

TRAINING

How should paediatric examiners be trained?

N Khera, H Davies, H Davies, T Lissauer, D Skuse, R Wakeford, J Stroobant

Arch Dis Child 2005;90:43–47. doi: 10.1136/adc.2004.055103

Analysis of both the content and process of examinations is central to planning the appropriate education and training of examiners in paediatric clinical examinations. This paper discusses the case for developing training, reviews the current literature, and suggests the desirable attributes of examiners and the components of a training programme. Potential areas of further research are also considered.

an optional postgraduate examination for doctors in primary care or other medical specialties. There is a written paper (part 1A) and a clinical examination. In addition to paediatricians, examiners are general practitioners, child psychiatrists, and paediatric surgeons.

Examiners have historically been selected on the basis of clinical experience as a consultant (or general practitioner for DCH), personal recommendation by the Principal Regional Examiner, Paediatric Regional Adviser, or college tutors, or by self-proposal. Potential examiners submit a structured curriculum vitae, concentrating on their expertise in teaching and education. Selection is made by the Examinations Committee and ratified by the College Council, ensuring that examiners have sufficient clinical and teaching experience and that there is appropriate distribution by specialty, hospital, and region. Until recently, the post-selection training of examiners has consisted of written information about the nature and conduct of the examination, observation of a clinical examination, followed by examining for the first time paired with the senior examiner. Subsequent monitoring of performance has sometimes been by a “hawk: dove” index⁴ devised by the Royal College of Physicians but has been inconsistently fed back to examiners. Until October 2004 marking was by pairs of examiners. Although marks were awarded independently, this system allowed examiners to assess their own level of expectation of a candidate's performance against each other's marks. There is also a Senior Examiner Review of the marks when discrepant marks between examiners affecting a pass/fail decision are scrutinised and these may be amended. Feedback is given to examiners in these occasional cases. All candidates who fail are offered copies of all mark sheets and examiner comments.

More recently, two day examiner training conferences have been introduced to teach both the practicalities of the examination process, as well as generic skills including discrimination, the principles of assessment of clinical reasoning, decision making, and communication skills. All examiners for the new MRCPCH clinical examination have received specific training, including simulated or videoed examination encounters as a basis for discussion about conduct of the encounter and standard setting.

As assessment drives learning, analysis of the content and process of examinations will allow the development of more robust systems of assessment and is central to the training and professional development of tomorrow's paediatricians. In clinical examinations the conduct of the examination is dependent on examiners who are able to maintain set standards and they, therefore, need defined attributes and expectations, as well as training and monitoring of their performance as examiners.

The published literature relating to the training of examiners is limited. However the Royal College of General Practitioners,^{1,2} the Royal College of Surgeons, and the Resuscitation Council³ have developed programmes for selecting and/or training examiners and instructors. This paper arose from an initiative of the Royal College of Paediatrics and Child Health (RCPCH) to improve examiner training and describes the components of a new programme for training paediatric examiners in the UK and the desirable attributes of examiners. The paper draws on current literature relating to examinations, both undergraduate and postgraduate, to illustrate the principles of paediatric examiner training, although not all the contexts presented necessarily relate equally to its development.

THE STATUS QUO

The RCPCH assesses doctors' readiness to become middle grade paediatricians with the MRCPCH (Membership of the Royal College of Paediatrics and Child Health) examination. It consists of written papers in basic child health (part 1A), the scientific basis of child health (part 1B), and clinical problem solving (part 2 written). There is also a clinical examination (part 2 clinical), which, from October 2004, will consist of 10 stations, including communication and consultation skills, and clinical examination and child development. All apart from one station with clinical video clips will still rely on the assessment of clinical ability by examiners. The RCPCH also appoints and trains examiners for the Diploma in Child Health (DCH) which is

See end of article for authors' affiliations

Correspondence to:
Dr J Stroobant, Consultant
Paediatrician, University
Hospital Lewisham,
London, UK; stroo@
biscituk.biz

Accepted 15 August 2004

PROBLEMS WITH THE STATUS QUO

Necessary competencies

Assessment methods and practices that have been unchanged for many years need to be re-evaluated and refined. There is evidence that examiners' current knowledge of technical

aspects of assessment, including robust approaches to standard setting, are not yet optimal and require more development.⁵ The competencies of a well trained satisfactory examiner have not yet been adequately defined.

Examiner selection

The process needs to be developed to ensure that it is transparent, fair, and objective. This should be done building on existing models of good practice.

Concerns about reliability and validity

Agreement between examiners can be poor because of undue emphasis on examinee factors such as apparent confidence, appearance, and use of language, particularly in oral examinations.⁶ Decision making by examiners during oral examinations, for example, is frequently a process of some complexity, and includes initial impressions, hypothesis generation, and hypothesis testing.⁷ In order to achieve a highly consistent or reliable result from a multiple station clinical examination, about 10 hours of testing time per candidate is needed.⁸ Clearly this is not feasible, so strategies to increase reliability while maintaining implementation feasibility will be necessary.

Management of change

The nature, content, and design of clinical examinations are in evolution. Examiners need to be equipped with new, relevant skills which require ongoing monitoring.

Conduct of examiners

Examiners must approach the examination with a standardised set of competencies to apply in their assessment of candidates, and they must use these techniques in a consistent way. It is not up to them to introduce idiosyncratic practices of assessment and rating. Because of the potential magnitude of examiner variance many educators have expressed concerns about pass/fail decisions being made on the basis of clinical examinations.⁹ Candidates and examiners must have a clear definition of the clinical competencies required to be demonstrated and a clear understanding of what constitutes suitable evidence of these in order for the examination to be judged to be fair.¹⁰

Managing candidate diversity

A fair examination requires equal opportunity for candidates. Examiner awareness about a candidate's disabilities will assist in making reasonable accommodation in examinations where required.¹¹ While studies show no evidence of explicit examiner behaviour that might disadvantage students from ethnic minorities in comparison to white students in oral examinations, there is evidence that some ethnic minority candidates tend to use communicative styles that are poorly rated. This may influence the assessment of the candidates' communication skills.¹²

The cost of legal challenge

In an environment of increasing awareness of clinical governance and an expectation that those responsible for appraisal and assessment will be trained and accountable for their decisions, complaints or legal challenges relating to decisions are possible. In addition, candidates have access to data documenting their performance in examinations and this highlights the need for documentation of decisions to be rigorous and transparent.

Limited research

Research that underpins the process of the examination of clinical competence has been limited. There is a need for closer collaboration between research and practice.¹³

DOES TRAINING IMPROVE VALIDITY, RELIABILITY, OR FAIRNESS OF EXAMINATIONS?

Validity

Examiners need to be conversant with the concepts of content and construct validity.¹⁴ Content validity is a measure of the ability of the specific assessment systematically and representatively to assess the full range of activities it is designed to reflect. In other words, if the examination is supposed to evaluate potential clinical competence, does the test measure an adequate range of abilities from which to draw a conclusion that the candidate is indeed competent. Construct validity is the extent to which performance on a test fits into a theoretical scheme about the attribute the test seeks to measure. In clinical terms the assessment must measure skills that do, in fact, reflect the abilities theoretically characteristic of a competent clinician at this stage in their training. Although these issues are the responsibility of those designing the examination and are not the primary responsibility of examiners, a better understanding of such concepts is required for examiners to develop uniformity in assessment. Validity is a measure of how completely an assessment tool measures what it purports to. Table 1 presents the key features of the various type of validity.

Reliability

This is the degree to which an assessment method produces reproducible outcomes. The impact of examiner training on reliability remains to be clarified. The literature is sparse, and where available, is divided on this issue.¹⁶ In a study of undergraduate examiners, the most significant improvements of reliability were achieved by identifying and eliminating the most inconsistent examiners, and, surprisingly, training as such had no significant impact on reliability.¹⁷ In another study, the training of doctors as examiners provided no significant benefit in terms of reliability over the use of medical students and lay staff as examiners in objective structured clinical examinations (OSCEs).¹⁸ There is also a debate around whether long examinations are needed to achieve higher levels of reliability in assessment by OSCEs.¹⁹ Studies on oral examinations show that unless examiner behaviour is extreme, it can be changed by training.¹

Variability

This is the degree to which differences exist between the scores of different examiners. Variability is an important contributor to overall error in examinations. Previously, significant inter-examiner variability in scoring was adjudicated by the senior examiner and the Examinations Committee; in the restructured clinical examination to be introduced in October 2004, examiners will no longer assess candidates in pairs but will work alone to judge performance on a specific parameter of clinical competence. While this strategy will allow a more detailed assessment by increasing the components of the clinical examination, it is important to reduce the problem of inter-examiner variability while maintaining the validity of individual examiner input.²⁰ Influences on variability include differences in the training of examiners and their involvement in station construction.²¹ Another can be a lack of explicit rating criteria.²² In any situation there are several potential sources of error, all of which will contribute to the overall reliability of the assessment. These include in addition to inter-observer error, intra-observer error, and test-retest reliability, as well as complex interactions between examiners, candidates, and station. Examiner involvement and commitment are more associated with inter-examiner reliability in examining, than experience in clinical medicine or as an

Table 1 Validity

Type of validity	Key features
Face validity	How well does the test <i>appear</i> to measure what it is supposed to? May be important in terms of doctors' willingness to accept assessment tools. Doesn't actually mean that it does measure what it appears to.
Content validity	Refers to the concept that we can infer from a given task or series of tasks how well a doctor functions in a given domain.
Criterion validity	Refers to the correlation of a measure of attribute or skill of interest with some other measure of that attribute or skill, ideally a "gold standard" which has been used and accepted. Comparison may be contemporaneous (concurrent validity) or with a measure sometime in the future (predictive validity). Gold standard often not available for performance assessment.
Construct validity	Refers to how well a test measures attributes that are not directly observable and is also useful where there is no gold standard to measure an attribute. Establishing construct validity is an ongoing process as it often involves testing a number of constructs—for example, "based on my theory of construct X, people who score high on a test of X differ from people who score low in terms of attributes A, B, and C, where A, B, and C can be other instruments, behaviours, diagnoses and so on". ¹⁵
Consequential validity	Refers to the effect that assessment has on learning—while less well described than other types of validity it is highly relevant because it recognises the close link between assessment and learning and the need to use assessment strategically.

examiner.²³ Increasingly, therefore, a more sophisticated approach to the determination of reliability is being used, known as generalisability which allows all sources of error to be taken into account. In addition it is possible to take the analysis one step further and model the circumstances necessary to achieve a given level of reliability.²⁴

COMPETENCIES OF EXAMINERS AND THEIR SELECTION

Work is currently being undertaken by the RCPCH to define the competencies of an appropriately trained examiner. We suggest the desirable attributes of an examiner should include:

- Ability to use defined techniques to elicit the best performance from candidates
- Keep abreast of current developments and issues in the profession²⁵
- Have knowledge of the principles and practicalities of the examination
- Have an understanding of educational theory and practice in relation to assessment
- Be able to make consistent and unbiased judgements
- Have an understanding of reliability and validity
- Have the ability to make and justify pass/fail decisions and develop the skill of marking candidates using the full marking spectrum
- Be active clinically²
- Act as an effective member of a small team²
- Possess effective interpersonal skills²
- Be dedicated to respect, fairness, and courtesy towards candidates while maintaining an appropriate level of enquiry
- Be willing to accept training and regular monitoring of performance
- Be objective in analysing and comparing a candidate's performance against defined levels of competence
- Be able to manage the diversity of candidates, incorporating the adaptation of examining style to candidate needs
- Have the commitment and professionalism to examine and host the examination regularly, actively participate in regular examiner technique updates, and provide questions for the written examination
- Be appropriately qualified with respect to degree requirements, level and length of general paediatric experience, professional credentials, revalidation, and accreditation
- Be involved with junior staff training to be conversant with the standard expected of them.

In order to maintain the credibility of the examination, the selection of examiners should be rigorous, open, and well defined. Consistency in evaluation has been shown to be crucial, and prior assessment of this ability improves the selection of appropriate examiners.^{20 26 27} A model of selecting generic examiners (for both MRCPC and DCH) adapted from the comparable model of the RCGP² might be:

- Self-proposal or recommendation/nomination
- Applicants must be formally supported by at least two colleagues (e.g. Principal Regional Examiner, Regional Adviser in Paediatrics, Member of the College Examining Board, or a current examiner)
- Applicants must also (in concordance with the competencies of the examiner listed above):
 - Hold FRCPC or equivalent postgraduate qualification
 - Be clinically active in either general paediatrics or an appropriate specialty
 - Be up to date with a continuous professional development programme (CPD)
 - Demonstrate their ability to judge performance by ranking candidates in an order that correlates well with other examiners
 - Have experience in managing and supervising junior and middle grade staff
 - Have completed some appraisal and assessment training
 - Demonstrate to referees evidence of prior experience/training in communications and managing diversity awareness
 - Be prepared to commit to a training programme which will include preliminary assessment (examples could be sitting the part 2 written paper, creating questions, and possibly being observed performing during a mock clinical examination)
 - Be prepared to commit to the ongoing requirements of being an active examiner
 - Demonstrate the ability to receive and act on feedback.

Final ratification as an examiner would depend on satisfactorily completing examination training. Should an examiner not be selected in spite of completing a training programme, it is essential that this can be justified. An area that merits further evaluation is allowing middle grade doctors such as specialist registrars (SpRs) to nominate and serve as referees for consultants who are potential candidates for the post of examiner. SpRs have been shown to have considerable insight into the components of both clinical competence and good training.²⁸

COMPONENTS OF A TRAINING PROGRAMME

The components of a formal training programme have yet to be determined but should be based on the principles discussed above.

An indicative (not exhaustive) list of areas to be covered might include:

- Principles of assessment
- Examination design
- Interviewing skills
- Communication skills
- Ethics, accountability, discrimination, fairness
- Understanding cultural differences and expectations
- Handling different learning and working styles in candidates
- Planning and running structured medical examinations
- Giving feedback (for counselling)
- Revision on clinical examination skills (what to look for and what to ask)
- Writing questions
- Role play (by examiner trainees) and feedback (videod sessions)
- Observation of actual examiners by experienced supernumerary examiners for an exchange of ideas and approaches and a refinement of content and style of questioning⁶
- Assessment of clinical reasoning and clinical decision making
- Critique of videod clinical examinations and evaluation of candidates' performance
- Final assessment of examiners skills.

The final elements of selection and training need collaboration between clinicians and those with educational experience. Training of examiners is a necessary but not a sufficient condition for the conduct of examinations. Hospital Trusts will have to commit to providing protected time for courses and assist with hosting of examinations, and additional research will have to be conducted to assess whether what has been learnt is applied in regulatory assessments.²⁹ Examiners must be seen to be transparent in decision making in order to ensure fairness.³⁰

AREAS FOR FURTHER RESEARCH

- *Managing diversity of candidates:* Cultural and language differences can lead to unfair assessment.^{31 32} Further studies are needed to discover how such bias can be avoided.
- *Reliability and validity:* Further studies are needed to continue to establish reliability of postgraduate examinations. It is necessary to discover what reliability indicators are intended to measure and what conclusions can be claimed from their use. The use of generalisability to determine reliability of existing examinations would allow the sources of error contributing most to the overall error to be determined. Training of examiners and any subsequent modification of the examinations could be targeted at those sources of error contributing most to the overall error of measurement. Further work on all aspects of validity is important. There is some evidence to suggest that measurement of face validity has proved feasible and valuable and will assist in the further development of RCPCCH examinations themselves,³³ with obvious implications for relevance in examiner training.
- *Further exploration of the variation in marking among current examiners:* The effects of examiners' gender, experience,

academic rank, regional affiliation, and country of qualification on examiner behaviour have been studied and have proved useful in eliminating deviant patterns of grading and presenting candidates with a balanced panel of examiners, in addition to improving the standardisation of content.²² Factors influencing examiner behaviour in College examinations could be investigated.

- *Evaluation of the training programme envisioned here:* The training programme for examiners should be evaluated periodically so that it can continue to be fine tuned and updated. Data collection and analysis should focus on outcome measures such as examiners' views on the training process, candidates' confidence in the MRCPCH examination for fairness, and educationalists' analysis of changes in validity and reliability.

CONCLUSIONS

The training of examiners is an essential component of a process to ensure validity and reliability in UK paediatric clinical examinations. It is crucial, however, that training is rigorous, evidence based, contains formative and summative assessment of examiners, is regularly evaluated, and has a defined curriculum directed towards achieving the required competencies of examiners.

Authors' affiliations

N Khera, Senior Research Fellow, Lewisham NHS Trust, Senior Research Fellow, Sheffield Children's Hospital, Sheffield, UK

H Davies, Consultant Paediatrician, Sheffield Children's Hospital, Sheffield, UK

H Davies, Consultant Paediatrician, Central Middlesex Hospital, London, UK

T Lissauer, Consultant Paediatrician, St Mary's Hospital, London, UK

D Skuse, Institute of Child Health, London, UK

R Wakeford, University Staff Development Officer, University of Cambridge, Cambridge, UK

J Stroobant, Consultant Paediatrician, University Hospital Lewisham, London, UK

REFERENCES

- 1 **Wakeford R**, Southgate L, Wass V. Improving oral examinations: selecting, training and monitoring examiners for the MRCP. *BMJ* 1995;**311**:931-5.
- 2 **Neighbour R**. *Becoming an examiner for the MRCP*. London: Royal College of General Practitioners, 2000.
- 3 **Mackway-Jones K**, Walker M, eds. *Pocket guide to teaching for medical instructors*. London: BMJ Publishing Group, 1998.
- 4 **Dacre J**, Besser GM, White P. MRCP (UK) part 2 clinical examination (PACES): a review of the first four diets (June 2001-July 2002). *J R Coll Physicians Edinb* 2003;**33**:285-92.
- 5 **Fowell SL**, Maudsley G, Maguire P, et al. Student assessment in undergraduate medical education in the United Kingdom 1998. *Med Educ* 2000;**34**(suppl 1):1-78.
- 6 **Thomas CS**, Mellsoy G, Callender J, et al. The oral examination: a study of academic and non-academic factors. *Med Educ* 1993;**27**:433-9.
- 7 **Yaphe J**, Street S. How do examiners decide? A qualitative study of the process of decision making in the oral examination component of the MRCP examination. *Med Educ* 2003;**37**:764-71.
- 8 **Van der Vleuten CMP**, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. *Teaching and Learning in Medicine* 1990;**2**:58-76.
- 9 **Wilson GM**, Lever R, Harden RM, et al. Examination of clinical examiners. *Lancet* 1969;**i**:37-40.
- 10 **Tate P**, Foulkes J, Neighbour R, et al. Assessing physicians' interpersonal skills via videotaped encounters: a new approach for the Royal College of General Practitioners Membership examinations. *J Health Commun* 1999;**4**:143-52.
- 11 **Little D**. Learning disabilities, medical students, and common sense. *Acad Med* 1999;**74**:622-3.
- 12 **Wass V**, Roberts C, Hoogenboom R, et al. Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study. *BMJ* 2003;**326**:800-3.
- 13 **Black N**. Evidence based policy: proceed with care. *BMJ* 2001;**323**:275-9.
- 14 **Jolly B**. The professor has submitted her questions and thus we may proceed to set the examination paper. *Med Educ* 2000;**34**:7-8.
- 15 **Hutchinson L**, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med Educ* 2002;**36**:73.

16 **Raymond MR**, Webb LC, Houston WM. Correcting performance-rating errors in oral examinations. *Evaluation and the Health Professions* 1991;**14**:100-22.

17 **Newble DI**, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. *Med Educ* 1980;**14**:345-9.

18 **Van der Vleuten CPM**, Van Luyk SJ, Van Ballegooijen AMJ, et al. Training and experience of examiners. *Med Educ* 1989;**23**:290-6.

19 **Newble D**. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ* 2004;**38**:199-203.

20 **Weingarten MA**, Polliack MR, Tabenkin H, et al. Variations among examiners in family medicine residency board oral examinations. *Med Educ* 2000;**34**:13-17.

21 **Hinchcliffe R**. Medical examiner variability. *J Laryngol Otol* 1997;**1**:8-14.

22 **Leichner P**, Sisler GC, Harper D. A study of the reliability of the clinical oral examination in psychiatry. *Can J Psychiatry* 1984;**29**:394-7.

23 **Wilkinson TJ**, Frampton CM, Thompson-Fawcett M, et al. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad Med* 2003;**78**:219-23.

24 **Streiner D**, Norman G. *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press, 1995.

25 **Shekelle PG**, Kahan JP, Park RE, et al. Assessing appropriateness by expert panels: how reliable? *J Gen Intern Med* 1995;**10**(suppl):81.

26 **DesMarchais JE**, Jean P, Delorme P. Training in the art of asking questions in oral examinations. *Annals of the Royal College of Physicians and Surgeons of Canada* 1989;**22**:213-16.

27 **DesMarchais JE**, Jean P. Effects of examiner training on open-ended, higher taxonomic level questioning in oral certification. *Teaching and Learning in Medicine* 1993;**5**:24-8.

28 **Khera N**, Stroobant J, Primhak R, et al. Training the ideal hospital doctor: the specialist registrars' perspective. *Med Educ* 2001;**35**:957-66.

29 **Preston-Whyte ME**, Fraser RC, McKinley RK, et al. Hospital clinicians' views on training as examiners for undergraduate regulatory examinations. *Med Educ* 2000;**34**:964.

30 **Roberts C**, Sarangi S, Southgate L, et al. Oral examinations—equal opportunities, ethnicity, and fairness in the MRCGP. *BMJ* 2000;**320**:370-5.

31 **Roberts C**, Sarangi S. Hybridity in gatekeeping discourse: issues of practical relevance for the researcher. In: Sarangi S, Roberts C, eds. *Talk, work and institutional order: discourse in medical, mediation and management settings*. Berlin: Mouton de Gruyter, 1999:473-503.

32 **Manzar S**. The English language and Arabic medical students. *Med Educ* 1999;**33**:394-5.

33 **Tweed M**, Cookson J. The face validity of a final professional clinical examination. *Med Educ* 2001;**35**:465-73.

