occurred in medical education. The focus of curricula has shifted from the acquisition of knowledge to the achievement of competence.[1,2] Competence has been defined as 'the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individuals and communities being served'.[3] The challenge has been to find instruments that formatively support the development of competence in an integrated, coherent and longitudinal fashion and summatively assess whether competence is being achieved.[4,5] The portfolio is acclaimed as such an instrument.[6] The past 10–15 years have seen the introduction of portfolios in all stages of the medical education continuum: in undergraduate medical education;[5,7]

Overview

What is already known on this subject

Despite claims that portfolios are an excellent instrument with which to enhance integrated, self-reflective, self-directed, longitudinal learning, they have met with mixed success.

What this study adds

Inter-rater reliabilities contradict the presumed 'subjectivity' of portfolio assessment. Portfolios can be used simultaneously for summative and formative purposes.

Effective portfolios require:

- a proper introduction and mentoring;
- integration within context and procedures;
- provision of information to students and teachers;
- provision of clear guidelines that do not curtail students' freedom,
- user-friendliness that includes limited time demands on students and mentors.

Suggestions for further research

Studies addressing the effectiveness and user-friendliness of portfolios, the merits of holistic assessment procedures, and the competences of effective portfolio mentors should be encouraged.

in postgraduate specialist training,[8–10] and in the continuing medical education (CME) of practising doctors.[6,11–14]

Portfolios that are used in education contain evidence of how trainees fulfil tasks and how their competence is progressing. Portfolios may be digital or paper-based and content may be prescribed or left to the students' discretion. Despite variations in content and format, portfolios basically report on work done, feedback received, progress made, and plans for improving competence. Additionally, portfolios may stimulate reflection, because collecting evidence for inclusion in a portfolio requires looking back and analysing what one has accomplished.

Reflection can be defined as the mental process of trying to structure or restructure an experience, a problem, or existing knowledge.[15] This can help learners to understand their development[16] and plan their learning.[15] Reflecting on task performance and development of competence implies self-assessment or self-rating:[17] learners have to compare their own performance with (external) standards. Reflection and self-assessment are essential skills for lifelong learning, but the literature on self-assessment is quite clear in showing that students and doctors have a limited ability to self-assess their competence and learning needs.[17] Hence, it has been suggested that self-assessment should be supported by other (external) sources of information.[17,18] Portfolios may have the potential to improve self-assessment, by combining external assessment, mentoring and self-assessment.[19]

Since their introduction into medical education in the early 1990s, portfolios have been the subject of educational research. The evidence to date suggests that their introduction has met with mixed success.[20–22] There is little explanation for these differences, which may relate to many factors. The aim of this study was to conduct a systematic review of the literature on portfolios to seek evidence and clarify why in some contexts portfolios appear to be largely ineffective, whereas in others they are successful.

METHODS

Data sources

PubMed (1966–May 2007) and EMBASE (1989–May 2007) were searched using the keyword 'portfol*'. The databases PsychInfo (1970–May 2007) and ERIC (1966–May 2007) were searched using the keywords 'portfol*' and 'medical education'.

The searches were limited to publications in English and Dutch because it was not feasible to translate non-English or non-Dutch articles. To identify studies not picked up in the initial search, we contacted experts in the field and checked the references of the papers retrieved by the initial search.

Selection of studies

We used broadly defined inclusion criteria to ensure all aspects of the research question were addressed. We included studies that:

1 focused on portfolio use for educational purposes in medical training;
2 were performed within the context of undergraduate, postgraduate or continuing medical education, and
3 reported empirical data.

We excluded studies concerning: portfolios for other health professions (nursing, dentistry, dietetics, veterinary medicine), administrators, managers, teachers and trainers in hospitals, management, finance, education, teaching, specialist trainers and academic portfolios; portfolio-related instruments, such as logbooks, personal digital assistants, and personal development plans, and descriptive articles without evaluative data.

Data abstraction

The literature search was performed by 2 of the authors (ED, JvT) and an information specialist. Three of the authors (ED, JvT, CvdV) determined the inclusion criteria. Two of the authors (ED, JvT), supported by a third author (CvdV), reviewed the titles and abstracts of retrieved publications and selected relevant articles for possible inclusion. Data abstraction methods were developed by 3 of the authors (ED, JvT, CvdV) and were applied by 2 of the authors (ED, JvT). Disagreements about search criteria, data abstraction and classification of study results were resolved by consensus. The reviewers were not blinded to any portion of articles. The authors of 1 of the studies were contacted and asked to clarify some points, which they did.

The articles fulfilling the selection criteria referred to a wide range of studies where portfolios were used for different purposes in a variety of contexts within medical education, and methods and quality varied. Most of the selected studies used a variety of measurement methods and surveyed a range of portfolios which differed in purpose, content and format. With the exception of inter-rater reliability, statistical pooling of the results proved impossible. We made narrative descriptions of the findings and quality of the studies according to the criteria suggested by the Best Evidence Medical Education Collaboration (BEMEC).[23] Assessment of the quality of the studies was based on the study design, questionnaire validation, sampling frame and size, response rate and outcome measures.[24] Data were synthesised and reported where possible in relation to the influence of context and portfolio goals to address the aims of the review.

We used a modified version of the BEMEC coding sheet for data abstraction.[23] The form included details of the research question, number of subjects, study design, setting, findings and study limitations. Further details of the intervention (i.e. the goal, contents and structure, and the mentoring and assessment of the portfolio[25]) were also included. The impact of the intervention was rated using a modified version of Kirkpatrick's hierarchy to analyse outcomes such as learner satisfaction, learning outcomes, performance improvement and patient or health outcomes.[26]

The inter-rater reliability across all studies was estimated by averaging domain-referenced reliability coefficients or kappas. The Spearman–Brown prophecy formula was used to estimate the projection of inter-rater reliability for the use of multiple raters.[27]

RESULTS

Search results

The search revealed 1939 publications. After reading titles and abstracts, we excluded 1853 articles that failed to meet the inclusion criteria. More detailed review of the remaining 86 publications yielded 30 articles that met the inclusion criteria (Fig. 1). Of these articles, 9 related to the use of portfolios in pre-clinical undergraduate medical education, 7 addressed undergraduate clinical clerkships, 9 concerned postgraduate medical training, and 5 dealt with CME (see Tables S1 to S4 published online as supplementary material).

Many of these studies had methodological limitations. With the exception of 2 studies,[21,28] all had a single-group design. The majority were conducted in a single institution. In 5 studies, participants were self-selected volunteers.[20,28–31] Many studies lacked a detailed description of the portfolio, how it was introduced to its users, the sampling frame, the study method, data analysis or outcomes, which limited our ability to fully appraise the quality of the study or generalise the findings.

A total of 19 studies evaluated outcomes at Kirkpatrick level 1 (i.e. surveying the satisfaction level of the users). Only 2 studies reported outcomes in terms of performance improvement (level 3).[22,30] None of the studies measured patient outcomes.

We report the results in relation to the 2 broad portfolio goals: learning and assessment.

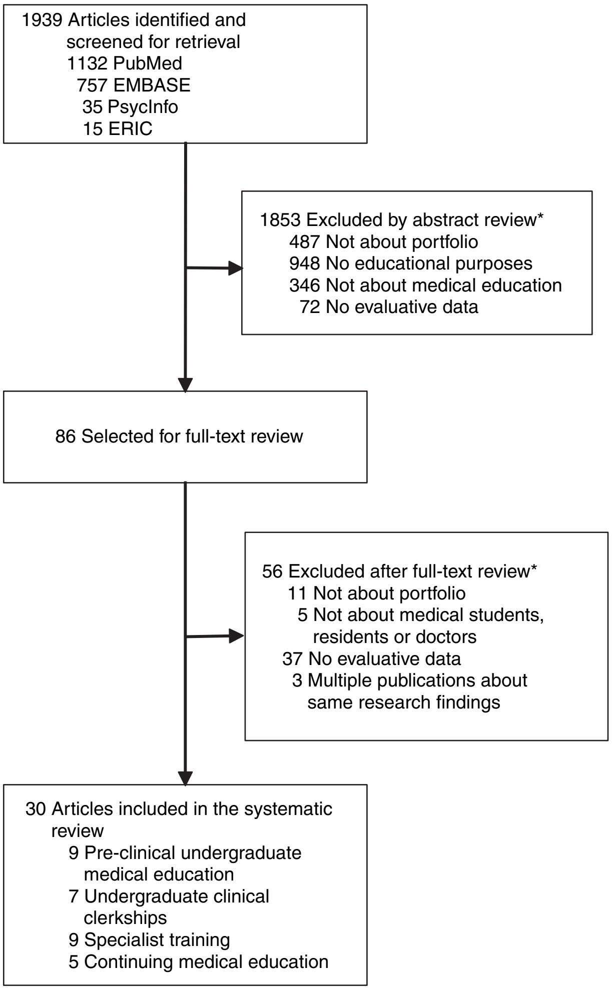

Figure 1 Selection of articles for review
*Some articles were excluded for multiple reasons

Goal 1. Learning

Two studies performed in the pre-clinical phase of medical school reported that portfolios contributed to Year 1 students' reflective learning[32,33] and 1 study reported that portfolios contributed to students' personal and professional development.[22] Two studies reported successful use of portfolios in organising, monitoring and evaluating a pre-clinical oncology programme and helping students understand the impact of malignant disease on patients.[21,34] Studies where portfolios were used in CME yielded mixed results. Some reported that portfolios could stimulate reflective learning[28,30] and support the planning and monitoring of CME.[30]

Many studies across a range of contexts reported problems related to the poor preparation and introduction of portfolios by the institution. Examples of this claimed either that the purpose of the portfolio was not clearly defined[35] or that learners and teachers were poorly or insufficiently informed about the portfolio and what it entailed.[35–38] In 1 study this resulted in poor commitment from both residents and their trainers and limited use of the portfolio.[35] However, in 5 studies where portfolio design centred on informing, training and gaining commitment from both trainers and trainees, portfolios were found to be suitable for graduate training.[39–43] One study demonstrated that hands-on introduction with a proper briefing of students by staff on the portfolio's purpose and procedures had a positive effect on portfolio scores and student satisfaction with the portfolio.[44]

The use of the portfolio in undergraduate education was more successful when portfolios were not used in isolation but were part of other educational activities.[21,22,32,34] These educational activities included pairing students with oncology patients,[21,34] organising tutorial groups[35] or mentoring,[32] or linkage to an interview.[22]

Results of the use of portfolios in CME often suggested that portfolios were not used by doctors to their full potential. This was related mainly to time constraints imposed by high daily workloads[20,31] and the perception that maintaining the portfolio was time-consuming.[20,28,45] On occasion, studies referred to extraneous issues, including difficulties with information technology, such as problems with downloading necessary software[20] or lack of IT skills.[31] Lack of time was also an issue for postgraduate training.[38,41,46] Trainees and their supervisors were concerned that the portfolio might be too time-intensive and for this reason avoided using it.[38,41,46] Mathers *et al.* made a plea for portfolios to be 'smarter' (less paperwork) to aid feasibility.[28] A study investigating the use of such an efficient portfolio supports this supposition, as undergraduate clerks did not find the portfolio labour-intensive.[47]

The format of the portfolio also influenced the contribution it made to learning. An effective portfolio had a clear but flexible structure, allowing learners opportunities to describe their own unique development.[32,35,48] Clear instructions were important. Most users wanted to know what kind of information they were expected to provide.[29,44,49] In clinical contexts where the content of a portfolio was often highly prescribed, portfolios were experienced as bureaucratic instruments.[35–37,45,46] Portfolios were more highly appreciated when learners were given a certain amount of freedom to determine their content.[32,50]

Many studies reported the lack of adequate support from mentors.[20,31,35] Other studies confirmed that mentoring by teachers, trainers or educational supervisors made an important contribution to the success of the portfolio.[21,28,32,37,39–41,48] Mentors included teachers, trainers, supervisors or peers.[31] General practice trainees made more use of their portfolios when they had a supportive trainer.[35,38,41] Because of the significant impact of mentoring, it was difficult sometimes to discriminate between the effects of the mentor and the practicalities of completing the portfolio itself.[35] Obviously, mentoring requires teacher and supervisor time.[32] However, mentoring aimed at stimulating the development of reflective ability[32] and deep learning strategies focused on comprehension and understanding[28,43] merited the effort.

Goal 2. Assessment

A study investigating the validity of portfolio assessment[51] demonstrated it was indeed a valid test of reflective ability. Quality of reflection was the strongest predictor of the final assessment grade. Other criteria, such as lay-out and writing style, had negligible effect. Six studies estimated the inter-rater reliability of portfolios.[45,46,48,51–53] The average reliability across these 6 studies was 0.63, representing the estimated reliability if one assessor were to be randomly replaced by another. However, with 2, 3 or 4 raters, the reliability would increase to 0.77, 0.84 and 0.87, respectively. A value of 0.80 is usually required for high-stakes tests.[27] The studies suggested that a number of measures had a positive impact on inter-rater agreement: use of a small group of (trained) assessors;[45,46,51–53] discussion among the raters before the actual assessment[46,51–53] and after assessing part of the portfolio,[46,53] and the use of global criteria with rubrics.[45,46,53]

In general, there was more support for the formal assessment of portfolios from teachers and examiners than from students themselves.[36,44,45] For example, in a study in which examiners were positive about the use of portfolios for assessment, final year undergraduate students reported that a comprehensive portfolio with prescribed content involved far too much paperwork and, if they were to be formally assessed, they needed more advance information about how to construct the portfolio.[36]

The use of the portfolio for assessment and learning is often seen as conflicting: students may be less open in their reflections when their portfolios are to be assessed. However, 2 studies examining the combination of formative mentoring and summative assessment in 1 portfolio reported that this was not an issue. One study described mentors' reports that portfolio assessment had no effect on students' openness[32] and another claimed that the combination of support and assessment did not appear to be problematic for general practitioners and their CME tutors.[28] Two studies showed that if portfolios were not formally assessed, other summative assessment instruments were prioritised and the use of portfolios tailed off.[35,43]

DISCUSSION

To our knowledge, this is the first systematic review of the literature on portfolio use in medical education. We found many descriptive articles, opinion papers and commentaries on portfolio use. Only 30 of the retrieved articles reported empirical data. The available evidence demonstrated that portfolios can support both the learning and assessment of more general, yet essential, competencies in pre-clinical undergraduate education, such as reflective ability, personal and professional development, communication skills, and empathy towards terminally ill patients and their families. This finding is consistent with the evidence from a recent literature review of portfolio use in nursing.[54] Portfolios also have potential as tools to organise workplace learning during clerkships and postgraduate specialist training. Here, a more mixed picture emerged of contrasting poor and successful examples of portfolio introduction at all stages of training. Our review is in agreement with earlier literature, showing that several key issues are decisive in the successful use of a portfolio,[55] and is consistent with findings from studies in other disciplines. For example, in 1996 a study in teacher education showed that, for a portfolio to stimulate reflection, certain conditions had to be met, including: a thorough introduction of the portfolio and its intended use; student ownership; a clear structure, and appropriate use of the portfolio in discussions with mentors or trainers.[56] McMullan concluded in a recent study in nursing that portfolios can be very effective as an assessment and learning tool, but only if both students and mentors receive clear guidelines and support for their use.[57] McMullan noted that, without support and clear guidelines, students and mentors became increasingly stressed and demoralised about the use of portfolios in practice.[57] The studies highlight several success factors for portfolio use.

Success factors

The various goals of working with a portfolio need to be clear but can be successfully combined.[28,32] Portfolios can be used concurrently in both the formative promotion of learning and summative assessment. This is in contrast to previous debate in the literature, where the use of portfolios for assessment and learning was seen as conflicting on the grounds that students may be less open in their reflections when their portfolios are to be assessed.[58–60] From a systematic review of the literature, combining the 2 goals of learning and assessment does not appear to cause problems. On the contrary, summative assessment was found to be important to ensure that portfolio learning maintained its status alongside other assessed subjects.[35,43]

It is advisable to regard a portfolio not as a separate, independent instrument but as an activity that can be integrated with other educational activities.[21,22]

Effectiveness of learning is enhanced by providing a mentor to support the portfolio. Mentorship requires a substantial time investment, but appears to be crucial to successful portfolio use.[21,28,32,35,37,39,41,43,48] The effectiveness of assessment can be enhanced by combining the portfolio assessment with an interview.[22,36,47]

A major challenge for the integration of a portfolio into medical education is that its status must be maintained in the eyes of assessment- driven students. This review suggests that it must be part of the institutional assessment procedures.[32,35,43] We found surprisingly high levels of inter-rater reliability in the studies.[45,46,48,51–53] This contrasts with findings in other domains, such as for the teaching portfolios of general practice trainers.[61,62] The results of our review suggest that assessment panels may be limited to 2 or 3 assessors depending on the stakes of the assessment. Part of the success in achieving high reliabilities appears to be attributable to the use of a small group of trained assessors,[45,46,51–53] specific assessor training exercises,[46,51–53] including benchmarking, assessor discussion (before and intermediate) and use of holistic scoring rubrics (global performance descriptors).[45,46,53] In her review of portfolio assessment in nursing, McCready[54] also calls for experienced assessors, explicit guidelines for portfolio construction and a holistic assessment procedure. The good news seems to be that putting these appropriate measures in place makes adequate assessment of portfolios possible, without the need to prescribe the content and structure of the portfolio in detail.[63]

Another issue that impacts on portfolio success is a flexible learner-centred format. A rigid structure in which every detail of portfolio content is prescribed elicits negative reactions from portfolio users and is regarded as counterproductive.[28,32,35,36] Findings in this review and other literature[56,57] appear to indicate that too much structure implies a greater risk than too little structure. This does not deny the fact that learners do need clear directions and guidance to support the development and assessment of broad competencies.[29,57] However, direction should be achieved through clear guidelines and well defined portfolio goals rather than minute directives for every detail of the portfolio.[57] Striking the right balance is crucial here.

Time, or rather lack of it, is another key issue. Many learners who are asked to create a portfolio, and their supervisors or mentors, are concerned that building and judging portfolios will be exceedingly time-consuming or downright impossible.[20,28,31,38,41,45,46,64] The finding that time constraints appear to be less of a problem for preclinical students may indicate that these students have relatively more time at their disposal. For learners in clinical settings, it is clearly difficult to find time amidst the pressures of clinical practice. Many of the portfolios described in the studies we reviewed were not user-friendly and involved huge amounts of paperwork, forcing portfolio users to comply with strict and detailed guidelines.[35,36,45,46] Too much specific obligatory content makes portfolios bureaucratic, with the result that they both fail to serve any educational purpose and force learners to search for content outside their direct and lived experiences.[32,57]

Table 1 summarises the factors promoting portfolio success that emerged from this review.

Study limitations

Several limitations in this review should be considered. Firstly, the label 'portfolio' refers to a broad range of instruments. The purpose, context, structure and content of the portfolios described in the literature reviewed here differed considerably. Because of these differences and the variety of study methods and study quality, it was not possible to use a statistical meta-analytic approach. We attempted to overcome this limitation by synthesising the data as much as possible per context and per goal. Secondly, the literature in medical education often lacks the use of extensive medical subject headings, which could have contributed to our non-retrieval of some

Table 1 Summary of factors promoting portfolio success

Factor	Recommendation
Goals	Clearly introduce the goals of working with a portfolio
	Combine goals (learning and assessment)
Introducing the portfolio	Provide clear guidelines about the procedure, the format and the content
	Be cautious for problems with information technology
Mentoring/interaction	Provide mentoring by teachers, trainers, supervisors or peers
Assessment	Use assessment panels of 2–3 assessors depending on the stakes of the assessment
	Train assessors
	Use holistic scoring rubrics (global performance descriptors)
Portfolio format	Use a hands-on introduction with a briefing on the portfolio's purpose and the procedures
	Keep the portfolio format flexible
	Avoid being overly prescriptive about the portfolio content
	Avoid too much paperwork
Position in the curriculum	Integrate the portfolio into other educational activities in the curriculum

studies. In addition, different labels were sometimes used in the text of articles. Because we excluded studies of instruments like logbooks, appraisals, personal digital assistants and personal development plans, we may have missed studies in which they were used similarly to portfolios if the authors did not use the term 'portfolio'. Although we manually searched reference lists to overcome these subject heading and label limitations, we may have missed some studies. Thirdly, many studies lacked a full description of the actual portfolio, the portfolio introduction, the study method, data analysis and outcomes. This limited our ability to describe the studies more fully or to generalise more. Finally, in some studies it was not possible to distinguish whether the observed outcomes were the result of working with a portfolio or of mentoring[35] or other educational activities.[21,22,28] We believe, however, that future studies should not try to solve this limitation, as this review showed the crucial importance of integrating portfolios and mentoring in the curriculum.

Implications for research and practice

The results of this review show that many questions regarding portfolio use are still unanswered and this has important implications for both research and practice. We found many studies where the description of the portfolio structure and its implementation were inadequate. In view of the wide variation in portfolio formats, researchers and peer reviewers should insist that details of portfolio structure (purpose, content, mentoring and assessment)[25] are given, along with the context in which the portfolio was implemented, to ensure that papers can be critically appraised by others in an adequate fashion. Although the literature indicates that portfolios are not always successful, many studies did not examine how they were implemented and why they failed. We found no studies investigating the influence of the context in which a portfolio is introduced. To claim success for an educational intervention, such as the portfolio, researchers need to look carefully at the intervention in practical settings.[65]

The implementation of a portfolio requires greater rigour than we encountered in many papers. This lack of scientific rigour may account for our disappointing finding that there was no trend toward improvement in portfolio delivery over the time-span represented by these studies.

Future portfolio research could focus on the user-friendliness or feasibility of portfolios and address time constraints (e.g. by ensuring that portfolios are supported by curriculum arrangements, such as protected time for learning),[66] the merits of holistic assessment procedures, and the competences of effective portfolio mentors.

Contributors: ED and JvT conceived the study. ED and JvT, helped by an information specialist, performed the literature search. ED, JvT and CvdV determined the inclusion criteria, were responsible for the review of the titles and abstracts and selection of the relevant articles, and developed the data abstraction methods, which were applied by ED and JvT. All authors contributed to the analysis of the results. ED and JvT wrote the first draft of the paper, which was extensively revised by VW. All authors participated actively throughout the conduct, analysis and writing of the study.
Acknowledgements: the authors acknowledge Karlijn Overeem MD for her advice on the study design and for

reading and commenting on various drafts of the manuscript. The authors also wish to thank Victor Burger MSc our information specialist, for his invaluable help in performing efficient and comprehensive literature searches and Mereke Gorsira BA for help in preparing the paper and improving the English language of the first draft.
Funding: none.
Conflicts of interest: none.
Ethical approval: not required.

REFERENCES

1 Rabinowitz HK, Babbott D, Bastacky S *et al.* Innovative approaches to educating medical students for practice in a changing health care environment: the National UME-21 project. *Acad Med* 2001;**76**:587–95.
2 National Health Service. *Modernising Medical Careers Foundation Programmes.* http://www.mmc.nhs.uk/pages/foundation. [Accessed 29 August 2006.]
3 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;**287** (2):226–35.
4 Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;**65** (9 Suppl):63–7.
5 Epstein RM. Assessment in medical education. *N Engl J Med* 2007;**356**:387–96.
6 Royal College of General Practitioners. *Portfolio-based Learning in General Practice: Report of a Working Group on Higher Professional Education: December 1993.* Occasional paper 63.
7 General Medical Council. *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education.* http://www.gmc-uk.org/education/undergraduate/tomdoc.pdf. [Accessed 29 August 2006.]
8 Accreditation Council for General Medical Education, American Board of Medical Specialties. *Toolbox of Assessment Methods*, Version 1.1. http://www.acgme.org/Outcome/assess/Toolbox.pdf. [Accessed 29 August 2006.]
9 Hayden SR, Dufel S, Shih R. Definitions and competencies for practice-based learning and improvement. *Acad Emerg Med* 2002;**9** (11):1242–8.
10 Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing practice-based learning and improvement. *Teach Learn Med* 2004;**16** (1):85–92.
11 Alliance for Academic Internal Medicine. *Portfolio for internal medicine residency programmes.* http://www.im.org/AAIM/Tools/Docs/Forms/competencies.pdf. [Accessed 29 August 2006.]
12 Bennett NL, Davis DA, Easterling WE Jr *et al.* Continuing medical education: a new vision of the professional development of physicians. *Acad Med* 2000;**75**:1167–72.
13 Lewis M, Elwyn G, Wood F. Appraisal of family doctors: an evaluation study. *Br J Gen Pract* 2003;**53**:454–60.
14 Zeiger RF. Toward continuous medical education. *J Gen Int Med* 2005;**20** (1):91–4.
15 Korthagen FAJ, Kessels J, Koster B, Lagerwerf B, Wubbels T. *Linking Theory and Practice: the Pedagogy of Realistic Teacher Education.* Mahwah, NJ: Lawrence Erlbaum Associates 2001.
16 Klenowski V. *Developing Portfolios for Learning and Assessment: Processes and Principles.* London: Routledge/Falmer 2002.
17 Davis DA, Mazmanian PE, Fordis M, van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;**296** (9):1094–102.
18 Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;**80** (10 Suppl):46–54.
19 van Tartwijk J, Driessen E, Stokking K, van der Vleuten C. Factors influencing the successful introduction of portfolios. *Qual Higher Educ* 2007;**13** (1):69–79.
20 Dornan T, Carroll C, Parboosingh J. An electronic learning portfolio for reflective continuing professional development. *Med Educ* 2002;**36** (8):767–9.
21 Finlay IG, Maughan TS, Webster DJ. A randomised controlled study of portfolio learning in undergraduate cancer education. *Med Educ* 1998;**32** (2):172–6.
22 Gordon J. Assessing students' personal and professional development using portfolios and interviews. *Med Educ* 2003;**37** (4):335–40.
23 Best Evidence Medical Education Collaboration. *Information for prospective topic review groups.* http://www.bemecollaboration.org/TRGInfo.htm. [Accessed 19 June 2006.]
24 Reed D, Price EG, Windish DM *et al.* Challenges in systematic reviews of educational intervention studies. *Ann Intern Med* 2005;**142**:1080–9.
25 Zeichner K, Wray S. The teaching portfolio in US teacher education programmes: what we know and what we need to know. *Teach Teach Educ* 2001;**17**:613–21.
26 Curran VR, Fleet LA. A review of evaluation outcomes of web-based continuing medical education. *Med Educ* 2005;**39**:561–7.
27 Magnusson D. *Test Theory.* Stockholm: Addison-Wesley 1967.
28 Mathers NJ, Challis MC, Howe AC, Field NJ. Portfolios in continuing medical education – effective and efficient? *Med Educ* 1999;**33** (7):521–30.
29 Vink S, Cools HJM, Elsen GMF, Vermunt JD. Reflectie van eerstejaars studenten geneeskunde: wat willen we zien en wat doen de studenten? [Summary in English.]*Dutch J Med Educ* 2005;**24** (1):4–13
30 Campbell CM, Parboosingh JT, Gondocz ST *et al.* Study of physicians' use of a software program to create a portfolio of their self-directed learning. *Acad Med* 1996;**71** (10 Suppl):49–50.
31 Dagley V, Berrington B. Learning from an evaluation of an electronic portfolio to support general practitioners' personal development planning, appraisal and revalidation. *Educ Prim Care* 2005;**16**:567–74.
32 Driessen EW, van Tartwijk J, Overeem K, Vermunt JD, van der Vleuten CPM. Conditions for successful reflective use of portfolios in undergraduate medical education. *Med Educ* 2005;**39**:1230–5.

33 Driessen EW, van Tartwijk J, Vermunt JD, van der Vleuten CPM. Use of portfolios in early undergraduate medical training. *Med Teach* 2003;**25** (1):14–9.
34 Maughan TS, Finlay IG, Webster DJ. Portfolio learning with cancer patients: an integrated module in undergraduate medical education. *Clin Oncol* 2001;**13** (1):44–9.
35 Pearson DJ, Heywood P. Portfolio use in general practice vocational training: a survey of GP registrars. *Med Educ* 2004;**38** (1):87–95.
36 Davis MH, Friedman BDM, Harden RM *et al.* Portfolio assessment in medical students' final examinations. *Med Teach* 2001;**23** (4):357–66.
37 Elango S, Jutti RC, Lee LK. Portfolio as a learning tool: students' perspective. *Ann Acad Med Singapore* 2005;**34** (8):511–4.
38 Kjaer NK, Maagaard R, Wied S. Using an online portfolio in postgraduate training. *Med Teach* 2006;**28** (8):708–12.
39 Ryland I, Brown J, O'Brien N *et al.* The portfolio: how was it for you? Views of F2 doctors from the Mersey Deanery Foundation Pilot *Clin Med* 2006;**6** (4):378–80.
40 Webb TP, Aprahamian C, Weigelt JA, Brasel KJ. The Surgical Learning and Instructional Portfolio (SLIP) as a self-assessment educational tool demonstrating practice-based learning. *Curr Surg* 2006;**63** (6):444–7.
41 Snadden D. Portfolio-based learning and general practice vocational training. *Med Educ* 1996;**30**:148–52.
42 Fung MKF, Walker M, Fung KKF *et al.* An Internet-based learning portfolio in resident education: the KOALA-super(TM) multicentre programme. *Med Educ* 2000;**34** (6):474–9.
43 Snadden D, Thomas ML. Portfolio learning in general practice vocational training – does it work? *Med Educ* 1998;**32** (4):401–6.
44 Duque G, Finkelstein A, Roberts A, Tabatabai D, Gold SL, Winer LR. Learning while evaluating: the use of an electronic evaluation portfolio in a geriatric medicine clerkship. *BMC Med Educ* 2006;**6**:4.
45 Tate P, Foulkes J, Neighbour R, Campion P, Field S. Assessing physicians' interpersonal skills via videotaped encounters: a new approach for the Royal College of General Practitioners membership examination. *J Health Commun* 1999;**4** (2):143–52.
46 O'Sullivan PS, Reckase MD, McClain T, Savidge MA, Clardy JA. Demonstration of portfolios to assess competency of residents. *Adv Health Sci Educ* 2004;**9** (4):309–23.
47 Burch VC, Seggie J. Portfolio assessment using a structured interview. *Med Educ* 2005;**39** (11):1169.
48 Grant AJ, Vermunt JD, Kinnersley P, Houston H. Exploring students' perceptions on the use of significant event analysis, as part of a portfolio assessment process in general practice, as a tool for learning how to use reflection in learning. *BMC Med Educ* 2007;**7**:5.
49 Rees CE, Sheard CE. Undergraduate medical students' views about a reflective portfolio assessment of their communication skills learning. *Med Educ* 2004;**38** (2):125–8.
50 Amsellem-Ouazana D, van Pee D, Godin V. Use of portfolios as a learning and assessment tool in a surgical practical session of urology during undergraduate medical training. *Med Teach* 2006;**28** (4): 356–9.
51 Overeem K, Driessen EW, Drenthe A, van Tartwijk J, van der Vleuten CPM. Kwaliteitscriteria voor portfoliobeoordeling van geneeskundestudenten. [Summary in English.]*Dutch J Med Educ* 2005;**24** (2):64–71.
52 Melville C, Rees M, Brookfield D, Anderson J. Portfolios for assessment of paediatric specialist registrars. *Med Educ* 2004;**38**:1117–25.
53 Rees CE, Sheard CE. The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: the Nottingham experience. *Med Educ* 2004;**38** (2):138–44.
54 McCready T. Portfolios and the assessment of competence in nursing. *Int J Nurs Stud* 2007;**44** (1): 143–51.
55 Pinsky LE, Fryer-Edwards K. Diving for PERLS – working and performance portfolios for evaluation and reflection on learning. *J Gen Int Med* 2004;**19**:582–7.
56 Wade RC, Yarbrough DB. Portfolios: a tool for reflective thinking in teacher education. *Teach Teach Educ* 1996;**12** (1):63–79.
57 McMullan M. Students' perceptions on the use of portfolios in pre-registration nursing education: a questionnaire survey. *Int J Nurs Stud* 2006;**43**:333–43.
58 Snyder J, Lippincott A, Bower D. The inherent tensions in the multiple uses of portfolios in teacher education. *Teach Educ Q* 1998;**25** (1):45–60.
59 Tigelaar DEH, Dolmans DHJM, Wolfhagen HAP, van der Vleuten CPM. Using a conceptual framework and the opinion of portfolio experts to develop a teaching portfolio prototype. *Stud Educ Eval* 2004;**30** (4):305–21.
60 Gallagher P. An evaluation of a standards-based portfolio. *Nurse Educ Today* 2001;**21**:409–16.
61 Pitts J, Coles C, Thomas P, Smith F. Enhancing reliability in portfolio assessment: discussions between assessors. *Med Teach* 2002;**24** (2):197–201.
62 Pitts J, Coles C, Thomas P. Educational portfolios in the assessment of general practice trainers: reliability of assessors. *Med Educ* 1999;**33**:515–20.
63 Frohna JG, Kalet A, Kachur E *et al.* Assessing residents' competency in care management: report of a consensus conference. *Teach Learn Med* 2004;**16** (1):77–84.
64 Lonka K, Slotte V, Halttunen M *et al.* Portfolios as a learning tool in obstetrics and gynaecology undergraduate training. *Med Educ* 2001;**35** (12):1125–30.
65 Collins A, Joseph D, Bielaszyc K. Design research: theoretical and methodological issues. *J Learning Sci* 2004;**13** (1):15–42.
66 Mucklow JC. Continuing medical education in clinical pharmacology and therapeutics: report of a questionnaire survey. *Br J Clin Pharmacol* 2001;**52** (1):9–16.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Table S1. Studies examining the use of portfolios in pre-clinical undergraduate medical education.

Table S2. Studies examining the use of portfolio in undergraduate clinical clerkships.

Table S3. Studies examining the use of portfolio in specialist training.

Table S4. Studies examining the use of portfolio in Continuing Medical Education (CME).

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365.2923.2007.02944.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received 14 August 2007; accepted for publication 20 August 2007

letters to the editor

Is there a relationship between where students sit in lectures and their performance in examinations?

Fayaz Roked & Paul Aveyard

Editor – The majority of teaching in the first 2 years of medical school in the UK is delivered in lecture format. During our first 2 years at the University of Birmingham, we realised early on that most students tend to sit in the same place for every lecture and that certain types of students sit in particular places; stereotypically those who are less committed to their studies are thought to sit further back.

We believed that people's attitudes and their academic drive may be reflected by their choice of seating position. Based on these observations, we carried out an ecological study to determine if there was a relationship between student seating position in lectures and examination performance.

We devised and distributed a questionnaire to all Year 2 medical students. It asked students to give their student identification number, from which we accessed their mean examination score from the previous academic year. Students indicated the seats in the lecture theatre where they most often chose to sit. Distance was quantified as the number of rows from the front of the lecture theatre. Students also indicated their reasons for choice of seat and the lowest examination grade they would be satisfied with in any given module. Results were analysed using Spearman's rank correlation coefficient.

A total of 288 students of a cohort of 351 (82%) attended the lecture at which we distributed the questionnaire. Of these, 264 (92%) completed questionnaires. There was no relationship between student seating position and examination performance (Spearman's $r = -0.12$, $P = 0.05$). There was a weak association between the lowest grade with which a student would be satisfied and seating position ($r = -0.22$, $P < 0.01$), indicating that students who sit closer to the front aim for higher grades than students who sit further back, but do not perform better academically. 'Being able to see the lecturer and/or screen clearly' was the most important factor in students' choice of seating.

Our study therefore shows partial support for the assumption that those who take the back rows are the least attentive and successful students. Such students are slightly more likely to express satisfaction with a bare pass than to aspire to higher grades. However, they are as likely as those at the front to achieve good grades. Medical students may believe that sitting at the back signals a devil-may-care attitude, but medical educators now know this is of social and not educational significance.

Correspondence: Fayaz Roked,
E-mail: fxp309@bham.ac.uk

doi: 10.1111/j.1365-2923.2007.02916.x

Surgical e-learning: validation of multi-media, web-based lectures

Breda Memon[1] & Muhammed Ashraf Memon[2–4]

Editor – We read with great interest the paper by Ridgway *et al.*[1] We would like to raise the following points.

- What was the rationale for restricting the study to only 5 weeks? Was this because the web-based, 5-week, surgical lecture course was related to a particular surgical rotation for the medical students?
- Lectures were presented via an educational web portal, Blackboard™. Why was this specific web portal chosen? Was it because it was cost-effective, easy to use, preferred by the university, already in use, etc?
- 'The lecture series was balanced specifically to contain paired topics in order to reduce pre-existent knowledge bias.' How did the authors pair the various topics? Was there a consensus amongst the authors of this study regarding such pairing? What topics were chosen and why? This relevant information is missing from the paper. Yet another important question that remains unanswered concerns the basis for choosing only 10 topics.
- 'A subgroup analysis was performed to estimate web usage which was estimated utilising surrogates such as the total number of page hits per student as information regarding time of log-in.' The biggest drawback of the Blackboard™ web portal (as acknowledged by the authors) was that it was unable to provide data on the mean sessions per student or total duration of log-in of the various sessions per student. If the authors were already aware of this weakness, why did they not use a different web portal which had this sort of information built into the system? (There are a number of such portals available commercially.) This information might have been very useful and might have easily tested further hypotheses regarding the direct impact of the total duration of web usage and achievement levels (i.e. marks achieved in multiple-choice question examinations).
- 'Face validity was assessed by a standardised anonymous questionnaire. The questionnaire was based on Likert scales and assessed the use, accessibility, relevance and content of the web portal.' Face validity refers simply to whether 'on the face of it' the construct seems to be an appropriate measure of what it is supposed to measure. A test that lacks face validity would not be credible, but the notion of face validity lacks rigour and is not useful beyond this point. 'Content' validity and 'construct' validity are more helpful ideas. The former concerns the extent to which the construct reflects the quality, depth and breadth of the attributes it seems to measure, and the latter is defined as 'the extent to which a construct may be said to measure that which it has been designed to measure'. Why the authors did not test the content and construct validity is puzzling as the argument would have been stronger.
- A total of 88 medical students participated in this study. What year did they belong to (e.g. Year 1 or Year 2)? What was their gender distribution? What was their age range? Demographic data are missing. Furthermore, a subgroup analysis of achievement versus usage of the web portal by gender may have provided us with some useful information.
- The authors concluded that web-based lectures that mimic real-life lectures (i.e. have an aural component) led to better transference of data, based on consistently higher scores compared with the text-only group. Were these scores significantly higher? If not, and based on a small sample and limited topics, over such a short period of time, is this conclusion valid? Secondly, a comparison with a cohort of students only attending the *real-life lecture series* based on exactly the same topic would have given a better insight into the impact of web-portal e-learning.

[1]Department of Higher Education, East Lancashire Institute of Higher Education, Blackburn, Lancashire, UK
[2]Department of Surgery, Ipswich Hospital, Ipswich, Queensland, Australia
[3]Department of Surgery, University of Queensland, Brisbane, Queensland, Australia
[4]Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Queensland, Australia

Correspondence: M A Memon FRCS, Department of Surgery, Ipswich Hospital, Chelmsford Avenue, Ipswich, Queensland, Australia.
Tel: 00 61 7 3810 1140;
Fax: 00 61 7 3810 1592;
E-mail: mmemon@yahoo.com

doi: 10.1111/j.1365-2923.2007.02926.x

We think the introduction of hybrid programmes would be of major benefit, especially in countries where there are shortages of instructors and instructional material (provided they are not too expensive!). Improvements in audio and video technologies have certainly revolutionised the way teaching is now conducted and such methods will continue to evolve. This article demonstrates the use of hybrid methodology to improve learning outcomes and student satisfaction and we thank the authors for their insightful study.

REFERENCE

1 Ridgway P, Sheikh A, Sweeney K *et al.* Surgical e-learning: validation of multi-media, web-based lectures. *Med Educ* 2007;**41**:168–172.

Reply to Memon and Memon: Surgical e-learning: validation of multi-media, web-based lectures

Paul F Ridgway[1], Athar Sheikh[1], Denis Evoy[1], Enda McDermott[1], Patrick Felle[2] & Arnold D Hill[3]

Editor – We thank Memon and Memon for their interest in our study on web-based learning, as well as for allowing us the chance to comment further in this letter. We believe validation of specific adjuncts to educational web-based programmes should be carried out in a synchronous fashion relative to implementation.

The study sought to answer a specific question on whether the addition of voice-overs to web-based lectures was worthwhile. Would it result in better outcomes? The study design was kept simple to reduce any confounding factors. The study was carried out in normal term-time during a normal first surgical clinical attachment (to reduce pre-existent knowledge bias) and lecture rotation. It involved the entire block of 88 students. Five weeks (and therefore 10 topics) were selected as a result of the duration of the rotation. Subjects were paired by author consensus and the topics for adjunct aural files were randomly selected by closed envelope. Demographic details were not presented in this internally controlled study as subgroup analysis was not expected to yield any useful comparative differences and would serve only to cloud the relevant findings. In particular, there were no statistical differences between the sexes in examination results or questionnaire answers.

The authors were aware of other web-based educational portals (such as moodle™; http://moodle.org/) that would allow more detailed user activity data, but it was university policy to utilise Blackboard™. We recognised ahead of the study that this limited aspects of user sub-analysis. These were not essential data and changing the standard in-use portal would have been unnecessarily destructive to the standard student routine.

Validity measures can be puzzling. In the first instance, identifying the suitability of the model to various measures of validity is paramount. The primary outcome was to see if a web-based lecture with voice-over format was better than that of web-based lecture alone in terms of examination results. (The significantly higher results seen in the former group are discussed at length in the paper.) This investigation was not well suited to the use of construct validation as oral voice-overs are not a potential measure in themselves. The secondary question concerned whether the perception of the web-based lecture and voice-over format compared favourably with that of conventional non-web-based lectures. Had this been a study where *paired* conventional and web-based lectures were compared, then content validity would have been an appropriate measure. As it was not, *face* validity was best used in this study. Once face validity had been established, the generalisation that the voice-supplemented lectures approximate conventional lectures was allowed.

[1]Department of Surgery, St Vincent's University Hospital, Dublin, Ireland
[2]Conway Institute of Bimolecular and Biomedical Science and Centre of Healthcare Informatics, University College Dublin, Dublin, Ireland
[3]Department of Surgery, Beaumont Hospital, Dublin, Ireland

Correspondence: Paul F Ridgway MD, MMedSc, FRCSI, Department of Surgery, St Vincent's University Hospital, Elm Park, Dublin 4, Ireland. Tel: +353 1 8093760; Fax: +353 1 8093758; E-mail: p.ridgway@ic.ac.uk

doi: 10.1111/j.1365-2923.2007.02925.x

Memon and Memon suggest that 'a comparison with a cohort of students only attending the *real-life lecture series* based on exactly the same topic would have given a better insight into the impact of web-portal e-learning'. Certainly this would be interesting if confounding biases could be controlled. It would not answer our primary question, however. We sought to investigate whether a specific adjunct to the web-based programme added benefit. We believe that this was best achieved by simplifying design to reduce biases, thus asking a simple question in a simple environment.

We also acknowledge that Memon and Memon are enthusiastic about the 'major' effect that hybrid education programmes would have on knowledge transfer. We add that they will also be of benefit in medical schools such as ours, where students are sent to geographically distant hospitals for clinical training. However, we support the maintenance of a critical approach to any new adjunct to web-based learning. As in this study, validation of a specific aspect of the web-based lecture is useful in justifying resource allocations when setting up such web-based programmes.

The need to improve the teaching of assessment of psychiatric symptoms at undergraduate level

Andrew Teodorczuk, John Paul Taylor, Joaquim M Soares Cerejeira & Elizabeta B Mukaetova-Ladinska

Editor – As liaison psychiatrists with an interest in educating our medical colleagues, we found the study by Wilson *et al.* highly informative.[1] Unfortunately, the relatively low rating of empathic history-taking and interview skills reported was not unexpected. Our clinical experience and the literature both indicate that it is clear that psychiatric illness is under-recognised in the hospital setting.[2] Further evidence suggests psychiatric disorders are missed because some medical doctors are less skilled in allowing patients to express significant verbal and visual cues than other doctors.[3] Importantly, empathy allows the expression of a patient's real thoughts and feelings and hence those cues which are central to detecting psychiatric illness. As specialists, we argue that the doctor–patient relationship should not rely solely on innate individuals skills, but must be supported by solid communication techniques that need to be taught in medical school. Perhaps failure to recognise these key skills by non-specialists contributes to the widespread under-recognition of psychiatric illness in the hospital setting.

The authors acknowledge that the Delphi method 'is only a single part of the overall process of curriculum development'.[1] We complement their findings with the results of a pilot study of 13 undergraduate students at Newcastle University. We administered confidence questionnaires to Year 3 medical students in June 2007. All students had previously received teaching in the assessment of psychiatric and physical symptoms and were asked to rate their confidence in each (on a scale of 1–10). We found that students were significantly (Wilcoxon signed rank test, $P < 0.001$) less confident in the assessment of psychiatric symptoms (median: 6; range: 5–9) than physical symptoms (median: 8; range: 6–9).

Our results suggest that there is a need to improve the teaching of assessment of psychiatric symptoms at the undergraduate level. This important need has been recognised in the USA, where the Education Committee Writing Group of the American Geriatric Society recommends that at least 5 psychiatric topics or psychiatrically related symptoms and problems be included in the undergraduate medical training.[4]

In conclusion, although the Delphi method has considerable benefits in addressing the theoretical core of the curriculum, it appears to have limitations with respect to recognising the practical psychiatric assessment skills of good clinical practice. By combining our findings with those of Wilson and colleagues, together we can inform and develop a more focused psychiatric undergraduate curriculum.

Newcastle upon Tyne, UK

Correspondence: Dr Andrew Teodorczuk, Academic Specialist Registrar, Liaison Old Age Psychiatry, Institute for Ageing and Health, Wolfson Research Centre, Newcastle General Hospital, Newcastle upon Tyne NE4 6BE, UK. Tel: 00 44 191 256 3321; Fax: 00 44 191 219 5071; E-mail: andrew.teodorczuk@ncl.ac.uk

doi: 10.1111/j.1365-2923.2007.02917.x

REFERENCES

1 Wilson S, Eagles JM, Platt JE, McKenzie H. Core undergraduate psychiatry: what do non-specialists need to know? *Med Educ* 2007;**41** (7):698–702.

2 Wancata J, Windhaber J, Bach M, Meise U. Recognition of psychiatric disorders in non-psychiatric hospital wards. *J Psychosom Res* 2000;**48** (2):149–55.

3 Davenport S, Goldberg D, Millar T. How psychiatric disorders are missed during medical consultations. *Lancet* 1987;**2** (8556):439–41.

4 Eleazer GP. The challenge of measuring quality of care in PACE. Program of All-inclusive Care for the Elderly. *J Am Geriatr Soc* 2000;**48** (8):1019–20.

list of reviewers for 2007

The Editor is grateful to the following, who have generously given their time and expertise in reviewing the papers submitted to the journal in 2007.

Abdulla, Aza
Abramovitz, Ruth
Adibi, Iman
Agius, Steven
Ahmad Hassali, Mohamed Azmi
Akabayashi, Akira
Akkad, Andrea
Albanese, Mark
Alder, Judith
Alexander, Heather
Allen, Isobel
Allen, Justin
Alston, Sebastian
Altman, Michael
Amaral, Carlos
Ambrose, Lucy
Andersen, Rikke
Anderson, Elizabeth
Anderson, Kirsty
Anderson, Marshall
Anderson, William
Andreatta, Pamela
Andriole, Dorothy
Anis, Rafik
Archer, Julian
Ark, Tavinder
Armstrong, Alicia
Arnold, Louise
Arnold, Richard
Arora, Amindra
Attena, Francesco
Aukes, Leo
Avdi, Elma
Baars, Gerard
Baca, Vaclav
Baer, Alan
Baerheim, Anders
Baker, Peter
Baker, Raymond
Banda, Sekelani
Bandaranayake, Raja
Bansal, Payal
Barach, Paul
Barbosa, Rosilu
Barnes, Rebecca
Barnett, Mandy
Baron-Maldonado, Margarita
Barson, Petrina
Barton, Peter
Bashook, Philip
Baumann, Martin
Beagan, Brenda
Beard, Jonathan
Bearman, Margaret
Beckman, Thomas
Bell, Erica
Bellini, Lisa
Bellodi, Patricia
Benson, John
Bergus, George
Berlingieri, Pasquale
Bested, Kirsten
Beullens, Johan
Biddle, Lucy
Bin-Jaliah, Ismaeel
Biro, Gabor
Blackmore, David
Blake, Kim
Blasco, Pablo
Bleakley, Alan
Blitz, Julia
Block, Robert
Bloomfield, Leah
Blue, Amy
Bobo, Linda
Boehler, Margaret
Boggis, Caroline
Bokken, Lonneke
Bolanowski, Wojciech
Bond, William
Booth, Jeremy
Bordage, Georges
Bore, Miles
Boshuizen, Henny
Bosse, Hans Martin
Boulet, John
Boursicot, Katharine
Boyd, Kenneth
Brainard, Andrew
Braunack-Mayer, Annette
Brera, Giuseppe
Brown, Ian
Bruce, David
Bruin, Veralice
Bryce, Robert
Bulstrode, Christopher
Bundred, Peter
Bundy, Chris
Burdick, William
Burke, Joanne
Busari, Jamiu
Busuttil, Anthony
Butler, Gary
Cahill, David
Caldicott, Catherine
Calimag, Maria Minerva
Campbell, Alastair
Campbell, Malcolm
Campion, Peter
Candy, Geoffrey
Cantillon, Peter
Carroll, Austin
Carroll, Mark
Carter, Fiona
Carter, Mary
Carter, Sarah
Carvalho, Paula
Case, Susan
Cave, Judith
Centeno, Angel
Chan, L C
Chao, Jason
Chapman, Dane
Charlin, Bernard
Chaturvedi, Sanjay
Chesser, Amy
Chretien, Katherine
Chur-Hansen, Anna
Clayden, Graham S
Cleland, Jennifer
Clyman, Stephen
Cohen-Schotanus, Janke
Colliver, Jerry
Conroy, Simon
Cook, David

Cook, Vivien
Cooper, Jeffrey
Cordingley, Lis
Cornford, Charles
Corrigan, Gerry
Corrigan, Mark
Cotton, Philip
Crites, Gerald
Crossley, Jim
Curran, Andrew
Curran, Vernon
Custers, Eugène
Dahlin, Marie
Dangerfield, Peter
Danoff, Deborah
Darnton, Richard
David, Tim
Davies, Helena
Dawson, Elizabeth
De Bruin, Anique
De Champlain, Andre
De Jong, Peter G M
De Vries Robbé, Pieter
De'Ath, Henry
Dennick, Reg
Dennis, Ian
Dent, John
D'Eon, Marcel
Dickinson, Claire
Dinis-Ribeiro, Mário
Dogra, Nisha
Dohn, Helge
Dornan, Tim
Downing, Steven
Duits, Aj
Egan, Tony
Eley, Diann
Elizondo-Omaña, Rodrigo
Elzubeir, Margaret
Enarson, Cam
Ephgrave, Kimberley
Espin, Sherry
Esposito, Thomas
Eva, Kevin
Evans, Alison
Evans, Darrell
Evans, Dason
Ewen, Shaun
Faresjö, Tomas
Farmer, Liz
Feldstein, David
Fenwick, Angela
Ferguson, Kristi
Fernandez-Liria, Alberto
Filak, Jr, Andrew
Finucane, Paul
Fitzpatrick, Patricia
Flett, Andy
Ford, Jason
Forrest, Frances
Foxall, Peta
Frank, Blye
Friedland, Joan
Fry, Heather
Frye, Ann
Fung, Cha-Chi
Furman, Gail
Gallen, Derek
Gambadauro, Pietro
Garden, Anne
George, A T
Geppert, Cynthia
Gibbins, Ian
Gibbs, Trevor
Gilbert, James
Gilula, Marshall
Glasser, Michael
Glavin, Ronnie
Glicken, Anita
Goldenhar, Linda
Goldie, John
Gordon, Jill
Gordon, Penelope
Govaerts, Marjan
Grady-Weliky, Tana
Grant, Andy
Graves, Lisa
Gray, Jean
Greengrass, Colin
Greveson, Gabrielle
Griffith, III, Charles
Groves, Michele
Gruppen, Larry
Gude, Tore
Gunderson, Anne
Gupta, B K
Gutiérrez, Javier
Hafler, Janet
Haig, Alex
Haig, Andrew
Hammick, Marilyn
Hamstra, Stanley
Hann, Mark
Haq, Inam
Harris, Peter
Harrison, Chris
Harrison, Van
Hart, Jo
Hastings, Adrian
Hatala, Rose
Hauer, Karen
Hawkey, Roy
Hays, Richard
Henriksen, Ann-Helen
Hermann, Christian
Hernandez-Fuentes, Maria
Herzberg, Joe
Heylings, D
Himmel, Wolfgang
Hirshbein, Laura
Hitchcock, Maurice
Ho, Ming-Jung
Hobma, Sjoerd
Hoffman, Steven
Homenko, Donna
Hosokawa, Michael
Howie, Linsey
Howley, Lisa
Hrisos, Susan
Hsiao, Mei-Chun
Hudson, Nicky
Hulsman, Robert
Humphris, Gerry
Hurley, Katrina
Hutton, Ingrid
Hyde, Catherine
Iblher, Peter
Illing, Jan
Ireland, Bob
Ishikawa, Hirono
Issenberg, Barry
Jacklin, Rosamond
Jackson, Cathy
Jaffery, Tara
James, William
Jamieson, Susan
Jobson, Naomi
Johnsen, Kyle
Johnson, Carl
Johnson, Jane
Jolly, Brian
Jones, Alison
Jones, Claire
Jones, Ian
Jones, Roger
Jouquan, Jean
Kadish, Stacey
Kalishman, Summers
Kandiah, David
Kane, Elizabeth
Kannai, Ruth
Kassab, Salah
Kaufman, David
Keane, Aisling
Kelly, Diane R

Kennedy, Tara
Kerridge, Ian
Khalil, Mohammed
Khan, Hassan
Khang, Peter
Kiessling, Anna
Klamen, Debra
Kneebone, Roger
Knight, Lynn
Knutson, Doug
Kogan, Jennifer
Kon, Alexander
Korin, Tatum
Kraft, Hans
Kreiter, Clarence
Krupat, Edward
Kuper, Ayelet
Lai, Nai Ming
Lambert, Trevor
Lane, Claire
Langendyk, Vicki
Langille, Donald
Laurence, Caroline
Lawson, Mary
Lazarus, Cathy
Leblanc, Vicki
Leinster, Sam
Leite, Maria Teresa
Lempp, Heidi
Levine, Rachel
Levinson, Anthony
Lewis, Clare
Lewis, Mark
Lillicrap, Mark
Lim, Erle
Linssen, T.A.W.
Lischka, Martin
Littlefield, John
Lloyd-Jones, Gaynor
Lohfeld, Lynne
Lombarts, Kiki
Löwe, Bernd
Lumma, Antje
Lunn, Brian
Lyon, Patricia
Macdougall, Colin
Macdougall, J
MacLeod, Rod
MacLeod, Una
Macnaughton, Jane
Macpherson, Cheryl
Magrane, Diane
Maia, Jose
Majoor, Gerard
Malling, Bente
Mamede, Silvia
Mandin, Henry
Mann, Karen
Manning, Gillian
March, Gail
Margolis, Stephen
Marshall, Robert
Martin, Anne
Martin, Ashleigh
Martino, Steve
Marz, Richard
Matheson, David
Matheson, Ruth
Matveevskii, Alexander
Maudsley, Gillian
Mavis, Brian
Mazor, Kathleen
McCaig, Dorothy
McCrorie, Peter
McHanwell, Stephen
McKendree, Jean
McKenzie, Kenneth
McKimm, Judy
McKinley, Danette
McKinley, Robert
McKinstry, Brian
McLaughlin, Kevin
McLean, Michelle
McWhorter, David
Mehraj, Vikram
Melville, Colin
Metcalf, Jane
Mettananda, Sachith
Michels, Martina
Miflin, Barbara
Miles, Jon
Miller, Grant
Minamiya, Yoshihiro
Mistretta, Martin
Monajemi, Alireza
Moon, Robert
Morahan, Page
Morcke, Anne
Moreira, Juan
Moriarty, Helen
Moss, Shannon
Moyer, Cheryl
Muijtjens, Arno
Murdoch, John
Musick, David
Musselman, Laura
Nadelson, Carol
Nagaj, Robert
Nenner, Frederick
Nestel, Debra
Nguyen, Diem-Quyen
Nicholls, John
Niemi-Murola, Leila
Nieminen, Juha
Nomura, Kyoko
Norcini, John
Norman, Geoff
Nothnagle, Melissa
O'Keefe, Maree
O'Neill, Bernadette
O'Neill, Paul
Osorno Covarrubias, Lorenzo
Owen, Cathy
Palmer, Robert
Pangaro, Louis
Papa, Frank
Pape, Sarah
Parikh, Rakesh
Patel, Leena
Patil, Nivritti
Paulus, Isabelle
Pawlikowska, Teresa
Peile, Eward
Peters, Sarah
Petersen, Stewart
Peterson, Ray
Petrusa, Emil
Phillips, Barbara
Plack, Margaret
Platt, Mary Jane
Ponnamperuma, Gominda
Popadiuk, Cathy
Postlethwaite, Keith
Postma, Cornelius
Powis, David
Prescott-Clements, Linda
Prins, Frans
Pugnaire, Michele
Pugsley, Lesley
Puljak, Livia
Punja, Zahra
Rabe, Milagros
Rajasekaran, Anand
Ranmuthugala, Geetha
Rao, Rahul
Reece, Ashley
Reed, Darcy
Rees, Charlotte
Rees, John
Regehr, Glenn
Reid, Steve
Reiter, Harold
Renshaw, Anthony
Rentmeester, Christy
Rhodes, Lesley

Richards, Boyd
Richards, Melanie
Riddle, Janet
Ridgway, Paul
Rikers, Remy
Ringsted, Charlotte
Roberts, Chris
Roberts, Jane
Roberts, Lesley
Roberts, Trudie
Robertson, Margaret
Roff, Sue
Rosenfeld, Gary
Round, Jonathan
Røvik, Jan
Rubin, David
Rudland, Joy
Runciman, Sue
Rushforth, Bruno
Sadoski, Mark
Salinas Sánchez, Antonio
Sanchis-Gimeno, Juan A.
Sanghrajka, Anish
Sargeant, Joan
Satterfield, Jason
Saxena, Anurag
Scallan, Sam
Scheele, Fedde
Scherpbier, Albert
Schijven, Marlies
Schmidt, Henk
Schnatz, Peter
Schneider, Heide
Schönrock-Adema, Johanna
Schuwirth, Lambert
Schwartz, Lisa
Schwartz, Peter
Scott, Craig
Seabrook, Mary
Searle, Judy
Seeberger, Astrid
Sefton, Ann
Selby, Georgina
Sen Gupta, Tarun
Serodio, Aluisio
Shaikh, Faisal
Shaikh, Nader
Shankar, Ravi
Shea, Judy
Silverman, Jonathan
Simon, Frank
Simsek, Sukran
Skakun, Ernest
Smail, Simon
Snadden, David
Sockalingam, Nachamma
Soethout, Marc
Somers, George
Southgate, Lesley
Spence, William
Spencer, John
Spike, Neil
Stalmeijer, Renee
Stansfield, R.
Stebbing, Justin
Steele, David
Steinert, Yvonne
Sterling, Gerald
Steven, Alison
Stewart, Derek
Stewart, Jane
Supino, Phyllis
Tabari Khomeiran, Rasoul
Taherian, Kasra
Tan, Edward
Taylor, David C M
Tekian, Ara
Teunissen, Pim
Thakore, Hemal
Thistlethwaite, Jill
Thompson, Britta
Tilburt, Jon
Till, Hettie
Tiwari, Agnes
Tolsgaard, Martin
Torppa, Auri
Toumi, Zaher
Trumble, Stephen
Tsai, Li
Turato, Egberto
Turnbull, Deborah
Van De Ridder, Johanna
Van Der Vleuten, Cees
Van Hell, Elisabeth
Van Mook, Walther
Van Nuland, Marc
Van Sandijk, Sebastiaan
Veloski, Jon
Vernon, Bryan
Vieira, Dorice
Viljoen, Margaretha
Von Fragstein, Martin
Vu, Nu
Wagner, Peggy
Wallerstedt, Sven
Walters, Kate
Warburton, Katharine
Warnock, Fay
Warren, Fiona
Wass, Val
Watson, M
Wayne, Diane
Wear, Delese
Wee, Bee
Westwood, Olwyn
Wilkinson, David
Wilkinson, Tim
Willett, Timothy
Williams, Brett
Williams, Reed
Wilson, Ian
Wilson, Sam
Wimmers, Paul
Wing, Lindon
Wittekindt, Boris
Wong, Samuel
Wood, Timothy
Woods, Nicole
Woolf, Katherine
Worley, Paul
Wright, Sarah
Wright, Scott
Wulff, Sherry
Yates, Janet
Yi, Michael
Younie, Mai Louise
Ziv, Amitai
Zuberi, Rukhsana