

Making the best of the "long case"

There is no fundamental difference between a clinical test in clinical medicine and an educational test for assessment of competence. Both need to be proved effective in their prediction of outcome. Whereas sensitivity and specificity are used as quality indicators in clinical research, reliability and validity are used in educational testing. Reliability reflects the accuracy of the scores obtained with the assessment; validity refers to the extent a test measures what it was intended to measure.

The "long case" is used to assess candidates' clinical competence in the UK, Ireland, and elsewhere because of validity arguments. Its strength is that it requires the judgment of candidates carrying out real-life clinical tasks. The professional authenticity of the long case and the direct contact between the examiner and the candidate make this examination better than many other assessment instruments. Its weakness is its unreliability. Reliability is usually linked to the objectiveness of the judgment—in this instance, the extent of agreement among examiners in the long case. Although examiner disagreement in most oral examinations tends to be quite poor,¹ it is not the most profound influence on reliability. Numerous empirical studies have shown that candidate performance is highly variable across different situations, because performance depends on the specific content or characteristics of the situation.²⁻⁴ We intuitively believe that when we have measured someone's competence with one patient we can predict how competent that person will be with another patient. Unfortunately, this prediction tends to be poor, and it is this factor that leads to serious unreliability. In other words, two long cases in the assessment usually disagree substantially more than two examiners will. This makes the outcome of the examination highly dependent on the "luck of the draw".

To remedy this problem it is necessary to increase the sample size of cases in the assessment as we would do for the number of examiners. Gleeson⁵ has lately suggested that reliability will also be improved if the number of judgments within a case is increased. For this purpose he has developed the objective structured long examination record, or OSLER (figure). However, substantial improvement can be achieved only through an increase in the number of cases. The use of a single long case inevitably leads to unreliable pass/fail decisions. With the same number of examiners, the reliability of OSLER would be improved by presenting the candidate with two long cases, each with a different single examiner.⁶ The exact number of long cases required is not entirely clear, since appropriate reliability studies have not been carried out with long cases themselves.

There is one important aspect in which an educational test differs from a clinical test—in the effect that the measurement has on the person (or training programme) being measured. Educational tests not only serve the purpose of making pass/fail decisions but also provide the necessary feedback for remedial action. Examinations drive learning to a great extent; the examination programme actually *is* the curriculum for the student. With a few exceptions,^{7,8} however, the driving force of examinations has hardly been studied. This is the strength of Gleeson's study. Despite the (probably) poor reliability of the long case, this study indicates that OSLER has a profound effect on student learning in all areas of

Objective structured long examination record (OSLER)			
Examiners are required to GRADE each of the ten items below and assign an overall GRADE and MARK concerning the candidate PRIOR to discussion with their co-examiners as follows:		EXAMINER
GRADES P+ = VERY GOOD/EXCELLENT P = PASS/BORDERLINE PASS P- = BELOW PASS		MARKS (60-80+) (50-55) (35-45)	CO-EXAMINER
PRESENTATION OF HISTORY			
PACE/CLARITY	→		
COMMUNICATION PROCESS: (history, eg CVS, investigation, eg endoscopy, management, eg patient education)	→		
SYSTEMATIC PRESENTATION	→		
CORRECT FACTS ESTABLISHED	→		
PHYSICAL EXAMINATION			
SYSTEMATIC	→		
TECHNIQUE (including attitude to patient)	→		
CORRECT FINDINGS ESTABLISHED	→		
APPROPRIATE INVESTIGATIONS IN A LOGICAL SEQUENCE (Communication Process option)	→		
APPROPRIATE MANAGEMENT (Communication Process option)	→		
CLINICAL ACUMEN (Problem identification/Problem solving ability)	→		
ADDITIONAL COMMENTS			
Please tick (✓) for CASE DIFFICULTY		INDIVIDUAL EXAMINER	PAIR OF EXAMINERS
Standard	<input type="checkbox"/>	Overall grade	Agreed grade
Difficult	<input type="checkbox"/>	Mark	Agreed mark
Very difficult	<input type="checkbox"/>		

Figure: First page of the OSLER

competence. Thus the educational value of OSLER is strong despite other measurement weaknesses. Similarly, the study indicates how valuable competence assessments may be for the educational programme itself. In all areas assessed, the performance of the candidates was poor—probably, as Gleeson noted, because the teaching of clinical skills is neglected in the undergraduate setting.

OSLER seems to be a powerful tool for providing feedback and therefore has great potential to increase clinical competence. This being so, we should continue to study the long case as a decision tool. After all, we cannot afford to take the wrong decisions in passing and failing candidates. Moreover, the provision of the wrong feedback, as a result of imprecision, may similarly not be fair to students.

Cees van der Vleuten

Department of Educational Development and Research,
University of Limburg, Maastricht, Netherlands

- Muzzin LJ. Oral examinations. In: Shannon S, Norman G, eds. Evaluation methods: a resource handbook. Hamilton: Program for Educational Development, McMaster University, 1995: 37-45.
- Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990; 2: 58-76.
- Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1994; 23: 23-30.
- Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educ Res* 1995; 24: 5-11, 35.
- Gleeson F. The effect of immediate feedback on clinical skills using the OSLER. In: Rothman AJ, Cohen R, eds. Proceedings of the Sixth

Ottawa Conference of Medical Education. Toronto: University of Toronto Bookstore Custom Publishing, 1994: 412-15.

- 6 Swanson DB. A measurement framework for performance-based tests. In: Hart IR, Harden RM, eds. Further developments in assessing clinical competence. Montreal: Can-Heal, 1987: 13-45.
- 7 Newble D, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983; 17: 165-71.
- 8 Jolly BC, Newble DI, Chinner T. The learning effect of repeating content over time in a clinical skills based examination. *Teach Learn Med* 1993; 5: 66-71.

Mastocytosis: for better and worse

Of all the endocrine and metabolic itches,¹ mastocytosis is the most logical, mysterious, and intimidating. Logical, because it is no surprise that the release of excessive amounts of histamine and other mast-cell mediators should cause itching, flushing, diarrhoea, and a related panoply of symptoms and signs.² Mysterious, because that panoply can be so variable: to know mastocytosis is to enter a stream of information that cuts across many fields. Intimidating, because the diagnosis can be difficult to make,³ the course hard to predict,⁴ and the treatment challenging.⁵ Systemic forms can occur without obvious cutaneous involvement, and cutaneous mastocytosis can occur with or without systemic involvement, and even without a rash.^{6,7}

Mastocytosis is one of those diseases that physicians fail to think of when they should, or consider only to be proven wrong: ten patients with recurrent unexplained flushing referred to the National Institutes of Health over a nine-year period, mostly with the diagnosis of mastocytosis or idiopathic anaphylaxis, did not earn either diagnosis after intensive evaluation.⁸ Perhaps they had idiopathic recalcitrant facial flushing syndrome.⁹ But there are caveats: histamine release may be intermittent. Diagnostic tests for histamine, its metabolites, and other mast-cell products such as tryptase¹⁰ are not yet always readily available. Small wonder that the failure to establish a diagnosis of mastocytosis leaves some of us persistently uneasy with the idiopathic retreat. Think of the possible fearful outcomes: gradual personality changes; 20 years of episodic headache, flushing, lacrimation, hypotension, and syncope;¹¹ or 40 years of mastocytotic diarrhoea without a diagnosis.¹² And so-called *indolent* mastocytosis can occasionally become aggressive, as in mast-cell leukaemia. How occasional is occasionally and how does one watch for it?

A Dutch group¹³ with a longstanding interest in mastocytosis and skilled in its chemical evaluation has now followed 16 adult patients with indolent systemic mastocytosis for a median 90 months (range 13-135). Sequential measurements of urinary N'-methylhistamine (MH) and N'-methylimidazoleacetic acid (MIMA) reveal that adult-onset indolent mastocytosis is not always progressive: chemical and clinical evidence of disease can stay the same or decline substantially. None of the 16 developed a myeloproliferative disorder. In two patients, both MH and MIMA excretion decreased, as did their enlarged spleens. In several patients, the changes in MH and MIMA were not concordant. Overall, assay of MH seemed to be a better monitor of the course of indolent mastocytosis than did MIMA, partly because of dietary influences on MIMA concentrations. Until now, the general opinion has been that most children with mast-cell disease get better whereas most adults get worse. The Dutch evidence that adult mastocytosis is not always

progressive is welcome—except, perhaps, for one disconcerting implication. Those of us who took some comfort in the notion that worsening over time would ultimately help establish a diagnosis in unclear cases will just have to worry that much longer.

Jeffrey D Bernhard

Division of Dermatology, University of Massachusetts Medical Center, Worcester, MA, USA

- 1 Bernhard JD. Endocrine itches. In: Bernhard JD, ed. Itch: mechanisms and management of pruritus. New York: McGraw-Hill, 1994: 251-60.
- 2 Longley J, Duffy TP, Kohn S. The mast cell and mast cell disease. *J Am Acad Dermatol* 1995; 32: 545-61.
- 3 Duffy TP. Getting the story right. *N Engl J Med* 1993; 328: 1333-36.
- 4 Metcalfe DD. Classification and diagnosis of mastocytosis: current state. *J Invest Dermatol* 1991; 96: 2s-4s.
- 5 Austen KF. Systemic mastocytosis. *N Engl J Med* 1992; 326: 639-40.
- 6 Kendall ME, Fields JP, King LE. Cutaneous mastocytosis without clinically obvious skin lesions. *J Am Acad Dermatol* 1984; 10: 903-05.
- 7 Legrain V, Taieb A, Bioulac-Sage P, Maleville J. Mastocytose cutanée diffuse sans lésion permanente. *Ann Dermatol Venerol* 1994; 121: 561-64.
- 8 Friedman BS, Germano P, Miletti J, Metcalfe DD. A clinicopathologic study of ten patients with recurrent unexplained flushing. *J Allergy Clin Immunol* 1994; 93: 53-60.
- 9 Tur E, Ryatt KS, Maibach HI. Idiopathic recalcitrant facial flushing syndrome. *Dermatologica* 1990; 181: 5-7.
- 10 Schwartz LB, Metcalfe DD, Miller JS, et al. Tryptase levels as an indicator of mast-cell activation in systemic anaphylaxis and mastocytosis. *N Engl J Med* 1987; 316: 1622-26.
- 11 Kuter I, Harris NL. Systemic mastocytosis. *N Engl J Med* 1992; 326: 472-81.
- 12 Mahood JM, Harrington CI, Slater DN, Corbett CL. Forty years of diarrhoea in a patient with urticaria pigmentosa. *Acta Derm Venerol* 1982; 62: 264-65.
- 13 Kors JW, Van Doormaal JJ, Breukelman H, Van Voorst Vade PC, De Monchy JGR. Long-term follow-up of indolent mastocytosis in adults. *J Intern Med* 1996; 239: 157-64.

From dendritic cells to tumour vaccines

One of the chemotherapist's greatest successes has been the treatment of the lymphomas, yet low-grade lymphomas, for all their chemosensitivity, remain incurable. This continued dismal outlook has prompted oncologists to look for other means of treatment. Immunotherapy is attractive because B-cell lymphomas carry on their surfaces a tumour-specific target. The antibody-combining site of the surface immunoglobulin has a unique molecular structure. This structure, when seen as an antigen, is known as the idiotype. Monoclonal antibodies directed against idiotypes have had some success in treating low-grade lymphomas,^{1,2} but they are cumbersome and expensive to produce and are unlikely to become a serious treatment option. Active immunisation with idiotypic vaccines overcomes many of the objections to passive serotherapy, and in mice can suppress lymphoma for the lifetime of the animal.³ In patients, a clinical trial of idiotypic vaccines successfully induced anti-idiotypic immune responses in 20 of 34 cases and in some instances led to tumour regression.⁴

Unfortunately, idiotype is a weak antigen, and the tumour has already grown with the tolerance of the patient's immune system. Attempts to increase the immunogenicity of vaccines have employed foreign carrier proteins, adjuvants, cytokines, and genetically engineered viruses.⁵⁻⁷ All these approaches seek to stimulate the processing of antigen and the presentation of antigenic peptides to T cells. Of all the dedicated antigen-presenting cells, the bone-marrow-derived dendritic cell is