Medical Education

In this issue

in this issue

pages 623–623

New journal Editor required from January 2008

New journal Editor required from January 2008

pages 624–624

Commentaries

The Academy of Medical Educators: a professional home for medical educators in the UK

John Bligh & Julie Brice

pages 625–627

The Richard Farrow Gold Medal

Allan Cumming

pages 627–628

My Story

Standing in the river: a personal statement

Maria A Said

pages 629–631
Student Attitudes

Self-report of quality of medical student health care

Susan J Henderson, Lisa Elon & Erica Frank
pages 632–637

High-quality learning: harder to achieve than we think?

Karen Mattick & Lynn Knight
pages 638–644

Perceptions of and attitudes towards medical errors among medical trainees

David Muller & Katherine Ornstein
pages 645–652

Becoming a new doctor: a learning or survival exercise?

Jeremy Brown, Tom Chapman & David Graham
pages 653–660

Assessment

The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year

George R Bergus & Clarence D Kreiter
pages 661–666

Factors in faculty evaluation of medical students' performance

Andrew R Pulito, Michael B Donnelly & Margaret Plymale
pages 667–675

Comparison of two instruments for assessing communication skills in a general practice objective structured clinical examination

Marc Van Nuland, Wim Van Den Noortgate, Jan Degryse & Jo Goedhuys
pages 676–683

Empathy

A preliminary study of empathy, emotional intelligence and examination performance in MBChB students

Elizabeth J. Austin, Phillip Evans, Belinda Magnus & Katie O'Hanlon
pages 684–689
The role of empathy in establishing rapport in the consultation: a new model

Tim Norfolk, Kamal Birdi & Deirdre Walsh
pages 690–697

Other Research

Core undergraduate psychiatry: what do non-specialists need to know?

Sam Wilson, John M Eagles, Julie E Platt & Hamish McKenzie
pages 698–702

A comparison between senior medical students' and pharmacy pre-registrants' knowledge and perceptions of generic medicines

Mohamed Azmi Hassali, David C M Kong & Kay Stewart
pages 703–710

Electronic discussion forums in medical ethics education: the impact of didactic guidelines and netiquette

Herman Buelens, Nicole Totté, Ann Deketelaere & Kris Dierickx
pages 711–717

Letters to the Editor

Identifying the aspects of rural exposure that influence medical students to take up rural practice

Geetha Ranmuthugala, John Humphreys, Barbara Solarsh, John Wakerman, James A Dunbar & Geoff Solarsh
pages 718–718

Letters to Editor

Response to Blake & Gusella on the impact of simulation on people who act as simulated patients

L Bokken & J-J Rethans
pages 719–719
What is a medical error?

Little is understood about how ‘medical error’ is defined by doctors and trainees. Did an error occur even if there is no adverse outcome? Who feels responsible for an error if it was made by a team? Do definitions of error change with increasing clinical responsibility? Ornstein and colleagues found that trainees had increasing feelings of guilt as outcomes worsened, and decreasing guilt when errors were made by a team. These and other factors will have to be explored if doctors are to interpret studies on medical errors accurately and adequately prepare trainees for the errors they will inevitably make.

See pages 645–652

Clinical performance ratings

Pulito and colleagues analysed the global scale ratings of 211 clerkship students to determine the relative importance of various performance characteristics in grading decisions made by faculty. At least 75% of the students’ grades could be predicted from the rating of any 1 of the 10 performance characteristics. The addition of a second or third variable added no more than 4% to the best predictive model. Faculty appear to make a simple, undifferentiated judgement regarding the quality of a student’s performance, which is then reflected to a high degree in the rating of each of the performance characteristics.

See pages 667–675

Communication assessment instruments

Research into communication skills relies on valid and reliable assessment instruments. Many communication assessment instruments have been developed during recent decades, but they seem to differ in many ways making it a troublesome task to make a choice. This research, by Van Nuland and colleagues at the Catholic University of Leuven, evaluates two such instruments that have been rated as among the best. Both instruments were used simultaneously in a final-year general practice OSCE. Results show that although the instruments have convergent validity, they differ in several ways. These differences may assist researchers in making a choice.

See pages 676–683

Making the curriculum relevant

Topics for inclusion in a syllabus for a medical specialty area are typically decided by the specialists themselves. But what use is such learning to most medical students, given the majority of them will end up working outside that specialty? Wilson and colleagues used Delphi methodology to ask non-psychiatrist hospital consultants, trainees and general practitioners what elements of psychiatry were relevant to their day-to-day practice. This approach informed the development of a core curriculum in psychiatry, and the authors propose that the same methods can be used for any medical specialty.

See pages 698–702
New journal Editor required from January 2008

Medical Education is an international monthly peer-reviewed journal, with distribution to readers in more than 80 countries. It is the pre-eminent journal in the field of education for health care professionals with a strong track record of publishing cutting-edge research, lively and provocative commentary and practical, high-quality information on best practice in teaching and medical education research. Its impact factor for 2005 was, at 2.232, the highest in the field of education for scientific disciplines (1/21).

The journal’s owners, the Association for the Study of Medical Education (ASME) and Wiley-Blackwell, are now seeking an academic editor to maintain the journal’s high reputation and to continue to develop its content and presence, making it increasingly useful and educationally important for its readers. Suitable applicants will have extensive experience of research and publishing in academic journals. This is an excellent opportunity to lead a well-established journal and to contribute significantly to the development of medical education through its research literature.

The main functions of the post will be: taking responsibility for the content of the journal; developing a strong academic policy and vision and communicating this effectively; maintaining an overview of the manuscript review and submission process; offering effective leadership to the editorial team; working closely with the Editorial Office to ensure that production targets are met.

The ideal editor will be expected to show evidence of the following:

- Strong team leadership and management skills and the ability to inspire and motivate a high-powered team of editors and professional staff;
- An understanding of publication and research ethics and an ability to translate this into practical policy and action;
- The drive and vision to make the journal responsive to its readers’ needs, while maintaining a clear overview of the field as a whole;
- The ability to represent the journal at the highest levels.

In addition applicants should possess:

- Excellent written and verbal communication in English, together with the edge, style and wit that will make the journal stimulating to read;
- Sound academic judgement on the quality of papers;
- An ability to offer clear advice to authors, coupled with the ability to respond fairly and effectively to appeals and enquiries;
- Excellent knowledge of the field of international research in medical education;
- An ability to work to tight deadlines;
- An awareness of the commercial field in which the journal operates and an ability to advise on marketing priorities.

It is expected that the workload will require a maximum of 16 hours per week, although this is likely to be highly variable. Attendance at the key international meetings in medical education will be required.

In return the Editor will be offered full support from an experienced professional staff (the journal office is based in Plymouth UK, although applicants need not be UK-based and applications from individuals based in other countries are welcomed). An honorarium and travel allowance will be payable.

The appointment is for five years in the first instance, subject to annual appraisal.

Expressions of interest should be conveyed by 30th September 2007. For informal preliminary discussions about the post please contact:

Dr Graham Buckley, Chair, ASME, 12 Queen Street, Edinburgh, EH2 1JE, UK, Telephone: +44 131 225 9111, E-mail: graham@asme.org.uk

To obtain an application pack and for further information, please contact:

Mrs Julie Brice, Managing Editor, Medical Education, Room 102, ITTC Building, Tamar Science Park, Davy Road, Plymouth PL6 8BX UK, Telephone: +44 1752 764412, E-mail: Julie.brice@mededuc.com

doi:10.1111/j.1365-2923.2007.02808.x
The Academy of Medical Educators: a professional home for medical educators in the UK

John Bligh & Julie Brice

A new national Academy of Medical Educators recently sprang into existence in the UK. A Transition Council has already established it as a permanent incorporated body and the first applications for membership are currently being received.1 But why was it considered necessary to form a new body for medical educators and what is to be the Academy’s purpose?

In many ways, medical education as a field of research and professional practice is more threatened than it has ever been. Time available for teaching has been seriously curtailed with the new consultants’ contract. There is a consensus in the medical education community that the teaching of doctors needs to be taken more seriously as a field of practice and research. Medical teaching does not attract the career status it deserves and the result is that promising young medical educators may be deterred and disheartened almost before they have begun their teaching careers.2

There are many reasons for this, chief among which is that, traditionally, teaching medical students has been seen as one task among the many that doctors perform, unpaid and unrecognised, as part of their professional role. It takes time for attitudes to change.3,4 But the reasons for medical education’s lack of support and recognition go deeper than that. There are a number of structural, organisational and cultural challenges that affect funding and support at institutional, regional and national levels.

Nonetheless, there are grounds for optimism. It is only recently that medical education research has been included in the UK national research assessment exercise,5 new training posts for future generations of academic clinical medical educators have been centrally funded,6 and a growing number of non-clinical educators are becoming involved in preparing future generations for clinical practice in an increasingly open and inclusive atmosphere within medical schools. One flourishing organisation in the field, the Association for the Study of Medical Education (ASME), celebrates its 50th anniversary in 2007. As a Learned Society of members committed to the study of medical education, it plays a prominent part in fostering and supporting medical education research through its meetings, conferences, awards and small grants programmes and, of course, through its journals, Medical Education and The Clinical Teacher.

How can medical education in the UK build on these foundations and draw together that ‘critical mass’ which is key to establishing an ‘effective community of medical educators’?7 In the belief that medical educators need a professional home where their contributions and achievements may be formally recognised and their careers fostered and supported, organisers hope that the new UK Academy will provide the critical mass needed to move medical education on to the next stage of its development.

In the USA, some institutions have, at a local level, attempted to respond to threats to the educational mission of medical schools by setting up academies of medical educators. Irby and colleagues8 studied 8 US institutional academies and identified 4 defining characteristics – mission, membership, organisational structure and resources – that distinguish academies from other types of initiatives. These characteristics are worth discussing here, because they reflect on the thinking behind the UK Academy, which, although a national initiative, is, like its American namesakes, explicitly designed to represent a structural approach to bringing about change.

MISSION

Like the US academies, the UK Academy has a clear sense of mission, summed up in its motto ‘Improving clinical care through..."
teaching excellence’. Specifically, its 3 main aims are:

1 to develop and sustain medical education as an academic discipline;

2 to support academic and professional leadership in medical education, and

3 to develop and support a transparent career structure for specialist teachers in medical education (both medical and non-medical) and to develop and support all those involved in medical education, including senior students.¹

What makes the UK Academy unique is that, for the first time in the UK, there is a body specifically formed to offer opportunities for medical teachers to demonstrate their expertise and achievements in medical education through a formal accreditation and re-accreditation process. Such accreditation may be essential in future professional revalidation processes, and to this end the initiative has been warmly welcomed by the General Medical Council and the Postgraduate Medical Education Training Board. As a standard-setting body, the new Academy will be ideally placed to lobby for clear and formal career structures for medical educators, something which has been lacking within the UK for too long.⁹

MEMBERSHIP

Every active participant in the enterprise of medical education (and this includes teachers, assessors, managers, researchers, curriculum developers and senior students) is a potential member of the UK Academy. As with the US academies, there are criteria for membership and an application process, and there may be different levels of membership as standard setting is an important part of the Academy’s role. Nevertheless, as a reflection of the emerging culture of inclusiveness in the field, everyone, medically qualified or not, who is actively involved in and committed to the development of medical education will be eligible to apply for membership. Once the Academy is established, the achievement and maintenance of full membership will be subject to standards of teaching and scholarship as laid down in the Curriculum for Medical Educators, which is still under development.

STRUCTURE

Among the most important components of academy structure in the USA are management committees, which, in addition to safeguarding the missions of the organisations they govern, offer opportunities to develop wider mission functions. The relationships these academies develop with other connected organisations and institutions allow for strong advocacy roles that ‘go beyond department boundaries’.⁷ The UK Academy likewise hopes to contribute to policy making within the UK at the highest level, along with existing bodies such as the Council of Heads of Medical Schools and the medical Royal Colleges.

At present, the UK Academy is governed by a temporary Transition Council, with Professor Dame Lesley Southgate and Professor Sir Kenneth Calman as its joint chairs. All members of the Transition Council will stand down towards the end of 2007, when the first full Council will be democratically elected by the membership. One of the newly formed Council’s first jobs will be to appoint its own officers (including its foundation president) to replace the current volunteers, none of whom will continue in post past the first elections.

RESOURCES

Although the UK Academy has so far been able to attract a small amount of seed funding, it is anticipated that its main source of income will come from membership subscriptions. These will be kept to a reasonable level, with clinical, non-clinical and student rates.

THE WAY FORWARD

It is, of course, too early to predict what effect the new Academy will have on existing structures and cultures within the UK context, but in terms of encouraging a new sense of mission, professional confidence and career focus for medical educators, most observers agree that it is likely to be both profound and valuable. Much depends on the large numbers of medical teachers in the UK who, it is hoped, will constitute the future membership of the Academy. Such members could provide the impetus to effect lasting and profound change on the status of medical education in the UK – it just requires enough of them to form a critical mass.

Irby et al. argue that if change in the professional status of medical education is to be successful, it must occur at the structural level.⁸ The new Academy is a structural innovation designed to bring about change at a national level within the UK, but concerns with the formal structures surrounding the education of health professionals are not unique to the UK or even to the northern hemisphere. They are part of the day-to-day issues faced...
Professor John Simpson is to receive the 2007 Richard Farrow Gold Medal. This inaugural award, for outstanding contribution to medical and healthcare education, celebrates the Golden Jubilee year of the Association for the Study of Medical Education (ASME). It is made in commemoration of the life of Richard Farrow, who made a significant contribution to medical education in the UK before his untimely death in 2004.

John Simpson was born in Toronto, Canada, but undertook his secondary and university education in Aberdeen, in the UK, graduating in medicine in 1965 and gaining a PhD in 1975. He trained in pathology, and had a particular interest in kidney disease and transplantation. He has published over 100 original papers in this area, has held 19 grant awards and became head of the Department of Pathology in Aberdeen prior to his retirement in 2006.

The award to John of the Richard Farrow Gold Medal recognises his achievements in medical education, locally, nationally and internationally, over the last 20 years. Starting as an interested and motivated teacher, John followed a path within his own university that led him to a position of responsibility at curricular level, to associate dean status and, subsequently, to a Personal Chair in Medical Education. As Chair of the Curriculum Committee in Aberdeen from 1990, John led the introduction of a new MBChB curriculum, which embodied many of the principles that subsequently emerged in the General Medical Council document *Tomorrow’s Doctors*. It is clear that changes in Aberdeen informed the framing of that document. John went on to pioneer other important developments in education in Aberdeen, including the establishment of a clinical skills teaching programme and the early adoption of e-learning to support the medical curriculum.

John was among the first teaching deans to negotiate the use of Additional Cost of Teaching funds to support clinical fellowships in medical education. This scheme aimed to professionalise medical education by creating a cohort of young doctors with training and experience in the field. Several of the original holders of these posts have now been appointed to senior National Health Service or university posts with key educational responsibilities. As such schemes become increasingly common around the country, the benefits of this work are coming to the fore.

In 1997, John initiated regular meetings of the teaching deans of the Scottish medical schools, a group that subsequently received funding and was formalised as the Scottish Deans Medical Curriculum Group. John chaired this group from its inception in 1997 until 2005. His wise and sensitive chairmanship was vital to enabling this collection of disparate individuals to not only meet and discuss, but to actually work together creatively to produce concrete outputs. The best known example is *The Scottish Doctor – Learning Outcomes for the Medical Undergraduate in Scotland*, and the subsequent document dealing with assessment of those outcomes. *The Scottish Doctor* represented the first nationally agreed set of structured and systematic learning outcomes for undergraduate medical degrees, and has influenced developments at national, European and global levels. The group remains active in areas such as shared assessment items and quality standards, and John remains a key member.

John has also filled a demanding international role as Chair of the Academic Council of the International Medical University, based in Malaysia. He retired from his university position in 2006, but remains active and committed in the field of medical education. He is a strong supporter of ASME, the Association for Medical Education in Europe (AMEE) and other relevant organisations.

The award of the Richard Farrow Gold Medal recognises a man who, without seeking to promote his own...
profile, has worked quietly and tirelessly to advance the cause of medical education. Many younger medical educators have benefited from his kind but authoritative mentorship. On a personal level, I recall attending my first ASME meeting in London in 1996. I sat next to John at dinner, and can remember thinking that if he was a typical medical educator, then medical education was a field I would be happy to enter. Others have had similar experiences. The characteristics John embodies – energy, intellect, political awareness, sensitivity, humanity and kindness to others – are very much at the heart of health care education. If John ever does decide to retire ‘for real’ to rural Aberdeenshire, we will all miss his presence and influence greatly.

REFERENCES


Standing in the river: a personal statement

Maria A Said

After my last call night of internship, I fell into a dark, exhausted sleep. Fourteen hours later, I woke and found myself a second year resident and already late in applying for fellowships. All around me the bustle had begun: my fellow residents were signing up for electives, doing research and meeting with attending doctors. I started to bustle as well, dashing this way and that way until I realised I had no idea where I was going.

Woody Allen said that 80% of success is about showing up. I had learned to show up most of the time, and thus had done fairly well in my burgeoning medical career. There was no question of where I should be every day. Medical training is remarkably linear. Anatomy, physiology, clinical clerkships, step 1, step 2, step 3... But what was step 4?

What are you doing? is a common conversation opener among residents. At the time, of course, we’re all doing the same thing. We are sitting nervously around the cafeteria table, pushing chewy, cold French fries into our mouths and waiting for pagers to start vibrating. We are memorising the Brugada Criteria, struggling with central lines that will not wire properly, and trying to find the computer order set for the buprenorphine taper. But when we talk, there is an understanding that we will not be doing this forever. We know that, in the next few months, we will face the dreaded Personal Statement required by applications, in which we must craft a story of how the events in our lives have culminated in our desire for a particular fellowship or job.

In the past, I had written many such essays, describing my trajectory as I hopped from stone to stone across a river; gaining experience, awards and knowledge at each stone until I was only a leap away from the bank on the other side, just needing someone (usually an admissions officer) to hold out a hand and bring me to shore. This time, however, I found myself hopping from stone to stone with no clear view of the shore on the other side. For the first time, I wrote a personal statement for myself so that I could examine the stones I had alighted on, establish which direction I was facing, determine whether I was even in the right tributary, and look for an inkling of firm land. As I wrote, I saw that my momentary perches had become successively more solid, each stone a mineralised and slowly developing amalgamation of who I was, my experiences, the skills I had learned, and the ideals I returned to when the current got swift.

These are 3 of the ‘stones’ that led me to medicine.

Firstly, I liked to solve mysteries. I have found medicine to be reassuringly objective. I had worked as a journalist and freelance writer before returning to school to take pre-medical classes. When I began studying science, I felt an enormous amount of relief that there was a single right answer. While writing, I had been frustrated that what was considered ‘good’ was debatable and might vary according to politics and trends. In the sciences, the facts do not change. The composition of nitrogen remains the same. The earth continues to revolve around the sun. A disease’s mechanism exists and is concrete, and it is our job as doctors to discover its truth. In the classic British mystery, there is a crime, a victim, a murderer, a detective, and 12 or so suspects, all of whom can be implicated in a crime. There is evidence for and against each person, and it is the job of the detective (and the reader) to piece the story together. My enjoyment of medicine is in many ways akin to my enjoyment of mystery stories. There is a disease, a patient, an aetiology, a doctor, and 12 or so possible causes of the disease. The differential diagnosis is the puzzle, and I derive a good deal of satisfaction in sitting with this mystery and trying, in the most logical and thorough way, to solve the crime.
Secondly, I wanted to be a doctor so I could learn how to listen to people better. I had been a Peace Corps volunteer in Eritrea and had spent 2 years teaching large, chaotic, secondary school classes. The days were hot and exhausting, but the times when I felt most gratified were often on the way home from school, when I could walk alongside a student and learn about his life and what filled him with fear or hope. One day, a student, ‘Aisha’, did not appear at the morning flag ceremony. I saw what could happen when someone cannot share their secrets. She was a school leader, a top student, a star in school plays, and a Muslim girl who was secretly seeing a Christian boy. When she became pregnant, she did not turn to her parents or teachers or the doctors in the town to ask for advice or help. She turned to the local medicine woman, underwent a procedure and, 24 hours later, was dead. I returned to the USA hoping to find a way to become someone people could go to: in medicine, I have found a job that allows me to listen to people’s stories and sometimes even help. I will not forget a patient, ‘Juanita’, whom I met on my psychiatry rotation in medical school. When I first met her, she was enveloped in the stained industrial sheets of the inpatient lock-in unit. One day, she emerged and shuffled out of the bedroom into the community room. As we talked, she transformed for me from a ‘psych patient’ to a person I came to respect and admire. She had been born in the Dominican Republic, had been abused as a child, and at 9 years old had escaped from home to live on the streets, disguising herself as a boy to protect herself. Her hardships had continued through her immigration to the USA, but somehow, she had emerged as a powerful community leader in her neighbourhood in Boston. She campaigned for the rights of Latinos and women, until every now and then, when her past caught up with her, she would find herself back on the third floor of the hospital, overwhelmed. I never ordered any blood tests or X-rays on Juanita. I never cured her. But I felt glad that I could be there when she talked.

Thirdly, I wanted to be a doctor because I saw it as a way to work toward justice in the world. I was 14 years old, sitting on a white concrete roof in a slum in the Dominican Republic when I first thought of myself in relationship to the rest of the world. I had grown up in suburban Virginia, attended an all-girls private school, gone to church youth group and been generally protected until my service project to Santo Domingo. Looking back on that project, I am not sure how useful we were to the community we served. There is no doubt, however, that the project served me, revealing a much larger world in which I was immensely privileged. The first day we arrived in the barrio, the local pastor took us on a walking tour. Eight of us followed him down the mud path toward the river, the path growing slicker and wetter as we moved further down. The smell was strong and different to anything I had smelled before. Bottles and plastic bags floated by in little furrowed rivers in the mud path. We entered a tin-roof shack, bare except for 2 metal chairs, a mattress, some cooking utensils, and a piece of red decorative cloth nailed to the wall. The pastor told us that when the river waters rose, the family members would grab all their belongings, stand on the chairs, and wait for the waters to recede. We walked further down toward the river as a drumbeat approached. We turned the corner and encountered a funeral march: 20 people were wailing around a coffin. We marched further on, and I noticed a little boy. He was shy and sweet and wanted to hold somebody’s hand. Like many more children that I would see during that 2-week period, one of his eyes was clouded over from what I now know to be *Chlamydia trachomatis* infection. The feeling then, at 14, that I must work to correct this inequality between us was almost visceral, but sometimes, when I am tired or the day has been hard, I can only barely hold on to that memory.

In writing to find my path forward, I found myself asking another question: ‘What is so great about always having a direction?’ I wondered how healthy or normal it was to be constantly planning the strategy for the next step and navigating the waters to ensure that I keep moving forward. Was it okay to stand still? I thought about the 2 years I had lived in the small, desert town in Eritrea, where I spent a great deal of time doing ‘nothing’. I would sit with my next-door neighbour, Rahwa, roasting coffee beans, pounding them, boiling them, and taking slow sips of sweet, strong coffee. Hours in the late afternoon would pass like this. I am not sure what I accomplished. But I am also sure that this was time well spent. Annie Dillard wrote, ‘I would like to learn, or remember, how to live... I might learn something of mindfulness, something of the purity of living in the physical senses and the dignity of living without bias or motive.’

To live without bias or motive seems impossible in a system in which we are asked to apply for fellowships early in our second year, before we are even halfway through our residency training. The overwhelming pressure to establish a career, publish papers, do research and be accepted to the next level can overshadow the ability to live in the day-to-day. I still don’t know where the stones in the river are going.
and whether I will apply this year for a fellowship. But I have regained the memory of why I started on this path and of what sustains me. As I go through my training, I want to be deliberate as I stand on my stones in the current, and enjoy, not fear, the river as it passes by.

Acknowledgements: I wish to thank Dr Scott Wright for his mentorship and support in completing this article.

Funding: none.
Conflicts of interest: none.
Ethical approval: not required.

Received 16 November 2006; accepted for publication 13 February 2007
Self-report of quality of medical student health care

Susan J Henderson,1 Lisa Elon2 & Erica Frank1

OBJECTIVES To summarise survey results for the quality of medical students’ personal health care, characterise the results according to the demographics and career orientations of the students, and evaluate the relationship between the perceived quality of health care received and the degree of emphasis on prevention in the health care provided.

METHODS We carried out a cross-sectional study with 2316 medical students in the class of 2003 from 16 medical schools, surveyed at 3 points during their training. We used a self-administered questionnaire designed to assess personal health care and related variables in medical students.

RESULTS The majority (92%) reported receiving health care that was at least good, but only a minority (23%) said they received excellent health care. Half had a regular doctor. Health care quality was rated more highly at Year 1 orientation than at later time-points by students who had a regular personal doctor, and especially by those with personal doctors who emphasised prevention.

CONCLUSIONS The majority of medical students perceived that they had received health care that was good or better, but most did not believe it was excellent. As the provision of preventive care is important to students, increasing the amount of preventive care provided to students may both increase their personal satisfaction with their health care and model good clinical preventive practices for them.

INTRODUCTION

In 2005, 67 000 medical students were enrolled in 125 Liaison Committee on Medical Education (LCME) accredited medical schools in the USA. These students will undergo a rigorous course of training, affecting their physical and psychological health and shaping their perceptions of the American health care system. According to the LCME, ‘Medical students must have access to preventive and therapeutic health services.’

Some of the first research on medical student health was conducted at the University of New Mexico and identified the issues of substance abuse and difficult access to care. To expand upon this work, the Collaborative Research Group on Medical Student Health was formed. In its survey of over 1000 students at 9 medical schools, the group found that 90% of medical students felt they needed health care, and nearly half had trouble accessing care despite having health insurance. Reasons for seeking medical care included preventive care needs (81%), mental health (49%), infections (51%) and substance abuse (10%).

The LCME does not define what would constitute high-quality health care for medical students, but the Institute of Medicine has identified the need to improve health care quality in all populations. In a Kaiser Family Foundation and Agency for Health Care Quality and Research survey of the general public, Americans identified the quality of their...
health care as their biggest concern in choosing a health plan. The top 3 components associated with quality health care were doctors’ qualifications, choice of doctors, and the patient–provider relationship.7

The personal health care of doctors has only recently been studied: the Women Physicians’ Health Study assessed female doctors’ ratings of their health care. The majority rated their health care as very good, but not excellent.8 To our knowledge, no studies have addressed medical students’ perceptions of the quality of health care received during their training.

Our research questions were:

1 How do medical students rate the quality of their personal health care visits to doctors?
2 How does the reported quality of care relate to demographics, career orientation and emphasis on prevention?

METHODS

All medical students in the Class of 2003 at 16 US schools were eligible to complete 3 questionnaire administrations during their medical training: at Year 1 orientation (1999); at orientation to wards (2001; typically at the transition between Years 2 and 3 of medical school), and in their senior year (2002–03). School participation was encouraged by offering the summary use of school-specific data (in aggregate and without student identifiers). A 17th school was dropped in 2002 for non-adherence to protocol that resulted in unacceptable response rates of 59% and 35% on the first 2 questionnaire administrations; data from this non-complying school were not included in any analysis.

Our sample of schools was designed to reflect all US medical schools in terms of age (Year 1 average = 24 years [national average = 24 years]), school size (average number of students per school = 563 [national average = 527]), National Institutes of Health (NIH) research ranking (school average = 64 [national average = 62]), private/public school balance (51% private schools [national average = 41%]), under-represented minorities (Blacks, Hispanics and Native Americans = 13% [national average = 11%]), gender (45% women [national average = 43%]) and geographic distribution.9–12

The confidential questionnaires were administered to students outside of formal classroom or training time. Students were informed that their participation was voluntary and that they could choose to withdraw from the study at any time. When the need arose, we used Dillman’s 5-stage mailing process13 to maximise response rates.

Student responses were linked across time using a unique identifier consisting of the student’s mother’s initials at her birth and the student’s father’s first 2 initials. Among the 16 protocol-compliant schools included in the final analyses, a total of 2080 students were eligible to complete the survey at freshman orientation, 1846 of whom responded; 1982 were eligible at entry to wards, 1630 of whom responded; and 1901 were eligible during their senior year, 1469 of whom responded. The overall response rate to our surveys was therefore 83% (4945 responses of 5963 eligible participants), with school response rates ranging from 48% to 98% (median = 85%). A conservative survey response rate calculated with the 17th school was 80%. Not all individuals were
available to respond at all 3 survey points (e.g. because of students leaving temporarily to pursue a complementary degree or returning from such studies). A total of 2316 individuals responded at some point to our survey (971 responded at 3 timepoints, 687 at 2 timepoints (366 at T1 and T2, 161 at T2 and T3, 160 at T1 and T3), and 658 at 1 timepoint (349 at T1, 132 at T2, and 177 at T3), making a total of 4945 observations. Questionnaire items had a median non-response rate of 3%.

Among other health-related questions, students responded to several about their personal health care, including:

- ‘Do you have a regular personal physician? If yes, what gender? If yes, what specialty? (general or family physician, internist, obstetrician or gynaecologist, other)’
- ‘The health care you receive is usually excellent, very good, good, fair or poor?’
- ‘How much emphasis have your physicians placed on helping you stay healthy and preventing disease rather than on treating sicknesses? (A lot, some, not much, none).’

The non-response rates for these items were ≤2%.

We used the chi-square test to evaluate the associations between timepoint in medical school, intended specialty, gender and ethnicity with personal health care variables. As a result of multiple testing, results with a P-value of < 0.01 are considered significant.

There was a lack of independence between observations due to student clustering within schools and longitudinal data collection; in all analyses we adjusted variance estimates to account for this dependence using SUDAAN, treating each school as a cluster and each student’s multiple responses as subclusters.

RESULTS

The demographic characteristics of this population have been previously reported. Half of the surveyed medical students said they currently had a regular doctor, and the majority reported receiving excellent (23%) or very good (39%) health care (Table 1). Of those with a regular doctor, 4% saw someone other than a general or family practitioner, internist or gynaecologist, and 85% of males and 39% of females saw a male doctor (data not shown). Overall, 17% of medical students reported that their doctors placed a lot of emphasis on prevention, compared with treating sickness, and a third felt that their doctors placed little or no emphasis on prevention.

Reporting that one’s doctors placed a lot of emphasis on prevention was associated with Year 1 status (compared with later timepoints), being female, being Black (compared with all other ethnic groups), and having a regular personal doctor. Having a regular doctor was more likely at Year 1 orientation, among females, and among those intending to specialise in primary care. Health care quality was rated more highly at Year 1 orientation rather than later, by those with a regular personal doctor, and especially by those with personal doctors who emphasised prevention: of students who reported that their personal doctors emphasised prevention a lot, only 1% reported fair or poor quality health care, whereas 35% did so if their doctors placed no emphasis on prevention. This relationship was even more pronounced among those without a regular doctor, where 2% reported fair or poor quality health care if their (non-regular) doctors emphasised prevention a lot, whereas 43% reported fair or poor quality health care if their doctors placed no emphasis on prevention (data not shown).

DISCUSSION

In this longitudinal survey of medical students at 16 American medical schools, most students reported their personal health care as being at least good, yet few rated it as excellent. Whereas 30% of Year 1 students reported excellent quality health care, this percentage dropped significantly by the start of clinical rotations to 18% and remained the same at the final year of medical school. Factors other than Year 1 status that were associated with quality health care were having a personal doctor and having a doctor who emphasised prevention.

The link between health care quality and prevention was especially striking: when their doctors placed a lot of emphasis on prevention, only 1% of students rated their health care quality as poor, compared with 35% of those whose doctors placed no emphasis on prevention. This corresponds to a general trend in health care quality ratings in which prevention is included in quality measures: in an Agency for Health Care Research and Quality/Kaiser Family Foundation Survey, the percentage of patients offered preventive services by health plans was an important quality measure. Previous studies of medical student health have identified preventive services such as
vaccinations and health maintenance as the top reasons for office visits.\(^5\)

Over 60% of medical students stated that their doctors placed at least some emphasis on helping them stay healthy as opposed to only treating illnesses. This percentage is fairly low, given that most medical students are in their 20s\(^9\) and are free from chronic illness. In the general population, preventive services also fall behind more acute needs and most doctors do not adhere to recommended screening and counselling guidelines.\(^16\) According to the National Ambulatory Medical Care Survey, preventive visits comprised only 15.7% of office visits in 2003.\(^17\) In addition to this being normative clinical behaviour for US doctors, medical students may be over-scheduled, or perceive themselves as being too busy (or be perceived by clinicians as being too busy) for preventive care.\(^4\) Our data support this: as students move from a pre-medical school life to a highly demanding schedule, the reported focus on prevention in their visits decreases. This may also be because their perception of the importance of prevention in medicine decreases after some time in medical school.\(^18\)

Clinicians’ reported emphasis on prevention differed according to student gender (\(P = 0.0018\)), with more female students reporting that their doctors emphasised prevention. This is not surprising, given that young women have reproductive health care needs that include contraception counselling and prescriptions, and preventive services such as pap smears. Doctor gender was not related to the stronger emphasis on prevention reported by female students. This finding conflicts with several studies, which have reported that female doctors provide more adult\(^19–21\) and adolescent\(^22\) preventive services than do male doctors. One possible explanation might be that female medical students, with their busy schedules,
request appointments based on needs for specific services, such as pap smears, rather than by provider gender. Another unexpected finding in this study was that primary care-orientated medical students were not more likely to report having doctors who emphasised prevention. In a survey of 599 Year 1 medical students, those who planned to go into primary care had more positive attitudes towards prevention. We had hypothesised that such students would be drawn towards prevention-orientated doctors, but students may have limited choices or knowledge about providers. Alternatively, primary care-orientated students may have higher standards for what constitutes an orientation towards prevention.

Both greater reported emphasis on prevention and higher health care quality were associated with having a regular doctor. In this sample of medical students, half the students had a regular personal doctor. The vast majority of the doctors were family doctors, internists and obstetricians or gynaecologists. This association between prevention and doctor continuity is supported by national data: in the National Medical Expenditure Survey of 30 000 adults, individuals with regular doctors were more likely to have received recommended preventive and therapeutic medical services. Likewise, in older adults, having a personal doctor has been associated with more colon cancer screening, and in women, continuity of care has been found to be associated with increased breast and cervical cancer screening. Of concern was the finding that medical students were less likely to have a personal doctor over time: 59% of incoming students had a regular doctor, but only 41% of seniors did. Contributing reasons may include an increase in self-care or in ‘kerbside consults’ in which students seek diagnostic or therapeutic advice informally from medical residents or attending physicians. This is a common practice in medical students: in 2 separate studies, 63% of US students and 43% of British students were found to consult doctors informally who were colleagues, friends or family members.

How can the quality of health care in this population be improved? According to the students, high-quality health care was associated with an emphasis on prevention. Medical students frequently do not seek the help they need because of concerns about confidentiality or repercussions on career decisions. Doctors who care for medical students should be proactive about screening for mental and physical illness and should reassure students that their medical records will remain confidential. Augmenting the amount of preventive care provided to students could both increase their personal satisfaction with their health care and model good clinical preventive practices, as doctors tend to preach the health behaviours they practise.

An important strength of this study is its longitudinal design: students were followed over 4 years of medical school. A large number and a nationally representative variety of schools were included, and student response rates, both overall and on specific items, were high. However, this study has some limitations. First, visits to health care providers such as physician assistants and nurse practitioners were not assessed, although there is an increasing trend to see health professionals other than medical doctors for primary care. Second, the rating of health care quality was subjective on the part of the students rather than based on specific preventive services or performance measures. Third, recall bias is possible in any retrospective study. Although students were surveyed at 3 different points in their training, the questions were posed in such a way that their answers were not restricted to any specific time frame. Therefore, when asked about their experience with health care, they may have responded about their current, recent or previous doctors. Finally, although the medical schools were diverse in size and geographic location, the study was not randomised.

In summary, although the majority of medical students rated their health care as good, only a minority said it was excellent. High-quality care was associated with a focus on prevention. Balancing medical students’ acute health care needs with preventive health screening and counselling is challenging but vitally important because achieving such a balance will have a potentially positive effect on students’ health and on their care of future patients.

Contributors: SJH and LE contributed to data analysis, and writing and editing the article. EF contributed to project design, data analysis, and writing and editing the article. All authors approved the final version.

Acknowledgements: none.

Funding: this project was funded by the American Cancer Society.

Conflicts of interest: none.

Ethical approval: this study was approved by Emory University Institutional Review Board.

REFERENCES


Lambrew JM, Defriese GH, Carey TS, Ricetts TC, Biddle AK. The effects of having a regular doctor on access to primary care. Med Care 1996;34:138–51.


Received 21 June 2006; editorial comments to authors 14 December 2006; accepted for publication 24 January 2007
High-quality learning: harder to achieve than we think?

Karen Mattick1 & Lynn Knight2

context

High-quality learning in the context of medical education can be defined by current conceptions of a deep approach to learning and studying, in combination with metacognitive skills such as personal organisation and reflection on learning. Modern undergraduate education aims to provide an environment that will promote high-quality learning, but this is not as easy to achieve as it might at first seem. Part of the difficulty arises because it is student perceptions of the learning and assessment environment that determine the adopted approach to studying and these are notoriously hard to predict.

Objective

To generate a detailed understanding of aspects that facilitate and inhibit high-quality learning within an innovative, undergraduate medical programme.

Methods

We carried out semi-structured interviews with Year 2 undergraduate students.

Results

Self-directed, problem-based and vocationally relevant activities appeared to promote high-quality learning. Unanticipated barriers to high-quality learning in this setting included a perceived lack of useful feedback on learning, the assessment of applied medical knowledge for a subset of underperforming students, anatomy as a curricular topic and the quantity of information to be assimilated in medicine.

Conclusions

Only by understanding the barriers as they are perceived by students can we design evidence-based modifications to curricula that are likely to be successful in promoting high-quality learning.

Keywords

humans; female; male; students, medical/*psychology; *curriculum; education, medical, undergraduate; problem-based learning; perception; Great Britain; attitude of health personnel.

Medical Education 2007: 41: 638–644
doi:10.1111/j.1365-2923.2007.02783.x

Introduction

Modern undergraduate education aims to provide an environment that promotes high-quality learning in its students. In medical education, high-quality learning can be defined by the latest conceptions of a deep approach to studying1 (learning through interest with the intention of understanding the material covered), combined with broader metacognitive attributes that include aspects of the strategic approach described previously2 (e.g. personal organisation, effort monitoring, effort regulation) and characteristics perceived as desirable in trainee doctors3,4 (e.g. reflective learning). Medical schools can strive to achieve a deep approach through careful curriculum design that anticipates student perceptions of the learning environment, but metacognitive attributes have received less emphasis so far. Both are important to produce graduates with deep understanding of and interest in medicine, knowledge that is transferable to vocational contexts and an intrinsic motivation to study.

The fact that it is student perceptions of the learning and assessment environment that influence the approach adopted, rather than the intentions of curriculum designers or teachers, is crucially important, as these remain notoriously hard to predict. Indeed, even small changes to teaching or
Overview

What is already known on this subject

Desirable approaches to learning and studying predict academic success and probably also good practice beyond graduation.

Student perceptions of the learning and assessment environment determine the approach adopted, but remain notoriously hard to anticipate.

What this study adds

This paper gives a detailed insight into student perceptions of factors that facilitate and inhibit approaches to learning and studying within a contemporary curriculum.

Suggestions for further research

Future research might aim to develop understanding of the impact of evidence-based modifications to the learning and assessment environment.

assessment can sometimes bring about qualitative changes in student learning that are the opposite of that intended. Furthermore, research into undergraduate medical education describes an additional set of largely unexplored influences beyond the formal curriculum, collectively described as the ‘hidden curriculum’, which will also impact on the approaches to learning and studying adopted and therefore on high-quality learning.

Given the great complexity of factors influencing high-quality learning, the presence of potential and actual barriers to achieving it are inevitable, even in the most successful undergraduate programme. Exploring and coming to understand student perceptions of curricula is a vital prerequisite to making targeted, informed developments. This study aims to do this within a contemporary medical programme. Although there have been several quantitative studies of approaches to learning and studying in medical student populations, there has been little to develop a deeper understanding of the issues or of the metacognitive aspects important to high-quality learning.

METHODS

Aims

The data presented here identify specific aspects of an undergraduate medical programme that facilitated and inhibited high-quality learning within a new medical school curriculum. They form part of a larger study to explore approaches to learning and studying and epistemological development in Year 2 medical students.

Context

A number of new medical schools have opened in the UK in recent years in order to produce more home-grown doctors. New schools have the opportunity to design curricula based on educational theory and the latest guidance and evidence base. The context for this study was 1 such school, the 5-year programme of which strives to promote desirable approaches to learning and studying through the constructive alignment of curriculum design, strategies for teaching and learning, timetabling and assessment.

Whereas the first 2 years are spent in a predominantly academic environment, the final 3 are spent almost exclusively in the clinical arena. The academic year is divided into 2-week case units, which include problem-based learning (PBL) sessions, clinical placements and opportunities to reflect on these, biomedical science tutorials, clinical skills training and a small number of lectures. Learning opportunities are designed around a ‘spiral curriculum’, with topics revisited longitudinally to provide reinforcement for learning and increasing complexity. Self-directed learning (SDL) is encouraged by allowing a significant amount of the timetable to remain unstructured. Interspersed amongst the case units are 2-week special study units (SSUs), providing an opportunity to research non-core topics in depth and culminating in the production of a publication-style report.

Assessment is characterised by a ‘little but often’ approach. Applied medical knowledge is assessed by a progress test sat quarterly, featuring clinical vignettes and questions at graduation level, with any topic covered in any test. Clinical skills are tested via regular competency assessments, with an additional integrated structured clinical examination (ISCE) (similar to an objective structured clinical examination (OSCE) but with longer stations combining several activities) at the end of Years 2 and 4. Team-working
and professional behaviours are assessed by tutor and peer judgements, and by reflective portfolios.\textsuperscript{13}

**Semistructured interviews and data analysis**

Full details of the interviews and data analysis are reported elsewhere.\textsuperscript{14} Briefly, 44 Year 2 medical students were invited to participate in the study. They were selected if their Approaches to Learning and Studying Inventory scores\textsuperscript{1,15} had changed significantly between the beginnings of Years 1 and 2. After 15 interviews, data saturation had occurred and no further interviews were performed. The participants varied in their assessment performance: 1 achieved an ‘excellent’ grade in all 6 progress tests taken, whereas 12 received an ‘unsatisfactory’ grade in at least 1 test. Although participants were not selected to be representative of the year group, perusal of the information in Table 1 will aid interpretation of the data. During interviews, we explored students’ previous studying experiences and their perceptions of teaching and assessment at medical school. Interviews were recorded and transcribed verbatim.

Framework analysis\textsuperscript{16} was performed independently by the authors using ATLAS.ti software. The authors worked alone initially, familiarising themselves with the data and identifying key issues, concepts and themes. They met subsequently to discuss and negotiate a thematic framework or index. The data were then removed from their original context and rearranged according to the thematic framework (charting). These charts were then mapped and interpreted across the data as a whole. Five main themes and various subthemes were identified: ‘approaches to learning and studying’; ‘epistemological thinking’;\textsuperscript{14} ‘school to medical school transition’; ‘transition to SDL’, and ‘development of professionalism’. This paper focuses on data from charts that identified curricular factors across the various themes, which appeared to promote or inhibit high-quality learning.

**RESULTS**

**Facilitating factors: curriculum**

Studying at university had more features of higher quality learning than learning in secondary education. At school, students reported having been generally motivated to achieve high grades for entry into medical school, using strategies such as copying and reducing notes to a core that could be memorised. Very few participants described significant interest in the subject matter before university. Once at medical school, students talked about taking on responsibility for learning, moving away from rote memorisation towards achieving personal understanding, linking learning to vocational experiences and introducing reflection on learning.

Specific aspects of the undergraduate medical programme responsible for promoting this desirable transition and those that acted as barriers to high-quality learning are listed in Table 2. Participants reported the need for the vocational relevance of information to be inherently clear or made explicit for high-quality learning to occur. Without this, participants found it difficult to engage with learning because they were unsure if and why they would need to understand or be able to apply this information as a doctor. Seeing the application of information made it easier to understand and remember. Indeed, vocational relevance and the spiral curriculum proved motivating by activating prior knowledge, triggering interest and reinforcing learning.

<table>
<thead>
<tr>
<th>Study participants</th>
<th>Year group</th>
<th>Matteic and Knight, unpublished data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature students*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree level</td>
<td>20% (n = 3)</td>
<td>38%</td>
</tr>
<tr>
<td>PhD level</td>
<td>7% (n = 1)</td>
<td>2%</td>
</tr>
<tr>
<td>Female</td>
<td>46% (n = 6)</td>
<td>51%</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>66% (n = 10)</td>
<td>73%</td>
</tr>
<tr>
<td>Asian</td>
<td>27% (n = 4)</td>
<td>10%</td>
</tr>
<tr>
<td>Mixed</td>
<td>7% (n = 1)</td>
<td>5%</td>
</tr>
</tbody>
</table>

* Mature students are classified as being > 21 years of age on entry to this higher education establishment; school-leavers were classified as being ≤ 21 years of age

| Table 2 Curricular features facilitating and inhibiting high-quality learning |
|-------------------------------|-----------------------------|
| **Facilitating high-quality learning** | **Inhibiting high-quality learning** |
| Learning in context | Perceived quantity of information |
| Self-directed learning | Perceived lack of feedback from assessment |
| Spiral curriculum | Disease frameworks |
| Progress test of applied medical knowledge | Anatomy |
| Summative ISCE | Progress test of applied medical knowledge |

I’ve never had that kind of learning experience before where ... you see things happening that you’ve read up about ... and then you kind of have the knowledge and you can build on it while you’re at placement. That really encourages me to keep reading and things and I think that’s where it all gets more exciting.’ (Participant 14, female, school-leaver)

Self-directed learning appeared daunting to many participants, who were uncertain about the appropriate amount of individual study, what to cover and how its success could be evaluated. The concept of SDL was generally first encountered at medical school. However, SDL gradually encouraged students to take responsibility for their study, as they had to make decisions about what, when and how much to do. Participants perceived this to be closer to the way in which a professional works and to require intrinsic motivation.

The SSUs involved in-depth independent study and generally promoted learning for interest through the wide variety of research topics available and the greater time allocation. Participants described how critical appraisal skills developed through the analysis of published literature could be applied directly to evidence-based decision making in medical practice.14

Problem-based learning promoted high-quality learning by encouraging independent study, interest in topics and an ability to relate ideas. It gave students the independence to choose and prioritise what to study. Having to articulate knowledge in a PBL tutorial was perceived to require deep understanding at a qualitatively different level than that achieved through previous study. Some participants described their growing confidence within PBL and their increasing comfort in questioning contributions and critically appraising the information presented.

Facilitating factors: assessment

Inhibiting factors: curriculum

The perceived quantity of information within medicine provoked considerable anxiety, with many participants unsure about what constituted an appropriate depth of knowledge. An integrated curriculum could sometimes seem to comprise a daunting, amorphous mass of information. Anxiety, combined with a workload perceived as high, was cited as a barrier to high-quality learning:

‘There’s so much to learn you cannot go so deep, so I found I was going very deep to try and understand the science of it and then ... I’m not getting the other stuff done.’ (Participant 6, male, mature entry student)

Some participants focused on studying biomedical science topics, despite apparently clear messages that assessment would test their application to clinical scenarios. This may indicate resistance to deviate from learning strategies that had been successful prior to university. The study of anatomy, distinct from other biomedical sciences, seemed to prompt students to resort to memorising information, despite the faculty’s intention of promoting deep understanding. Participants described a large amount of vocabulary and ‘facts’ in anatomy, which prompted undesirable learning strategies. Even closely related topics were considered to be qualitatively different to anatomy in this respect. This viewpoint was unexpectedly discussed by several participants:

‘Now I don’t actually memorise apart from anatomy, which you just have to memorise... but like physiology and stuff I don’t memorise.’

(Participant 1, female, school-leaver)

Although these are not provided by the school, 2 students described the use of disease frameworks in their SDL, which appeared to promote a formulaic approach to learning at the expense of understanding. Disease frameworks are available via the Internet and, once identified, had been distributed widely amongst the student body:

‘We have these disease frameworks. All we got to do is fill in this framework so it’s become like a, we don’t think about it anymore, it’s like filling in the box... I don’t think they really help me understand the disease.’ (Participant 6, male, mature entry student)

Facilitating factors: assessment

Participants described desirable changes compatible with high-quality learning occurring in response to the progress test. Students reported that understanding and application of information was needed to succeed in the progress test as a result of its use of clinical vignettes. Students believed the test rewarded those who worked consistently. The fact that it could test any topic at any time deterred short-term memory strategies and ‘cramming’:

‘I think that’s what makes progress tests so difficult is because you’ve got to be learning constantly, you...
Many students were looking to intrinsic sources of motivation because they felt they could no longer rely on assessment to motivate them to study. This was partly a result of its frequency, as each round is ‘low stakes’ and only the cumulative effect is important to progression decisions.

Inhibiting factors: assessment

The progress test was sometimes seen as a barrier to high-quality learning, however, when students became excessively anxious about it. Anxiety seemed to drive students to do more work in terms of quantity, but not necessarily the right kind of work in terms of quality:

‘If you’re tensed up you need to really rote learn a lot of it.’ (Participant 8, male, school-leaver)

Participants talked about students who studied hard but did not excel in assessment. Many could not explain this and described it as unfair. Some went on to observe that perhaps these individuals were concentrating on memorising at the expense of understanding, were less aware of the modes of assessment and their rationale, or were failing to tailor their learning to assessment. Thus participants seemed to be describing a subset of students with poor metacognitive skills who did not excel in assessment:

‘There are a lot of people who aren’t doing very well who work a lot harder than me. They do a lot more work. Really clever people. But they just can’t do well in the progress tests.’ (Participant 12, male, school-leaver)

Interestingly, the ‘little but often’ approach to assessment was sometimes perceived as representing an absence of useful feedback on progress. This was perceived as anxiety-inducing for many participants, and it also deterred a focus on assessment:

‘I don’t know if I’m doing enough now. I seem to be doing all right, so I think I’m, think I’m learning enough. But there’s nothing that tells you whether you are or not, I find that quite difficult.’ (Participant 13, female, school-leaver)

‘I kind of had to change the way I motivate myself… I’ve had to do it by myself rather than let other people motivate me I think.’ (Participant 14, female, school-leaver)

Whereas assessment was generally ‘little but often’, the ISCE constituted a single examination which must be passed to progress to Year 3. As such, it had high significance for students and provoked anxiety. Despite this, many participants felt a final examination with high significance such as the ISCE was a good motivator for learning and consolidating the skills learned and practised over the previous 2 years:

‘I ended up studying a couple of things that weren’t in the, in the ISCE because of the general panic that was out there. It’s hard not to be influenced by that.’ (Participant 10, male, mature entry)

Participants felt a cramming-type approach was an appropriate way to prepare for the ISCE. It was unclear if this was because it tested practical protocols for which they could practise by repetition or if it was a response to the high levels of anxiety it invoked. Participants perceived the advice from teachers about preparing for skills-based examinations (‘practise, practise, practise’) to conflict with that for knowledge-based assessment (‘avoid rote learning’). Although these are distinct competencies and guidance was probably appropriate in both contexts, some students did not seem to make the distinction and felt mixed messages were sent by teaching and assessment environments:

‘The ISCE is like a final exam, as such, so I think if you wanted to you could really cram for that.’ (Participant 8, male, school-leaver)

DISCUSSION

In this setting, learning that was self-directed, problem-based, contextualised and project-based was compatible with our definition of high-quality learning. This makes conceptual sense and fits with the existing evidence base. Students using a deep approach spend more time in SDL as a result of their greater interest, and freedom of learning promotes a deep approach. Problem-based learning is an instructional strategy that promotes SDL and understanding. Medical schools using PBL tend to achieve more desirable approaches to learning and studying in students than do traditional curricula. Learning in context is a key principle of adult learning and promotes understanding in medical students. The application of knowledge...
to new contexts and problem-based questions also promotes a deep approach.\(^2^2\) Thus, whilst these curricular factors are working as predicted by the literature, the complexity of students’ perceptions of teaching and assessment are such that this should not be taken for granted.

Barriers to desirable learning were perhaps more interesting as they were unintended consequences of the curriculum. Here we will focus on 2 key issues emerging from this study: feedback on learning, and assessment anxiety and remediation.

Participants felt they did not receive much feedback on progress, despite being given results from 4 progress tests, 4 project reports, numerous skills competencies and 10 personal and professional development judgements per academic year. This may reflect the format, timing or level of detail of feedback and constitutes an important finding as the formative functions of assessment can promote high academic standards.\(^2^3\) Ideally, feedback on learning should be both explicit, from assessment, and implicit, as part of the student–tutor interactions within teaching and learning activities.\(^2^4\) The former is likely to be particularly important within a curriculum that promotes SDL and features reduced contact time with teachers. Interestingly, although the perceived lack of assessment feedback sometimes caused concern and frustration amongst students, it generally promoted high-quality learning in the longer term. In the perceived absence of assessment feedback, participants became motivated by intrinsic factors (where something is done because it is inherently interesting or enjoyable\(^2^5\)), a desirable development for lifelong learning in which there will be no formal assessment. Such consequences must be considered when supporting students in their SDL and the approach must depend on the student in question.

It is important to understand the impact of assessment on what is learned and how, and use this to promote high-quality learning\(^2^6\) and constructive alignment between teaching and assessment.\(^1^0\) The progress test is known to reward students with a deep approach when starting medical school\(^9\) and the current study provides further evidence that it can promote high-quality learning during medical school. Students who were highly sensitive to task demands and able to adapt to the learning environment tended to succeed in assessment. However, receptiveness to factors that should promote high-quality learning was variable, both in this study and elsewhere,\(^2^7\) and previous work indicates that underperforming students often adopted or retained undesirable approaches to learning in response to poor grades (Mattick, Dennis and Bligh 2004, unpublished data). The anxiety associated with poor performance seems to drive a focus on assessment that is incompatible with high-quality learning\(^1^8\) and students may work more intensively without reconsidering their learning conceptions or approaches.\(^2^8\) This situation is potentially compounded by the inherent higher levels of anxiety associated with SDL.\(^2^2\)\(^2^9\) Underperforming students clearly need help and support to become accomplished learners and be successful in higher education and beyond. Difficulties encountered with the transition from school learning to medical school learning might be partially addressed by early study skills workshops targeting appropriate learning strategies and student anxieties. Early identification of underperforming students is essential and remediation should be tailored to meet the student’s needs and delivered by someone with a good understanding of the learning and assessment environment and the student’s future career options. Furthermore, remediation should aim to challenge why, what and how individuals study, the very conception of learning, whilst bearing in mind the premise that a student’s existing conception of learning will shape how he or she interprets guidance.\(^2^8\)

The key recommendation emerging from this study is that medical educators should revisit the ways by which high-quality learning is promoted in their respective programmes and, especially, the ways by which study skills are introduced and underperforming students are remediated. We suspect there is still much that can be improved in these areas. This study confirms that understanding student perceptions of barriers to high-quality learning is vital for designing evidence-based modifications to the programme. The findings are important in our specific context and also more widely, as the lessons learnt in our context will raise awareness of the sorts of barriers to high-quality learning that undoubtedly exist elsewhere.

Contributors: both authors contributed to the initial conception and design of the study, and the data collection and analysis. KM took the lead in writing the paper and LK reviewed subsequent drafts of the manuscript. Both authors approved the final manuscript.

Acknowledgements: we thank our participants, who gave freely of their time to be involved in this study.

Funding: none

Conflicts of interest: none.
REFERENCES

26 Spencer J, Jordan R. Educational outcomes and leadership to meet the needs of modern health care. *Qual Health Care* 2001;10 (Suppl 2):38–45.

Received 17 May 2006; editorial comments to authors 20 June 2006, 14 December 2006; accepted for publication 26 January 2007
Perceptions of and attitudes towards medical errors among medical trainees

David Muller1,2 & Katherine Ornstein2,3

OBJECTIVES The purpose of this study was to develop a better understanding of how medical trainees define medical errors and what factors influence medical trainees’ perceptions of medical errors.

METHODS We surveyed 423 medical students and house staff at an urban academic medical centre to learn about how they defined medical errors, their experiences with medical errors, their beliefs about when a patient should be informed of an error, and their attitudes towards medical errors with differing severity of outcomes.

RESULTS Trainees stated that an event could be considered an error regardless of outcome, negligence, intention or consent. Definitions did not vary according to gender or level of training. Trainees had increasing feelings of guilt and fear as the outcomes related to errors worsened. Respondents were more likely to feel guilty and angry at themselves, and be afraid of accusations of malpractice, losing their licence, damaging their reputation, or losing confidence when errors were made while working individually versus in a team setting. Female trainees were more likely than male trainees to feel guilty and angry at themselves, and were afraid of losing confidence if they made an error.

CONCLUSIONS Trainees’ perceptions and attitudes towards errors vary depending on whether they are in their clinical years, the severity of outcome, and whether the error is attributable to an individual or a team. These factors will have to be explored in greater depth if we are adequately to prepare young doctors for the errors they will inevitably make.

KEYWORDS humans; male; female; students, medical/*psychology; *education, medical, undergraduate; medical staff, hospital/*psychology; perception; *attitude of health personnel; *medical errors.

Medical Education 2007; 41: 645–652
doi:10.1111/j.1365-2923.2007.02784.x

INTRODUCTION

Since the first studies addressing medical errors were published in the early 1990s, and certainly since the Institute of Medicine’s report To Err is Human; Building a Safer Health System was published in 1999, medical errors have held a prominent place in discussions of clinical practice and, to a lesser extent, medical education. Aside from the focus on how often, and why, errors occur,1,2 there has been a great deal of exploration around issues such as the definition of error:2–4 the impact on doctors and how they cope;3,5–8 whether errors should be disclosed to patients and their families,4,9–14 the perceptions of patients,8,10,15 practising doctors6,10,12,13,16 and trainees,2–4,10,11,17,18 and proposed educational programmes.5,19 For example, Mizrahi found that 47% of the 74 house staff she interviewed had committed ‘grave’ errors during the course of training.3 Newman reported that 77% of the 30 family practitioners he surveyed reported having made a mistake,7 and Wu et al. reported that 45% of 254 house staff had also made an error.2 These widely varying figures are in part a result of different definitions of error: ‘an unanticipated negative consequence of a medical intervention’; an act or omission for which the house officer felt responsible...
that had serious or potentially serious consequences for the patient and that would have been judged wrong by knowledgeable peers at the time it occurred, or failure of a planned action to be completed as intended, or the use of a wrong plan to achieve an aim. The steep learning curve of medical school and residency creates an environment that is error-prone. However, the literature about errors encountered during training deals almost exclusively with coping skills and the ethical dilemma of disclosure. Only 1 report addresses the types, perceived causes of and responses to errors among house staff. There has been no published report describing how medical trainees define ‘medical mistake’, what their attitudes and perceptions are of different types of mistakes, what factors influence their perceptions, and whether their perceptions vary by level of training.

We sought to answer these questions by surveying medical students and house staff at different milestones in their medical training. We hypothesized that the factors which most strongly influence trainee perceptions are whether or not the error has an adverse outcome and whether the error is made by an individual or by a team of providers. We believe that perceptions of error probably change with exposure to clinical work and that the major transition takes place once trainees have begun their clinical training.

METHODS

Participants

The study was conducted at an urban tertiary care academic medical centre. At the transition between the academic years 2003 and 2004, we distributed surveys to medical trainees ranging from incoming medical students to graduating internal medicine house staff. Participants completed self-administered questionnaires anonymously.

Data collection

We did not find a validated survey instrument in the literature, so a survey was developed in collaboration with other medical educators. The survey instrument was designed to ascertain medical trainees’ definition of medical errors, personal and professional experience with medical errors, opinions about when patients should be informed of medical errors, and attitudes towards medical errors. The survey was reviewed by the authors for face validity and was pilot-tested on a sample of medical residents and revised accordingly.

Overview

What is already known on this subject

Previous research has mostly addressed the ethics of disclosure and coping skills for trainees. There has been scant focus on how trainees learn to define errors and what determines how they respond to them.

What this study adds

This study elucidates how trainees with various levels of experience define errors and how attitudes towards errors vary, depending on gender, level of training, severity of outcome, and whether the error is attributable to an individual or a team.

Suggestions for further research

Better understanding of what doctors think constitutes an error will help educators plan interventions that can clarify and reinforce the importance of documentation, disclosure and reflective practice.

Table 1 Definitions of medical mistakes

<table>
<thead>
<tr>
<th>Number</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A medical mistake is made only if the doctor is negligent</td>
</tr>
<tr>
<td>2</td>
<td>A medical mistake is made even if the doctor is diligent</td>
</tr>
<tr>
<td>3</td>
<td>A medical mistake is made only if the doctor has negative intentions</td>
</tr>
<tr>
<td>4</td>
<td>A medical mistake is made even if the doctor is well intentioned</td>
</tr>
<tr>
<td>5</td>
<td>A medical mistake is made only if there is a negative outcome</td>
</tr>
<tr>
<td>6</td>
<td>A medical mistake is made even if there is no negative outcome</td>
</tr>
<tr>
<td>7</td>
<td>A medical mistake is made only if informed consent was not obtained</td>
</tr>
<tr>
<td>8</td>
<td>A medical mistake is made even if informed consent was obtained</td>
</tr>
</tbody>
</table>

Definitions of medical errors were measured using 8 items (Table 1) rated on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. Each item was reverse-coded when appropriate. We combined the 8 items into a summary score ranging from 8 to 40, in which a lower score indicates a broader definition of medical errors and a higher score represents an extremely narrow view of medical errors (i.e. a medical error only occurs in extreme situations, such as the doctor being negligent). Cronbach’s \( \alpha \) for the scale was 0.84, suggesting high internal consistency.

Experience with errors was assessed by asking respondents if they or anyone close to them had been the victim of a medical error and whether or not they had made any errors with varying degrees of adverse outcomes (none, reversible, irreversible).

Respondents were asked in a multiple-choice format to indicate when they would inform a patient of an error and when it was acceptable not to do so. Free text comments were also allowed for these questions. Trainees were also asked whether they felt their attitudes about errors had changed over time. Attitudes towards medical errors were measured using 7 statements to which (dis)agreement was measured on a 5-point Likert scale. Respondents were asked about their attitudes concerning errors that occurred when they were working independently versus as the doctor on an interdisciplinary team, as well as when the medical error had:

1. no adverse outcome;
2. a reversible adverse outcome, or
3. an irreversible adverse outcome.

Data on gender, ethnicity and level of training were also collected.

Data collection

The data were analysed for any significant differences across demographic information, definition of medical errors, experience in being the victim of medical errors, and committing medical errors according to training level and gender using \( t \)-tests and chi-square analyses. Respondents were divided into 2 groups for level of training based on clinical experience: pre-clinical training (Years 1 and 2), and post-clinical training (Year 3 students and house staff).

Paired \( t \)-tests were performed to test between differences in attitudes towards medical errors considered to have been committed by a team versus those committed by an individual. Paired \( t \)-tests were also used to compare differences between attitudes when a mistake is made with:

1. no adverse outcome compared with a reversible outcome;
2. no adverse outcome compared with an irreversible adverse outcome, and
3. a reversible adverse outcome compared with an irreversible adverse outcome.

Gender and training differences in attitudes were measured using \( \chi^2 \)-tests. Adjustments were made for other variables using linear regression analysis. \( p \)-values of <0.05 were used to determine level of significance.

The data from the surveys were tabulated and analysed using Stata Version 8.0 for Windows.

RESULTS

A total of 423 trainees, 219 with pre-clinical training only and 204 with post-clinical training, completed the study, for an overall response rate of 87%. The response rate for those with only pre-clinical training was 92%, compared with 82% for those with post-clinical training. Table 2 provides a detailed breakdown by year of training. Medical students were recruited in a classroom setting, which accounts for their high response rate (95%). Residents were asked to fill out the survey in their own time and send it back to staff, resulting in a lower response rate.

As shown in Table 3, the pre-clinical group included a greater percentage of females (63%) compared with the post-clinical training group (49%). With
regard to ethnicity, 62% of the population surveyed were White, 18% were Asian, 6% were Hispanic and 5% were African-American. There were no significant ethnic differences between the training groups. Fully 30% of the sample (n = 127) stated that they or someone they were close to had been the victim of a medical mistake. Although being a victim had no association with level of training, women were more likely than men to make this claim (P < 0.001).

As expected, those with post-clinical training were more likely (P < 0.001) than those with pre-clinical training to admit to having committed medical errors, regardless of the outcomes. Surprisingly, a small percentage of the pre-clinical training group reported having committed medical errors across all outcomes. This is presumably related to clinical encounters prior to medical school or early encounters in the first 2 years. As a separate subgroup, 91% of interns and residents reported having committed errors with no adverse outcomes, 71% reported errors with reversible adverse outcomes, and 22% reported errors with irreversible adverse outcomes.

The overall mean definition of medical error score was 13.8 ± 5.7. There were no significant differences in mean definition scores between men and women or between pre-clinical and post-clinical training groups. When asked when they should inform patients of medical errors, 44.5% stated always, and 1.7% stated never. Those respondents with only pre-clinical training were more likely to state that patients should never be informed about errors (P < 0.05). A total of 27% of all respondents said they would inform patients of an error only if they had been harmed and 17% said they would inform patients only if the patient would benefit from knowing. When asked if their perceptions of medical errors had changed over time, respondents with clinical training were more likely to say yes (P < 0.001). No gender differences were found with regard to committing any type of medical error, informing patients regarding errors, or change in perception of errors over time.

**Attitudes**

**Attitudinal differences based on outcomes of mistake**

Trainees were asked about their attitudes towards medical errors across 3 different types of adverse outcomes: none, reversible, and irreversible. As illustrated in Fig. 1, the overall trend was towards increasing feelings of guilt, anger and fear as the outcomes worsened, with the change from reversible to irreversible being the most extreme. This trend can be seen in both individual and team situations. All pairwise comparisons were significant at the P < 0.001 level, with the exception of being angry at the patient in the team situation (P < 0.05).
Individual versus interdisciplinary team differences

When individuals made errors, they were more likely to feel guilty, be angry at themselves, be afraid of accusations of malpractice, afraid of losing their licence, afraid of damaging their reputation, and afraid of losing confidence than if they made an error as part of a team (Table 4). The significant findings

![Figure 1 Agreement with attitudes about errors statements by varying outcomes in individual and interdisciplinary team situations. *P-value < 0.001 for all pairwise comparisons (no adverse outcomes versus irreversible adverse outcomes, no adverse outcomes versus reversible adverse outcomes, reversible adverse outcomes versus irreversible outcomes); †P-value < 0.05 for all pairwise comparisons (no adverse outcomes versus irreversible adverse outcomes, no adverse outcomes versus reversible adverse outcomes, reversible adverse outcomes versus irreversible outcomes); ‡P-value < 0.001 for no adverse outcomes versus irreversible adverse outcomes and reversible adverse outcomes versus irreversible outcomes only.](image-url)

**Table 4** Differences in attitudes about medical errors committed by an individual and those committed in an interdisciplinary team situation across 3 types of outcome

<table>
<thead>
<tr>
<th></th>
<th>No adverse outcome</th>
<th>Reversible adverse outcome</th>
<th>Irreversible adverse outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel guilty</td>
<td>3.93</td>
<td>4.35</td>
<td>4.80</td>
</tr>
<tr>
<td>Angry at patient</td>
<td>4.00</td>
<td>4.22</td>
<td>4.74</td>
</tr>
<tr>
<td>Afraid of malpractice accusations</td>
<td>3.23</td>
<td>3.39</td>
<td>4.53</td>
</tr>
<tr>
<td>Afraid of losing licence</td>
<td>2.91</td>
<td>3.01</td>
<td>4.43</td>
</tr>
<tr>
<td>Afraid of reputation damage</td>
<td>3.55</td>
<td>3.61</td>
<td>4.55</td>
</tr>
<tr>
<td>Afraid of losing confidence</td>
<td>3.79</td>
<td>3.85</td>
<td>4.54</td>
</tr>
<tr>
<td>Team</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feel guilty</td>
<td>3.08†</td>
<td>3.31‡</td>
<td>4.10‡</td>
</tr>
<tr>
<td>Angry at patient</td>
<td>2.60†</td>
<td>2.77‡</td>
<td>3.56‡</td>
</tr>
<tr>
<td>Afraid of malpractice accusations</td>
<td>2.63†</td>
<td>2.77‡</td>
<td>3.61‡</td>
</tr>
<tr>
<td>Afraid of losing licence</td>
<td>2.36†</td>
<td>2.54‡</td>
<td>3.34‡</td>
</tr>
<tr>
<td>Afraid of reputation damage</td>
<td>2.71†</td>
<td>2.85‡</td>
<td>3.57‡</td>
</tr>
<tr>
<td>Afraid of losing confidence</td>
<td>3.57†</td>
<td>2.64‡</td>
<td>3.26‡</td>
</tr>
</tbody>
</table>

* Individual = mean score for statement with 1 = strongly disagree and 5 = strongly agree when respondent is acting individually
† Team = mean score for statement with 1 = strongly disagree and 5 = strongly agree when respondent is acting as part of an interdisciplinary team
‡ P < 0.001 based on 2-tailed paired t-test
§ P < 0.05 based on 2-tailed paired t-test

in the individual versus the team situation were found across all outcomes. Despite what appear to be some significant differences between groups in the ‘angry at patient’ category, the very low overall Likert scores (none above 1.53) indicate that no respondents actually felt angry at their patients.

Gender and training differences in attitudes

After controlling for level of training (pre-clinical versus post-clinical), female trainees were more likely than male trainees to be afraid of losing confidence if they made an error, regardless of outcomes or whether or not they were working in a team situation (Table 5). Female trainees were more likely to feel guilty and fear all repercussions in nearly every scenario.

After controlling for gender, no differences were found among training groups in attitudes towards errors in team situations, with the exception that those with pre-clinical training were more likely than those with post-clinical training to believe they would lose confidence if they committed errors with no adverse or irreversible outcomes. When working independently, those with pre-clinical training were more afraid than those with more training of accusations of malpractice, losing their licence and damage to their reputation when outcomes were reversibly or irreversibly adverse.

DISCUSSION

During medical training, students and residents are under extreme pressure to learn, perform and conform to the culture of clinical practice. Among the things they struggle with most is the concept that medical errors occur in all settings, to all doctors, and with disturbing frequency. Like much of medical education, they will learn these lessons by watching their role models make mistakes, cope with them on a personal level, and confront them on a professional level. Examples include, at the upper end of the spectrum, the surgeon who sits at the patient’s bedside, with the family present, and explains that he sutured the wrong end of the colon to the colostomy and that the patient will require another surgical procedure. An example of the other extreme may be the interventional cardiologist who uses euphemisms (such as ‘enzyme leak’) to describe complications of cardiac catheterisation, or simply does not confront them at all. Housestaff are keenly aware of both these extremes of coping with error, and everything in between.

Mizrahi explores how house staff acquire their perceptions of mistakes and how they cope with them. She describes 3 different forms of defence – denial, discounting and distancing – that trainees utilise to come to terms with, or rationalise, a mistake. Her data were acquired by observations, interviews and questionnaires, and she laments that there has not been more extensive scrutiny of house staff beliefs and behaviour. Other studies have approached the topic by eliciting case reports from medical students and house staff and using these cases as a platform for discussing root causes, disclosure, outcomes and the emotional toll on trainees.

Our study returns to fundamentals by first trying to clarify what trainees mean when they think of error. Despite the fact that ‘error’ has been defined in the literature, there are variables related to outcome and to intention that may sway a particular doctor’s definition. In fact, among Mizrahi’s first-line defence mechanisms (under denial) is redefinition, which she believes is very common when there has not been a bad outcome. Trainees in our survey defined a negative act as an error regardless of
outcome, yet their responses to hypothetical errors were significantly different depending on outcome. This dual understanding seems to imply that there is variability in the way in which error is identified based on the circumstances surrounding the error. The concept of ‘circumstantial error’ may have a significant impact on our understanding of error rates and research based on rates found in the literature. It also supports Mizrahi’s belief that redefinition is a commonly used defence mechanism. Finally, our finding that guilt, fear and anger increase with severity of outcome provides a plausible basis for Mizrahi’s notion of escalating lines of defence against error.

Whether studying errors among trainees or practising clinicians, it is universally recognised that common responses to making a medical mistake include self-doubt, self-blame, disappointment, shame, fear (of litigation, of losing reputation, of losing trust), remorse and anger. In their study of errors made by house staff, Wu et al. reported that 81% felt remorse, 79% were angry, 60% felt inadequate and 28% feared negative repercussions. We further refined reactions such as anger and fear by specifically addressing what repercussions trainees feared and who the objects of their anger were. We believe that these more refined responses provide a better measure of the complexity of doctors’ reactions when they make mistakes. For example, we found that the patient was almost never the object of a trainee’s anger.

Some studies have shown that responses to error among patients and clinicians change depending on the severity of the outcome, with response increasing in severity as the outcome worsens. These studies also demonstrate clearly that the relationship between response and outcome is much more striking among patients than doctors. We were able to demonstrate that this adverse outcome—response relationship is true of house staff perceptions across a large number of personal and professional reactions. Factors that we hypothesised might have a major influence on response to error, for example, having been victim of a medical error, turned out not to be of great consequence, whereas gender clearly had a major impact, with women feeling much more guilty, angry at themselves, and fearful of losing confidence than their male counterparts. This gender difference has never been explored in the literature, yet it may have a major impact on how we help our trainees understand error and how we train them to better cope with the errors they will inevitably make.

We were unable to find any discussion in the literature of the doctor’s role when the team the doctor is leading makes a mistake. Do doctors accept responsibility for these mistakes? Should they? and how do their responses to a team error compare with their responses to an error they have made on their own? Our data show that, although trainees feel personally responsible for team errors, the degree to which they internalise or accept this obligation appears to be significantly different from that they apply to errors they have committed themselves. As medicine is increasingly practised in the setting of interdisciplinary teams, this distinction on the part of doctors may signal a diffusion of responsibility and accountability that is not compatible with best practices. It is an area that requires further exploration, and possibly needs to be addressed more explicitly in our formal and informal curricula, as we train young doctors to work in, and lead, interdisciplinary teams.

Limitations of our study include the validity of the survey instrument and the number of trainees completing the survey. Our questionnaire was designed by medical educators, piloted on house staff and reviewed for face validity, but it is essentially a new instrument and has not been validated by prior research. In addition, despite good response rates, our total n was low for trainees in internship and residency because the class sizes are so much smaller than medical school classes. This particular limitation precluded us from making some of the potentially more interesting observations about the evolution of perceptions of error as trainees are indoctrinated into the world of clinical medicine. Finally, our results may not generalise to students and house staff not training in an urban academic medical centre that is closely affiliated with a school that has an ethics and medical errors curriculum.

Although we were unable to show evolving trends in perceptions over time, we were able to clearly demonstrate that trainees believe their perceptions of error change with time. Further research will be required if we are to answer questions about how clinical experiences during training may change the way we define, respond to and cope with errors in clinical practice.

Contributors: DM contributed to the study conception and design, and interpretation of data. KO contributed to the study design, and analysis and interpretation of data. Both authors drafted the article and approved the final version.
Acknowledgements: none.
Funding: none.
Conflicts of interest: none.
Ethical approval: this study was approved by the Mount Sinai School of Medicine Institutional Review Board.

REFERENCES

10 Rajendran PR. Ethical issues involved in disclosing medical errors. JAMA 2001;286:1078.
18 Fost N. Ethical issues in whistleblowing. JAMA 2001;286:1079.

Received 1 June 2006; editorial comments to authors 2 January 2007; accepted for publication 12 February 2007
Becoming a new doctor: a learning or survival exercise?

Jeremy Brown,1,2 Tom Chapman1 & David Graham2

OBJECTIVE This study set out to establish why some new doctors view their training as a valuable period in their professional development, whereas others see it as a year to be endured and survived.

METHODS This multi-method case study focused on the interaction of key participants within 1 deanery, sampling the 237 pre-registration house officers (PRHOs) and 166 educational supervisors populating the associated 12 National Health Service trusts at the time (2001). The design of the case study was predicated on gathering the views of both teachers and learners in a way that allowed each stage of the data collection process to inform and influence the next phase.

RESULTS Lack of formal guidance and support were common characteristics associated with the first few days in post. The first day in post as a doctor is, for most, an experience that is hard to prepare for, even after a useful induction period. Those PRHOs who felt they were not guided or advised on how to undertake their new professional responsibilities tended to feel undervalued and under-recognised as individuals.

CONCLUSIONS Without the support of senior colleagues who can help the new doctor reflect on quite difficult and uncertain situations, new doctors will almost certainly perceive the first year of the new Foundation Programme as a survival exercise. If new doctors are working in an environment where their learning is properly facilitated, they are more likely to recognise their progress in their professional development and be more proactive in addressing concerns about professional expectations.

KEYWORDS multicentre study [publication type]; humans; attitude of health personnel; medical staff/hospital/ *psychology/ education; *education, medical, graduate; professional practice; clinical competence/ standards; interprofessional relations; self-concept.

Medical Education 2007; 41: 653–660
doi:10.1111/j.1365-2923.2007.02785.x

INTRODUCTION

Reforms in postgraduate medical education have gathered momentum since the publication of The New Doctor1 in 1997 and Implementing the New Doctor2 in 2002. The recommendations and guidance on general clinical training in these documents were produced by the General Medical Council in the light of accumulating evidence that the quality of educational supervision and support for new doctors varied considerably across the UK.3–6

The study reported here spans the period between 2001 and 2002 when The New Doctor recommendations were bedding down and postgraduate training reforms were under discussion.7,8 Both before and since the recent reforms, postgraduate training has rested on the bedrock role of the educational supervisor (ES) (first established in 1987) as the single most powerful influence upon the standards of conduct and practice of doctors in training.9

Despite the very considerable expenditure on the educational support and training of doctors in the early stages of their medical careers, no systematic evaluation reported in the public domain has
attempted to assess the quality of supervision available and the effectiveness of current practices. This case study attempts to address the deficit, in part, by exploring the views and perceptions of both new doctors and their ESs in a single, arguably typical, deanery. The study was designed to gather the basic knowledge required to improve the overall quality of the educational experience at the critical juncture during which doctors acquire and embed their clinical and professional skills.10

METHODS

This case study focused on the interaction of key participants within a single deanery, sampling the 237 new doctors and 166 ESs who populated the 12 National Health Service trusts at the time (2001). A case study approach, grounded in the interpretive tradition, was adopted, combining quantitative and qualitative aspects of data collection in order to provide both factual and representative data. An important aspect of the research design related to the need to prioritise the issues and considerations raised by the subjects of the research, rather than imposing constructs deemed to be important from the perspective of the researchers. To this end the study proceeded through 6 phases, with each phase building on the knowledge and data gathered previously.11 This inductive approach was based upon the principle that initial observations are built up through enquiry to form ideas that are in turn investigated and explored.12–14

There were 6 phases in the data collection process. Questionnaire items were primarily based on *The New Doctor* recommendations (Fig. 1).

Recruitment to the study

All PRHOs (n = 237) and ESs (n = 166) were invited to take part in the study in a covering letter that accompanied the questionnaire. Every PRHO and ES in the deanery was given the opportunity to take part at any stage of the study. The covering letter explained the further stages of the study and asked for volunteers to identify themselves for taking part in these other stages. A total of 55 (23.2%) PRHOs and 29 (17.5%) ESs volunteered to take further part in the study.

Consent and confidentiality

All study participants took part in this study voluntarily and were regularly reminded that they could withdraw their participation at any stage. All data were anonymised and confidentiality was assured. Interviews and focus groups were audio-recorded with the consent of participants.

The relationship between each step in the data collection process

The first PRHO questionnaire was constructed to collect baseline data to inform the qualitative data collection process. The aim of the focus groups was specifically to inform the design of the PRHO interviews. The PRHO interviews aimed to investigate in more detail the issues raised in the previous phases of the study. The ES questionnaire and interviews aimed to add another perspective of the first year in professional practice. The second PRHO questionnaire aimed to gather the views of all PRHOs on themes raised during the interview phases of the study.

Data analysis

Qualitative data were analysed using thematic framework analysis, a matrix-based method for analysing

This framework enables a rigorous and transparent data management process that allows for the systematic analytical hierarchy to be constructed. The analytical hierarchy describes the process of conceptual 'scaffolding' as the data are analysed. A separate framework was developed for each phase of data collection. To ensure inter-rater reliability, an independent researcher reviewed 10% of the transcripts to identify any discrepancies in the themes highlighted. Emerging themes were refined until the researchers judged that the framework was comprehensive (Table 1).

RESULTS

The main findings will be reported under key themes identified. (Table 2) outlines the relationship between each phase of data collection and identifies where key findings emerged. Table 2 also highlights the response rates to the questionnaires and the number of focus groups and interviews undertaken.

**Shadowing**

A total of 105 (97.2%, n = 108) respondents who had experienced some form of shadowing felt that it was a particularly valuable element of the preparation process for their post (data collection, phase 6 [P6]). Ninety (89.1%, n = 101, P4) ESs similarly valued shadowing as an important and constructive element of the preparation process.

**Adapting to professional working life**

Adapting to working life as a doctor often involves undertaking a post that has been established for many years to meet a significant service commitment. The hierarchical framework of medical practice can often lead to the most junior doctors feeling they are in a mundane post with little recognition:

'I think there's a theory in medicine in general though that everyone's been through their PRHO year and everybody knows it's basically a year for you to be a dogsbody.' (Focus group 3, P2)

At the start of their posts, new doctors had to be prepared to ask their colleagues on the wards many practical questions that could not have been answered in the induction process. Nurses were often relied upon in the first few days to give advice and guidance:

'I just remember having to ask the nurses to tell me exactly what I needed to write and where.' (Focus group 1, P2)
Table 1 Pre-registration house officer (PRHO) interviews: thematic framework sample page. Service pressures and the impact on PRHO learning

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Paperwork/clerking</th>
<th>Impedes on education/reflection</th>
<th>Time management</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>You can sometimes miss the consultant ward rounds or sometimes the planned admissions (TCIs) who come in for chest drains which you would love to do just a few more times but because you are on-call you essentially throw away the chance. There are times in the day, especially in medicine, when you are constantly seeing new patients. Once you have seen a patient you see another one</td>
<td>Learn to organise your time.</td>
</tr>
<tr>
<td>2</td>
<td>They can’t do without us. We have a bigger role in medicine.</td>
<td>Sometimes you have got so much to be getting on with in the day that you don’t have a chance to think about why you are doing it. Not encouraged to attend theatre. If I do questions are asked if there is nothing to do on wards. Most of the job is ward-based.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>I am so tired that I want to go to bed. I have to remind myself that a patient is there.</td>
<td>Time is the biggest pressure to quality learning. Speed up. You haven’t got time to do all those fancy posh histories that you learn at medical school.</td>
</tr>
<tr>
<td>4</td>
<td>Surgery doesn’t use your brain at all. You are just doing clerking regime.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>We spend a lot of time on secretarial work.</td>
<td>I think less pressure of work, more time really to spend on the job and allow us to take part and to learn really.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Overview of the data collection process (2001–02)

<table>
<thead>
<tr>
<th>Data collection phase</th>
<th>No. of responses/ interviews</th>
<th>Aim of data collection phase</th>
<th>Brief summary of findings from each data collection phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1 (P1) PRHO questionnaire (Jan–Feb)</td>
<td>n = 148 62.4%</td>
<td>Questionnaire items based on The New Doctor</td>
<td>‘Shadowing’ was perceived as a valuable element to the preparation process. Good and bad educational practice within the Mersey PRHO year is reflected in the finding that 196 PRHOs respondents (72.1%) confirmed that they developed a learning plan with their ESs soon after their post began but 41 PRHOs (27.9%) had not. PRHOs described the difficulties they had to overcome at the start of their post including the shock of practising as a doctor for the first time. The main themes that were raised in the focus groups were: preparation for the PRHO role; adapting to professional practice; educational supervision; feedback; service commitment; learning plan; and support.</td>
</tr>
<tr>
<td>Phase 2 (P2) PRHO focus groups (March)</td>
<td>4</td>
<td>To inform the interview schedules</td>
<td>The nature and quality of educational supervision was a critical factor impacting on PRHOs. Being able to work within a supportive team with regular feedback was viewed as very important.</td>
</tr>
<tr>
<td>Phase 3 (P3) PRHO semistructured interviews (March–May)</td>
<td>20</td>
<td>To investigate further the issues raised in questionnaire and focus groups</td>
<td></td>
</tr>
<tr>
<td>Phase 4 (P4) ES questionnaire (June–July)</td>
<td>n = 127 55.9%</td>
<td>To gather the opinions of ESs on the issues raised in phases 1–3</td>
<td>Only half the respondents reported that they had received formal training in educational supervision. ‘Shadowing’ was recognised as a valuable element in the preparation process for professional practice. PRHOs were generally perceived as not generally taking an active part in clinical meetings.</td>
</tr>
<tr>
<td>Phase 5 (P5) ES semistructured interviews (July 2001–January 2002)</td>
<td>20</td>
<td>To investigate the individual opinions of ESs on the issues raised in phase 4</td>
<td>Team support was perceived as being a vital element in developing and facilitating experiential learning. ESs recognised that feedback often focused on the negative aspects of performance and that more constructive feedback should be regularly given. Time and service pressures impacted heavily on their commitment to the ES role and this limited their ability to reflect on cases with their junior staff.</td>
</tr>
<tr>
<td>Phase 6 (P6) Second PRHO questionnaire (July)</td>
<td>n = 108 65.1%</td>
<td>To gather the views of the full cohort of PRHOs on issues raised in the semistructured interviews</td>
<td></td>
</tr>
</tbody>
</table>

PRHO = pre-registration house officer; ES = educational supervisor
New doctors described feelings of anxiety that they did not always know what was expected of them at the start of their post:

‘It’s still a shock. Nothing prepares you for what you are to do.’ (PRHO interviewee 17, P3)

‘I didn’t know how the system worked when I first started.’ (PRHO interviewee 16, P3)

Settling in to the new role involved learning how to fit in to the ward team without necessarily being guided in what was expected of them as PRHOs:

‘No-one sits you down and talks you through anything.’ (PRHO interviewee 2, P3)

Having to fit in to a system where one’s knowledge of the job is almost invariably assumed by others has potentially detrimental implications for both self-confidence and patient care. The following response was not atypical:

‘It is expected you understand the system. It is expected that when you become a senior house officer you appreciate that you are responsible to your house officer. How many other businesses work on: “This is your job I assume you know what you are going to do”.’ (PRHO interviewee 9, P3)

All the ESs interviewed recognised that the transition from student to medical practitioner had been difficult in their day, and remained so:

‘It is physically a hard job. It is mentally a very hard job and I also think it is a hell of a shock when you have been a student for 5 years.’ (ES interviewee 14, P5)

‘When house officers start their job it is such an overwhelming change.’ (ES interviewee 19, P5)

Meeting with the ES to develop a learning plan

A total of 106 PRHO respondents (72.1%) confirmed that they had developed a learning plan with their ES. However, 41 PRHOs (27.9%) stated they had not (P1). Ultimately, it is the responsibility of the PRHO to arrange this initial meeting. It was evident that meetings were sometimes not taking place. In some cases they had taken place but no learning plan had been developed. Eighty respondents (64.5%, P6) confirmed that they had met their ES very soon after coming into post and developed a learning plan that set realistic and achievable objectives. However, 37 (29.8%) respondents felt that this scenario did not match their own experience.

There were no significant differences in experience between PRHOs in district general or teaching hospitals ($\chi^2(1) = 0.567, P > 0.05$).

Feeling valued as part of the team

Every post should provide the PRHO with the sense of being a valued member of a team.1 (p 16) Being valued should encourage junior members of the team to seek advice from more senior colleagues and reflect on clinical experiences. Although only 10 respondents (6.8%, P1) reported that they did not feel valued as part of the team, 31 (21.2%) felt unsure.

General feelings of insecurity

Feelings of insecurity in the first few days and weeks of the job seemed to be exacerbated by the fact that many new doctors were not given any guidance on how they should practise or what was expected of them:

‘Nothing can really prepare you for the PRHO job – you just get on with it and learn along the way...’ (Respondent 47, P6)

‘Five years at medical school should prepare you for [the] job but [it] doesn’t.’ (Respondent 59, P6)

However, what may seem like simple gestures by senior colleagues can have positive effects on new postholders as they learn to practise:

‘My consultant came and introduced himself to me which was an encouraging start.’ (Respondent 1, P1)

Support

Many ESs recognised the importance of supporting their junior doctors and being approachable:

‘I see part of my job is to be approachable, and that he or she could phone me up if they weren’t sure.’ (ES interviewee 15, P5)

One ES explained how much he had always valued a supportive figure throughout his career:
‘I have always needed a strong supportive figure, especially in my first consultant post.’ (ES interviewee 19, P5)

Many ESs actually pinpointed their own transition from specialist registrar to consultant when considering the notion of support:

‘I think in medicine there is a hierarchy you steadily climb up and even when you have reached what you think is the top there is always some sense of insecurity.’ (ES interviewee 18, P5)

‘Everybody needs security. Consultants are the same.’ (ES interviewee 15, P5)

It was significant that those new doctors who felt they were making good progress were often those who felt they were part of a well supported team. This allowed the new doctor to take on an appropriate amount of responsibility:

‘I think that all of the people that I work with treat me with respect and value my opinion and listen to what I have to say.’ (PRHO interviewee 5, P3)

Those new doctors who worked in a well supported team were very appreciative of being recognised not just as doctors, but also as individuals:

‘It is just nice to be part of a team and for someone to be interested in you as a person.’ (PRHO interviewee 4, P3)

Formal training in educational supervision

Although 51 (49%, P4) respondents reported that they had received formal training as an ES, 52 (51%) had not.

PRHO participation in clinical meetings

Sixty-five (62.5%, P4) ESs felt PRHOs in general played an active role in ward rounds, whereas 14 (13.5%) felt they rarely did (Table 3). However, respondents considered that PRHOs played a much less active role in the other 4 settings listed in The New Doctor. Only 4 (5.8%) ESs felt PRHOs played an active role in pathology or histology demonstrations.

Feedback

A total of 72 (57.1%, P6) respondents felt they were receiving regular informal feedback. However, over a third of PRHOs who responded to the questionnaire felt they were not receiving regular feedback. There was no evidence to suggest that working in a district general hospital or teaching hospital affected whether they received regular informal feedback or not ($\chi^2(4) = 2.060, P > 0.05$).

DISCUSSION

At face value, questions may be raised about the relevance of a study that investigated the pre-registration year, prior to the introduction of the new Foundation Programme. The first year of the Foundation Programme equates to the old PRHO year and, in our view, the issues investigated in depth here, are, in many cases, more relevant now than they were in 2001 when data collection began. Despite the numerous changes introduced into the new Foundation Programme, many of the underlying problems related to the transition into professional practice are still to be addressed. A considerable number of them reflect the underlying tension between service and training.17

The first year in professional practice as a doctor has historically been a challenging experience, which
tests the individual’s fortitude and resolve to work in his or her chosen profession. It is therefore a period where the stakes are high for the new doctor. It is also a crucial period in the wider context of the profession and society as a whole as more doctors are needed to provide the high-quality health care currently demanded. Findings demonstrate that lack of support from colleagues can result in many new doctors feeling they are merely fulfilling a service rather than progressing in a training post. A general lack of recognition and the repetitive nature of some duties can also contribute to feelings that the first year in practice is a year simply to be endured, prior to moving on and up the career ladder.

Transition often raises feelings of anxiety and insecurity arising from the fear of not fully knowing how to interpret or adapt to a new professional role and working environment. At such junctures individuals tend to be acutely sensitive to the fact that their actions are being scrutinised by their peers and senior colleagues, who are perceived as making continuous judgements of their capabilities. In practice, of course, such judgements are most likely to be fleeting, sporadic and circumspect. Wilkie and Raffaelli state that all transitions involve a fundamental re-examination of who and what we are, even if this processing occurs at a largely unconscious level. If this is the case, any fundamental re-examination of who and what he or she is by any individual will rely to a considerable extent on the support, guidance, judgements and feedback of senior colleagues.

Whereas the public perception of professional practice is that it is grounded in rational and dependable expert knowledge, it is in fact based on what Eraut describes as ‘wise judgement under conditions of considerable uncertainty’. Wise judgement depends on the personal attitudes and beliefs held by the person occupying the professional role. For these personal qualities to develop, the new professional requires support and nurturing as he or she develops the skills that actually help to tackle the considerable uncertainties of practice.

At its very core, medicine’s foundation is based on human values and complex judgements. Without the support of senior colleagues who can help the new doctor reflect on quite difficult and uncertain situations, they will almost certainly continue to regard the first year of the Foundation Programme as very much a survival exercise. The educationalist, Schön, compared learning to practise as a professional with working in the swampy lowlands, symbolising the risky and undervalued nature of the practice setting. Schön argued that feelings of vulnerability are often hidden and frequently not articulated. Assumptions that new doctors know what they should be doing are common. This implies that few doctors actually talk about, or try to explain, how they practise. By contrast, if new doctors work in an environment where their learning is clearly structured and facilitated, they are more likely to recognise that they are progressing in their professional development and be more proactive in addressing the concerns they have about professional expectations. These concerns relate to the notion of acceptance, to one’s own adequacy or inadequacy, and to relationships with patients, relatives and colleagues.

The findings reported here help to identify the components that positively contribute to new doctors feeling able to make material progress in the very early stages of their professional development. Crucially, they should be given the guidance necessary to make a confident start in interpreting and delivering their role, and feel reassured from the outset that they are valued members of the team. Unfortunately, for a sizeable minority this has not always been the case. Appropriately trained ESs are key to providing an optimum teaching and learning environment which encourages new doctors to reflect on, and articulate, their experiences as openly as possible. Although the importance of developing skills in reflection is increasingly recognised in practice, we conclude that more systematic training for both trainees and supervisors is required.

The new Foundation Programme builds very much on the previous system of PRHO training, which provided variable support. The time seems right to attempt to disseminate the best standards of educational supervision experienced by some new doctors to all of them. This arguably requires a more radical review and development of the current mechanisms for supervision and guidance. Unless this is carried out in a systematic fashion, a substantial proportion of new doctors will continue to perceive their foundation period as an experience in survival, rather than a valuable term in their professional development.

Contributors: This work was undertaken by JB as part of his PhD study. TC acted as supervisor and DG as adviser. JB and TC conceived the design of the study. JB undertook

Acknowledgements: We would like to thank all the pre-registration house officers and educational supervisors who took part in this study.

Funding: None.

Conflicts of interest: None.

Ethical approval: In Spring 2001, there was no requirement for this study to go through a formal ethical approval submission process. If the study had been undertaken 12 months later, the researcher (JB) would have been required to submit the study to the Central Office of Research Ethics Committees (COREC) for full ethical approval as the data collection process involved the participation of NHS staff.

REFERENCES


Received 26 April 2006; editorial comments to authors 18 August 2006, 24 January 2007; accepted for publication 20 February 2007
The reliability of summative judgements based on objective structured clinical examination cases distributed across the clinical year

George R Bergus1,2 & Clarence D Kreiter1,3

CONTEXT Objective structured clinical examinations (OSCEs) can be used for formative and summative evaluation. We sought to determine the generalisability of students’ summary scores aggregated from formative OSCE cases distributed across 5 clerkships during Year 3 of medical school.

METHODS Five major clerkships held OSCEs with 2–4 cases each during their rotations. All cases used 15-minute student–standardised patient encounters and performance was assessed using clinical and communication skills checklists. As not all students completed every clerkship or OSCE case, the generalisability (G) study was an unbalanced student × case × clerkship design. After completion of the G study, a decision (D) study was undertaken and phi (ϕ) values for different cut-points were calculated.

RESULTS The data for this report were collected over 2 academic years involving 262 Year 3 students. The G study found that 9.7% of the score variance originated from the student, 3.1% from the student–clerkship interaction, and 87.2% from the student–case nested within clerkship effect. Using the variance components from the G study, the D study suggested that if students completed 3 OSCE cases in each of the 5 different clerkships, the reliability of the aggregated scores would be 0.63. The ϕ, calculated at a cut-point 1 standard deviation below the mean, would be approximately 0.85.

CONCLUSIONS Aggregating case scores from low stakes OSCEs within clerkships results in a score set that allows for very reliable decisions about which students are performing poorly. Medical schools can use OSCE case scores collected over a clinical year for summative evaluation.

KEYWORDS clinical clerkship/*standards; programme evaluation; validity and reliability; clinical competence/*standards; judgement.

Medical Education 2007: 41: 661–666
doi:10.1111/j.1365-2923.2007.02786.x

INTRODUCTION

The objective structured clinical examination (OSCE) is frequently used to provide summative and high-stakes judgements about clinical skills. Since its introduction in 1979,1 the OSCE has been widely adopted by medical schools and licensing boards. At present, all students seeking medical licensure in the USA and Canada are required to successfully pass national clinical skills examinations based on the OSCE. In addition, comprehensive clinical skills testing using OSCEs is carried out by many medical schools. In a report published in 2005, investigators found that 76 of 91 (84%) Liaison Committee on Medical Education (LCME)-accredited schools responding to their survey reported using a comprehensive assessment OSCE at the end of Year 3 or during Year 4.2

Whereas licensing boards and medical schools both wish to identify students with substandard skills, medical schools also have the objective of helping
their students acquire these skills. From a learning perspective, having students complete multiple, small, formative OSCEs integrated into clerkships and distributed across time is more likely to facilitate mastery of clinical skills than is a single high-stakes testing session at the end of the course of study.3 One solution could be for medical schools to stage both types of OSCEs concurrently. A series of small OSCEs distributed throughout the clinical years could be used to maximise formative feedback, whereas a larger, summative OSCE at the end of the clinical years could inform a medical school about which of its students had not adequately mastered basic clinical skills. However, although this dual approach would provide both formative and summative evaluations, it would require considerable resources that many medical schools could not afford.

A possible alternative is to aggregate the students’ OSCE scores generated from formative cases distributed across the clinical year into single scores, which could be used for summative evaluations. However, there is little research addressing whether these aggregated scores allow reliable summative assessments regarding the clinical skills of students. The goal of the present study was to determine the generalisability of an aggregated score derived from OSCEs distributed across the clerkship year.

**METHODS**

Year 3 medical students at the University of Iowa College of Medicine complete OSCEs during 5 major clerkships in surgery, paediatrics, psychiatry, obstetrics and gynaecology, and outpatient internal medicine or family medicine, respectively. Each clerkship-based OSCE involves between 2 and 4 cases, which are selected by clerkship directors and directly reflect clerkship content. All cases are designed for 15-minute student–standardised patient (SP) interactions. Student performance is assessed by clinical (25 items) and communication (15 items) skills checklists completed by the SP. The clinical skills checklists are case-specific but the same communications skills checklist is used throughout. The SPs for all the OSCE cases are recruited and trained by a single group of trainers.

After each OSCE, students receive information about their performances on each case. In addition to their own scores, students receive information about the highest, lowest and mean scores in their testing group. At least 1 case is observed by faculty members during the majority of clerkship OSCEs and the students are given immediate formative feedback. Additionally, students receive written feedback on each case and are provided with suggestions about how they can improve their performances.

Although the OSCEs are required by the clerkships, performance on these OSCEs has generally contributed <10% to students’ clerkship grades. Two of the clerkships hold their OSCEs during the last week of each rotation. Three others hold the testing in the last 2 weeks of the clerkship, thereby allowing these clerkships to provide remediation to poorly performing students. Most, but not all, students complete all 5 major clerkships during the academic year. The order in which the clerkships are completed is variable.

Student data from the 2004–05 and 2005–06 academic years were used for this study. We performed generalisability (G) analysis and decision (D) studies to determine the reliability of an aggregated score. The aggregation of the scores was a multistep procedure. As a first step, all scores collected for
individual OSCE cases over the year were converted to standardised scores with a mean of 80 and a standard deviation (SD) of 5, so that all cases were given approximately equal weighting in computing a composite score summarising a student’s performance across cases. Then the standardised scores were entered into urGENOVA® software (Robert L. Brennan, CASM, University of Iowa, Iowa City, IA, USA). As not all students (p) experienced every clerkship (d), and as some cases (c) within a clerkship changed during the year, there was variability between students in which particular cases a student experienced. Hence, the data used in the G study represented an unbalanced p × (c : d) design.

After completion of the G study using each year’s data, the variance components were then pooled across academic years by averaging the variance components weighted by those undertaken to examine the effect on generalisability of manipulating the number of cases and clerkships.4 As has been recommended, pooled variance estimates that were negative were set to 0 for the D studies, although the negative component estimates were used in calculations of mean variance components.5 The D studies provided estimates of the reliabilities of aggregated scores from various numbers of clerkships and cases for generalisations about student performance across the universe of cases that might be administered within the 5 clerkships. In these analyses, the clerkship facet was treated as fixed. Phi (ϕ), the dependability index at a score cut-point, was calculated to estimate the reliability of decisions about whether students’ true scores fell below a pre-selected cut-point set to indicate that a student needed additional clinical skills practice.6

The student data were collected as a routine part of the evaluation of student clinical skills performance.

RESULTS

During the 2004–05 academic year, OSCE case performance data were collected on 130 Year 3 students. Data for 5 students were dropped from this analysis because the students completed fewer than 6 OSCE cases. For the remaining 125 students, the mean number of cases per student was 10.9 cases (range 6–13 cases). Over half the students (51.5%) completed at least 12 cases. During the 2005–06 academic year, OSCE data were collected on 139 Year 3 students. Data for 2 students were dropped because the students completed fewer than 6 OSCE cases. The remaining 137 students completed a mean of 12.8 cases per student (range 6–15 cases) and 112 students (81.7%) completed at least 12 cases. Thus, data for a total of 262 students were used in this G study.

The mean OSCE case score in 2004–05 was 72.2% (SD 11.1); in 2005–06 it was 73.0% (SD 10.7). The mean score for each student aggregated from all cases completed by a student during the 2004–05 year was 72.8% (SD 4.4); in 2005–06 it was 72.9% (SD 4.6). The aggregated scores did not significantly correlate with the number of OSCE cases students completed over the 2004–05 academic year (r = 0.06, P = 0.48) or the 2005–06 academic year (r = 0.04, P = 0.61).

Table 1 displays the G study results and shows that 9.7% of the score variance originated from the student, 3.1% from the student–clerkship interaction, and 87.2% from the student–case nested within clerkship. Because the OSCE scores were standardised with a mean of 80 and an SD of 5, the clerkship and case : clerkship facets contributed only minimally to the variance. The student–clerkship interaction facet contained confounded variance

| Table 1 The generalisability analysis using urGENOVA estimated the sources of score variance* |
| Facets                          | Variance components |
|                                | 2004–05 | 2005–06 | Overall | Percentage of total variance |
|                                | n = 125 | n = 137 | n = 262 |                                 |
| Student                        | 2.163   | 2.696   | 2.442   | 9.7%                             |
| Clerkship                      | −0.110† | 0.016   | 0.003   | 0.0%                             |
| Case : clerkship               | −0.252‡ | −0.206‡ | −0.256‡ | 0.0%                             |
| Student–clerkship              | 1.829   | −0.204‡ | 0.766   | 3.1%                             |
| Student–(case : clerkship)     | 20.929  | 22.746  | 21.879  | 87.2%                            |

* Some facets contribute negatively to score variance because all the scores from each objective structured clinical examination case were standardised with a mean of 80 and a standard deviation of 5
† When the variance component is negative it is set to 0 in the D study and in calculating the percentage of total variance arising from the different facets
from the clerkship and occasion (the time that students completed each of the 5 clerkship OSCEs was separated by several weeks). The G study produced several small negative variance components. This can occur when the true value of the variance is close to 0 because these estimated components are the algebraic product of several different model-related mean square calculations for the sample.

Using the variance components from the G study, a series of D studies were undertaken (Table 2). For examinees completing 3 OSCE cases embedded into each of the 5 different clerkships, a reliability of 0.63 was obtained. The reliability would be approximately 0.54 if the same group of examinees completed 15 cases embedded into only 1 randomly selected clerkship. In addition, expanding the total number of cases to 35 (7 cases per each of 5 clerkships) would result in a reliability of 0.80.

The reliability of decisions about whether a student’s performance was below a selected cut-point, expressed as $\phi$, was also determined using the variance components from the G study. For a cut-point of 1 SD below the mean, the $\phi$ would be approximately 0.85 for students completing 3 cases in 5 clerkships. This cut-point is equivalent to a mean score of 77.5 because the SD of the mean across OSCE cases is 2.5. A cut-point of 0.72 SD below the mean (standardised score of 78.2) would result in a $\phi$ reliability of 0.80. The $\phi$ associated with other cut-points for students completing 15 cases distributed over 5 clerkships are shown in Fig. 1.

### DISCUSSION

Implementation of the OSCE within a medical school curriculum can result in several benefits. The most obvious is the reliable assessment of students’ clinical skills. The OSCE has also been shown to impact on how students approach the task of learning clinical skills. The OSCE motivates students to actively work to improve their clinical skills and provides feedback to students so that they can better assess which clinical skills need additional work. The OSCE also motivates faculty staff to provide additional observation and feedback to students who are in the process of acquiring clinical skills. Another benefit is that the OSCE can be used not only for student assessment, but for evaluation of the curriculum. When many students perform poorly on an OSCE case, the problem may not lie with the students, but rather with a mismatch between the skills and knowledge required to perform competently on a case and those covered by the formal curriculum.

Although high-stakes, gateway OSCEs are widely used in LCME-accredited medical schools, experts in medical education have questioned whether this is the best way to use the OSCE. In an attempt to focus on learning, at least 1 school has shifted some SP encounters from summative to formative assessment. We questioned whether we could make use of the same OSCEs for both types of assessment.

Our research suggests that aggregating student scores from OSCE cases distributed across the academic year can provide summative scores with moderate reliability. In addition, these aggregated scores can be used to reliably identify those students performing at least 1 SD below the mean of their peer group.

<table>
<thead>
<tr>
<th>Number of clerkships with OSCEs</th>
<th>OSCE cases per clerkship</th>
<th>Total OSCE cases completed</th>
<th>Reliability of aggregated scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>15</td>
<td>0.54</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>14</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>15</td>
<td>0.61</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>15</td>
<td>0.63</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>35</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 2. A decision study, using the variance components from the generalisability analysis, modelling the effect on reliability of the aggregated scores arising from manipulating the number of cases and clerkships.

$\phi$ = objective structured clinical examination.
Comparing the reliability of scores from our distributed OSCE with that of scores on a single gateway assessment is challenging, although both appear to be somewhat lower than reliability reported for the same number of cases completed during a single testing session. From our data we conclude that aggregated scores based on a total of 15 cases distributed over 5 clerkships (with a testing time of 45 minutes per clerkship) provide scores with a reliability similar to that reported for 2 hours of testing time in a gateway examination.13

The lower reliability of the scores from the distributed OSCE may be the result of several factors. As cases are completed over the course of the year, performances may be affected by different learning trajectories and prior rotations,13 although these influences have not been documented by all studies.14 In addition, students may have different levels of motivation within the different clerkships as a function of their career plans and this might add variation to their performances in the OSCEs. The impact of these factors is not defined, although we note that in our data the student–clerkship interaction (which also included occasion influences) had only a small impact on the D study results. It is also important to note that a G study of a gateway examination administered on a single occasion cannot estimate the influence of an occasion facet as a gateway examination is typically conducted during a single session. The variance attributable to this hidden occasion facet is by default included as part of the object of measurement variance (p). Hence, D studies using variance components from a single occasion gateway examination will overestimate reliability.

In this study, D study results in which the clerkship facet (d) is fixed provide an estimate of what the computed G coefficient might have been for a gateway examination consisting of similar cases. Specifically, the G coefficient from the D study mixed model with 5 observed clerkships provides an approximation of a gateway OSCE examination consisting of a similar mix of cases. In addition, the G study, shown in Table 1, lets us comment on how reducing the number of clerkship OSCEs would impact the reliability of the aggregated scores for making generalisations across clerkships. Although student–clerkship interaction contributes 3.1% of the observed score variance, this is a third of the score variance contributed by the student facet. As a result, as Table 2 shows, if we have observations on at least 3 clerkships, the reliability is only minimally decreased.

We are aware of 1 previous study which investigated the generalisability of OSCE scores obtained across clinical clerkships.15 This study found low levels of correlation between the scores from OSCEs from different clerkships. However, formal G and D studies were not undertaken with the data and it was not possible to calculate ‘true score’ correlations between clerkship performances. In addition, the numbers of checklist items used for cases in the different clerkships in this earlier study varied considerably, which may have impacted on correlation across clerkships. When the correlations between clerkship scores were calculated based on a uniform instrument, used by faculty to assess student performances in all clerkships, the correlations were much higher.

Our study has several limitations that should be taken into account when interpreting our findings. The first is that the data are from a single medical school. How these generalise to other schools is unknown. However, in designing all our cases, we used a format similar to that used by the National Board of Medical Examiners for the United States Medical Licensing Examination Clinical Skills (USMLE CS) examination. Another limitation is that these analyses are based on data from only 2 academic years. Lastly, although we are able to report on the reliability of an aggregated set of scores, assessing the validity of these scores requires additional research.

Before a medical school adopts the distributed OSCE approach, it should carefully consider the disadvantages of doing so. The first is related to the lower reliability of this approach compared with that of having students complete all cases on a single testing date. In addition, not all students completed all OSCE cases, although most completed at least 12. As the cases are distributed within clerkships, collecting a complete set of performance data requires all students to complete all the major clerkships in the same year. Although most students do this, some do not. Additionally, we had several students complete all 5 clerkships but miss individual clerkship OSCEs as a result of schedule conflicts. Clerkships were not always insistent that students completed a missed OSCE. We also worry that a disadvantage with the distributed approach is that students are cued to the case content by the clerkship in which the OSCE case is used.15 A final concern is that distributed OSCEs do not provide students with an experience that approximates the much longer USLME CS examination, although the importance of providing a similar experience is not clear.16
There are several readily obvious benefits to distributing OSCE cases across the clinical year. The first is that this approach allows assessment to be structured so as to enhance instruction.12 By distributing the cases across clerkships, students are provided with focused feedback about their skills in these different clinical domains. In addition, the distribution of cases allows students greater practice of their clinical skills. Deliberate practice (practice with feedback on performance) is essential to skill acquisition.17 Another advantage is that clerkships may take greater ownership of and have greater involvement with OSCEs that are embedded within them. We have found that most of our clerkship directors create cases for their OSCEs and recruit faculty observers to participate in the OSCE sessions. Clerkship ownership of the OSCE might also help link performance assessment to the clerkship curriculum and give clerkships an incentive to guide remediation if it is needed.18 Lastly, by distributing the OSCE cases across the year, SP resources can be used more uniformly over the year. Whether this results in greater cost-effectiveness deserves formal study.

In conclusion, our results show OSCE cases distributed within clerkships over Year 3 of medical school can produce aggregated scores of moderate reliability. Although the reliability of these scores is lower than that recommended for high-stakes judgements, the distributed OSCE can reliably ($\phi > 0.80$) identify underperforming students using a cut-point of 1 SD below the mean of a Year 3 class. To our knowledge, this is the first time that the generalisability for a panel of OSCE cases distributed over an entire academic year in multiple clerkships has been reported.

Contributors: GRB contributed to the conceptualisation of the study and data collection, and wrote the first draft of the paper. CDK contributed to the conceptualisation of the study and data analysis. Both authors revised and approved the manuscript.

Acknowledgements: none.

Funding: none.

Conflicts of interest: none.

Ethical approval: this project was approved by the University of Iowa Institutional Review Board.

REFERENCES


Received 24 August 2006; editorial comments to authors 14 December 2006; accepted for publication 23 January 2007
Factors in faculty evaluation of medical students’ performance

Andrew R Pulito, Michael B Donnelly & Margaret Plymale

CONTEXT Faculty members often use global rating scales as a method of assessing various characteristics of medical students’ clinical performance. The purpose of this study was to determine if some performance characteristics are more highly associated with the overall faculty grade than others.

METHODS The clinical performance of 211 surgery clerkship students was evaluated by 2 or 3 faculty preceptors. Faculty rated students on 10 specific performance characteristics, using a 5-point scale. Faculty then assigned a numerical grade summarising the faculty’s view of the student’s performance. Reliability of the ratings was estimated by the intraclass correlation, and 1-way (analysis of variance) ANOVA was used to test for differences among the students’ mean ratings. Logistic regression was employed to determine the accuracy of each performance measure in predicting students’ grades (A or B). Stepwise logistic regression was used to determine if there was a combination of performance characteristics that best predicted students’ grades.

RESULTS The inter-rater reliabilities were low (≤ 0.33). Four of the 10 performance characteristics were each found to predict the assigned grade most accurately (80%). However, the least predictive characteristic achieved 75% accuracy. The stepwise logistic regressions indicated that combinations of 2 or 3 of the 10 performance characteristics increased prediction accuracy by ≤ 4%.

CONCLUSIONS The rating on any single characteristic predicts a student’s overall grade very accurately. Additional measures do not greatly increase the accuracy of prediction. Our results suggest that faculty make a relatively undifferentiated judgement in assigning student grades, and there is little to be gained by using multi-item forms in assessing medical students’ clinical performance.

KEYWORDS humans; *students, medical; *clinical clerkship; surgery/*education; clinical competence/*standards; faculty, medical/*standards; educational measurement/*methods; ANOVA.

Medical Education 2007; 41: 667–675
doi:10.1111/j.1365-2923.2007.02787.x

INTRODUCTION

Assessment of medical student performance is a complex undertaking. According to Edelstein et al.: ‘A growing body of research suggests that a combination of evaluation methods is necessary to properly assess the complex skills that make up the practice of medicine.’

Evaluation by faculty preceptors using a global rating scale is among the most common methods of rating medical students’ clinical performance.

Cowles and Kubany (1959)2 reported the use of global rating scales to evaluate medical student performance. For their purposes, 12 experienced faculty members were asked to determine the most important characteristics for a student to possess or acquire in preparing for general practice. Eight categories emerged from these interviews. Students were rated on a 3-point scale. The 8 categories identified were:

1 knowledge of medical information;
2 the ability to gain and maintain the patient’s confidence;
Overview

What is already known on this subject

There are several difficulties involved in the use of clinical performance ratings. One is a strong tendency for the ratings on single characteristic to be positively correlated with ratings on all other characteristics.

What this study adds

We examined the validity of the ratings of 10 performance characteristics of surgery clerkship students in predicting the overall clinical grade. Results indicated that each of the 10 performance characteristics predicted the grade at similar levels of accuracy, and that only 2 or 3 variables were necessary to optimally predict the overall grade.

Suggestions for further research

Future research needs to determine if faculty evaluators can be trained to judge and rate the various components of clinical performance more independently.

3 assumption of responsibility;
4 observing, recording and reporting skills;
5 the ability to develop and verify hypotheses from clinical data;
6 stability in difficult situations;
7 integrity, honesty and ethics, and
8 interest in the profession and self-improvement.

The global rating scales used in medical education today often consist of some variation of Cowles and Kubany’s 8 characteristics, along with more sophisticated scaling techniques. Although employed extensively, this type of evaluation has both advocates and critics.

Neufeld and Norman (1985) suggested that subjective evaluation through global rating scales has the advantages of low cost, flexibility and unobtrusiveness, as well as the ability to assess domains not suited to objective measurement. Eisner (1979) believed that areas where science and art mixed were best evaluated by subjective expert judgement.

Others have questioned both the precision (reliability) and validity of global rating scales. However, as mentioned earlier, subjective evaluation by faculty using a global rating scale is an extremely popular method of medical student evaluation.

At the University of Kentucky College of Medicine, cumulative faculty evaluations in this format account for roughly 50% of the surgery clerkship course grade. In keeping with prior research, we employ a letter grade system. Reznick et al compared interrater reliabilities of faculty ratings of surgical students using 5 different grading systems: pass or fail; pass, fail or honours; letter grades; number grades of 1–10, and number grades of 1–100. The authors found that the letter grade system provided good discrimination and had high to moderate reliability, and, as a result, they favoured it over all other systems.

The primary purpose of this study was to determine the relative importance of various performance characteristics in determining the grade faculty members assign medical students during a junior surgery clerkship. Is there evidence that some performance characteristics are more highly associated with the assigned grade than others or are they all weighted the same? A secondary purpose was to determine whether the patterns of correlations for relatively brief and relatively long rotations were similar.

METHODS

During the surgery portion of a joint medicine and surgery clerkship, the clinical performances of 211 medical students were evaluated by 2 or 3 surgical faculty preceptors, resulting in 585 evaluations suitable for study. Six students received a preceptor grade lower than B. These evaluations were excluded, as they represented too small a cohort to study formally. The study was conducted over 2 consecutive academic years.

The combined medicine and surgery clerkship lasts 16 weeks; the surgery portion of the clerkship is 8 weeks long. The surgery rotation consists of two 2-week general surgery rotations and one 4-week subspecialty rotations. A few students elect to remain on a single subspecialty service for all 4 weeks. At the end of each of the 3 (or 2) rotations, the faculty from each service evaluate the student’s performance by completing a single evaluation form.

The clerkship evaluation form is presented in Table 1. As can be seen, faculty evaluate the
students descriptively on 10 specific performance characteristics, using a 5-point scale. The anchors for each scale describe performance for ratings of 1, 3 and 5, respectively, where 1 = performance needs improvement, 3 = performance at expected level, and 5 = outstanding performance. Corresponding descriptive anchors are provided for each particular skill. For example, for the skill Fund of Knowledge, the general anchor ‘performance needs improvement’ is defined descriptively as ‘usually has difficulty answering questions correctly on rounds’. Each skill at levels 1, 3 and 5 has both a general and specific definition of the trait for that level. At the end of the form, the faculty evaluator assigns a numerical grade summarising his or her view of the student’s performance. Guidelines for assigning grades are presented on the form (Table 1). The numbers use the same scale as the National Board of Medical Examiners (NBME) surgery shelf-examination in the sense that a score of 66 on the NBME shelf-examination is interpreted in the same way as a preceptor rating of 66. Both would be a low B.

Table 1 Form for evaluation of Year 3 medical students at the University of Kentucky College of Medicine, Department of Surgery

<table>
<thead>
<tr>
<th>Performance characteristic</th>
<th>Performance needs improvement</th>
<th>Performance at expected level</th>
<th>Outstanding performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rating = 1</td>
<td>Rating = 3</td>
<td>Rating = 5</td>
</tr>
<tr>
<td>Fund of Knowledge</td>
<td>Usually has difficulty answering questions correctly on rounds</td>
<td>Usually answers most questions correctly on rounds</td>
<td>Consistently answers questions correctly on rounds</td>
</tr>
<tr>
<td>Diagnostic Ability</td>
<td>Usually has difficulty developing a thorough differential diagnosis</td>
<td>Usually can develop a thorough differential diagnosis</td>
<td>Consistently develops a thorough differential diagnosis</td>
</tr>
<tr>
<td>Evidence of Study</td>
<td>Very seldom volunteers and occasionally, when asked, provides information that demonstrates outside reading</td>
<td>Occasionally volunteers and usually, when asked, provides information that demonstrates outside reading</td>
<td>Consistently volunteers information that demonstrates outside reading</td>
</tr>
<tr>
<td>Clinical Skills</td>
<td>Usually does not demonstrate correct techniques and accuracy in physical examination</td>
<td>Usually demonstrates correct techniques and accuracy in physical examination and mechanical skills</td>
<td>Consistently demonstrates correct techniques and accuracy in physical examination and mechanical skills</td>
</tr>
<tr>
<td>Professionalism</td>
<td>Is not consistently dependable</td>
<td>Dependable and punctual</td>
<td>‘ Goes the extra mile’ for team and patients</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>Usually verbal presentation is not clear and concise</td>
<td>Usually verbal presentation is clear and concise</td>
<td>Verbally presents consistently clear and concise</td>
</tr>
<tr>
<td></td>
<td>History of present illness is disorganised</td>
<td>Physical examination is occasionally lacking in information or organisation</td>
<td>Physical examination is clear, organised and concise</td>
</tr>
<tr>
<td>Write-ups/Progress Notes</td>
<td>Write-ups and daily notes are disorganised and not thorough</td>
<td>Write-ups and daily notes are organised and thorough</td>
<td>Write-ups and daily notes are extremely well organised, thorough and reflect outside reading</td>
</tr>
<tr>
<td>Interpersonal Skills with Team and Staff</td>
<td>Poor interpersonal skills</td>
<td>Solid interpersonal skills</td>
<td>Truly outstanding interpersonal skills; a joy to work with</td>
</tr>
<tr>
<td></td>
<td>Difficult to get along with Contentious and sullen</td>
<td>Pleasant, amicable</td>
<td>Uplifting for team morale</td>
</tr>
<tr>
<td>Interpersonal Skills with Patients</td>
<td>Poor interpersonal skills</td>
<td>Enjoyable to work with</td>
<td>Truly outstanding interpersonal skills; enjoyed by patients and families; looked to as the doctor</td>
</tr>
<tr>
<td></td>
<td>Awkward with or unpleasant towards patients</td>
<td>Solid interpersonal skills</td>
<td>Consistently takes charge (with direction)</td>
</tr>
<tr>
<td>Patient Management Skills</td>
<td>Able to take charge only occasionally</td>
<td>Takes charge (with direction) at least 50% of the time and is able to independently determine diagnosis and treatment plans</td>
<td>Able to independently determine diagnostic and treatment plans</td>
</tr>
<tr>
<td></td>
<td>Usually able to independently determine diagnostic and treatment plans</td>
<td></td>
<td>Almost at intern level</td>
</tr>
</tbody>
</table>

Please grade this student using the following standards:

<table>
<thead>
<tr>
<th>Letter</th>
<th>Numerical grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>77-86*</td>
</tr>
<tr>
<td>B</td>
<td>66-76</td>
</tr>
<tr>
<td>C</td>
<td>61-65</td>
</tr>
<tr>
<td>U</td>
<td>56-60</td>
</tr>
<tr>
<td>E</td>
<td>20-55</td>
</tr>
</tbody>
</table>

* Grades cannot be > 86 or < 20. These numerical ranges are employed so that the preceptor grade will be on the same scale as National Board of Medical Examiners shelf-examination, which is weighed equally in determining student grade for the clerkship.
Numerical grade is ultimately converted into a letter grade that varies from A to E (Table 1). Grading judgements are not usually separated from ratings on specific performance factors on evaluation forms of this type. Often, the students' grades are obtained by averaging the ratings on the specific factors. When assigning grades to students, our faculty members are instructed that: 'Listed below are 10 important performance characteristics. Please evaluate this student on these characteristics. You may consider factors other than the 10 performance characteristics, and weigh aspects of performance as you think appropriate.'

The intraclass correlation with Spearman–Brown correction for 3 raters was used to estimate reliability for each of the performance measures and grade. One-way analyses of variance (ANOVA) were used to test for the significance of the difference among students' mean ratings ($n = 3$). If the ANOVA was not statistically significant, it was concluded that the corresponding reliability was also not significant.

Logistic regression was used to identify how accurately each performance rating differentiated students who received A and B grades. If an item does not discriminate among levels of performance, one can question its validity and importance. Of course, some items can be very important but not discriminating, such as 'integrity'. It is expected that this is a characteristic of almost all students, and therefore will not be discriminating. Nevertheless, an item like this should be included in any clinical evaluation form.

Stepwise logistic regression was used to determine if there was a combination of performance factors that best differentiated the students who received preceptor grades of A from those who received grades of B.

Principal axis factor analysis (varimax rotation and roots > 1.00 rule) was used to determine how many factors underlie the judgements made on 10 descriptive items.

The unit of analysis in this study was the individual evaluation made by each faculty member, with the exception of the reliability and ANOVA analyses, where the unit of analysis was the mean rating given to the student. As rotation length may have an effect on the ratings made, results for the 4-week rotations were initially analysed separately from those for the 2-week rotations.

**RESULTS**

The 585 student evaluations included 395 (68%) A and 190 (32%) B grades. As noted earlier, 6 evaluations were excluded from the study because they constituted a sample size that was too small to analyse meaningfully.

We first determined if there were systematic differences in the ratings obtained from the 2- and 4-week rotations by means of a 2-way ANOVA with 1 repeated measure. The between-groups factor was evaluations from 2-week versus 4-week rotations, and the repeated measure reflected the 10 performance ratings variables. The overall mean for the 2-week rotations was 4.2 (0.01); that for the 4-week rotations was 4.1 (0.02). The difference between these means was not statistically significant ($P = 0.13$). The interaction between rotation length and scale items was significant ($P < 0.01$); however, the Tukey b post hoc test indicated a significant difference between the mean ratings of the 2- and 4-week rotations for only 1 of the scales. For this reason, evaluations from the 2- and 4-week rotations were combined for the purposes of the statistical analyses.

Table 2 presents the intraclass correlation and the reliability of the mean of 3 ratings for the 10 specific performance measures, for grade (A or B) and grade score (a number that can vary from 20 to 86). As can be seen from this table, reliabilities were $\leq 0.33$, implying that there were low levels of agreement among faculty raters both in terms of performance ratings and the grades they assigned. To determine if the reliabilities were statistically significant (i.e. whether it was possible to differentiate the mean ratings among the students), 1-way ANOVAs were used to test for differences among the mean ratings. The fourth column of Table 2 presents the $P$ values from those ANOVAs (4-fold table chi-square analysis for the grade). There were significant differences among the students in how positively they were evaluated on the performance characteristics, with the exception of Oral Presentations. The last column of Table 2 presents the skewness index for each of the scales. These values range from $-1.00$ to $-0.04$. However, although we might expect higher reliability for those scales with less extreme skewness values, this was not the case. The correlation between skewness and reliability of the measures is 0.17; descriptively this correlation is very modest and is not statistically significant ($P = 0.59$).
Table 3 summarises the results of the logistic regression of each performance characteristic rating on grades. The table lists the percentages of students whose grade could be predicted from their rating on a particular performance characteristic. Professionalism, Patient Management Skills, Oral Presentations and Clinical Skills best predicted grade (80% accuracy). Interpersonal Skills with Patients was the least accurate (75%). Overall, each performance characteristic predicted grade quite well.

Stepwise logistic regressions were performed to determine which combinations of the performance measures best identified A and B students, respectively. In performing the logistic regression, 2 criteria were used to add variables to the equation. Firstly, each performance measure that was added to the equation had to be statistically significant, and secondly, it had to increase the accuracy of the prediction by ≥1%. Table 4 presents the 5 stepwise logistic equations obtained from our analyses. Model 1 includes Professionalism and Fund of Knowledge. These 2 performance measures together predicted grade with 84% accuracy. No additional performance measures met our criteria for entry into this equation. Another logistic equation was generated by dropping Professionalism and Fund of Knowledge from the pool of measures considered for inclusion. Model 2 includes Oral Presentations, Clinical Skills, and

### Table 2 Intraclass correlation, reliability of the mean of 3 raters and P-values of the significance of differences among students

<table>
<thead>
<tr>
<th>Scale</th>
<th>Intraclass correlation</th>
<th>Reliability of mean of 3 raters</th>
<th>P from ANOVA comparing students</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Skills</td>
<td>0.14</td>
<td>0.33</td>
<td>0.001</td>
<td>−0.08</td>
</tr>
<tr>
<td>Fund of Knowledge</td>
<td>0.14</td>
<td>0.32</td>
<td>0.001</td>
<td>−0.35</td>
</tr>
<tr>
<td>Patient Management Skills</td>
<td>0.13</td>
<td>0.30</td>
<td>0.005</td>
<td>−0.14</td>
</tr>
<tr>
<td>Evidence of Study</td>
<td>0.12</td>
<td>0.28</td>
<td>0.003</td>
<td>−0.70</td>
</tr>
<tr>
<td>Diagnostic Ability</td>
<td>0.11</td>
<td>0.27</td>
<td>0.005</td>
<td>−0.04</td>
</tr>
<tr>
<td>Professionalism</td>
<td>0.11</td>
<td>0.27</td>
<td>0.004</td>
<td>−0.92</td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>0.08</td>
<td>0.22</td>
<td>0.022</td>
<td>−1.90</td>
</tr>
<tr>
<td>with Team and Staff</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write-ups/Progress Notes</td>
<td>0.08</td>
<td>0.20</td>
<td>0.034</td>
<td>−0.19</td>
</tr>
<tr>
<td>Interpersonal Skills with Patients</td>
<td>0.08</td>
<td>0.29</td>
<td>0.034</td>
<td>−0.60</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>0.04</td>
<td>0.11</td>
<td>0.159</td>
<td>−0.37</td>
</tr>
<tr>
<td>Grade (A or B)</td>
<td>0.12</td>
<td>0.28</td>
<td>0.002</td>
<td>−0.76</td>
</tr>
<tr>
<td>Grade score</td>
<td>0.12</td>
<td>0.28</td>
<td>0.003</td>
<td>−0.39</td>
</tr>
</tbody>
</table>

### Table 3 Accuracy of group identification from the logistic regression of each performance rating on grade

<table>
<thead>
<tr>
<th>Scale</th>
<th>Overall percentage of groups correctly classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professionalism</td>
<td>80</td>
</tr>
<tr>
<td>Patient Management Skills</td>
<td>80</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>80</td>
</tr>
<tr>
<td>Clinical Skills</td>
<td>80</td>
</tr>
<tr>
<td>Fund of Knowledge</td>
<td>79</td>
</tr>
<tr>
<td>Evidence of Study</td>
<td>78</td>
</tr>
<tr>
<td>Diagnostic Ability</td>
<td>77</td>
</tr>
<tr>
<td>Interpersonal Skills with Team and Staff</td>
<td>76</td>
</tr>
<tr>
<td>Write-ups/Progress Notes</td>
<td>76</td>
</tr>
<tr>
<td>Interpersonal Skills with Patients</td>
<td>75</td>
</tr>
</tbody>
</table>

* All logistic regression equations are statistically significant (P < 0.01)
Diagnostic Ability. These 3 measures predicted grade with 83% accuracy. Two other models were generated that predicted grade with accuracies of 82% and 79%, respectively. The only variable not to enter any of these equations was Interpersonal Skills with Patients.

From these analyses, it appears that the addition of a second or third measure to the model leads to a rather modest increase in the accuracy of the prediction. Additionally, these data strongly suggest that virtually any combination of 2 or 3 of these performance measures can be used with similar results.

Do faculty make a series of independent judgements about the student’s performance in each area or are the judgements correlated? The previous data suggest that faculty do not form a differentiated view of student performance in that about 80% of A-grade students can be distinguished from B-grade students based on any single performance rating. The data suggest that a unitary, general dimension may underlie faculty’s evaluation of students as there is relatively little variability left to explain.

To test this hypothesis, the 10 specific performance items were analysed using principal axes factor analysis (roots > 1.00 rule) and varimax rotation. There was 1 eigenvalue > 1 and the corresponding factor explained 65% of the total variance. All the factor loadings were ≥0.75 (range: 0.75 for Clinical Skills to 0.86 for Patient Management Skills), indicating that a 1-factor solution fitted the data well.

In order to further explore the implications of the factor analysis, we attempted to determine if the various performance characteristics measured anything more than could be explained by the general factor. We estimated this by determining the point-biserial correlation between each performance measure and grade, holding the other 9 measures constant (i.e. ninth order partial correlations). Table 5 presents the results of these analyses. It shows the partial correlations and the percentage of variance explained by those measures whose partial correlations were significant. Only 4 of the measures had a statistically significant relationship with grade after the variance explained by the other measures had been described. Of these 4, Professionalism correlated best with grade, although even this correlation was very modest.

The preceding analyses suggest that the inclusion of any 2 or 3 of these performance measures in a clinical rating form would be about as good as including any other 2 or 3 measures. However, before concluding that it doesn’t much matter which items are included in a rating form, at least in terms of establishing a grade, we need to explore an additional consideration: are all performance measures scored at the same level or are some rated higher or lower than others? Table 6 presents the means, standard deviations and standard errors for each of the 10 performance measures. A repeated measures ANOVA indicated that there were significant differences in the mean ratings ($P < 0.0001$). Fisher’s plausible least significant difference post hoc test was used to identify the pattern of specific differences among the

<table>
<thead>
<tr>
<th>Scale</th>
<th>Partial correlation</th>
<th>Percentage of variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professionalism</td>
<td>0.25*</td>
<td>6%</td>
</tr>
<tr>
<td>Clinical Skills</td>
<td>0.16*</td>
<td>3%</td>
</tr>
<tr>
<td>Fund of Knowledge</td>
<td>0.11*</td>
<td>1%</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>0.10*</td>
<td>1%</td>
</tr>
<tr>
<td>Evidence of Study</td>
<td>0.06</td>
<td>–</td>
</tr>
<tr>
<td>Pt Management Skills</td>
<td>0.04</td>
<td>–</td>
</tr>
<tr>
<td>Diagnostic Ability</td>
<td>0.03</td>
<td>–</td>
</tr>
<tr>
<td>Interpersonal Skills with Team and Staff</td>
<td>0.03</td>
<td>–</td>
</tr>
<tr>
<td>Interpersonal Skills with Patients</td>
<td>0.06</td>
<td>–</td>
</tr>
</tbody>
</table>

* $P < 0.05$

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal Skills with Team and Staff</td>
<td>4.46</td>
<td>0.69</td>
<td>0.03</td>
</tr>
<tr>
<td>Professionalism</td>
<td>4.44</td>
<td>0.70</td>
<td>0.03</td>
</tr>
<tr>
<td>Interpersonal Skills with Patients</td>
<td>4.37</td>
<td>0.68</td>
<td>0.03</td>
</tr>
<tr>
<td>evidence of Study</td>
<td>4.20</td>
<td>0.77</td>
<td>0.03</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>4.13</td>
<td>0.76</td>
<td>0.03</td>
</tr>
<tr>
<td>Write-ups/Progress</td>
<td>4.11</td>
<td>0.74</td>
<td>0.03</td>
</tr>
<tr>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Skills</td>
<td>4.03</td>
<td>0.76</td>
<td>0.03</td>
</tr>
<tr>
<td>Fund of Knowledge</td>
<td>3.99</td>
<td>0.77</td>
<td>0.03</td>
</tr>
<tr>
<td>Level 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Management Skills</td>
<td>3.92</td>
<td>0.82</td>
<td>0.04</td>
</tr>
<tr>
<td>Diagnostic Ability</td>
<td>3.91</td>
<td>0.76</td>
<td>0.03</td>
</tr>
</tbody>
</table>
mean (Table 6). The means could be separated into 4 groups. The items rated highest were Interpersonal Skills with Team, Professionalism, and Interpersonal Skills with Patients. Diagnostic Ability and Patient Management Skills were rated least positively. Overall, measures associated with the personal characteristics of students were rated highest, communication skills were rated in the middle and items related to cognitive skills were rated lowest.

DISCUSSION

The results of our analyses of faculty’s evaluation of students can be summarised as follows.

1 There was apparently poor agreement among faculty preceptors in their evaluation of medical students (low inter-rater reliability).

2 The rating on any individual performance characteristic predicted a student’s overall grade very accurately (75–80%). The addition of a second or third variable added ≤4% to the best predictive model.

3 The reason why any single performance characteristic predicted a student’s assigned grade so well was because faculty tended to make a simple, undifferentiated judgement on the quality of the student’s performance. They formed an overall impression, which was then reflected to a high degree in the rating of each of the performance characteristics. However, not all the performance characteristics were equally affected by the generalised view of the student.

4 The mean rating of students on non-cognitive aspects of clinical performance was significantly higher than that on cognitive aspects.

Other studies have reported low reliabilities similar to those we obtained.7–9 Increasing the number of raters is a recognised way to improve the reliability of ratings.10,11 However, that would probably not solve the problem of faculty’s generalised view of a student’s performance.12

The solution to the validity problem does not appear to reside in solving the low reliability problem. More probably it is the other way around. If we solve the validity problem (generalised view of student performance), the reliability of ratings will improve.

Perhaps the best way to address the validity issue is to ask faculty to rate those characteristics they have ample opportunity to observe. Evaluation forms should probably be restricted to behaviours that are directly witnessed by raters (faculty). Other modalities (e.g. standardised patients) should be employed to assess important performance characteristics not directly measured by faculty during the clinical clerkship.13

The reliability analyses raised another question: is there sufficient agreement among raters to differentiate the performance of 1 student from another? One way to address the issue is to determine if an ANOVA can detect significant differences among the mean evaluations (of the 5 rotations) of the students. If the ANOVA is not significant, there would appear to be little basis for assigning students different grades. In our study, we found significant differences for all the variables, with the exception of Oral Presentation (Table 2). Thus, to some degree, we can validly assign different grades (A or B) to students. Our data strongly suggest that even mean ratings with low reliabilities can differentiate among students.

The primary purpose of this study was to identify the factors that underlie the grading decisions of faculty. One view would hold that all the items on our evaluation form were important as each allowed about the same accuracy of grade prediction as any other (Table 3). Interpersonal Skills with Patients correctly predicted assigned grades 75% of the time, giving the lowest accuracy rate of any of the 10 performance characteristics. Four of the performance characteristics (Professionalism, Patient Skills, Oral Presentations and Clinical Skills) tied for the highest at 80%. However, the logistic equations reveal that there is a small, but statistically significant, increase in the accuracy of the models between the least and most predictive (Table 4). After the first variable enters the equation (differentiating A students from B students with 80% accuracy), subsequent variables each add modestly to the differentiation between the 2 groups. No selected combination of performance characteristics exceeds 84% in predictive ability, which interestingly is the accuracy of prediction if all 10 performance characteristics are employed. Thus, almost all the explainable variance relates to 1 measure and, to some degree, any measure will do. The explanation is that ratings by faculty reflected 1 factor, a unitary, undifferentiated view of the student. Others have appreciated the same reality.14,15 Simply put, faculty make an overall decision as to whether a particular student deserves an A or a B, and then tend to score the 10 performance variables to reflect that grade.

Principal axis factor analysis of the 10 performance characteristics was carried out and led to the same
conclusion. This statistical methodology identified that a simple, undifferentiated judgement factor explains 65% of the total variance. In addition, each of the 10 individual performance ratings correlates well with the common judgement (factor loading \( \geq 0.75 \) for each item).

The logistic regression analyses suggested that 2 or 3 variables underlie the overall grade assignment. By contrast, the factor analysis indicated that the various scales measure 1 common factor. Our interpretation of these analyses is that there is 1 general factor as well as 1 or 2 specific factors underlying these ratings. The 1 or 2 specific factors have much less impact on the grading decision than does the general factor.

Appreciation of the implications of low reliabilities for each of the scales and the high correlations among the scales is important. The former tell us that the judgement of 1 attending doctor does not correlate highly with the perceptions of other faculty; the latter indicate that those judgements by individual faculty tend to be undifferentiated and 1-dimensional.

To address whether this simple common factor explains all the variance, we then asked this question: if the other 9 performance characteristics are held constant, are there any cases in which the addition of the 10th performance characteristic will meaningfully increase the accuracy of grade prediction? Ninth-order partial correlation analyses indicate there are 4 such situations. When any of Professionalism, Clinical Skills, Fund of Knowledge or Oral Presentations is added as the 10th factor to the other performance ratings, accuracy of grade prediction increases by 1–6%, which is a small but statistically significant improvement. This would indicate that faculty ratings of these 4 characteristics are not totally predicted or explained by the general factor and, in these areas at least, attending doctors may make slightly more sophisticated rating decisions. These 4 variables appear to have some small specific validity, but in practical terms the effect of this reality is minimal.

We found it interesting that, although all performance characteristics correlated well with assigned grade, the means for the non-cognitive characteristics were all significantly higher than the means for the other 7 characteristics (Table 6). Several explanations are possible. Faculty work closely with students on a variety of surgical rotations, and presumably get to know and like most of them. As a result, they rate students highly on personal qualities such as Professionalism and Interpersonal Skills. It is easier to point out to a student the anatomy questions he or she answered incorrectly during an operation (Fund of Knowledge) than it is to explain why he or she was not judged highly for Professionalism and Interpersonal Skills. It is also quite likely that early in their clinical exposure, most students are better ‘people’ than they are doctors or surgeons. Fund of Knowledge, Patient Management Skills and Diagnostic Ability, for example, have room for improvement, whereas a developed work ethic and the ability to relate to others reflect more innate traits and are less likely to change as training and learning progress. Lastly, evidence suggests that surgeons consider Professionalism, in particular, a vitally important characteristic. A student might be grade A calibre with a Fund of Knowledge rating of 4.0, for example, but a similar score for Professionalism might not be commensurate with an A grade.

Our study has at least 3 implications. Firstly, at a practical level, it makes little sense to have long multi-item forms. Our data suggest that an evaluation tool containing 2 or 3 items is as informative as a tool containing 10 or more. Certainly, shorter forms are easier for faculty to use, but careful consideration should be paid to the subjects included. Following this study, we revised our faculty evaluation-of-students form. It now consists of 5 items. We surveyed faculty to determine which performance characteristics they had an opportunity to observe directly and we chose these:

1. Fund of Medical and/or Surgical Knowledge;
2. Clinical Evaluation and Judgement;
3. Interactions with Patients and Staff;
4. Professionalism, and
5. Communication of Medical Information.

Secondly, it could be argued that the ratings on the performance characteristics listed on the evaluation form should be added together to form a composite clinical performance score. The content validities of the specific item ratings are questionable.

The final implication of this study is that we do not understand fully why faculty evaluation of students apparently represents such a simple judgmental process. Future research should address this issue. Because the decisions made on these forms have serious professional implications, is important to better understand their low reliability and questionable differential validity.
Contributors: AP and MP contributed to the study design and the write-up of the paper. MBD contributed to the study conceptualisation and design, data analysis and the revision of the paper.

Acknowledgements: none.

Funding: none.

Conflicts of interest: none.

Ethical approval: not required.

REFERENCES

7 van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. Teach Learn Med 1990;2:58–76.

Received 22 June 2006; editorial comments to authors 10 October 2006; accepted for publication 12 February 2007
Comparison of two instruments for assessing communication skills in a general practice objective structured clinical examination

Marc Van Nuland,1 Wim Van Den Noortgate,2 Jan Degryse1 & Jo Goedhuys1

OBJECTIVE In recent decades, there has been increased interest in tools for assessing and improving the communication skills of general practice trainees. Recently, experts in the field rated the older Maas Global (MG) and the newer Common Ground (CG) instruments among the better communication skills assessment tools. This report seeks to establish their cross-validity.

METHODS Eighty trainees were observed by 2 raters for each instrument in 2 standardised patient stations from the final year objective structured clinical examination for Belgian trainee general practitioners. Each instrument was assigned 6 raters.

RESULTS Trainees showed the lowest mean scores for evaluating the consultation (MG7), summarising (MG11), addressing emotions (MG9) and addressing feelings (CG5). Inter-rater \( \kappa \) statistics revealed fair-to-moderate agreement for the MG and slight-to-fair agreement for the CG. Cronbach’s \( \alpha \) was 0.78 for the MG and 0.89 for the CG. A generalisability study was only feasible for the MG: it was more helpful to increase the number of cases than the number of raters. Agreement between the instruments was examined using \( \kappa \) statistics, Bland–Altman plots and multi-level analysis. Ranking the trainees for each instrument revealed similar results for the least competent trainees. Variances between and within trainees differed between instruments, whereas case specificity was comparable. Multi-level analysis also revealed a rater–item interaction effect.

CONCLUSIONS The 2 instruments have convergent validity, but the drawbacks of the CG, which has fewer items to be scored, include lower inter-rater reliability and score variance within trainees.

KEYWORDS comparative study [publication type]; clinical competence/ *standards; *communication; *physician–patient relations; family practice/ *standards; vocational education; Belgium.

Medical Education 2007; 41: 676–683
doi:10.1111/j.1365-2925.2007.02788.x

INTRODUCTION

The importance of competent doctor–patient communication as central to effective general practice is now widely accepted. It has also been emphasised in the European Definition of General Practice/Family Medicine.1 Accordingly, more attention has been given in recent decades to evaluating communication skills in the summative assessment of trainee general practitioners (GPs). This is also true in Flanders, the Dutch-speaking part of Belgium. Part of the Flemish objective structured clinical examination (OSCE) is focused on communication skills. Unfortunately, the scores on communication skills within this OSCE are disappointingly low. Interventions to improve communication skills during GP vocational training have had variable results, with some interventions apparently resulting in certain ameliorations, whereas others show no effect in improving certain aspects of GP trainee performance.6,7 The quality of the assessment instruments used is crucial to this type of research. The absence of a ‘gold standard’ in medical education,8 which also applies to the assessment of communication skills, and the issue of context specificity makes it
Overview

What is already known on this subject

The Maas Global (MG) and the Common Ground (CG) instruments are rated among the better communication assessment instruments.

What this study adds

The MG and CG have convergent validity. The CG measures a single major aspect of communication skills, whereas the MG takes a broader view.

The drawback of the CG, which has a lower number of items, is its lower inter-rater reliability and variance within trainees.

Multi-level analysis reveals an interaction effect between rater and item.

Suggestions for future research

The instruments could be compared in a higher number of stations.

Item-rates interaction could be researched qualitatively.

difficult to assess communication skills. Recent reviews of communication skills assessment instruments show that numerous instruments exist, but they differ in many ways.9–11

Instruments vary in the number of items to be scored, ranging from few (e.g. the Common Ground instrument12) to many (e.g. the LIV-MAAS scale13). Recently, a consensus panel of 6 experts in the field evaluated 15 instruments for assessing doctor–patient communication.9 They rated the Maas Global (MG)14,15 and the Common Ground (CG)12 instruments among the better tools with regard to their overall value, usability and psychometric properties. Both instruments are based on global ratings and earlier research has suggested that they are more reliable than detailed and lengthy checklists.15 Experience with the MG goes back about 20 years16 and is supported by many validation studies.14,15,17 The CG, by contrast, was developed only recently.12 It seems promising in that it focuses more on what are believed to be essential communication skills18 and in the lower number of items to be scored. This research report aims to compare the CG with the better known MG in terms of item scores, internal consistency, generalisability and agreement between raters and between instruments. By investigating the properties of both instruments, this report may produce some evidence to guide the choice between them.

METHODS

We selected 2 stations with a focus on communication from the final year OSCE for Flemish general practice vocational trainees and assessed all participating trainees. In the first station (P), the standardised patient (SP) impersonated a 36-year-old with an unplanned pregnancy. The second station (O) presented a 50-year-old obese man with ambivalence towards efforts to lose weight. Each of these 8-minute stations was directly observed by 4 raters: 2 using the Dutch MG19 and 2 using the translated CG. The MG contains 13 global ratings for the assessment of communication skills, scored on a 7-point Likert scale of 0–6. For both stations, only 11 of the 13 items from the MG instrument were applicable; the scores for Follow-up Consultation and Physical Examination were irrelevant in the context of these stations. The CG consists of 7 global ratings scored on a 5-point Likert scale of 1–5. In order to compare both instruments in the way they evaluate trainee competence, we chose the middle of both scales (3) as the cut-off for distinguishing between insufficient and sufficient competence. Twelve raters were recruited by advertisement among final-year psychology students and recently graduated psychologists. All raters were given a 3.5-hour training session on 1 of the assessment instruments, 6 on the MG and 6 on the CG. The SPs had participated in previous OSCEs and were trained by faculty members for 3 hours in performing the selected stations (O or P). Both raters and SPs were paid.

We used SAS software for most of the statistical analyses (descriptive statistics, factor analysis, Cronbach’s α, inter-rater κ calculation, multi-level analysis), GENOVA for generalisability analysis and MedCalc to construct Bland–Altman plots. We tested null hypotheses using 2-sided significance tests with a 5% level of significance.

© Blackwell Publishing Ltd 2007. MEDICAL EDUCATION 2007; 41: 676–683
RESULTS

In total 80 vocational trainees were observed, yielding $4 \times 80$ total scores per instrument. Raters differed in the number of trainees they observed, which ranged from 20 to 80, with a mean of about 53 trainees. Missing item scores were limited to 63 of 5760 (1.1%) for all trainees.

Comparing the instruments for item content and observed scores

Each instrument has its own set of items, shown in the first column of Table 1. The remaining columns summarise the distribution of item scores. On average, our cohort of trainees showed insufficient competency (item score < 3) when using the MG in Evaluating the Consultation (MG7), Summarising (MG11) and Dealing with Emotions (MG9). When using the CG, trainees showed insufficiency only in Addressing Feelings (CG5).

Internal consistency

In order to compare the instruments regarding internal consistency, we first calculated Cronbach’s $\alpha$, which equalled 0.78 for the standardised items of the MG and 0.89 for the CG items. This indicates acceptable internal consistency for both. As the last item in the CG (CG7) relates to Global Interview Performance, we recalculated Cronbach’s $\alpha$ without CG7, lowering its value from 0.89 to 0.85.

Secondly, internal consistency for each instrument was investigated by exploratory factor analyses. The extent to which major factors explain common variance within an instrument reflects its internal consistency. Factor analysis of the MG revealed 1 major factor (F1: Patient-oriented Problem Analysis and Explanation) with an eigenvalue of 3.09, which explained 60% of the common variance within the MG scores. By contrast, the 1 major factor we found for the CG (F2: Patient-centred Communication), with an eigenvalue of 3.35, explained a greater proportion (86%) of its common variance. Factor loadings for both major factors are shown in Table 2. Other factors in the analysis of both instruments were not retained as the first factor already explained a substantial portion of the common variance, and as the eigenvalues of the second and subsequent factors were considerably smaller than the eigenvalue of the first factor. This was clearly confirmed in a Cattell’s Scree Plot.20

Moreover, when more factors were retained, they appeared to be difficult to interpret, even after rotation.

<table>
<thead>
<tr>
<th>Item content</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maas Global</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section 1: Communication skills for each separate phase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG1 Opening</td>
<td>4.13</td>
<td>0.74</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>MG3 Clarification of the reason for encounter</td>
<td>3.02</td>
<td>1.49</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>MG5 Evaluation and diagnosis</td>
<td>3.98</td>
<td>0.73</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>MG6 Management and arrangements</td>
<td>4.13</td>
<td>0.97</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>MG7 Evaluation of consultation</td>
<td>1.67</td>
<td>1.78</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Section 2: General communication skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG8 Exploring</td>
<td>4.05</td>
<td>0.90</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>MG9 Dealing with emotions</td>
<td>2.89</td>
<td>1.52</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>MG10 Providing information</td>
<td>4.15</td>
<td>0.88</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>MG11 Summarising</td>
<td>2.66</td>
<td>1.77</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>MG12 Structuring</td>
<td>3.77</td>
<td>0.86</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>MG13 Empathy and flexibility</td>
<td>4.92</td>
<td>1.05</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td><strong>Common Ground</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG1 Rapport building</td>
<td>3.34</td>
<td>0.75</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG2 Information management</td>
<td>3.11</td>
<td>0.78</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG3 Eliciting all agenda items</td>
<td>3.01</td>
<td>0.80</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG4 Active listening for full understanding of ideas, concerns, and expectations</td>
<td>3.32</td>
<td>0.80</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG5 Addressing feelings with patients</td>
<td>2.87</td>
<td>0.92</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG6 Reaching common ground</td>
<td>3.21</td>
<td>0.82</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>CG7 Global interview performance</td>
<td>3.22</td>
<td>0.82</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

* Calculations of means and standard deviations are based on 4 observations, 2 in each station

SD = standard deviation
Generalisability

Generalisability (G) analysis showed acceptable reliability of the MG scores. Raising the number of stations increased reliability more than raising the number of raters per station. As a result of the low variance of some scores, it was technically impossible to complete a G analysis of the CG.

Agreement between raters

The mean inter-rater correlation was $r = 0.49$ for the MG item scores and $r = 0.26$ for the CG item scores. The mean weighted inter-rater $\kappa$ for the MG was 0.56 (range 0.31–0.74) in the O station and 0.35 (range 0.14–0.56) in the P station. Equivalent mean $\kappa$ statistics for the CG were, respectively, 0.23 (0.04–0.51) and 0.19 (0.03–0.46).

Agreement between the 2 instruments

To investigate how the 2 instruments agree, we looked at Bland–Altman plots and at their correlation matrix, calculated $\kappa$ statistics and made empirical Bayes estimates based on the results of multi-level analyses.

1 In order to construct Bland–Altman plots, we transformed the scores into percentages (Fig. 1). These plots give insight about whether agreement depends upon the magnitude of the measurement and show that for weak trainees (an average of < 3, or < 50%), scores tended to be higher with the MG than with the CG, whereas for stronger trainees, scores tended to be lower with the MG than with the CG.

2 The correlation matrix of the items of both instruments does not reveal any relevant correlation between MG items and CG items (all $r < 0.32$). The correlation between the factors computed in the factor analysis of the MG and CG measured $r = 0.36$ ($P < 0.0001$), suggesting that they assess rather different dimensions of communication.

3 As we had defined scores of < 3 as insufficient ability (fail) and scores ≥ 3 as sufficient ability (pass), the $\kappa$ between instruments on the pass or
Multi-level analysis (MLA) was used to further explore the relationship between the 2 instruments. In general, multi-level models are used to analyse clustered data. Characteristics of the clusters, as well as of the units within the clusters, can be used as independent variables in describing the variation in 1 or more dependent variables. In our case, we have item scores that are clustered in persons. We used a multivariate model, which treated the item scores of the CG and the MG as 2 dependent variables. Multi-level analysis showed that the 2 instruments had different mean scores (MG 3.50, CG 3.16), as well as a different variance within trainees (MG 0.10, CG 0.072). The MLA also revealed an interaction effect between trainees and items: trainees may score relatively strongly or weakly on particular items compared with other items.

Multi-level analysis confirmed case-specificity as MG scores on the obesity station were on average 0.15 lower than on the pregnancy station (P = 0.0015), whereas CG scores were only 0.06 lower (P = 0.07). Regarding this interaction effect between instrument and station, the difference between the MG and CG was not statistically significant (P = 0.12).

Multi-level analysis further revealed significant differences in the way raters gave scores for each instrument (P < 0.0001 for the MG and CG). It also revealed an interaction effect between rater and item (P < 0.0001), meaning that individual raters may tend to give higher or lower scores on particular items in comparison with other raters.

As in practice scores will usually be based on 1 instrument only, we also performed 2 univariate MLAs (1 for each instrument) in order to calculate the empirical Bayes estimates of trainees’ skills. A correlation of r = 0.59 (P < 0.0001) was found between the estimates of MG and CG scores (Fig. 2).

It was possible to rank the trainees based on the estimates of their total scores for the 2 stations. A Spearman’s rank correlation coefficient of r = 0.55 (P < 0.0001) was found. A closer look at the data for this analysis reveals that of the 8 trainees in the first decile of the MG, 5 were in the first decile and 7 were in the first 2 deciles of the CG. Similarly, of those trainees in the first decile of the CG, 5 were in the first decile and 7 in the first 2 deciles of the MG.

Conversely, only 3 of the trainees from the last decile of the MG were found in the last 2 deciles of the CG and 3 from the last decile of the CG were found in the 2 last deciles of the MG. These findings suggest that the 3 instruments agree more in selecting less competent trainees than in selecting the best.

DISCUSSION

In this study, we cross-validated 2 communication assessment instruments in a standardised setting (OSCE). Bland–Altman plots and interinstrument k calculation revealed a slight-to-fair agreement between instruments when used in 1 station. Multi-level analysis showed a correlation of 0.59 between the estimates of individual trainees’ total scores on both instruments, based on the observations in 2 stations. These are arguments for convergent validity of the 2 instruments. To our knowledge, this study is the first to compare 2 communication assessment instruments in a simultaneous direct observation setting. To us, this experimental method of comparing assessment instruments seems to be more valid than a comparison based on ratings by experts, such as in the study by Schirmer et al. Three aspects will be considered in the further discussion of our study: study design; study results, and suggestions for further research.

Study design

Choosing a standardised setting for our study allowed us the advantage of avoiding the bias caused by the patient variance that would occur in real patient consultations, but the disadvantage of measuring...
competence rather than performance. Considering the purpose of our study, however, this disadvantage is of minor importance. Furthermore, by applying both instruments simultaneously in direct observation and by deploying 2 raters per instrument, we maximised standardisation and minimised bias resulting from patient variance. Bias resulting from missing data seems rather improbable as only 1.1% of item scores were missing and a closer look at the missing scores did not reveal a link to a particular item, rater or trainee. Restricting our study to only 2 stations had the disadvantage of providing less information concerning generalisability and as such, the respective conclusions must be regarded with caution. Restrictions in budget, time and manpower made it impossible to expand our experiment over more than 2 stations.

We chose to engage psychologists to rate communication skills for 2 reasons. Firstly, as they are not doctors, their scores on communication will be less influenced by ratings on the medical content of the consultation, a potential source of bias. Secondly, psychologists are familiar with using rating scales when observing behaviour and human interaction. As such, the 3.5-hour experiential training sessions ought to have been sufficient. A disadvantage of engaging psychologist raters is that we remain uncertain whether doctors, the usual raters of OSCE stations, would have produced equal results and conclusions.

**Study results**

When comparing the MG and CG instruments, the different variances cannot be attributed solely to the different scales. If the 5-point scale from the CG were to be extended to a 7-point scale, the standard deviation (SD) would rise by a factor of $7/5$ and variance by a factor of $1.95 (= [7/5]^2)$. The variance within trainees for the CG ($= 0.61$) would rise to 1.20, which is lower than the MG variance of 1.99. However, the variance between trainees for the CG ($= 0.072$) would rise to 0.14, which is higher than the MG variance of 0.10. This lower variance between trainees in the MG may contribute to the decreasing slope of the Bland–Altman plots.

Ranking trainees according to their scores on 2 stations revealed highly similar results for the weakest trainees on both instruments, suggesting that the MG and CG are comparable in selecting trainees in need of additional support regarding their communication skills. Nevertheless, further research is needed because this study included only 2 OSCE stations and, as a result of case specificity, this number is too low to infer generalisability. Cushing’s overview mentions that 8–24 stations are necessary in order to obtain generalisability, whereas Lang et al. concluded in their study of the CG that only 5 cases were needed to achieve a generalisability coefficient ≥ 0.80. Our study could not confirm the results of the Lang study, yet the generalisability we found for the MG in our study was more consistent with findings in previous research.

The mean item scores we found on the MG for GP vocational trainees in their final year are very similar to those in the study by van Dalen et al., which are on average higher than those found by Kramer et al. (Fig. 3). It is striking that in both Dutch studies, as in ours, final year GP trainees had the weakest score on the item Evaluating the Consultation. This item also seems to have a low factor loading on the major

![Figure 3](image-url)
factor. This suggests that this item measures an aspect of low relevance to the rest of the scale.

Interobserver agreement for item scores measured by the weighted $\kappa$ ranged from 0.35 (P) to 0.56 (O) for the MG, coming into the category of fair-to-moderate agreement. For the CG these were 0.19 (P) and 0.23 (O), corresponding to slight (P) and fair (O) agreement. This rather low agreement was also found in the study by van Thiel et al., which used a former version of the MG. The inter-rater agreement we found for the CG was lower than that reported by Lang et al., who used only 2 strictly selected raters observing videos of 100 students carrying out 4 SP interviews. We, by contrast, used 6 raters per instrument and no pair of raters observed more than 20 of the same trainees.

An interesting result of the MLA was that it showed an interaction effect between raters and items: some raters tended to give relatively high or low scores on specific items, compared with others. This aspect may not be apparent in studies that rely on a small number of raters.

The 2 instruments seemed to differ in terms of internal consistency. The CG7 item (Global Interview Performance) could bias the Cronbach’s $\alpha$ calculation as it is not a ‘true’ item. As such, calculating $\alpha$ without CG7 may give a result that more accurately represents internal consistency. Nevertheless, Cronbach’s $\alpha$ remained higher for the CG than for the MG, although the number of items was smaller.

This result, as well as the higher proportion of variance explained by the first factor from our factor analysis, suggests that the CG measures 1 major aspect of communication skills in particular.

**Further research**

This study proved to be valuable in gaining better understanding of the MG and CG and showed the extent of agreement between them. Analogous studies with more and different stations (e.g. focusing on specific GP communication tasks, such as breaking bad news, counselling, handling specific conditions) are needed to confirm our results. Our study suggests that increasing the number of stations enhances generalisability better than increasing the number of raters per station. Qualitative research focused on raters may contribute to a better understanding of item–rater interaction.

**CONCLUSIONS**

This study cross-validated the MG and CG instruments for use in 2 OSCE stations. It revealed that they are comparable regarding case specificity and in tracing weak trainees.

By contrast, the instruments proved to be different in terms of item content, internal consistency (CG > MG), variance of scores within trainees (MG > CG), variance between trainees (MG < CG) and inter-rater reliability (MG > CG). These results may aid in choosing between the 2 instruments, depending on the aim or context of the assessment (e.g. formative or summative; global communication skills or specific skills such as addressing feelings, structuring the consultation or finding common ground).

Multi-level analysis showed a significant item–rater interaction effect for both instruments, shedding new light on the complexity of inter-rater agreement in communication assessment.

**Contributors:** MVN developed the study’s protocol, trained the raters, organised the data collection, carried out the data analysis, and prepared and edited the manuscript. WVDN carried out data analysis and critically revised the manuscript. JDG raised the original idea of the study design, carried out the generalisability study and revised the manuscript. JG contributed to the study protocol, trained the raters, carried out the data analysis and critically revised the manuscript.

**Acknowledgements:** we would like to thank the staff of the Centre for Postgraduate General Practice Vocational Training (Interuniversitair Centrum voor HuisartsenOpleiding, ICHO) and the trainees who took part in the objective structured clinical examination.

**Funding:** none.

**Ethical approval:** not required.

**REFERENCES**


Received 11 June 2006; editorial comments to authors 15 October 2006; accepted for publication 20 February 2007
A preliminary study of empathy, emotional intelligence and examination performance in MBChB students

Elizabeth J. Austin,1 Phillip Evans,2 Belinda Magnus1 & Katie O’Hanlon1

CONTEXT There is considerable interest in the attributes other than cognitive ability that medical students need in order to be professionally successful, with a particular focus on empathy and emotional intelligence (EI). Selection considerations have also motivated interest in such attributes as predictors of academic success. There are reports of declines in empathy in US medical students, but no comparative information is available for UK students.

OBJECTIVES This study aimed to compare empathy levels in medical students in Years 2, 3 (pre-clinical) and 5 (clinical), to examine gender differences in empathy and EI, and to investigate whether EI and empathy are related to academic success.

METHODS Questionnaires assessing EI and empathy were completed by students. Previous empathy scores for the Year 2 cohort were also available. Empathy trends were examined using ANOVA; trends for the Year 2 group for whom Year 1 scores were available were examined using repeated-measures ANOVA. Associations of EI and empathy with academic success were examined using Pearson correlation.

RESULTS A significant gender × cohort effect was found, with male empathy scores increasing between Years 1 and 2, whilst female scores declined. Peer ratings in Year 2 problem-based learning (PBL) groups were positively correlated with EI.

CONCLUSIONS Trends in levels of empathy differed by gender. The reasons for this require further investigation, particularly in relation to course content. Associations between academic performance and EI were sparse, and there were none between academic performance and empathy, but the effects of EI (and other characteristics) on PBL group functioning represent a promising area for future study.

KEYWORDS humans; male; female; students, medical/*psychology; *empathy; *education, medical, undergraduate; men/*psychology; women/*psychology; Scotland; cross-sectional studies; longitudinal studies.

Medical Education 2007: 41: 684–689
doi:10.1111/j.1365-2923.2007.02795.x

INTRODUCTION

Increasing attention is paid to the attributes over and above cognitive ability (as assessed by standard IQ tests) that are required for the successful pursuit of a career in medicine and to how these might be included in selection processes for medical courses.1–3 In this context, the effects of personality on academic performance in medical students have been studied, resulting in a general finding that the personality trait of Conscientiousness is a predictor of academic success in undergraduates.4 This finding has been replicated in UK and Flemish pre-clinical medical students,5,6 although Conscientiousness was found to be negatively associated with performance in UK clinical students.5 Within the specific context of UK medical student selection, interest in predictors other than examination scores is also motivated by considerations of fairness to all applicants, given concerns that medical school admission criteria and
Selection processes based mainly on academic attainment appear to disadvantage some applicant groups.\(^7\),\(^8\) There is also increasing recognition in the medical education literature that interpersonal skills are important for doctors, with discussion focused on the quality of the doctor-patient relationship. These considerations have led to some debate on, and implementation of, training in empathy and emotional intelligence (EI) skills as a medical degree component.\(^3\),\(^9\),\(^10\)

Emotional intelligence and empathy are related but distinct constructs. Although there is some divergence amongst EI researchers on the best model for EI, there is agreement that it covers the ways in which people differ in their emotional capabilities, in both the intrapersonal (mood regulation, stress management, perceiving one’s own emotions) and interpersonal (social skills, perceiving others’ emotions) domains. As suggested by this definition, EI measures normally assess a number of EI subcomponents as well as combining these into an overall EI score. Empathy overlaps with interpersonal EI and covers the ability to be aware of and understand another person’s feelings. Recent discussions of empathy in doctors have focused on the desirability of doctors being able to understand the feelings of their patients without becoming overly emotionally involved.\(^9\),\(^11\),\(^12\)

Research and discussion on EI and empathy in medical students has been paralleled by studies showing that higher perceived levels of doctor empathy and EI are associated with higher levels of patient satisfaction.\(^13\)–\(^15\) Although the interpersonal and empathic skills of medical students and doctors have received wider literature coverage than intrapersonal skills, there is also recognition that capabilities such as mood regulation and stress management are relevant to the work environment with which doctors are required to deal\(^1\),\(^3\) and that doctors’ personality traits are related to outcomes such as stress and burnout.\(^16\) Given the intensity of the debate on EI and empathy in medicine, there has been surprisingly little research on the extent to which they are associated with success in medical studies. Emotional intelligence was found to be positively associated with academic success in Year 1 Canadian (non-medical) students, a finding which is believed to reflect the role of EI in facilitating the transition from school to university.\(^17\) A recent study\(^18\) found some evidence for positive associations between EI and academic performance in Year 1 UK medical students. Although the general mechanism of EI facilitating the transition to university is a possible explanation for these findings, it is noteworthy that the association was found only for a course component for which EI skills were relevant and was absent for a basic science course taken by the same students. Another study found empathy to be positively associated with clinical competence in US students.\(^11\) Some longitudinal studies have shown evidence of a decline in levels of empathy and conceptually related measures, such as patient-centredness, in North American medical students as they progress through their training,\(^19\)–\(^22\) which is regarded as a cause for concern.\(^23\) There are no published findings on longitudinal trends in empathy in UK medical students (where, unlike in North America, medicine is taken as an undergraduate degree).

Whilst discussing issues related to EI and empathy in medical students, it is also relevant to mention gender effects. Female medical students generally show better academic performance than males.\(^24\) Higher scores on total EI are often found in females, although some studies report no gender difference. Subscales for EI show a more mixed picture. As with total EI, gender differences are not invariably found, but there are findings that females score higher on subscales.

**Overview**

**What is already known on this subject**

A decline in empathy during medical training has been reported in North American medical students and is regarded as a cause for concern. There is some evidence for associations between empathy and EI and academic performance in medical students.

**What this study adds**

In this UK sample, empathy trends differed by gender, with male scores increasing and female scores declining between Years 1 and 2. Problem-based learning (PBL) group peer ratings were related to emotional intelligence.

**Suggestions for further research**

The reasons for the observed gender differences in empathy trends and the effects of emotional intelligence on PBL group performance should be studied further.
relating to interpersonal capabilities, whereas males score higher on subscales relating to intrapersonal capabilities.\textsuperscript{25,26} For empathy, females are generally found to have higher scores than males.\textsuperscript{27} These findings raise the possibility, assuming that higher levels of empathy and interpersonal EI are helpful in medical studies, that EI and/or empathy levels may be a contributory factor to differences between men and women in academic success.

In the light of the above literature review, 1 objective of the present study was to compare empathy levels in medical students in Years 1, 2 (pre-clinical) and 5 (clinical), in order to determine whether the decline in empathy noted in North American students also occurs in UK students. Two additional objectives were to assess gender differences in student levels of empathy and EI, and to examine whether EI and empathy are related to academic performance.

METHODS

Participants

The participants consisted of 273 medical students (85 male, 188 female) in Years 1, 2 and 5 of the MBChB programme at Edinburgh University. Years 1 and 2 of the Edinburgh curriculum promote a multiple approach to teaching and learning, with 3 hours of problem-based learning (PBL) per week set within a timetable of blended learning (e-learning, lectures, clinical experiences, tutorials and practicals) that are constructively aligned.\textsuperscript{28} Year 5 students experience conventional clinical attachments in surgery, medicine and general practice.

Measures

Participants completed a 41-item EI scale\textsuperscript{29} and the 20-item Jefferson Scale of Physician Empathy.\textsuperscript{12} The EI scale has 3 subcomponents. Optimism/Mood Regulation provides a measure of intrapersonal EI; Appraisal of Emotions is a measure of ability to perceive emotions in others, and Utilisation of Emotions contains items relating to the use of emotions in problem-solving. The 3 subscale scores can be combined to give an overall EI score. The empathy scale was originally constructed for use with US medical students; a minor modification was applied to the scale for UK use, whereby the term ‘physician’ was replaced with ‘doctor’. A previous study carried out with UK medical students indicated this modified scale to be highly reliable and to show a pattern of correlations with other variables that would be expected for an empathy scale, suggesting its suitability for use with this student group.\textsuperscript{18} For the Year 2 subgroup, scores on the Scale of Physician Empathy obtained in Year 1\textsuperscript{18} were also available. The questionnaire also contained an optional section that allowed participants to indicate whether they were willing for their examination and coursework results to be accessed later, and to supply their student ID numbers to allow data matching. Information was available on end-of-year overall academic performance for all 3 year groups.

Procedure

Questionnaires were distributed at the start of lectures, when a brief explanation of the study was given. Completed questionnaires were either returned to an investigator at the end of the lecture or filled in later and returned to a collection box.

RESULTS

Internal reliabilities of the scales

Internal reliabilities (Cronbach’s $\alpha$) were good for the total EI and Physician Empathy scales (0.82, 0.88) and for the EI Optimism/Mood Regulation and Appraisal of Emotions subscales (0.71, 0.75). The Utilisation of Emotions subscale was slightly less reliable (0.66).

Gender differences

Table 1 shows EI full-scale and subscale scores and empathy scores for males and females in the whole sample and for each year group. Significant gender differences were found for the whole sample on EI, empathy and the Utilisation of Emotion subscale, with females scoring higher than males on all 3 scales ($t[268] = 3.23, P = 0.001$; $t[263] = 2.43, P = 0.016$; $t[268] = 3.52, P = 0.001$, respectively). There were no significant gender differences in end-of-year marks in any of the 3 year groups.

Comparison of Physician Empathy scores across year groups

In order to examine possible empathy differences amongst the Year 1, 2 and 5 groups, a 2-way (cohort × gender) ANOVA was performed. There was a significant main effect of gender ($F[1,259] = 6.44, P = 0.012$) but not of year ($F[2,259] = 1.06, P = 0.35$, not significant [NS]), and a significant gender–year interaction ($F[2,259] = 5.91$,...
These results are illustrated in Fig. 1, which shows increasing empathy levels in males between Years 1 and 2, whereas empathy in females declines. Post-hoc testing showed that the difference between male and female levels of empathy was only significant in Year 1 ($t_{(95)} = 4.02$, $P < 0.001$). The difference in empathy levels between Year 1 and 2 males was significant ($t_{(54)} = 2.18$, $P = 0.034$), and the difference for females showed a tendency towards significance ($t_{(135)} = 1.86$, $P = 0.065$).

A repeated-measures ANOVA was also performed on the data for Year 2 students, using their current and previous (Year 1) scores; scores for both years were available for 70 students (24 males, 46 females). There was no significant main effect for year ($F_{(1,69)} = 1.31$, $P = 0.26$ [NS]) or gender ($F_{(1,69)} = 0.90$, $P = 0.35$ [NS]), but gender × year was again significant ($F_{(1,69)} = 6.12$, $P = 0.016$). These results are illustrated in Fig. 2. Paired sample $t$-tests showed that the increase in empathy in males was marginally significant, but the decline in females was non-significant ($t_{(23)} = 1.95$, $P = 0.064$; $t_{(45)} = 1.18$, $P = 0.25$, respectively).

Taken together, these 2 analyses suggest different empathy trajectories for males and females between Years 1 and 2.

**Correlations of EI and empathy with academic performance**

There was little evidence of associations between total EI, EI subcomponents or empathy and academic performance, with no significant associations with end-of-year marks found for any of the year groups. For Year 1 students, the mark for an exercise involving communication skills (‘Talking with Families’) was, however, significantly correlated with total EI ($r = 0.20$, $n = 100$, $P = 0.05$). An interesting finding in Year 2 students was that the summed peer ratings for the 2 semesters (obtained when students rate each others’ contributions to PBL groups) was positively and significantly correlated with both EI and the Appraisal of Emotions subscale, with the correlations being numerically identical ($r = 0.23$, $P = 0.03$, $n = 91$), suggesting that students who score high on EI, and those who are good at reading the emotions of others, are perceived by their peers to be more effective in these groups.
CONCLUSIONS

In this study, overall academic performance was found not to be associated with either EI or empathy in any of the year groups. There was an interesting finding for peer ratings in Year 2 PBL groups, showing that students who scored higher on EI and on the Appraisal of Emotion subcomponent tended to receive higher peer ratings. An interesting possibility for future study would involve examining how mean levels of EI, and also of personality traits, affect actual functioning of PBL groups. There is some evidence that mean group EI is positively related with performance in small-group problem-solving tasks, but there have been no specific studies of medical student PBL groups in this context.

The cross-sectional and longitudinal analyses of empathy scores suggest different empathy trajectories for males and females, a result not found in studies with US cohorts. The results suggest that, for UK undergraduate medical students, a decline in empathy among females early in the course might be a cause for concern, but the early course experience seems to facilitate an enhancement of empathy in males. Further work would be required to establish the mechanisms underlying these results. One possible explanation, suggested by the convergence of male and female scores to a similar value as the course progresses, is as apparent in Fig. 1, is that students use the information acquired as they progress through the course to adjust their levels of empathy towards a perceived norm for effective functioning as a doctor. High empathy scorers may pick up from their course the message that they need to moderate their response to distress in others in order to be able to act effectively, whereas low scorers perhaps receive a different message about the need to take the patient’s perspective into account more than they were initially inclined to do.

That students were self-selected for this study represents a limitation of the study because students were required to fill in the questionnaire and to allow access to their examination results. This is a source of potential bias in the results as, for example, it is possible that the more able and/or confident students would be more likely to participate. The participation rate in the present study was 46%, which is good for a voluntary questionnaire survey, but in future studies it would be desirable to explore means of encouraging higher participation rates.

The limitations associated with the use of self-report scales mean that it is impossible to determine from the present findings whether the reported empathy changes are truly internalised by students, or whether it is just the questionnaire responses that are adjusted. Interviewing selected students would clarify this point and would also allow us to engage with the intriguing idea that what is perceived as the norm for the level of doctor empathy might be shifted upwards or downwards by changing the ways in which issues around empathy and emotional skills are covered in the curriculum.

The present study was limited by being restricted to medical students at a single UK university and by the limitations of participant self-selection and self-report data, as discussed above. Undertaking further studies at other UK universities would be of interest in order to establish if the longitudinal and cross-sectional effects we observed are robust, or if there is a ‘course’ effect on empathy trajectories. It would also be of interest to study the impact of EI combined with other relevant variables, such as personality, on effective performance in non-traditional forms of study such as PBL.

Contributors: EA identified the questionnaires and supervised the analysis. PL identified the areas of the curriculum and the students. BM and KO’H distributed the questionnaires and completed the initial analysis. All authors contributed to writing the paper.

Acknowledgements: none.

Funding: none.

Conflicts of interest: none.

Ethical approval: not required.

REFERENCES


Received 23 April 2006; editorial comments to authors 9 August 2006; accepted for publication 2 March 2007.
The role of empathy in establishing rapport in the consultation: a new model

Tim Norfolk,1 Kamal Birdi2 & Deirdre Walsh2

CONTEXT Considerable research has been conducted recently into the notion of patient-centred consulting. The primary goal of this approach is to establish a clear understanding of the patient’s perspective on his or her problem, and to allow this understanding to inform both the explanation and planning stages of the consultation. The quality of this understanding is largely determined by the empathic accuracy achieved by the doctor; the primary benefit is a therapeutic rapport between doctor and patient.

METHODS To highlight the role of empathy and communication skills in establishing rapport, we initially developed a model which seeks to draw the various motivational and skill elements identified in separate research papers into a comprehensive model of the journey towards shared understanding between doctor and patient. We then conducted an initial validation of the model via qualitative analysis involving general practitioners (GPs) and clinical psychologists.

RESULTS The validation offered encouraging support for the principal elements of the model. Specific suggestions for clarification and extension were then incorporated in a revised model.

CONCLUSIONS The model appears to capture the dynamic process of establishing a therapeutic relationship (rapport) between doctor and patient, defined by the quality of the doctor’s understanding of the patient’s perspective on his or her problem. Arguably, the most important contribution of the model is to highlight the fact that ‘empathy’ and consequent ‘rapport’ are not mystical or exclusive concepts but, rather, involve the use of specific skills accessible at some level by all.

KEYWORDS humans; *patient-centred care; *empathy; communication; physician–patient relations; clinical competence/*standards; motivation.

Medical Education 2007: 41: 690–697
doi:10.1111/j.1365-2923.2007.02789.x

INTRODUCTION

The quality of doctor–patient communication remains central to the effectiveness of the medical consultation, both in terms of immediate patient satisfaction and longer-term health outcomes.1,2 In this context, analysis of effective communication in the consultation has increasingly been focused on the patient-centredness of the encounter, where the patient’s perspective is specifically addressed alongside the presented symptoms.3,4 The central goal here is a professional rapport between doctor and patient, a therapeutic alliance based on trust and co-operation5 and established through a shared understanding of the patient’s perspective.

The strength of the rapport between doctor and patient is largely determined by the quality of the empathy at the heart of it, a notion this paper seeks to reinforce. The concept of empathy has been the subject of continuous debate for many years within the medical literature and beyond. The original use of the term referred to the cognitive skill of ‘interpersonal imagination’,6 used to establish an accurate understanding of the thoughts or words of the other. Since then, apart from occasional suggestions that it might be an exclusively affective or emotional response to another person,7 empathy has generally been seen as a multidimensional quality...

1Department of Psychology, City University, London, UK
2Institute of Work Psychology, University of Sheffield, Sheffield, UK
Correspondence: Tim Norfolk, 2 Brook House, Church Street, Bradwell, Derbyshire S33 9BJ, UK.
Tel: 00 44 1433 623919; Fax: 00 44 1433 623919;
E-mail: tim.norfolk@zen.co.uk

© Blackwell Publishing Ltd 2007. MEDICAL EDUCATION 2007; 41: 690–697
Overview

What is already known on this subject

Patient-centred consulting involves establishing a shared understanding of the patient perspective. Empathy plays a key role and encompasses both the desire and the ability to understand.

What this study adds

This study demonstrates a more comprehensive model of the journey towards shared understanding. It identifies the specific interaction between a doctor’s empathy and communication skills in developing rapport. It validates the model developed through exposure to experienced practitioners.

Suggestions for further research

Future research might focus on experimental validation of the model in the training environment and analysis of its impact on patient outcomes.

potentially informed by either or both affective and cognitive factors,\(^8\) but also involving a behavioural component (i.e. empathy demonstrated through behaviour).\(^9\)

The primary need for a doctor’s perceptions to be accurate, if shared understanding is to be reached, highlights the importance of cognitive empathy – which should itself be distinguished from sympathy, as the latter is characterised by a more generalised or globally supportive assessment of another person’s problem (‘feeling sorry’ for someone).\(^{10}\)

Whether the doctor’s empathy has an emotional dimension to it or not, however, the key is to maintain sufficient clarity of mind in search of empathic understanding of the actual thoughts and feelings of the patient.\(^{11,12}\)

This paper builds on specific recent attempts to develop a more practical understanding of the factors involved in relationship-building. This is epitomised by a recent review of the relevant empirical literature, which generates a summary model of 5 aspects of patient-centredness and 5 factors influencing its development.\(^{13}\) The review suggested, however, that identification of the ‘more complex and contextual dimensions of patient-centredness’ required a different approach.\(^{13}\) We have therefore chosen to direct the focus of this discussion more closely onto the active heart of the model, defined as ‘doctor behaviour’, looking at the specific motivation and skills required by doctors to establish the therapeutic relationship at the heart of patient-centred medicine.

Other recent approaches have involved looking specifically at verbal ‘opportunities’ for empathic engagement, which patients offer and doctors either respond to or miss,\(^{14}\) or to suggest specific verbal triggers that help develop empathic understanding.\(^{15}\) What this paper argues is that the verbal and non-verbal skills involved in relationship-building are interpersonal communication skills rather than empathic skills, because the latter are internal diagnostic skills: the doctor is picking up important signals from the patient and interpreting them, much in the same way as he or she internally assesses the patient’s presenting symptoms. This internal empathic journey, taken by the doctor alongside any clinical assessment, is then facilitated and reinforced by the use of communication skills (e.g. open question style, checking patient agreement).

This paper therefore takes the research forward in 3 distinct ways:

1 the proposed model (Fig. 1) sets out to integrate past research and models looking at the nature and power of empathic accuracy in helping to ensure that a patient’s experience of his or her illness is sufficiently explored;\(^{8,14}\)

2 the paper seeks to clarify the specific roles played by the doctor’s empathic motivation and skills, alongside verbal and non-verbal communication skills, in ensuring empathic understanding – and hence therapeutic rapport – and

3 the model is then tested, via applied research with relevant practitioners.

The specific objective of generating the model was to better inform the development of the relevant skills among practitioners.

Development of new model

The main components of the model emerged from a variety of sources – in part from the wider psychology literature and a body of related medical research (as indicated above), but also from more recent validated...
research into the competencies determining effective consulting in general practice. This highlighted the important distinction between empathy (as an internal process) and communication skills (as used in open expression between individuals).

Briefly, the model suggests that the essential therapeutic rapport between doctor and patient is derived from a particular interaction between the doctor’s empathic motivation, empathic skills and communication skills, potentially constrained by a range of other specific factors.

**Key stages in the new model**

**Empathic motivation**

The journey towards empathic understanding initially involves volition, the conscious or spontaneous commitment to engage fully with the patient’s individual experience. Having the initial desire to understand the patient’s perspective, in other words, is fundamental to achieving the goal.

There appear to be both innate (spontaneous) and consciously chosen (deliberate) sources of this motivation to listen or attend fully to a patient. Two particular personal attributes seem to trigger more spontaneous attention to the patient as a unique individual: innate interest or curiosity (i.e. being open to new possibilities) and innate warmth (i.e. being inclined to care for others). The curiosity about people is an intellectual trigger, linked to cognitive empathy; the warmth towards people is an emotional trigger, linked to affective empathy. Alternatively, if or when a doctor has low levels of both spontaneous triggers to this motivation (i.e. intellectual curiosity and affective warmth), a more conscious or deliberate focus on the patient can be generated. This is in effect drawn from established values rather than personality – representing a professional commitment to the patient – and is specifically intended to reach a sufficient understanding of the patient’s experience of his or her illness.

**Empathic skills**

The second element in the journey to empathic understanding highlights the important distinction between the desire to understand and the ability to understand. Motivation has no necessary association with ability: the former reflects an attitude of mind; the latter reflects skill, and the specific skills are internal diagnostic skills running parallel to those used to assess the patient’s clinical presentation, which allow a doctor to first identify significant clues to the patient’s thoughts and feelings, and then make constructive sense of what has been identified.

---

**Figure 1** Developing therapeutic rapport in the consultation (via empathic understanding of the patient’s underlying ideas, concerns and expectations)

© Blackwell Publishing Ltd 2007. MEDICAL EDUCATION 2007; 41: 690–697
Communication skills

A doctor might have strong empathic motivation and skills but, unless the patient can see or sense this, the consultation might in theory be no less dysfunctional than a meeting conducted by a doctor with little or no interest or skill in identifying the patient’s perspective. This is self-evidently because if the doctor is without the communication skills to articulate or demonstrate either empathic interest or, later, empathic skill, the patient has no reason to believe that the doctor wants either to listen or understand. This fundamental link has been recognised since Rogers’ early formulation of the empathic process, yet the clear distinction between empathy and communication skills is seldom made in either consultation models or assessment tools.

Communication skills perform two specific roles in helping establish a therapeutic relationship in the consultation. Firstly, they act as eliciting skills, encouraging patient disclosure. This involves both verbal skills (e.g., appropriate use of open questions, reflecting or echoing patient words, clarifying and summarising) and non-verbal skills (e.g., warmth of voice, appropriate use of silence, smiling, nodding, mirroring of posture). Secondly, these skills will determine the doctor’s empathic accuracy through testing or checking how well he has read the patient’s verbal and non-verbal behaviour in terms of the clues it offers to the patient’s thoughts and feelings.

Empathic understanding (and its impact on establishing rapport)

This defines the effectiveness of the empathic journey, or the degree to which the doctor has managed to accurately identify the patient’s perspective. The greater the level of understanding reached, the stronger the rapport between doctor and patient. The degree of empathic understanding achieved will be related to the doctor’s use of specific skills in search of that understanding. Empathy and associated communication skills are in this sense being employed or activated in pursuit of a deliberate professional goal, providing a sufficient sensitivity to the ‘moment’ in an individual consultation that allows for the establishing of a therapeutic alliance or rapport between doctor and patient.

It is important to recognise that the rapport established between doctor and patient is not a static moment or outcome, but rather a dynamic, iterative process in which the doctor attempts to reach an increasingly accurate understanding of the patient’s thoughts, feelings and expectations. The strength of the rapport can therefore fluctuate, and the pace at which it develops will often vary between patients.

Possible constraints on the development of therapeutic rapport

There are clearly various factors that will potentially influence the rapport established between doctor and patient. Outlined above are specific motivational and skill factors central to the process itself, but the quality of every interpersonal encounter is in part determined by momentary or stable characteristics of both the individuals involved and the particular environment in which they meet.

Doctor factors would potentially include, among other things, other personality characteristics (in addition to previously mentioned levels of innate curiosity and warmth), professional confidence, self-awareness, present mood and the impact of past experiences – both professional and personal.

Patient factors would potentially include, among other things, personality characteristics, self-awareness, health beliefs, present mood, communication skills and the level of complexity attached to the presented problem (which might, for instance, refer to an openly acknowledged complexity in the presented problem, or complex psychological elements in the patient’s perception of the problem).

Environmental factors would potentially include, among other things, stable aspects of the practice setting (e.g., waiting room and consulting room size, layout and lighting), available resources, time of day or week and number of patients waiting.

Content validation

To assess the conceptual or ‘content’ validity of the model, and in particular the role of empathy in the development of rapport, the researchers (occupational psychologists) needed to measure theory against the realities of a practice environment. They therefore conducted a separate piece of qualitative research across two populations for whom the therapeutic relationship is of primary practical importance: general practice and clinical psychology practitioners. The aim was to test the model, the appropriateness of its components and the
relationships between them, and then to revise the model accordingly.

METHODS

Twelve experienced practitioners, identified through opportunity sampling, took part in the study: 6 GPs (with an average of 16 years in practice) and 6 clinical psychologists (with an average of 19 years in practice). The client bases of both were comparable in that neither set of practitioners dealt solely with patients or clients with chronic problems.

Rather than start with direct analysis of the model and the role of empathy in generating rapport, each practitioner first took part in a 90-minute semi-structured interview based on the critical incident technique, in which he or she was initially asked to recall and describe separate specific patient sessions in which the development of rapport had:

1 gone well, and
2 proved difficult.

Practitioners were then asked to define and describe the concept of ‘rapport’, and the relationship between empathy and rapport. Finally, they were asked for their professional opinions on the validity of the model. (Inviting general reflection on the establishing of rapport allowed researchers to identify spontaneously the perceived role played by empathy before more specific prompting about its role.)

An average of 4 patient sessions were discussed with each participant, resulting in a total of 48 analysed sessions, which were audiotaped with full consent and later transcribed.

Analysis

Qualitative data for each section of the interviews were initially integrated for the purposes of analysis. Given that an existing conceptual rapport framework was being addressed and a significant amount of qualitative data had been collected, template analysis was chosen as a suitably systematic method of analysis. This defines and codes a priori a number of themes reflecting key components of the research topic. Five themes salient to the model were identified and used for coding transcripts: empathic motivation; empathic skills; communication skills; empathic understanding, and model specifics (sequencing and terminology).

RESULTS

Empathic motivation

Motivation was implicit in all participants’ descriptions of rapport-building. Its inclusion in the model as a discrete concept was viewed as a useful distinction by the GPs. More specifically, the idea that it was important to be aware of one’s natural motivational sources (and responses to this) throughout the session, whether driven primarily by curiosity or warmth, was found to be helpful by the GPs. The role of ‘professional dedication’ as a motivational source was seen by both groups as comprising many factors, including respect and integrity.

The influence of expectations and role clarity became apparent in the experiences of both groups, particularly those of the psychologists. ‘Expectations’ in this sense are what each individual brings to the interview: the client’s expectations, for instance, are based on previous experience of similar interviews, and beliefs about his or her health and the role that he or she plays in shaping it; the practitioner’s expectations are often based on previous experience of patient groups (e.g. related to age, gender or other demographics) or patterns of clinical presentations.

The individual’s particular expectations, whether client or practitioner, were seen to shape the motivation and initial dialogue in terms of information provided and how this was subsequently dealt with. It was felt therefore that reference to ‘expectations’ should be included in the model.

Empathic skills

Two levels of empathic skill emerged from the data, related to picking up clues and building perceptions from these clues, respectively. These were very much intertwined in practitioner descriptions of the development of rapport, but it was felt to be useful – certainly for training purposes – to distinguish clearly between them so that each step could be identified and modified where necessary.

All participants made reference to the notion of ‘attending’, either explicitly or implicitly. This was seen as being distinct from empathic skills. It did not involve picking up clues, but rather referred to the concentration on and interest shown in the patient or client, and served as encouragement to the patient to share his or her story. It was suggested that this be more explicitly recognised in the model.
**Communication skills**

The purpose and content of communication skills within the model were accepted as valid by all participants. Checking the accuracy of empathic perceptions was significantly apparent across both groups. The majority suggested that despite regularly using such empathic and communication skills, they had not previously been able to conceptualise them with such clarity themselves.

**Empathic understanding (and therapeutic rapport)**

Therapeutic rapport was defined by both psychologists and GPs as a shared understanding or connection between practitioner and patient or client, much in line with the model. Empathy was viewed as a key aspect in the development of rapport by all participants.

**Model specifics**

**Sequencing**

Overall the sequencing of the behaviours was felt to be appropriate, provided the iterative process of developing rapport was highlighted via more feedback loops, particularly from empathic understanding to empathic skill, so that some level of mutual understanding would then facilitate the gathering of further or more in-depth information.

**Terminology**

All the psychologists were very familiar with the language used. The majority of GPs initially found the terminology slightly unnatural but, once familiar with it, found it represented clear and useful language in the context of training and practice purposes.

**DISCUSSION**

There was strong evidence for the content validity of the model across and within both populations. All concepts contained in the model were thought to be valid and relevant, and its value as a potential training tool was commended. Few amendments or additions were suggested, and the majority of these related to issues of terminology or the expansion of existing concepts. Despite some specific differences in perception between the GPs and psychologists, findings suggested the model had a high degree of potential for generalisability as a model of medical relationship-building.

In response to the findings of the content validation study, a revised model was generated (Fig. 2),

---

**Figure 2** Developing therapeutic rapport in the consultation (via an empathic search for understanding of the patient’s dominant thoughts, feelings and expectations)
involving a number of clarifications to the initial model and the introduction of an additional stage to the process of developing therapeutic rapport (Empathic Attention).

Limitations

Reference should be made to certain limitations of this study. Firstly, most of the GPs had some knowledge of the model prior to the interview, because of their role in previous training activities that had either referred to or incorporated competencies used in the model. However, the only GP who was entirely unfamiliar with the competencies underpinning the model described the concepts and sequencing of the model very closely in reports of her experiences.

Further, no data were collected from the patient or client perspective as the content validity of the model was evaluated from the practitioner perspective alone. Inclusion of the patient or client perspective might have generated different conclusions.

CONCLUSIONS

We believe the revised model presents an accurate and transferable summary of the empathic journey towards therapeutic rapport. As the initial validation suggests, the model has immediate and ongoing practical relevance for selection and training in many areas of medicine, whether for GPs or other related practitioners (e.g. clinical psychologists, counsellors and nurses). The relationship-building skills at the heart of the model are certainly those targeted early in medical training and assessment and, because the emphasis is on demonstrating specific skills rather than achieving specific tasks, the model has obvious potential as a formative instrument that may help trainers and trainees to concentrate on developing specific core skills as well as measuring or ticking off the achievement of a list of tasks.

The authors have subsequently been involved in using the new model in training related to the development of rapport; the model certainly now requires a more rigorous experimental validation. Further research might look at the role and impact of the new model, both on other elements of the consultation (e.g. negotiation to establish a management plan) and on patient outcomes (e.g. satisfaction and adherence to treatment plans). The model’s generic potential might also be validated through the assessment of its impact on the performance of other related professional groups, such as nurses.

Contributors: TN conceived and designed the model. DW organised the acquisition and analysis of data. All authors contributed to interpretation of data, and the revision of the article. All authors approved the final manuscript.

Acknowledgements: the authors thank Dr Pat Lane (Director, Postgraduate General Practice Education, South Yorkshire and South Humber Deanery) for ongoing support, and Professor Fiona Patterson (City University) for ongoing advice.

Funding: this research was supported by the North Trent Workforce Confederation.

Conflicts of interest: none.

Ethical approval: this study was approved by Northern and Yorkshire Multi-Centre Research Ethics Committee.

REFERENCES


Received 30 August 2006; editorial comments to authors 18 December 2006; accepted for publication 7 February 2007
Core undergraduate psychiatry: what do non-specialists need to know?

Sam Wilson,1 John M Eagles,1 Julie E Platt1 & Hamish McKenzie2

OBJECTIVE The purpose of this study is to define the most relevant topics for inclusion in an undergraduate psychiatric curriculum by asking non-psychiatrists what knowledge, skills and attitudes related to psychiatry they need in their day-to-day practice.

METHODS A questionnaire study involving non-psychiatric doctors (based both in hospitals and general practice) was carried out using Delphi methodology in 2 waves. In the first wave, 408 doctors described the psychiatric competencies they required in their current posts. From this, a list of 101 psychiatric topics was generated. In the second wave, 867 doctors rated these topics according to the relevance of each topic to their practice.

RESULTS Depression, alcohol misuse and drug misuse were rated as most relevant. General practitioners found more topics relevant to their practice than did hospital doctors, and there were disparities in the relative importance that the 2 groups gave to topics.

CONCLUSIONS This study demonstrates a systematic method for developing core curricular undergraduate learning objectives in a specialty area by asking doctors outside that specialty to identify topics that are relevant to their practice. Similar methods could be used for a range of specialties other than psychiatry, and could provide a rational and transparent means of developing a core curriculum for medical students, when combined with perspectives from other sources.

KEYWORDS *education, medical, undergraduate; psychiatry/ *education; curriculum; clinical competence; Scotland; attitude of health personnel; family practice; medical staff; hospital.

Medical Education 2007: 41: 698–702
doi:10.1111/j.1365-2923.2007.02790.x

INTRODUCTION

The purpose of this study was to define topics for inclusion in the University of Aberdeen’s undergraduate curriculum, by asking doctors outside psychiatry what psychiatric knowledge, skills and attitudes (psychiatric competencies) are required in their day-to-day practice. Historically, the majority of doctors have received all their psychiatric training as undergraduates, with only 40% of general practitioners (GPs) gaining postgraduate experience.1 Apart from these, probably only a tiny minority of trainees in other specialties have any exposure to psychiatry after graduation, and there are no data regarding them. Changes in postgraduate education in the UK2 may increase the proportion of trainees who gain experience in psychiatry, but it will remain necessary for undergraduates to acquire a range of relevant psychiatric competencies.

The most recent edition of the General Medical Council document Tomorrow’s Doctors3 states: ‘The core curriculum must set out the essential knowledge, skills and attitudes students must have by the time they graduate.’ This reflected a widespread view that undergraduate curricula at the time the document was first published (1993) suffered from factual overload, but that it was possible to define a core curriculum containing essential elements that were generalisable to all medical practice. However, definitions of ‘core’ vary.1 In this study we are
primarily considering ‘core psychiatry’ not as the ‘essence’ of psychiatry from the viewpoint of a specialist, but in terms of the psychiatric competencies that are necessary for a broad range of medical practice. A number of curricula in psychiatry have been suggested, but they have been defined by experts with little reference to the needs of non-specialists for psychiatry competencies. Non-specialists have been asked to rate the importance of psychiatric topics in some studies, but the methods relied upon definitions of the original topics by psychiatrists, which may have adversely dominated the range of issues.

Delphi methodology has been used for analysing needs for curriculum development in a range of health care environments. Delphi studies canvass stakeholders in a series of waves. The first stages are open-ended and encourage participants to generate ideas. The latter stages ask participants to comment on the ideas that have been raised. The present study deployed Delphi methodology to establish the psychiatric competencies required by non-psychiatrists.

METHODS

Subjects excluded psychiatrists and comprised all GPs, consultants, junior doctors, clinical academics, staff grade and associate specialists working in the Grampian region, an area of Scotland with a mixed rural and urban population. They were contacted through details held by the University of Aberdeen Medical School and NHS Education for Scotland (the postgraduate deanery). Originally, there were thought to be 1502 doctors working in the region, but responses indicated that there were a number of duplicate entries in the data held by the medical school and the deanery and that many of the doctors listed had ceased practising or had moved out of the area. In the end, the best estimate of the number of possible participants was 1345. Response rates are based on this figure.

We developed a questionnaire for wave 1 of the Delphi process. It asked 3 questions:

- Have you required any psychiatric knowledge, skills or attitudes in the last 6 months? If so, please outline the nature of these.
- What psychiatric problems, if any, have had a significant impact on your patients in the past 6 months?
- What comments do you have about your own undergraduate psychiatry teaching? Was there anything which was not taught, but which would have been helpful for your chosen specialty or current post?

These were designed to be broad and open-ended, providing the opportunity for respondents to generate topics relevant to their current medical practice. The purpose of the study was described in a brief covering letter. Reminders were not sent to non-respondents.

There was considerable overlap in the topics recorded in the 408 responses to the wave 1 questionnaire. SW compiled a deliberately over-inclusive list of 157 items. SW and JME reduced these items by creating broader topics from narrower items, but no topics were removed even if, in the authors’ opinions, they were essentially unrelated to psychiatry. Following this, 101 topics remained, and these items formed the basis of the second questionnaire.

In wave 2 the same 1345 subjects were asked, in a forced-choice Yes/No format, whether each of the 101 topics were relevant to their current practice.
Only a few respondents took up the option of anonymity in the wave 2 questionnaire and these were colour-coded to permit broad identification of GPs, academics and grades of hospital doctors. A reminder was sent out to non-respondents (and anonymous respondents) after 1 month, along with another copy of the second questionnaire.

RESULTS

A total of 408 wave 1 questionnaires were returned, representing a response rate of 30%; 867 wave 2 questionnaires were returned, representing 64.5% of the population of doctors in the Grampian area. Table 1 shows the topics rated in order of the percentages of respondents who felt the topic was relevant to their current practice. Depression was rated as the psychiatric topic that most doctors (89.5%) considered relevant to their practice. The next most frequently rated were alcohol misuse and drug misuse, which 86.9% and 82.7% of respondents, respectively, felt were relevant to their practice.

General practitioners and hospital doctors exhibited differences in their ratings of the relevance of different topics to their practice. The highest-rated topics by both groups are also shown in Table 1. Hospital doctors rated fewer topics as relevant; the median-ranked topic for GPs was rated as relevant by 83.3% of respondents, whereas the median topic for hospital doctors was rated as relevant by only 37.9%.

DISCUSSION

This study provided the opportunity for a large group of non-specialists to generate possible topics for inclusion in the undergraduate curriculum and then for the same group to rate the relevance of those topics to their current practice. We suggest that this is a more logical and coherent manner in which to...

### Table 1 Percentage of doctors rating each topic as relevant to their clinical practice (n = 867)

<table>
<thead>
<tr>
<th>Overall rank</th>
<th>Percentage</th>
<th>Topic</th>
<th>Rank for GPs (n = 294)</th>
<th>Rank for hospital doctors (n = 573)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>89.5%</td>
<td>Depression</td>
<td>1 (99.5)</td>
<td>1 (84.3)</td>
</tr>
<tr>
<td>2</td>
<td>86.9%</td>
<td>Alcohol dependence and misuse</td>
<td>3 (99.0)</td>
<td>2 (80.6)</td>
</tr>
<tr>
<td>3</td>
<td>82.7%</td>
<td>Drug dependence and misuse</td>
<td>11 (95.6)</td>
<td>3 (76.1)</td>
</tr>
<tr>
<td>4</td>
<td>79.4%</td>
<td>Management of chronic pain</td>
<td>10 (95.9)</td>
<td>5 (70.9)</td>
</tr>
<tr>
<td>5</td>
<td>78.1%</td>
<td>Dementia</td>
<td>2 (99.3)</td>
<td>6 (67.2)</td>
</tr>
<tr>
<td>6</td>
<td>77.7%</td>
<td>Anxiety and panic disorders</td>
<td>4 (98.3)</td>
<td>7 (67.2)</td>
</tr>
<tr>
<td>7</td>
<td>75.2%</td>
<td>Dealing with death and dying</td>
<td>22 (91.8)</td>
<td>8 (66.7)</td>
</tr>
<tr>
<td>8</td>
<td>74.6%</td>
<td>Grief and bereavement</td>
<td>5 (98.0)</td>
<td>9 (62.7)</td>
</tr>
<tr>
<td>9</td>
<td>71.5%</td>
<td>Assessing ability to consent to, or refuse, medical and surgical treatment</td>
<td>79 (71.8)</td>
<td>4 (71.4)</td>
</tr>
<tr>
<td>10</td>
<td>69.9%</td>
<td>Psychological reactions to diagnosis of terminal or untreatable illness</td>
<td>28 (90.5)</td>
<td>12 (58.8)</td>
</tr>
<tr>
<td>11</td>
<td>69.7%</td>
<td>Suicide, deliberate self-harm, and assessing suicide risk</td>
<td>6 (96.9)</td>
<td>14 (55.7)</td>
</tr>
<tr>
<td>12</td>
<td>67.1%</td>
<td>Confidentiality</td>
<td>52 (83.0)</td>
<td>11 (59.0)</td>
</tr>
<tr>
<td>13</td>
<td>66.1%</td>
<td>Illness behaviour and sickness role</td>
<td>28 (90.5)</td>
<td>17 (53.6)</td>
</tr>
<tr>
<td>14</td>
<td>66.0%</td>
<td>Dealing with difficult relatives</td>
<td>56 (82.7)</td>
<td>13 (57.4)</td>
</tr>
<tr>
<td>15</td>
<td>64.4%</td>
<td>Simple counselling and brief therapy skills</td>
<td>18 (93.2)</td>
<td>23 (49.6)</td>
</tr>
<tr>
<td>16</td>
<td>64.0%</td>
<td>Learning disability or mental retardiation</td>
<td>48 (84.0)</td>
<td>16 (55.8)</td>
</tr>
<tr>
<td>17</td>
<td>63.8%</td>
<td>Choosing, starting and stopping antidepressants</td>
<td>8 (96.6)</td>
<td>29 (46.9)</td>
</tr>
<tr>
<td>18</td>
<td>63.6%</td>
<td>Assessment of cognitive function</td>
<td>20 (92.9)</td>
<td>25 (48.5)</td>
</tr>
<tr>
<td>19</td>
<td>63.6%</td>
<td>When to refer to psychiatrist</td>
<td>38 (88.4)</td>
<td>21 (50.8)</td>
</tr>
<tr>
<td>20</td>
<td>62.5%</td>
<td>Delirium, e.g. drug-induced, postop, 2 ° to trauma</td>
<td>80 (67.7)</td>
<td>16 (59.9)</td>
</tr>
<tr>
<td>21</td>
<td>62.5%</td>
<td>Empathic history taking and interviewing skills</td>
<td>58 (82.5)</td>
<td>19 (52.4)</td>
</tr>
<tr>
<td>22</td>
<td>62.2%</td>
<td>Chronic fatigue, myalgic encephalopathy (ME) or fibromyalgia</td>
<td>14 (95.2)</td>
<td>34 (45.2)</td>
</tr>
<tr>
<td>23</td>
<td>62.2%</td>
<td>Functional, somatic, or medically unexplained symptoms</td>
<td>35 (89.1)</td>
<td>26 (48.3)</td>
</tr>
<tr>
<td>24</td>
<td>61.4%</td>
<td>Presentation of psychiatric problems in old age</td>
<td>7 (96.9)</td>
<td>41 (43.1)</td>
</tr>
<tr>
<td>25</td>
<td>61.1%</td>
<td>Eating disorders</td>
<td>13 (95.6)</td>
<td>40 (43.5)</td>
</tr>
<tr>
<td>26</td>
<td>61.0%</td>
<td>Management of aggression and use of sedation</td>
<td>70 (75.9)</td>
<td>18 (53.4)</td>
</tr>
<tr>
<td>27</td>
<td>60.9%</td>
<td>Personality disorders</td>
<td>19 (92.9)</td>
<td>37 (44.5)</td>
</tr>
<tr>
<td>28</td>
<td>60.8%</td>
<td>Making sure the patient understands their diagnosis</td>
<td>57 (82.5)</td>
<td>22 (49.7)</td>
</tr>
<tr>
<td>29</td>
<td>60.7%</td>
<td>Schizophrenia and related illnesses</td>
<td>9 (96.3)</td>
<td>44 (42.4)</td>
</tr>
<tr>
<td>30</td>
<td>60.6%</td>
<td>Organic causes of psychiatric problems</td>
<td>46 (85.4)</td>
<td>27 (47.8)</td>
</tr>
</tbody>
</table>

Topics ranked from 31 to 101 are available from the authors.

GP = general practitioner

design curricular content objectively than the current specialty-centred approach.

**Delphi methodology in this study**

The Delphi technique draws on social research methodology to achieve consensus. It has been used for analysing curricular needs in a range of health care environments, such as in the cancer education requirements of GPs,\textsuperscript{12} in paediatric nursing education\textsuperscript{13} and in dermatology.\textsuperscript{14} The general strength of this methodology is that it surveys groups with an expert opinion, without the researchers seeking to influence the direction of the consensus, and gathers opinion without needing to consult respondents together, thus removing group-contaminating effects. It is transparent and, because it uses experts, has content validity. Within our study, we took the view that expert opinion resided in doctors practising in other specialties who had been undergraduate consumers of psychiatric education.

However, despite its current fashionable status, there are a number of criticisms of the methodology,\textsuperscript{15} including a lack of consistency within the spectrum of Delphi techniques. There must be at least 2 rounds of inquiry, but there is no clear consensus on how many more there should be. Focus groups may or may not be involved, and there is no consistent method for reporting the results. Finally, reaching consensus does not ensure that the conclusions are necessarily reliable and valid.

This large study defines core curricular topics useful to the non-specialist by using a rational, transparent and reproducible method. Other studies have used Delphi methodology to develop curricular goals in narrow areas, but this is the first study to do so for a major specialty with a large number of respondents. Nevertheless, there are some limitations to the study. Because of concerns regarding the Data Protection Act, it was not possible to access the regional database of junior doctors and thus, in wave 1 and the first round of wave 2, this group did not receive personalised questionnaires and its response rate was low. Because of this, individual departments were approached for names of trainees; once juniors received personalised reminders, their response rate improved. The low response rate to the wave 1 questionnaire was probably caused by a number of reasons, the chief being that the first questionnaire was designed to be open-ended and did not give examples or model answers so that respondents would not be led by the authors’ ideas. Nevertheless, both our questionnaires incorporated factors associated with increased response rates in other studies.\textsuperscript{10,17}

**Developing the undergraduate curriculum in psychiatry**

This study reports perceived needs for psychiatric competency, whether met or unmet. It does not look at unrecognised needs for psychiatric competency, that is, deficiencies in respondents’ practice of which they are unaware. For example, a doctor who is unaware of his or her lack of appropriate knowledge, skills or attitudes about depression cannot report this. Thus it is possible that areas of practice in which psychiatric competencies may be required may go unrecognised in this study design. It is notable that most of the reported competencies were in the knowledge domain, with few topics related to skills or attitudes. However, this method is only a single part of the overall process of curriculum development. Taking multiple views from many parties, including specialists and patients, will tell us more about the competencies non-specialists need from a specialty, whether or not these have been recognised or achieved.

Several areas, rated relevant by large numbers of doctors, are not specifically psychiatric (for instance, chronic pain, dealing with death and dying, bereavement and assessing the ability to give consent) but do have a major psychiatric or psychological component. This suggests psychiatrists should further extend their activities in collaborative teaching with other specialists. More integrated psychiatric teaching may have benefits in promoting students’ holistic views of medicine.\textsuperscript{6}

Clear differences in perceived needs emerged between GPs and hospital doctors. Although there were some similarities, unsurprisingly hospital doctors identified issues likely to present within hospitals, such as consent and bereavement, whereas GPs were more concerned with those likely to present in primary care. This raises questions about the purpose and scope of a core curriculum. Non-psychiatrists will vary in their need for psychiatric competency, depending upon the area in which they work. Whereas some of those junior doctors who go on to become GPs currently receive additional psychiatric training as postgraduates, and their numbers may increase as a result of the Modernising Medical Careers initiative,\textsuperscript{2} this is less likely for most other specialties. We should perhaps bear in mind the rather narrower needs of hospital doctors when designing the core curriculum, and
allow some of the areas highlighted by GPs to be taught during postgraduate training. Generic training on some of the non-psychiatric issues, such as dealing with death and dying, and gaining informed consent, could well fit into the Foundation Programme curriculum.

We present the methodology of this study as a transparent, rational means of developing core competencies for undergraduate medicine based on the needs of non-specialists, and as one that can be used for specialties other than psychiatry. Although the number of participants was large in this study, medical schools seeking to develop their core curricula in this manner could probably use fewer participants, provided the representation of practitioners was adequate. In fact, were they to choose to sample local doctors on multiple occasions for multiple specialties, smaller numbers would be desirable to avoid questionnaire fatigue. The method could be extended to encompass the views of specialist practitioners, patients and other professional groups within medicine and health care.

In conclusion, we believe this study shows it is possible to develop topics for inclusion in the core undergraduate curriculum on a rational and transparent basis. The perspectives of specialists, patients and other groups should also be taken into account, and these may be assessed in a similar manner.

**Contributors:** SW devised the study, contributed to data analysis and drafted the initial manuscript. JME advised on the study design and contributed to data analysis. JEP contributed to data analysis. HM advised on the study design. All authors contributed towards revisions of the manuscript.

**Acknowledgements:** the authors would like to thank Angela Girling and Lana Hadden for secretarial support, and NHS Education for Scotland for distributing questionnaires to junior doctors.

**Funding:** this study was funded by the Committee for Undergraduate Teaching and Education, Royal Cornhill Hospital, Aberdeen. SW held a Clinical Teaching Fellowship funded by Additional Cost of Teaching monies (ACT) from NHS Grampian.

**Conflicts of interest:** none.

**Ethical approval:** not required.

**REFERENCES**


Received 31 August 2006; editorial comments to authors 4 December 2006, 8 January 2007; accepted for publication 24 January 2007.

A comparison between senior medical students’ and pharmacy pre-registrants’ knowledge and perceptions of generic medicines

Mohamed Azmi Hassali,1 David C M Kong2,3 & Kay Stewart2

OBJECTIVE To ascertain any differences in knowledge and perceptions of generic medicines between senior (final year) medical students and pharmacy pre-registrants in Australia.

METHODS National web-based surveys containing 16 common questions were administered to each group. Responses were compared using Mann–Whitney U-test.

RESULTS Responses were received from 400 medical students (response rate 26.7%) and 289 pharmacy pre-registrants (response rate 30.5%). Both groups scored poorly on the allowable bioequivalence limits when comparing a brand-name medicine with a generic medicine, with pharmacy pre-registrants scoring better ($P<0.001$). Pharmacy pre-registrants were more likely ($P<0.001$) to report having been introduced to bioequivalence during their courses, and less likely ($P<0.001$) to desire more information on bioequivalence testing. Both groups correctly agreed that generic medicines are bioequivalent to corresponding brand-name medicines but not necessarily to each other, and that generic medicines are cheaper. Pharmacy pre-registrants were more aware that generic medicines must be in the same dosage form ($P<0.001$) and contain the same dose ($P<0.001$) as the comparator brand-name medicine. Both groups incorrectly believed that generic medicines are required to meet higher safety standards, are inferior in quality, are less effective and produce more side-effects. Both groups agreed that it is easier to recognise a drug’s therapeutic class from its generic name and that they needed more information about the safety and efficacy of generic medicines, with medical students feeling more strongly about this ($P<0.001$).

CONCLUSIONS Although there were some differences in responses, both groups had knowledge deficits about the quality, safety and effectiveness of generic medicines, which need to be addressed by educators in order to increase the future use of generic medicines in Australia.

KEYWORDS comparative study [publication type]; multicentre study [publication type]; humans; male; female; adult; clinical competence/ *students, pharmacy/ *students, medical; pharmacology, clinical/ *students, medical; pharmacology, clinical/ *students, medical; pharmacology, clinical/ *students, medical; pharmacology, clinical/ *students, medical; therapeutic equivalency; Australia.

Medical Education 2007; 41: 703–710
doi:10.1111/j.1365-2923.2007.02791.x

INTRODUCTION

In the context of escalating health care costs, most western countries are enhancing rational pharmacotherapy to achieve best value for their constantly increasing expenditure on medicines.1,2 In Australia, increasing costs to the Pharmaceutical Benefits Scheme (PBS), a comprehensive system for the subsidy of prescription medicines covering the whole population, has resulted in the Commonwealth Government’s institution of a number of strategies to
Overview

What is already known on this subject

In Australia, the prescribing of generic medicines and generic substitution is still a contentious issue among practising prescribers and pharmacists, arguably as a result of a lack of understanding of the bioequivalence, safety and quality of generic medicines.

What this study adds

Although there were some differences in responses from senior medical students and pharmacy pre-registrants, both groups of future practitioners have knowledge deficits about the quality, safety and effectiveness of generic medicines. Respondents identified gaps in their undergraduate education in this regard. Poor understanding of these issues is likely to contribute to negative attitudes towards the use of generic medicines.

Suggestions for further research

Educational interventions with future practitioners on issues relating to generic medicines need to be developed, implemented and evaluated.

hold down rising pharmaceutical costs. These include attempts to increase the use of generic medicines by encouraging medical practitioners to prescribe generically. Since December 1994, the brand substitution policy has allowed pharmacists, with the consent of the patient, to supply a ‘generic equivalent’ in place of the medication prescribed, as long as the prescriber has not disallowed substitution by indicating on the prescription that only the named product should be supplied. Although these strategies have been implemented, the concepts of prescribing generic medicines and generic substitution are still contentious among both prescribers and pharmacists. The debate has centred on issues related to the bioequivalence, quality and safety of generic medicines. Previous studies have shown that changing existing prescribing behaviour is difficult. Similarly, to change the dispensing behaviour of some practising pharmacists will be challenging. To overcome this problem, education about the benefits of generic prescribing and generic substitution should be focused on medical and pharmacy students, the prescribers and dispensers of the future. To date, no studies have been conducted to assess and compare the knowledge and perceptions of senior medical students and recent pharmacy graduates with regard to generic medicines and generic substitution. An understanding of the perceptions held by future practitioners, such as senior (final year) medical students and pharmacy pre-registrants (graduates undertaking pre-registration training prior to becoming eligible to register to practise as pharmacists), is important because their knowledge and attitudes at this stage of their careers will influence their future behaviours with respect to the prescribing and substitution of generic medicines. This insight would enable any misconceptions held by them about generic medicines to be addressed by early education interventions. This paper compares data from 2 independent web-based surveys conducted among Australian senior medical students and pharmacy pre-registrants about their perceptions and knowledge of generic medicines, generic prescribing and generic substitution.

METHODS

Two web-based surveys were developed and used to gather data from senior medical students and pharmacy pre-registrants, respectively. Permission was obtained from the relevant universities and pharmacy pre-registration training course co-ordinators across Australia to survey their medical students and pharmacy pre-registrants. Information about the web-based surveys was sent to potential respondents via their e-mail addresses, with help from their respective programme co-ordinators. Both surveys remained online for a period of 3 months and, in order to increase response rates, 4 follow-up reminder e-mails were sent to programme co-ordinators every 3 weeks asking them to remind their students or pre-registrants to participate. The initial survey items were developed using information from review of the literature and after e-mail consultation with several lecturers in medical and pharmacy schools across Australia. Both surveys were tested for face and content validity. The final surveys each comprised 20 items divided into 4 sections. The first part consisted of 4 demographic questions, about age, gender, university attended and other qualifications. The second part contained 5 items about knowledge of the bioequivalence of generic medicines, including 1 question asking for the allowable regulatory bioequivalence limits when
comparing generic medicines with brand-name medicines and 4 questions framed for responses on a 5-point Likert scale (1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree and 5 = strongly disagree). The third part of the survey contained 8 items evaluating understanding of brand-name medicines versus generic medicines, also using the 5-point Likert scale. The fourth part of the questionnaire consisted of 3 items to evaluate perceptions of current medical and pharmacy education. Responses were collected using the MySQL database, an open-source database which exists on the webserver, and exported to Microsoft Office EXCEL 2003 for Windows for data cleaning. The cleaned data were then exported to SPSS Version 12 for statistical analysis.

Data analysis

Responses to the 16 identical items on the 2 web-based surveys were merged for comparative statistical analysis in SPSS. The non-parametric Mann–Whitney U-test was used to compare the ordinal scale responses. This test was primarily chosen because the collected data did not meet the assumptions necessary for a parametric test, such as an independent samples t-test, in which the data collected must be in either interval or ratio scale and should be obtained from normally distributed populations. Statistical significance was set at a P-value of 0.05, based on a 2-tailed test.

RESULTS

During the 3-month period from 18 June 2004 to 18 September 2004, a total of 400 senior medical students responded to the survey. The response rate, based on 1497 students enrolled during the study period, was 26.7%. Demographic characteristics of respondents from each university are shown in Table 1.

The average age of the medical student respondents was 25.3 ± 3.6 years and 220 (55%) of them were female. A total of 234 (58.5%) respondents had entered medical programmes directly from high school, whereas 166 (41.5%) had entered medical school with a bachelor or higher degree. The first or higher degree qualifications held by the students were mainly in science-related subjects (87.4%).

By the end of the 3-month data collection period, 289 Australian university-qualified pharmacy pre-registrants had responded to the survey. The response rate, based on 948 Australian university-qualified pre-registrants enrolled in training courses during the study period, was 30.5%. Response by university is more relevant than response by pre-registration training course, as some questions referred to respondents’ undergraduate education. It was not possible to calculate response rates per university, as numbers were obtained per pre-registration training course, all of which are open to graduates from any university. Numbers of respondents by university of graduation are shown in Table 2.

The average age of the pharmacy respondents was 23.0 ± 3.0 years. A total of 194 (67.1%) respondents were female. Almost all the respondents (94.1%; n = 272) had no previous degree qualifications. The remainder held first degree qualifications in science-related fields.

<table>
<thead>
<tr>
<th>University</th>
<th>Respondents n</th>
<th>Median age (years) ± SD</th>
<th>Gender</th>
<th>Entry qualification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45</td>
<td>23.5 ± 1.1</td>
<td>M = 16</td>
<td>DE = 42</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>24.6 ± 3.0</td>
<td>F = 29</td>
<td>FDHD = 5</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>23.4 ± 2.8</td>
<td>M = 24</td>
<td>FDHD = 19</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>28.8 ± 8.5</td>
<td>F = 28</td>
<td>DE = 33</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>24.9 ± 5.0</td>
<td>M = 27</td>
<td>FDHD = 7</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>29.0 ± 3.9</td>
<td>F = 24</td>
<td>DE = 0</td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td>26.8 ± 3.1</td>
<td>M = 11</td>
<td>FDHD = 51</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>23.8 ± 0.9</td>
<td>M = 17</td>
<td>FDHD = 7</td>
</tr>
<tr>
<td>9</td>
<td>51</td>
<td>24.1 ± 4.2</td>
<td>F = 22</td>
<td>DE = 0</td>
</tr>
<tr>
<td>10</td>
<td>28</td>
<td>23.9 ± 1.3</td>
<td>M = 26</td>
<td>FDHD = 4</td>
</tr>
<tr>
<td>Total (mean)</td>
<td>400</td>
<td>25.3 ± 3.6</td>
<td>M = 180</td>
<td>FDHD = 166</td>
</tr>
</tbody>
</table>

* DE refers to students who directly entered medical school upon finishing their secondary education; FDHD refers to students who entered medical school after completing a bachelor or higher university degree

SD = standard deviation; M = male; F = female; DE = direct entry; FDHD = first and higher degrees

© Blackwell Publishing Ltd 2007. MEDICAL EDUCATION 2007; 41: 703–710
Respondents were asked to select the correct allowed bioequivalence limits when comparing a generic medicine with a brand-name medicine. In order to ensure common understanding of the concept of bioequivalence, the following statement was provided with the question:

‘In pharmacology, the term “bioavailability” refers to the rate (how fast) and the extent (how much) to which an active ingredient is absorbed and becomes available at the site of the drug action. The Therapeutic Goods Administration (TGA), which is the drug regulatory body involved in registering medicines in Australia, considers a generic product to be bioequivalent if its bioavailability is within an allowed range compared with the currently marketed brand product.’

After this statement, the following question was asked:

‘The regulatory limits applied are that the 90% confidence intervals for the ratios (generic product : brand-name product) of the areas under the plasma drug concentration versus time curves and the maximum plasma drug concentrations must fall between: ...’

Six answer options were given. The correct answer was 80–125%.

Responses are shown in Table 3.

The Mann–Whitney U-test showed a statistically significant difference between the 2 groups ($P < 0.001$). A total of 256 (64.0%) medical students but only 36 (12.5%) pharmacy students did not select a numerical option. Of those who did select a range, only 2.1% of medical students and 13.0% of pharmacy pre-registrants selected the correct answer.

Responses to other questions about bioequivalence are shown in Table 4.

Both groups agreed that products classified as ‘generic equivalents’ under the PBS are equivalent

---

**Table 2** Responses from pre-registrants according to university

<table>
<thead>
<tr>
<th>University</th>
<th>Pre-registrants (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>289</td>
</tr>
</tbody>
</table>

**Table 3** Knowledge of allowed bioequivalence limits

<table>
<thead>
<tr>
<th>Response</th>
<th>Senior medical students</th>
<th>Pharmacy pre-registrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n) (%)</td>
<td>(n) (%)</td>
</tr>
<tr>
<td>80–120%</td>
<td>10 (2.5)</td>
<td>37 (12.8)</td>
</tr>
<tr>
<td>80–125%</td>
<td>3 (0.8)</td>
<td>33 (11.4)</td>
</tr>
<tr>
<td>90–100%</td>
<td>38 (9.5)</td>
<td>11 (2.8)</td>
</tr>
<tr>
<td>95–100%</td>
<td>16 (4.0)</td>
<td>23 (8.0)</td>
</tr>
<tr>
<td>95–105%</td>
<td>77 (19.3)</td>
<td>149 (51.6)</td>
</tr>
<tr>
<td>Not answered</td>
<td>256 (64.0)</td>
<td>36 (12.5)</td>
</tr>
<tr>
<td>Total</td>
<td>400 (100.0)</td>
<td>289 (100.0)</td>
</tr>
</tbody>
</table>

**Table 4** Knowledge and perceptions about generic equivalence

<table>
<thead>
<tr>
<th>Survey question/statement</th>
<th>Group</th>
<th>Response (%)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>All generic products of a particular medicine rated as ‘generic equivalents’ are therapeutically equivalent to the brand-name product</td>
<td>Medical students</td>
<td>16.5 69.0 4.5 9.0 1.0</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>Pharmacy pre-registrants</td>
<td>22.8 63.3 6.2 6.2 1.4</td>
<td></td>
</tr>
<tr>
<td>All generic products of a particular medicine rated as ‘generic equivalents’ are therapeutically equivalent to each other</td>
<td>Medical students</td>
<td>1.0 12.3 5.8 65.8 15.3</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>Pharmacy pre-registrants</td>
<td>3.5 14.5 12.5 50.5 19.0</td>
<td></td>
</tr>
<tr>
<td>I have not been introduced to the issues of bioequivalence for generic drugs during my pharmacy/medical education</td>
<td>Medical students</td>
<td>30.0 38.3 11.8 17.3 2.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Pharmacy pre-registrants</td>
<td>0.7 3.9 2.8 48.1 42.6</td>
<td></td>
</tr>
<tr>
<td>I need more information on how bioequivalence tests are conducted for generic medicines</td>
<td>Medical students</td>
<td>18.3 53.5 18.5 8.5 1.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Pharmacy pre-registrants</td>
<td>10.4 42.9 21.8 16.6 8.3</td>
<td></td>
</tr>
</tbody>
</table>

* Mann-Whitney U-test

SA = strongly agree; A = agree; N = neutral; DS = disagree; SD = strongly disagree
to the brand-name product, but not necessarily to each other. Pharmacy pre-registrants were more likely ($P < 0.001$) than medical students to report having been introduced to issues of bioequivalence during their courses, and less likely ($P < 0.001$) to desire more information on the conduct of bioequivalence tests.

Responses to questions about the quality, safety and efficacy of generic medicines versus brand-name medicines are shown in Table 5.

A higher proportion of pharmacy pre-registrants than medical students understood that, under the PBS requirements, a generic medicine must be in the same dosage form ($P < 0.001$) and contain the same dose ($P < 0.001$) as the comparator brand-name product. Both groups ($\geq 90\%$ of respondents in each group) incorrectly believed that generic medicines are inferior in quality, less effective and produce more side-effects than brand-name medicines, with significant differences detected between groups for each of these aspects. In each case, pharmacy pre-registrants showed stronger agreement than medical students: quality, $P = 0.016$; efficacy, $P = 0.002$; side-effects, $P < 0.001$. Both groups also wrongly believed that brand-name medicines are required to meet higher safety standards than generic medicines.

Responses to questions evaluating perceptions of current medical and pharmacy education with regard to generic medicines are shown in Table 6.

Both groups largely agreed that they found it easier to recognise the therapeutic class of a drug from its generic name. The majority of both groups felt they

<table>
<thead>
<tr>
<th>Survey question/statement</th>
<th>Group</th>
<th>Response (%)</th>
<th>$A$</th>
<th>$N$</th>
<th>$DS$</th>
<th>$SD$</th>
<th>$P$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A generic medicine is bioequivalent to a brand-name medicine</td>
<td>Medical students</td>
<td>19.8</td>
<td>67.5</td>
<td>4.5</td>
<td>7.5</td>
<td>0.8</td>
<td>0.062</td>
</tr>
<tr>
<td>A generic medicine must be in the same dosage form (e.g. tablet, capsule) as the brand-name medicine as the brand-name medicine</td>
<td>Pharmacy pre-registrants</td>
<td>36.0</td>
<td>48.1</td>
<td>3.1</td>
<td>10.0</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>A generic medicine must contain the same dose to brand-name drugs</td>
<td>Medical students</td>
<td>11.8</td>
<td>42.8</td>
<td>11.0</td>
<td>31.5</td>
<td>3.0</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Generic medicines are of inferior quality</td>
<td>Pharmacy pre-registrants</td>
<td>45.3</td>
<td>44.6</td>
<td>2.4</td>
<td>4.8</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>Generic medicines are less effective than brand-name medicines</td>
<td>Medical students</td>
<td>37.0</td>
<td>56.5</td>
<td>5.0</td>
<td>3.0</td>
<td>0.3</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Generic medicines produce more side-effects than brand-name medicines</td>
<td>Pharmacy pre-registrants</td>
<td>47.4</td>
<td>42.2</td>
<td>7.3</td>
<td>3.1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Generic medicines are less expensive than brand-name medicines</td>
<td>Medical students</td>
<td>37.8</td>
<td>57.0</td>
<td>4.8</td>
<td>0.3</td>
<td>0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>Generic medicines are required to meet higher safety standards than generic medicines</td>
<td>Pharmacy pre-registrants</td>
<td>43.9</td>
<td>47.4</td>
<td>4.2</td>
<td>3.5</td>
<td>1.0</td>
<td>0.710</td>
</tr>
</tbody>
</table>

* Mann-Whitney U-test
$	ext{SA} =$ strongly agree; $A =$ agree; $N =$ neutral; $DS =$ disagree; $SD =$ strongly disagree

<table>
<thead>
<tr>
<th>Survey question/statement</th>
<th>Group</th>
<th>Response (%)</th>
<th>$A$</th>
<th>$N$</th>
<th>$DS$</th>
<th>$SD$</th>
<th>$P$-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I need more information on the issues pertaining to the safety and efficacy of generic medicines</td>
<td>Medical students</td>
<td>10.0</td>
<td>56.8</td>
<td>18.3</td>
<td>13.3</td>
<td>1.8</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>I find it easier to recall a medicine’s therapeutic class using generic names rather than brand names</td>
<td>Pharmacy pre-registrants</td>
<td>45.8</td>
<td>39.5</td>
<td>10.3</td>
<td>5.8</td>
<td>0.8</td>
<td>0.339</td>
</tr>
<tr>
<td>I have been taught how medicines are subsidised in the Pharmaceutical Benefits Scheme (PBS)</td>
<td>Medical students</td>
<td>33.2</td>
<td>40.1</td>
<td>17.0</td>
<td>8.0</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Pharmacy pre-registrants</td>
<td>31.5</td>
<td>58.1</td>
<td>4.5</td>
<td>3.8</td>
<td>2.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Mann-Whitney U-test
$	ext{SA} =$ strongly agree; $A =$ agree; $N =$ neutral; $DS =$ disagree; $SD =$ strongly disagree
needed more information on the issues pertaining to the safety and efficacy of generic medicines, with medical students feeling more strongly about this than pharmacy pre-registrants (P < 0.001). Pharmacy pre-registrants were more definite than medical students (P < 0.001) about having been taught how medicines are subsidised under the PBS.

**DISCUSSION**

One important responsibility of both medical and pharmacy educators is to teach future practitioners about the cost-effective use of medicines.\(^1\) In the context of escalating costs of pharmaceutical expenditure, the use of cheaper generic medicines has been encouraged by most policy makers around the world, including those in Australia.\(^3\) This action has been supported by policy makers as reducing the cost of pharmaceutical expenditure, not only to the health care system but also to consumers.\(^3\) To increase the rate of generic medicine use, it is vital that medical practitioners and pharmacists have a sound knowledge of the issues surrounding generic prescribing and substitution. Future medical and pharmacy practitioners need to be exposed to these issues early in their education so that they can be confident in prescribing and substituting generic medicines whenever applicable.

The majority of future practitioners in both groups were unable to select the correct bioequivalence limits allowed by the TGA for approval of generic medicines. That a higher proportion of medical students did not attempt to select a range may be explained by differences in the professional curricula; for example, medical students are less likely to have been exposed to the topic of biopharmaceutics, in which bioequivalence will have been taught in pharmacy education.\(^3\) Although it was expected that this topic would have been well covered in pharmacy curricula, only a small proportion (11.4%) of the pharmacy pre-registrants answered correctly. This low rate of correct response may reflect the non-exposure of pharmacy students to official bioequivalence limits, or it may demonstrate a lack of understanding of the complex concepts of bioequivalence testing.

Under the current Australian medicine regulatory guidelines, a medicine that been rated as a ‘generic equivalent’ will be bioequivalent or therapeutically equivalent to the brand-name product,\(^4\) but not all the ‘generic equivalents’ are therapeutically or bioequivalent to each other.\(^27\) In the Schedule of Pharmaceutical Benefits (SPB), when more than 1 generic version of a listed item is available, bioequivalent products are flagged with a superscript ‘a’ or ‘b’ to indicate which generic products are equivalent to each other.\(^28\) About 85% of respondents in each group agreed that a medicine rated as a ‘generic equivalent’ should be therapeutically equivalent to the comparator brand-name product. Similarly, 70–80% respondents in both groups were aware that not all ‘generic equivalents’ are therapeutically equivalent to each other. These results may be explained by possible similarities in the education received by both groups during their formal medical or pharmacy education or may result from familiarity with using the information provided in the SPB, although it is not unreasonable to think that pharmacy pre-registrants – particularly those in community pharmacies – would have had more experience with this than medical students.

Differences in the professional curricula might have affected the future practitioners’ responses to the questions about education on bioequivalence and bioequivalence testing. Statistically significant differences were found between groups in response to both these questions (Table 4). More than 68% of senior medical students indicated that they had not been introduced to the issues of generic medicine bioequivalence during their medical education, whereas under 7% of pharmacy pre-registrants made this claim. Although a significantly higher proportion of pharmacy pre-registrants than medical students said they did not need more information on the conduct of bioequivalence tests, more than half the students in each group indicated that they did need more information. This was also reported by Nandita and Sudip,\(^26\) but it may reflect a general tendency to be willing to accept more information when it is offered.

Questions in the third part of the survey evaluated whether the future practitioners were aware of the similarities and differences between generic and brand-name medicines (Table 5). More than 80% of each group believed that a generic medicine was bioequivalent to the corresponding brand-name medicine. By contrast, 53.8% of the senior medical students did not think that a generic medicine was required to take the same form of dosage as the corresponding brand-name medicine, whereas only 12.8% of the pharmacy pre-registrants shared this view. In Australia, a generic medicine must be manufactured in a similar dosage form to the brand-name medicine.\(^4\) The lack of knowledge of senior medical students on this issue is probably the result of their receiving less education on drug formulation (a subject not taught at medical schools) compared with
pharmacists. Medical students were also much less aware than pharmacy pre-registrants that, to be classified as ‘generic’, a medicine must contain the same dose as the brand-name medicine. As noted earlier, subjects such as pharmaceutics are not taught in medical schools and this would be expected to have some influence on medical students’ answers to this question.

Most participants in each group perceived generic medicines to be inferior in quality, less effective, produce more side-effects and be less safe than brand-name medicines. Pharmacy pre-registrants held stronger views on these issues than medical students, with the exception of the safety issue, but it is clear that neither group had a clear understanding of how generic medicines are approved by the TGA. Despite their lower prices, generic prescription medicines in Australia are required to meet the same quality standards as brand-name medicines.\(^4\) Within this context, all medicines are subject to a rigorous, multistep review process that includes review of scientific data on ingredients, product quality and performance. More than 90\% of respondents in each group agreed that generic medicines cost less than brand-name medicines. This knowledge cannot be attributed to their professional education as the difference in pricing is generally understood by the broader community.

More than 50\% of both medical students and pharmacy pre-registrants said they would like more information about the safety and efficacy of generic medicines (Table 6). These responses clearly indicated a perception of deficiencies in current medical and pharmacy education, more so in medical curricula. The majority of medical students and pharmacy pre-registrants said they found it easier to identify a drug’s therapeutic class using generic names rather than brand names. This is probably a result of the extensive use of generic names in medical and pharmacy education and the naming conventions for drugs, which encourage similar names for drugs in a chemical class (e.g. use of the suffix ‘-pril’ for angiotensin-converting enzyme inhibitors). With regard to whether these future practitioners had been taught about how medicines are subsidised, the majority of the pharmacy pre-registrants reported that they had been taught about this, compared with less than half the senior medical students. This again can be attributed to differences in curricula. Given that pharmacists work closely with the PBS in the supply of medicines, it is appropriate that they understand the system well.

The main limitation of this study is the potential for non-response error. This is a disadvantage of web-based surveys, but a web-based survey was the most practical approach to reach the population of interest at the time the survey was performed because the medical students were involved in clinical attachments and the pharmacy pre-registrants were working full time. Reasons for the low response rate may include lack of interest or lack of time among participants. Another possibility is that students who did not regularly access their university e-mail accounts would not have been aware of the survey. All reasonable attempts were made to maximise the response rate; however, the constraints of Australian privacy legislation meant that we did not have access to the contact details of the survey population and we were therefore restricted to relying on intermediaries to pass on reminders to students. Similarly, these constraints prevented us from comparing respondents with non-respondents to gauge potential bias.

CONCLUSIONS

Although there were some differences in the responses received from these 2 groups of future practitioners, overall both groups demonstrated knowledge deficits, especially on specific issues relating to the safety and effectiveness of generic medicines. These are areas that need to be addressed by both medical and pharmacy educators. For medicines to be used as effectively as possible requires them to be available at an affordable cost to both consumers and society. It might be beneficial if regulatory authorities such as the TGA were to be involved in planning education interventions for these groups of future practitioners to convince them of the value of generic medicines.

**Contributors:** at the time of this study, MAH was a PhD candidate at the Department of Pharmacy Practice, Victorian College of Pharmacy, Monash University. MAH undertook the study as a PhD candidate under the supervision of DCMK and KS. All authors made substantial contributions to the conception and design of the study and to the interpretation of data. MAH undertook data collection and analysis, and wrote the first draft of the paper. All authors contributed to the critical revisions of the paper and approved the final version.

**Acknowledgements:** none.

**Funding:** none.

**Conflicts of interest:** none.

**Ethical approval:** approval for the survey of medical students was granted by the Standing Committee on Ethics.
REFERENCES


Received 29 April 2006; editorial comments to authors 18 August 2006; accepted for publication 26 January 2007
Electronic discussion forums in medical ethics education: the impact of didactic guidelines and netiquette

Herman Buelens,1 Nicole Totté,1 Ann Deketelaere2 & Kris Dierickx3

OBJECTIVES The quality of computer-mediated asynchronous group discussion in educational settings often has potential for improvement. To this end, 2 categories of recommendations can be found in the literature. Some authors advise providing didactic guidelines to students. Others recommend informing students about the basics of (formal) network etiquette (netiquette). This paper investigates whether the type of guideline provided has an effect on either or both the quality of asynchronous group discussion and the appraisal of participants in the context of a course in medical ethics.

METHODS A total of 112 students graduating in the biomedical sciences were randomly assigned to electronic discussion groups of 18–19 participants each. Three treatment groups were created by varying the guidelines presented to participants upon entering the discussion forum. Guidelines contained trivial information, netiquette advice, or a combination of didactic guidelines and netiquette advice.

RESULTS Both the quality of discussion and the participants’ appraisal of the discussion were highest in the third group. However, contrary to what is widely suggested, it was observed that exclusively supplying netiquette guidelines had no impact on discussion quality.

CONCLUSIONS To improve the quality of asynchronous group discussion, it is advisable to provide not only netiquette rules, but also didactic guidelines. Future research should focus on the effects of netiquette guidelines on students of different levels and skills and should allow for disentangling the effects of netiquette and didactic guidelines. It might also seek to establish methods of disseminating guidelines that enhance their effect.

KEYWORDS randomised controlled trial [publication type]; *ethics, medical; Internet/ *ethics/ standards; practice guidelines/ *standards; professional practice/ standards; ANOVA.

Medical Education 2007: 41: 711–717
doi:10.1111/j.1365-2923.2007.02793.x

INTRODUCTION

Contemporary progress in medical knowledge and technology create a number of ethical challenges for health care workers and medical researchers.1 Advances in the fields of genetic and genomic research, genetic testing and manipulation, birth control and reproductive medicine, intensive care and the prolonging of life all confront health care workers and medical researchers with imperative ethical considerations in their daily practice. Not only must health care workers be sensitive to the ethical, legal and social implications that come with medical and technological advances, they also must be able to confidently discuss ethical issues in depth with patients, and critically contribute to the social debate on ethical issues.

Although medical education has responded to this evolution by incorporating medical ethics into the curriculum,2–7 the challenge remains to provide
Overview

What is already known on this subject

There is some evidence that didactic guidelines improve the quality of asynchronous discussion in education. It is also suggested that poor quality writing in discussion forums improves when participants are informed about ‘network etiquette’ (netiquette).

What this study adds

Contrary to previous recommendations, providing netiquette guidelines alone did not improve the quality of online discussions. However, supplying netiquette in combination with didactic guidelines was effective.

Suggestions for further research

Future research should focus on the effect of netiquette guidelines for students of different levels and skills, should disentangle the effects of netiquette and didactic guidelines, and should seek methods that enhance the effectiveness of guidelines.

students with active, meaningful learning opportunities without sacrificing theory and content. In particular, large class sizes make it difficult to include face-to-face training facilities that allow students to practise the intellectual and communicative skills required for discussing ethical issues.8,9

Looking for a solution, educational technologists have proposed the use of electronic discussion forums.9–11 Because the number of participants per forum can be limited, all students have the opportunity to actively participate. The asynchronous format of a discussion forum allows students to read and contribute when convenient. Furthermore, by writing down what they think, learners’ thoughts become visible, and thus subject to intervention and coaching. Therefore, electronic discussion forums might contribute to a shift from the passive learning of the traditional, large-class lecture-based format to small-group, student-oriented, active education.

Examples of this shift have been reported by several instructors in the context of medical education. In a class of nursing students, online discussion promoted greater participation and feedback during assignments.12 Involvement in asynchronous discussion was significantly correlated with higher mean grades in a final physiology examination for medical students.13 Lipman et al.14 observed an improvement in student knowledge in a course on clinical ethics. Weir et al.15 reported similar findings in a course on air pollution and health. Kamin et al.16 mentioned higher participation rates and more developed comments in online discussion compared with face-to-face debate during an evidence-based medicine course.

It thus appears that there is every reason to promote the use of asynchronous discussion forums in an educational context. However, as Guzdial and Turns17 have noted, there are many ways in which providing a discussion forum in an educational context might fail to reach its goal. For instance, students may simply not participate, they may start writing about something irrelevant, or they may present their opinions in an offensive way. Hence, many authors have composed lists of didactic recommendations to improve the effectiveness of electronic forums.18–20 Furthermore, the websites of universities and colleges that use e-learning platforms often include didactic tips for adequate use of forums, advising students that they should articulate messages as accurately as possible, underpin statements with sound arguments, stimulate debate and metacommunication, and compare personal contributions with those of peers. It is clear that more research is needed to unravel the effects of the instructional components and interventions that contribute to high-quality electronic discussion.21 However, in the context of a course on global environmental change, Sorensen22 presented some evidence that the quality of online dialogue can be improved by adhering to didactic guidelines that encourage reflection and metacommunication.

It has also been suggested that the probability of quality interaction increases when participants know and follow network etiquette rules (or netiquette).18,23,24 Netiquette refers to the dos and don’ts of online communication, and embraces both common courtesy and the informal rules of cyberspace. It is assumed that overlooking netiquette conventions can disrupt an ongoing discussion. For example, blaming someone for making a linguistic error can shift the focus of the discussion from the message to the messenger. In order to avoid such distractions, Ekeblad25 has suggested that participants in electronic forums should be informed of common netiquette rules. Although netiquette
rules seem to be face-valid, the effect of the actual implementation of these guidelines on the quality of discussions has not been documented within the context of an educational setting.

In this paper 2 principles of good practice concerning the use of electronic discussion forums are put to the test in the ecologically valid context of a medical ethics course. With the goal of creating clarity concerning the effect of both netiquette guidelines and didactic guidelines, 3 groups were created. Prior to participating in a discussion forum, students were presented with:

1. trivial instructions (control group);
2. netiquette guidelines (netiquette group), or
3. both netiquette and didactic guidelines (netiquette and didactic group).

Based on the above, 3 hypotheses were suggested.

**Hypothesis 1**

Discussion forums where netiquette guidelines are provided will be more effective than discussion forums where only trivial instructions are provided, in terms of both measurable quality indicators (hypothesis 1a) and participant appraisal (hypothesis 1b).

**Hypothesis 2**

Discussion forums where both didactic guidelines and netiquette guidelines are provided will be more effective than discussion forums where only netiquette guidelines are provided, in terms of both measurable quality indicators (hypothesis 2a) and participant appraisal (hypothesis 2b).

**METHODS**

**Subjects and design**

Following an introductory lecture (‘Ethics and Law in Biomedical Research’), 112 graduating students of biomedical sciences at the Katholieke Universiteit Leuven (Belgium) were invited by the instructor to participate in an online discussion concerning 1 of the main topics of the course. All students were randomly assigned to 1 of 6 discussion forums (18 or 19 students per forum). In each of the discussion forums, students were presented with the following statement: ‘Sometimes, the reproductive cloning of human beings can be ethically justified.’ Although contributors could be identified, participation was voluntary and no marks were awarded for contributing. All discussion forums remained open for 3 weeks and could only be accessed by their assigned students and by the instructor. After this period, all discussion forums were closed and were no longer accessible to students.

Three treatment groups were created by varying the guidelines presented to participants upon entry to the discussion forum. In the control group, it was indicated that ‘the instructor would refer to some of the contributions during the next lecture’. In the netiquette group, netiquette advice was presented in addition to the information given to the control group. Students were recommended ‘to phrase their postings politely, to pay attention to spelling, grammar, tone and style and to refrain from posting offensive contributions’. Finally, students were urged ‘to have a second look at the linguistic and formal correctness of their contribution before posting it’. In the netiquette and didactic group, a set of didactic guidelines was added to the guidelines given to the netiquette group. The didactic part of the guidelines emphasised that ‘ethics is not just a matter of personal opinion’. Students were encouraged ‘to build strong reasoning, to bring in good arguments and to stress the (causal) relationships between these elements’. Furthermore, interaction was facilitated by asking students ‘to first read the contributions of others, to ask for clarifications or to respond to the writings of others, thereby starting by summarising or paraphrasing the gist of the initial posting’.

The guidelines were presented to students each time they entered the discussion forum. The 6 discussion forums were randomly allocated to 1 of the treatment groups (Table 1). The resulting design has 1 factor with 3 levels and 2 discussion forums nested within each level.

**Dependent measures**

One week after the discussion forums were closed, all students were invited to complete an online questionnaire with 2 sections. In the first section, students were asked to indicate:

1. how many times per week they visited the discussion forum, and
2. how many times per week they actually read what was posted.

In the second section of the questionnaire, the participants’ appraisal of the discussion was assessed by asking for levels of agreement to 13 items, such as
I found the discussion forum interesting and the discussion forum gave me new insights.

Besides collecting questionnaire data, an attempt was made to map out the intrinsic quality of the individual postings. Treatment-blind hard copies of the transcripts were made, and all contributions were independently analysed by all 4 authors according to 2 existing models of content analysis for discussion groups.26,27 Complete postings were used as the unit of analysis.28 As it turned out, interrater agreement was poor. Moreover, this problem was not resolved when transcripts were re-analysed according to a simplified (ad hoc constructed) version of the original classification systems.

The raters’ disagreement strongly contradicted the course instructor’s confidence that he (a co-author and an expert in medical ethics) could easily distinguish between postings that contributed significantly to discussions and those that did not, which he did by selecting contributions that were discussed during the course. In order to identify relevant characteristics of quality contribution, a set of 6 postings (1 from each forum) regarded as meaningful by the course instructor was compiled. The set was examined in detail by all 4 authors together, and it was agreed that quality contributions were characterised by 3 parameters. Firstly, interaction was evoked by posing questions (e.g. ‘What do you mean by ...?’; ‘Do I understand you correctly if ...?’; ‘Does everybody agree with ...?’). Secondly, good discussions were characterised by the bringing in and building of arguments (e.g. ‘Because...’, ‘You said that ... but contrasting your opinion with that of...’, ‘Referring to...’. ‘As was demonstrated by...’, ‘Confronting this reasoning with...’). Thirdly, there was a relative absence of statements that were neither preceded nor followed by an argument underpinning the statement. Such ungrounded statements are incompatible with conjunctive adverbs such as ‘thus’, ‘therefore’, ‘consequently’ and ‘hence’, as these adverbs refer to a preceding argument. Often, ungrounded statements start off with: ‘It is said that...’, ‘It is believed...’ or ‘In my opinion...’.

To become familiar with the process of counting questions, arguments and ungrounded statements, all 4 authors took part in a targeted training session. Next, all postings were randomly allocated to 2 researchers who independently counted the total number of:

1 questions;
2 arguments, and
3 ungrounded statements.

Again, researchers were blind to treatment groups. Interrater agreement was very satisfactory and subsequent discussions among the 4 authors resolved the few remaining inconsistencies. In addition to these 3 quality indicators, the number of sentences per contribution was counted.

RESULTS

Quality indicators (hypotheses 1a and 2a)

Overall, 80 of 122 students (72%) posted at least 1 message. In the control group as well as in the netiquette group, 25 of 38 students (66%) contributed to the discussion. In the netiquette and didactic group, the number of contributing participants was higher, with 30 of 36 students (83%) commenting on the discussion. The maximum number of contributions made by 1 student was 4. No differences in contribution frequency were detected between the treatment groups. No one used the forum for anything other than the topic invited for discussion and it seemed that overall, many contributions were of good quality.

A 1-way, between-group analysis of variance (ANOVA) with discussion forums (A–F) nested within treatment groups as the independent variable and the
number of sentences per contribution as the dependent variable revealed a main effect of treatment group \((F_{2,3} = 10.4, P = 0.045)\). The mean number of sentences per contribution was 6.1, 7.2 and 8.6 in the control, netiquette, and netiquette and didactic groups, respectively.

The length of contributions (measured by the number of sentences) correlated significantly \((r > 0.50, P < 0.001)\) with all 3 quality indicators (the number of questions, ungrounded statements and arguments). Hence, the number of sentences per contribution was used as a covariate in the 3 analyses of covariance (ANCOVA) that looked for differences in discussion quality between the treatment groups. In a first ANCOVA, discussion forums \((A–F)\) nested within treatment groups were defined as the independent variable and the number of questions was defined as the dependent variable. In a second and third ANCOVA, the number of arguments and the number of ungrounded statements were defined as the dependent variables. Preceding each analysis, it was ensured that there were no violations of the assumptions of normality, linearity, homogeneity of variances and homogeneity of regression slopes. After adjusting for the number of sentences per individual contribution, the only significant difference between treatment groups concerned the number of ungrounded statements \((F_{2,3} = 15, P = 0.027)\). The estimated mean number of ungrounded statements per treatment group (controlling for the covariate) were 3.1, 3.3 and 2.1 in the control, netiquette, and netiquette and didactic groups, respectively. Contrary to what was expected (hypothesis 1a), the a priori contrast between the control group and the netiquette group was not significant. In line with hypothesis 2a, an a priori contrast indicates that the number of ungrounded statements was significantly lower in the netiquette and didactic group compared with the netiquette group. In addition, the a priori contrast between the control group and the netiquette and didactic group indicates that the number of ungrounded statements was lowest in the latter group.

**Student appraisal (hypotheses 1b and 2b)**

A total of 87 of the 122 students completed the questionnaire. Eighty students stated that they had consulted the discussion forum more than twice a week. Half the respondents said that during consultation they had ‘always’ read the contributions; the other half said that they had done so ‘sometimes’.

In order to simplify the dataset, all 13 items were subjected to a principal component analysis and a subsequent scale construction procedure. The resulting 8-item scale was labelled ‘subjective appraisal’ and had a reliability coefficient of 0.87. The minimum and the maximum scale values were \(-1\) (indicating not valuing the discussion) and \(+1\) (strong appraisal), respectively.

A 1-way between-groups analysis of variance (ANOVA) with discussion forums \((A–F)\) nested within treatment groups was conducted to explore differences in subjective appraisal. The analysis revealed a statistical difference in scale score for the 3 treatment groups \((F_{2,3} = 11.3, P = 0.04)\). Mean scores for the control group, the netiquette group, and the netiquette and didactic group were 0.34, 0.29 and 0.50, respectively. Contrary to hypothesis 1b, an a priori contrast revealed no significant difference between the control group and the netiquette group. In line with hypothesis 1b, an a priori contrast indicated that participant appraisal was significantly higher in the netiquette and didactic group than in the netiquette group. Moreover, the a priori contrast between the control group and the netiquette and didactic group indicated that participant appraisal was significantly higher in the latter group.

**DISCUSSION**

The present study addresses the issue of improving the quality of asynchronous group discussions in an educational context. Within the study, good discussion is characterised by the absence of ungrounded statements and by the presence of arguments and questions. In particular, an investigation of a supposedly positive effect of providing netiquette and didactic guidelines on the quality of contributions was conducted within the context of a course in medical ethics for graduating students. The main finding is that contrary to previous recommendations, providing netiquette guidelines alone is not effective. Compared with a control condition, providing netiquette guidelines has no impact on the number of questions, arguments and ungrounded statements made by students. However, supplying netiquette guidelines in combination with didactic guidelines had a substantial effect on the quality of asynchronous group discussion. Postings were longer and a significant decrease in ungrounded statements was observed. Moreover, participants’ appraisal of meaningful discussions correlated with a relative lack of ungrounded
statements, indicating that ‘good discussion’ is indeed characterised by a relative absence of ungrounded statements.

Although the findings from our study are straightforward, there are some limitations that both impose restraints on its generalisability and suggest avenues for future research.

Firstly, the graduating students within this study all knew each other well, were highly motivated to learn about medical ethics, were highly qualified, and possessed outstanding writing skills. The non-anonymity between students and their qualifications and skills may all have contributed to the overall high quality of postings. Moreover, these student characteristics may have interfered with treatment conditions. It is possible that writing skills were already in line with netiquette rules, even before students’ attention was focused on the existence of such rules. As a consequence, providing netiquette guidelines alone will not be effective. Therefore, future research should focus on the effect of providing netiquette guidelines for various groups of students of different levels and skills.

Secondly, the unifactorial design of our study does not allow for judging the effect of didactic guidelines alone. Therefore, it would be worthwhile conducting a future study using a 2 × 2 factorial design that allows for disentangling the effect of netiquette and didactic guidelines (no guidelines/netiquette alone/didactic guidelines alone/both netiquette and didactic guidelines).

Thirdly, it is our feeling that the impact of guidelines could be enhanced by making them more obvious. One way of doing so might be by not ‘providing’ guidelines, but by opening discussion forums with a contribution that is completely in line with the guideline it stands for. Thus, future research might look for the effect of prototypical examples.

Finally, this study was carried out in a single institution and involved a single course run by a single instructor. Therefore, we invite instructors to falsify our findings within their own contexts. By engaging with discipline-based research in teaching and learning, we can optimise our teaching beyond trial and error. By communicating and disseminating our findings, by grounding statements with data from our own research, we all contribute to ‘good discussion’ on education.

Contributors: all authors contributed to the design of the study and were actively involved in the implementation, analysis and write-up.

Acknowledgements: the authors express their sincere appreciation to Kevin Chin (McGill University) for linguistic advice and for sharing his stimulating and invaluable thoughts. We also explicitly want to thank 3 anonymous reviewers for their detailed and constructive comments on an earlier version of the manuscript.

Funding: none.

Conflicts of interest: none.

Ethical approval: none.

REFERENCES

13 Taradi SK, Taradi M. Expanding the traditional physiology class with asynchronous online discussions and collaborative projects. Adv Physiol Educ 2004;28:73–8.

Received 27 June 2006; editorial comments to authors 10 October 2006; accepted for publication 30 January 2007
Identifying the aspects of rural exposure that influence medical students to take up rural practice

Geetha Ranmuthugala, John Humphreys, Barbara Solarsh, John Wakerman, James A Dunbar & Geoff Solarsh

Editor – Identifying the factors that influence students to take up rural practice is crucial to ensuring the success and continuing funding of strategies designed to address rural medical workforce shortages. In Australia, strategies to encourage the uptake of rural practice include increasing the intake of students from rural backgrounds, providing students with greater exposure to rural medicine through rural placements, and an increased rural focus in medical curricula. Although evidence testifies to the success of programmes that selectively admit students from rural backgrounds and those who intend to practise as family practitioners (both independent predictors of practice location), the specific contribution of ‘rural exposure’ to increasing uptake of rural practice is less clear, largely as a result of the lack of empirical evidence controlling for such confounding variables.

In 2006, a pilot study of Monash University medical students in Years 1–4 was undertaken to explore the contribution of rural exposure to student intentions to take up rural practice. Students were surveyed before and after rural placements, and logistic regression methods were used to regress student intention to take up rural practice against demographic characteristics, stated career intention and rural placement activities. The greatest impact of the placement in terms of changing future practice intentions occurred in Year 1 students, suggesting the importance of early rural exposure. Among these Year 1 students \( (n = 105) \), those with rural backgrounds (odds ratio \([\text{OR}]\) 24.1, 95% confidence interval \([\text{CI}]\) 5.7–93.5) and women (OR 4.7, 95% CI 1.1–19.6) were significantly more likely to state intention of rural practice, but, when adjusted for these characteristics, there was no demonstrable association between particular placement activities and increased intention of rural practice. Possible explanations for this may include the early stage of training, the cross-sectional nature of the study, and inadequate power to detect an association.

Rural exposure is complex and varies in content and delivery. The costs of rural placements to medical schools and governments are significant. In order to ensure the success of rural clinical schools in encouraging the uptake of rural practice, it is necessary to identify the aspects of rural exposure (in terms of nature, timing, duration and frequency) that contribute to students developing a positive attitude towards rural practice and thereby influencing their return to rural areas. Given the many influences on ultimate practice location, there is a need for comprehensive longitudinal tracking studies of trainee doctors to identify the specific aspects of rural exposure that are most effective in influencing the uptake of rural careers.

REFERENCES


3 Rabinowitz HK, Diamond JJ, Markham FW, Paynter N. Critical factors for designing programmes to increase the supply and retention of rural primary care physicians. JAMA 2001;286:1041–8.


Response to Blake & Gusella on the impact of simulation on people who act as simulated patients

L Bokken & J-J Rethans

Editor – In response to the letter by Blake and Gusella we would like to make a few comments. In general we agree with the points made by the authors. However, we think that the issues raised depend on the type of simulated patient (SP) performance; for instance performances for educational purposes, performances with or without feedback, performances of prestructured roles (‘just play the script’) or more personalised roles, where indeed individual past experiences may influence negative effects.

In our paper, Bokken et al. (2006) we describe the mild negative effects SPs experience due to performing a patient role. The authors suggest that the SPs we report on in our paper perform patient roles as a full time job. However, they do not. We do not have full time SPs. The majority of our simulations are simulations of surgery-hours appointments which last 20–25 min as a maximum. Also, in our focus groups we did not ask adolescents. We do have simulations with adolescents, but only once a year and indeed in these cases we always provide careful debriefing opportunities and very careful preparation of the cases and the adolescents. We are currently writing a research paper on our experiences with adolescents.

We agree with the authors that knowledge of SP recruitment and debriefing procedures is important in understanding negative effects in SPs due to their performance. In our SP programme SPs are carefully recruited, with attention to their medical history. Although there is no formal debriefing session after each role performance, we have reasons to believe our SPs do debrief after a performance. For example, in our paper SPs mentioned coming out of their roles by giving feedback to the students. By doing this SPs are able to wind up the consultation and put it behind them. In addition, SPs seem to debrief informally, by talking to other SPs and staff about their experiences. As a result the mild negative effects of performing a patient role are only of short duration, but still these effects exist and need to be addressed. Therefore, attention to the recruitment and debriefing of SPs are valuable additional measures to reduce the impact of portraying a patient role.

REFERENCES


Correspondence: Jan-Joost Rethans, MD, PhD, Skilllab, University of Maastricht, PO BOX 616, 6200 MD Maastricht, the Netherlands, Tel: +31.43.388.1790; Fax: +31.43.388.4127; E-mail: j.rethans@sk.unimaas.nl
doi: 10.1111/j.1365-2923.2007.02812.x