The Royal College of Surgeons of England

Ann R Coll Surg Engl 2001; **83**: 69–73

Training

The assessment of surgical skills and a simple knot-tying exercise

MG Tytherleigh, TS Bhatti, RM Watkins, DC Wilkins

Department of General Surgery, Derriford Hospital, Plymouth, UK

Basic surgical skills courses are mandatory for all surgical trainees taking the MRCS examination. An important aspect of these courses is the level of practical skill achieved by junior surgeons attending them. We present a simple knot-tying exercise, which may be used to assess the baseline skill level of trainees at the outset of the course and against which their progress can be judged after tuition and practice.

Key words: Basic surgical skills – Trainee assessment – Knot-tying exercise

The basic surgical skills (BSS) courses, now a mandatory requirement of the new MRCS curriculum, provide a singular opportunity to study ways in which surgical techniques are taught and learned. These courses, held in a workshop environment, are designed for very junior trainees, *i.e.* senior house officers (SHOs) of one or two years' seniority. The timetable is standardised and a maximum of 18 trainees are enrolled on to each course, which is divided into three sections, *i.e.* basic surgical techniques, trauma plus orthopaedics and minimal access skills. In a carefully controlled environment, skills such as knot tying, bowel anastomosis, vascular suturing, bone handling and basic laparoscopic techniques are taught by experienced surgeons in classes having a teacher/pupil ratio as low as 1:3.

Notwithstanding the fact that attendance at an approved BSS course and completion to a satisfactory standard is obligatory for basic surgical trainees, the courses have been very well received. A substantial number of centres have undertaken to organise them on a regular basis; feedback from trainees and trainers alike shows that both are enthusiastic, and the Royal Colleges are increasingly being asked to facilitate such courses in other countries.

Clearly this is a great achievement, but these courses also provide other opportunities for the refinement of surgical skills teaching that are as yet unrealised. Consider the fact that we have here a sizeable number of 'pupils' being taught a common curriculum under standardised conditions. Given some forethought, this should surely provide an exciting opportunity to study the way in which young surgeons learn their practical skills and the ways in which this aspect of surgery can be taught. In this setting, where many of the parameters are controlled, different methods of teaching a particular skill can be examined and compared. Other important aspects of skills training such as the aptitude of trainees, the efficacy of simulators and the performance of

Correspondence to: Mr DC Wilkins, Department of General Surgery, Derriford Hospital, Plymouth PL6 8DH, UK
Tel: +44 1752 777111; Fax: +44 1752 768976

Table 1 Assessment of surgical knot tying ability

Marking	Marks possible	Attempts 1st	2nd	3rd
1. Grasps thread				
Correctly	2	–	–	–
Needs correction	1	–	–	–
2. Ties first hitch				
Perfectly and beds knot with forefinger	4	–	–	–
Does not bed knot	3	–	–	–
With fumbling	2	–	–	–
Crossed/capsized	1	–	–	–
Cannot complete	0	–	–	–
3. Ties second hitch				
Perfectly and beds knot with forefinger	4	–	–	–
Does not bed knot	3	–	–	–
With fumbling	2	–	–	–
Crossed/capsized	1	–	–	–
Cannot complete	0	–	–	–
4. Finished knot				
Perfect (reef, tight)	3	–	–	–
Loose	2	–	–	–
Not a reef	1	–	–	–
Total (maximum 13)		–	–	–

Table 2 Assessment of surgical knot-tying ability – notes for the trainer

Objective

To assess the ability of the trainee to tie a one-handed knot of the type used to commence a running suture. The knot should be a reef knot with a single hitch at each stage. At completion, it should be reasonably tight and not capsized.

Method

An Ethicon™ jig is used, with a coloured knotting cord replacing the suture material. A needle holder is attached to the cord at one end and set up in relation to the jig as per the diagram (Fig. 1). The jig and cord must be set up exactly as shown for each trainee and for each attempt. Explain the exercise with the trainee sitting alongside and facing the same way. Give a single demonstration and remember to allow for left-handed pupils.

Marking

Mark each of 3 attempts. The average of these will be used. Please do not forget to complete the details of the trainee and trainer at the top of the form.

trainers could also be researched in this 'teaching laboratory'.

Common to all such studies, however, is the need for reliable methods of assessment. It is, self-evident that any evaluation of 'process' is dependent upon a robust method of assessment by which 'outcome' can be measured; the teaching of surgical skills is no exception. Devising such assessments for application to these often complex activities is, however, extremely difficult. To be successful an assessment exercise must be reproducible, subject to minimal inter-observer error or bias and yet be straightforward to apply. This is indeed a tall order, but the increasing involvement of educationalists in surgical teaching has undoubtedly lead to a better understanding by surgeons of the principles involved.

With the foregoing in mind, we have devised a simple test of the ability of individual trainees to tie a surgeon's knot. We have applied this test to trainees at the outset of a BSS course and again after the relevant training. We describe our experience and discuss ways in which this type of exercise may prove useful in the future.

Aims

This study aimed to assess the inter-observer variability of scoring associated with a simple test of practical skill. In this paper we report on our preliminary experience in the use of this test during a BSS course.

Methods

A score sheet based on the tying of a single-handed surgical reef knot was devised (Table 1). A simple knot-tying jig (Ethicon, UK) and cord was used (Fig. 1) and the scenario, or 'set', was such as to simulate the tying of the initial knot in a running suture. This starting position was carefully standardised and explained to each trainer and trainee using a diagram, written notes and verbal reinforcement prior to the start of each test (Table 2). A maximum score of 13 points was possible for each attempt.

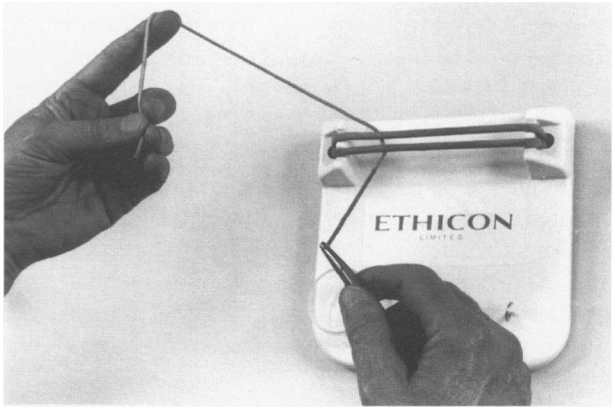

Figure 1 The knot-tying jig

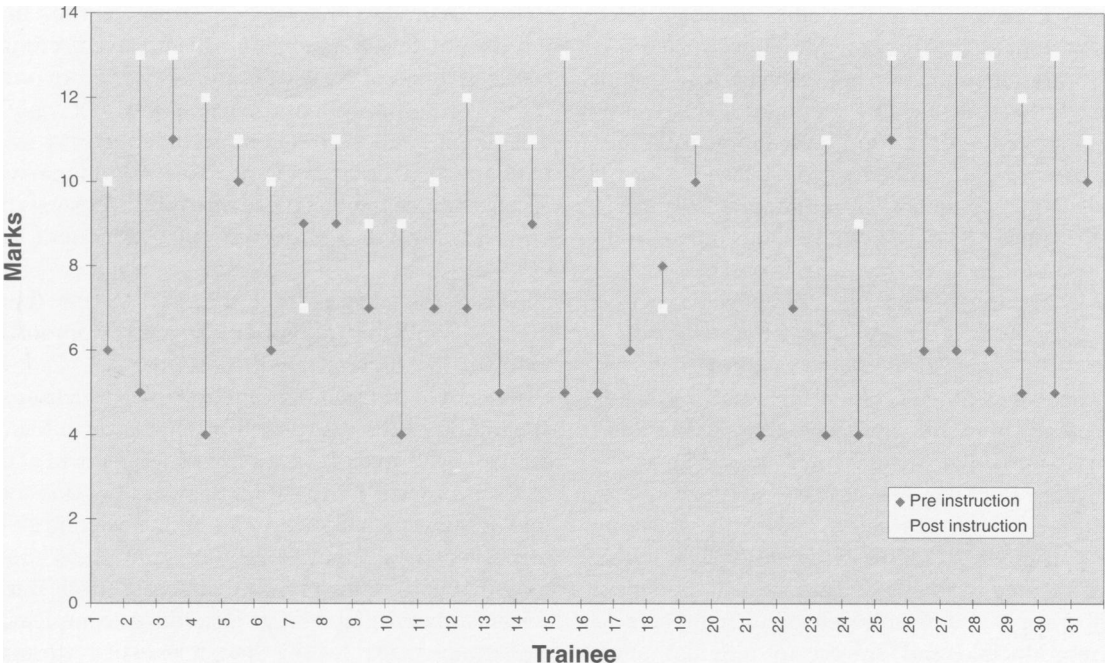

Figure 2 Change in scores pre-instruction versus post-instruction

Table 3 Results of 4 trainees tying 3 knots, examined by 6 trainers simultaneously

Trainees	Consultant trainers					
	A	B	C	D	E	F
1	12	11	10	11	10	10
	13	13	13	11	11	11
	12	11	13	12	10	11
2	6	4	4	4	6	6
	6	7	4	4	7	8
	6	7	4	4	6	9
3	13	13	13	13	13	13
	13	12	13	13	13	11
	13	13	13	13	13	12
4	13	12	12	12	12	13
	13	11	11	10	11	11
	13	13	13	13	13	13

In order to determine the inter-observer variability of the test, it was set first for four basic surgical trainees (senior house officers) who were watched by a panel of six consultant 'examiners' scoring simultaneously each of three attempts by each trainee in turn (Table 3). The test was then set for all trainees at the beginning and end of the first day of a basic skills course. In order to ensure that a true assessment of trainees' existing skill level was obtained, the initial test did not appear on the course timetable. No instruction or practice was permitted, but left-handed trainees were given the option of starting from a 'mirror image' position if they so wished. At the end of the first day's instruction, the test was performed under similar conditions by each trainee, but with marking carried out by a different trainer.

Results

The proforma was used by six consultant examiners marking four SHOs in general surgery, each having three attempts at tying a surgical reef knot. These results were analysed using the non-parametric Kruskal-Wallis test, which gave a P value of 0.82, *i.e.* there was good correlation between examiners (Table 3).

When applied to 31 trainees on two recent BSS courses, there was an improvement in individual scores, with the exception of two trainees who had a lower score at the end of the day's tuition (Fig. 2). The median improvement was 4.5 points, with a range of 1–8. The improvement was from a median initial score of 6.0 to a final score of 11. Using the Wilcoxon test, this difference was statistically highly significant, at the P <0.0001 level.

Discussion

The introduction of the Calman recommendations means that much more attention will be paid to the performance of trainees and trainers participating in

training programmes. The Royal Colleges have made attendance at a standardised BSS course mandatory for candidates wishing to take the new MRCS examination. It is also necessary for trainees to complete each section of the course to a 'satisfactory' standard. This raises problems of assessment of complicated practical skills exercises that surgical examiners have not faced before. It is accepted that assessments fall into one of two broad categories, namely 'summative' or 'formative'.[1,2] In simple terms, a formative assessment is one that gauges improvement in the performance of the trainee, provides feedback and sets targets. A summative assessment gauges the performance of the trainee against an agreed standard. There are strong arguments, at least in so far as basic skills acquisition are concerned, for making these assessments 'summative' in nature. This would be in line with the policy that requires trainees to meet a predetermined level of performance before they can be allowed to progress. It follows, therefore, that if this principle is accepted the method of assessment used must be practical to apply and reliable (reliability in this context meaning reproducible and subject to minimal inter-observer error or bias).

The existing structure of the BSS courses provides for an assessment exercise at the end of each section. The existing assessments do not discriminate between small differences in the level of performance and for the most part place trainees into one of only a few categories, *i.e.* satisfactory or unsatisfactory. More importantly, the criteria on which the assessments are made are sometimes poorly defined and very subjective. There is no doubt that course organisers are under great pressure to ensure that the maximum is taught in the time available and it is, therefore, paramount that any time spent on assessment should be used efficiently. It is very difficult, however, to design assessments that are straightforward to apply and yet do not take up an inordinate amount of the time available for teaching. With these constraints in mind, we focused on one specific area of BSS teaching and attempted to devise a test that fulfils the above criteria.

Variation in marking between examiners is a fundamental problem in any examination or assessment exercise. We were encouraged that several consultant trainers, after suitable briefing, arrived at comparable scores during simultaneous assessments. To facilitate this, we were careful to allocate precise scores to the levels of performance that were anticipated at each stage of the procedure and to define those levels very carefully (Table 1). It can be seen that the detail required to achieve this degree of specificity is considerable. Without this very rigid definition of the criteria, it is extremely unlikely that consistency between examiners will be achieved.

For an assessment to be considered fair, scoring must reflect accurately the level of performance. In other words, did this test provide satisfactory discrimination between different levels of performance? From our initial studies we found that there existed not only good agreement between different observers whilst watching the same manoeuvre being carried out, but there was also excellent agreement on the level of achievement when the same manoeuvre was carried out by different trainees. One can dispute the maximum number of points allocated for each part of this part of the (knotting) exercise, but this allocation is not as important as ensuring that each section is assessable.

A problem faced by any course organiser is the evaluation of the instruction on that course. Measuring the 'exit' skill levels by means of an assessment at the end of each course is fairly straightforward but does not take into account the wide variation in the existing skill level of trainees when they enrol. For example, some SHOs may be nearly 2 years into their basic surgical training whilst others may only just have completed pre-registration posts. Testing the exit level of performance of a particular skill will not, therefore, tell one much about the progress that an individual has or has not made during the course. The wide variation in the initial scores confirmed the prediction of a wide range of ability in the SHOs at the outset of teaching. We were encouraged, however, that 83% of trainees achieved an 'acceptable' level of performance (score 10 or greater) after instruction with one-third achieving 'perfect' scores. It is of interest that of this latter group nearly half had scored very poorly (4 points or less) at enrolment. Conversely, several trainees who exhibited an 'average' level of skill at the outset of the course did not improve their scores significantly following instruction and the performance of two apparently deteriorated! These variations raise intriguing questions regarding instruction, learning and aptitude. Course designers, course organisers and the teachers themselves will all have a vested interest in the effectiveness of a particular course and the teaching thereof. 'Before' and 'after' training tests offer the only objective way of measuring this. Using a suitable scoring system such as the one described in this paper, it should be possible to construct a database from which useful information on performance can be determined. For example, the methods of a teacher or teachers demonstrating a consistently superior performance in terms of his/her pupils' scores surely need to be analysed and emulated. Conversely, instructors whose pupils do less well need feedback and encouragement to improve. In so far as the trainees themselves are concerned, a well-constructed, reliable scoring system that gives good discrimination will act as a useful guide of progress and

aptitude – particularly for the small but problematic group who fail to improve despite adequate instruction.

There is no doubt that detailed assessment of practical skills takes a lot of time. It is likely that computer technology, which has not yet been widely employed in BSS courses, will be helpful in this respect. Simulators that not only provide training but also measure performance could be very useful. One such training/assessment machine (MIST™) has already been developed in connection with laparoscopic work. Further developments in this area will be of great interest to those involved with the design of many types of practical skills courses and the teaching thereof.

We have used our simple test to show that pre- and post-teaching assessments offer some interesting possibilities when applied to the development of skills teaching. This particular assessment was easy to apply, not time consuming, and required a minimum of briefing for the examiners and trainees. In our view there is a great need for the development of further reliable and discriminating tests of practical surgical skills in both the simulated and working environment.

Acknowledgements

The authors would like to thank the trainees and trainers who have attended The Royal College of Surgeons' basic surgical skills courses held in Plymouth, for their help in carrying out this work.

References

1. Newble D, Cannon R. *A Handbook for Medical Teachers*, 3rd edn. The Netherlands: Kluwer, 1994; 85–6.
2. Jolly B, Grant J. (eds) *The Good Assessment Guide*. London: Joint Centre for Education in Medicine, 1977.