

# Clinical skills evaluation of trainees in a neurology department

CM Wiles, K Dawson, TAT Hughes, JG Llewelyn, HR Morris, TP Pickersgill, NP Robertson and PEM Smith

**ABSTRACT** – Three to 12 evaluations of clinical performance using the mini-clinical evaluation exercise (Mini-CEX) (n=124) or direct observation of procedural skills (DOPS) (n=21) were performed on 27 trainees working in an NHS neurology department. The communications/counselling skills subdomain was scored in 64 evaluations. For Mini-CEX the focus was on gathering data (22%), diagnosis (31%), management (34%) and counselling (7%) (focus not recorded in 6%). For DOPS, lumbar puncture was the most common evaluated procedure (57%). Mini-CEX evaluations lasted 23.8 minutes (10.6) (mean, sd) and DOPS 25.9 minutes (12.6). Mini-CEX scores for overall competence and communication skills were mean 5.99 (sd 0.95, range 4–8) and 5.98 (sd 1.21, range 3–9) and for DOPS 5.71 (sd 0.90, range 4–8) both on scales of 1 to 9. Overall trainee competence and communication scores increased with year of training ( $p < 0.001$ ,  $p < 0.004$  univariate analysis). Assessors undertook up to three or four assessments in a session. Assessors and trainees considered that the observation and feedback had been ‘very’ or ‘quite’ useful in providing a relevant element of assessment. These assessments were feasible and useful in a neurology department and provided some evidence for increasing performance with trainee seniority. More assessor time (approximately one hour) than trainee time (24–26 min) was needed for each assessment undertaken.

**KEY WORDS:** direct observation of procedural skills, mini-clinical evaluation exercise, neurology, performance, trainees

## Introduction

Assessment of clinical competence and performance by direct observation of trainees are elements of the new methods of assessment currently being introduced to support foundation, basic and higher specialty training. The Royal College of Physicians (RCP) adopted the Mini-CEX for evaluation of clinical skills and developed its own assessment method, direct

observation of procedural skills (DOPS). The former had been developed in the United States and has been used by the American Board of Internal Medicine for trainees<sup>1,2</sup> and separately trialled for medical students.<sup>3</sup> The Joint Committee for Higher Medical Training (JCHMT) of the RCP undertook a pilot study which, inter alia, trialled the use of Mini-CEX and DOPS for specialist registrars (SpRs) in specialties which demonstrated feasibility and good reliability (J Wilkinson and W Wade, personal communication, 2007). In the current study the techniques were utilised as components of the evaluation of all trainees (pre-registration house officer (PRHO), senior house officer (SHO)/foundation year 2 (F2) and SpRs) since we wished to gain a ‘feel’ for how the use of such techniques would work in everyday clinical life and to gain practical experience of implementation.

## Methods

The Mini-CEX and DOPS documentation used was very similar to that of the RCP study:

- the range of scores for each domain was 1–9
- the descriptors for score bands on the Mini-CEX and DOPS documentation were unsatisfactory (1–3), satisfactory (4–6), above average/superior (7–9).

Assessors, who were consultant neurologists, were asked to devote 3.5–4 hours (one session) per month to evaluations. During this session assessors were asked to observe any element of the normal work which a trainee happened to be undertaking. Trainees were aware of, and had agreed to, the general process but were not specifically forewarned about possible evaluations on particular days. Assessors introduced themselves to patients, explained that they were there to observe the doctor and were not involved directly in their care, and asked for consent to observe. This scenario was appropriately modified if the consultant was clinically responsible for the patient’s care.

Guidance notes were provided to consultants and issues were discussed but no formal training was provided. The notes suggested, inter alia, that:

- it might be easier not to utilise patients under their direct care

**CM Wiles**<sup>1,2</sup>  
PhD FRCP, Professor of Neurology

**K Dawson**<sup>2,3</sup>  
FRCP, Consultant Neurologist

**TAT Hughes**<sup>2</sup>  
MD FRCP, Consultant Neurologist

**JG Llewelyn**<sup>2,3</sup>  
MD FRCP, Consultant Neurologist

**HR Morris**<sup>1,3</sup>  
PhD MRCP, Senior Lecturer in Neurology

**TP Pickersgill**<sup>2</sup>  
MRCP, Consultant Neurologist

**NP Robertson**<sup>1,2</sup>  
MD FRCP, Senior Lecturer in Neurology

**PEM Smith**<sup>2</sup>  
MD FRCP, Consultant Neurologist

<sup>1</sup>Department of Neurology, Cardiff University

<sup>2</sup>Department of Neurology, Cardiff and Vale NHS Trust

<sup>3</sup>Department of Neurology, Gwent NHS Trust

- 15–25 minutes should be the typical period for each evaluation
- a broad range of clinical activities should be observed.

It was suggested that no more than two evaluations should be undertaken with a trainee in any one session. The importance of positive and constructive educational feedback at each evaluation was emphasised. One staff member (CMW) briefed all trainees about the pilot and ensured that they were agreeable to participating in the process.

Immediately or soon after each evaluation the relevant one-page Mini-CEX or DOPS form was completed, the assessor documented any relevant feedback information and returned the form to the trainee. The trainee completed their own comments and retained the form, sending a copy to CMW for collating and analysis.

Each Mini-CEX and DOPS sheet was entered onto a spreadsheet for the trainee. This spreadsheet provided summary scores in each of the principle scored domains and allowed the syllabus area to be mapped. Prior to the record of in-training assessment or at the end of their post, trainees were offered a copy of the summary sheet. In this paper, only data on overall clinical competence and the communication subscore are presented. Analysis was with SPSS v11.0 for Macintosh.

## Results

Between February 2004 and January 2006 most trainees were evaluated with Mini-CEX and DOPS. For this paper, data have only been used from 27 trainees who had a minimum of three evaluations and by assessors who had undertaken a minimum of 10 evaluations. These trainees were  $\leq 1$  to 9 years post-primary medical qualification. Of the 145 evaluations (124 Mini-CEX, 21 DOPS) undertaken by seven consultants on 11 SpRs, 11 SHOs and five PRHOs, 79 were on SpRs, 47 on SHOs and 19 on PRHOs. The number of assessments and assessors per trainee are shown in Table 1.

Over the period, each assessor undertook between 10 and 37

**Table 1. The number of assessments and assessors per trainee.**

Assessments/trainee	Trainees
3	8
4	5
5	3
6	3
7	4
9–12	4
Assessors/trainee	Trainees
1	2
2	8
3	8
$\geq 4$	9

evaluations (mean 21) on between 4 and 20 trainees (mean 12). The time span over which evaluations were completed for each trainee was 84 (82) (8–516) days (median, interquartile range, range).

The principal clinical presentations, procedures and case features observed are shown in Tables 2a–2c. The mean (sd) duration for a Mini-CEX evaluation was 23.8 minutes (10.7) and for DOPS 25.9 minutes (12.6).

Mini-CEX and DOPS ‘overall clinical competence’ scores ranged from 4 to 8 with mean (sd) scores of 5.99 (0.95) and 5.71 (0.90) respectively (Table 3). The communication skills domain subscores (subscore 5 of Mini-CEX and consolidated subscores 5 and 10 of DOPS) ranged from 3 to 9 with mean (sd) scores of 5.92 (1.21). There was no ‘overall competence score’ of  $\leq 3$ : there were 5 domain subscores  $\leq 3$ . Overall, competence scores given by individual assessors varied somewhat ( $f=5.562$ ,  $p<0.001$ ) but not communication subscores ( $f=1.487$ ,  $p=0.215$ ) (Figs 1a and 1b): post hoc analysis indicated that assessor four scored significantly lower than assessors one and three (Bonferroni correction).

The duration of each evaluation, focus and setting did not systematically affect scores but there was a tendency for cases rated as being of higher complexity to be scored slightly higher for overall competence (borderline significance). Overall competence score and communication subscore increased significantly with the year of postgraduate training ( $f=2.491$ ,  $p=0.019$ ;  $f=3.459$ ,  $p=0.002$ ) (Figs 2a and 2b): in a stepwise regression model including year of training, trainee grade and assessor only, year of training was an independent predictor of both ‘overall competence score’ ( $t=4.38$ ,  $p<0.001$ ) and communication subscore ( $t=4.73$ ,  $p<0.001$ ).

There were five occasions in four trainees where any domain subscore of three was given; these related to an important missed physical sign, disorganised examination technique, need for more communication with patient during a ward round, insufficient knowledge on clinical matter in hand, and one score where no specific feedback about the event was written down. Five trainees had a lower 95% confidence interval for overall competency score  $< 4$  but all had had only three or four assessments and none had had any domain subscore or overall score of  $\leq 3$ .

## Discussion

Trainees had an evaluation of their performance and active feedback across a broad range of clinical topic areas. For a department with four or five SpRs at any one time the frequency of assessment fell comfortably within the four Mini-CEX/year recently recommended by the JCHMT. As expected with shorter posts SHOs and PRHOs had fewer evaluations but a higher proportion of these were DOPS, most commonly lumbar puncture. Forms were generally completed satisfactorily: there was a tendency for assessors, however, to forget to write the case diagnosis which may be important when mapping assessments against a syllabus. Occasionally the assessor did not provide any written feedback: on one occasion this was in the context of a domain subscore of three. Clearly all unsatisfactory scores do need specific reasoning and feedback to be annotated.

**Table 2a. Principal clinical presentations evaluated.**

Mini-CEX and DOPS clinical topics	Number of patients (n=145) (%)
Peripheral nerve, muscle, weakness	23 (16)
Consciousness, sleep, higher function	19 (13)
Epilepsy	15 (10)
Spinal cord, roots	14 (10)
Parkinsonism, movement disorder, MND	11 (8)
Headache	11 (8)
Infection/CSF	10 (7)
Cerebrovascular	8 (6)
Pain	7 (5)
Visual and cranial nerves	7 (5)
Dizzy, falls	7 (5)
Metabolic/toxic	5 (3)
Demyelination	4 (3)
Tumour	4 (3)

CSF = cerebrospinal fluid; DOPS = direct observation of procedural skills; Mini-CEX = mini-clinical evaluation exercise; MND = motor neurone disease.

Seven out of eight available assessors undertook  $\geq 10$  evaluations in the study: this differed from the large US study<sup>2</sup> where 32% of 316 examiners only undertook a single evaluation. In the system used, assessors designated their own time to undertake evaluations and sought out trainees on the ward or clinic. This approach is probably more time efficient for the assessor and from the trainee viewpoint the evaluation is more unselected. The disadvantages are that, by chance, some trainees are more likely to be seen by the same assessor (though in fact this was uncommon), that some waiting around was inevitable (eg on arriving in clinic and having to wait for the next patient) and that specific elements of practice may not be covered. Specifying minimum numbers of assessment on core elements will avoid the latter but create significant implementation difficulties. This problem would substantially increase if the evaluation of large numbers of syllabus domains became essential. Mini-CEX may be conceived of as reliably testing a generic clinical ‘ability’: it is an assumption that performance over 12–16 evaluations (ie the recommended four evaluations/year by JCHMT carried out over three to four years) will reflect performance in other clinical areas. On the other hand, evaluations to confirm ‘competence’ or ‘performance’ in many individual elements of a syllabus will entail implausibly large numbers of evaluations. Similar issues of generalisability from evaluations on the relevance of a specific procedure (using DOPS) to other procedures clearly apply. Further testing of these assumptions is needed in order to develop the most robust but economical use of these evaluations.

**Table 2b. Principal direct observation of procedural skills (DOPS) performed.**

DOPS	Number of procedures performed (n=21) (%)
Lumbar puncture	12 (57)
Botulinum toxin injections	2 (10)
Catheter/nasogastric intubation	2 (10)
Lying and standing blood pressure	2 (10)
Prescription	1 (5)
Discharge sheet	1 (5)
Blood sample	1 (5)

**Table 2c. Principal mini-clinical evaluation exercise (Mini-CEX) case features.**

Mini-CEX case features	%	
In/outpatient	55/45	
New cases	57	
Case complexity	low	16
	moderate	50
	high	28
	unscored	6
Case focus	data gathering	22
	diagnosis	31
	management	34
	counselling	7
	unscored	6
Communication (subscore 5)	48	

As indicated, assessors varied somewhat in score attribution with one in particular tending to score lower than the others. This emphasises the importance of having multiple assessors for a given trainee in order to achieve reliability – a point clearly demonstrated in the RCP pilot (Wilkinson and Wade, personal communication, 2007): the requirement for multiple assessors

**Table 3. Overall clinical performance scores and communication subscore for all mini-clinical evaluation exercise (Mini-CEX) and direct observation of procedural skills (DOPS) evaluations by trainee grade (upper part of table) and separate Mini-CEX and DOPS scores (lower part of table).**

Grade of trainee (n)	Overall clinical competence		Communication subscore	
	n evaluations	Mean (sd)	n evaluations	Mean (sd)
PRHO (5)	19	5.68 (0.89)	8	5.25 (1.28)
SHO (11)	47	5.68 (0.86)	21	5.38 (1.02)
SpR (11)	79	6.18 (0.96)	35	6.40 (1.12)
All (27)	145	5.95 (0.95)	64	5.92 (1.21)
Mini-CEX only	124	5.99 (0.95)	59	5.98 (1.21)
DOPS alone	21	5.71 (0.90)	5	5.00 (1.10)

n = number; PRHO = pre-registration house officer; SHO = senior house officer; SpR = specialist registrar.

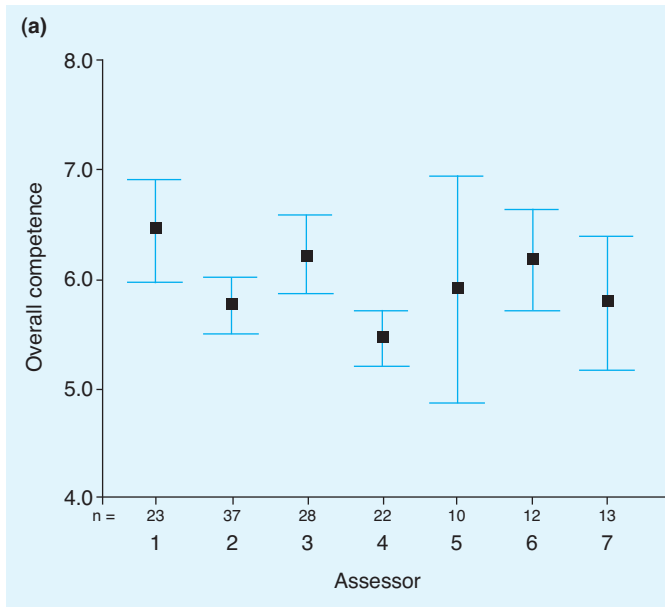


Fig 1a. Overall competence scores (mean 95% confidence interval) by assessor (1-7). n = evaluations/assessor.

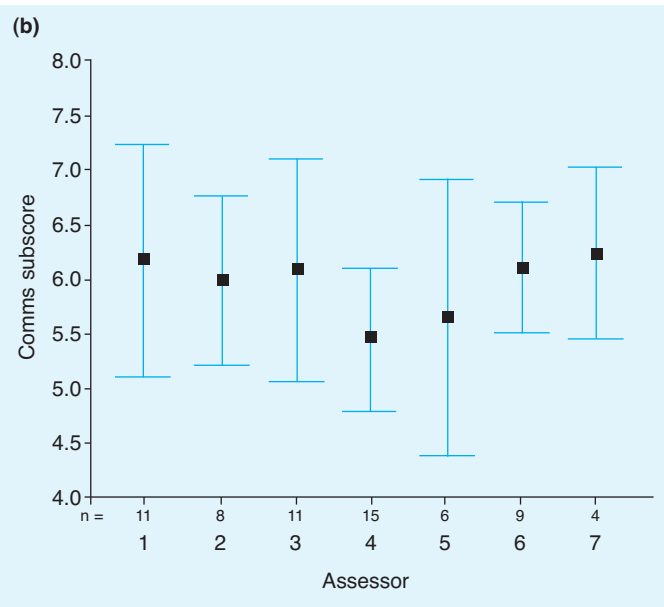


Fig 1b. Communication subscores (mean 95% confidence interval) by assessor. n = evaluations/assessor.

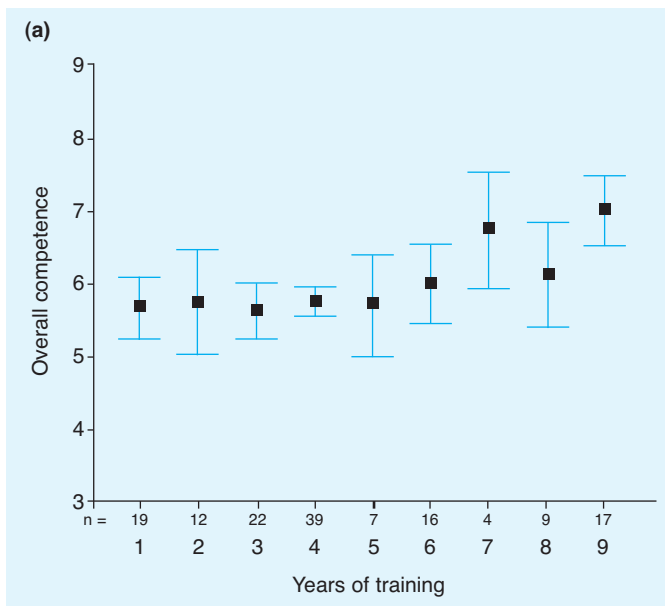


Fig 2a. Overall competence scores (mean 95% confidence interval) by year of training (1-9). n = number of evaluations/year of training.

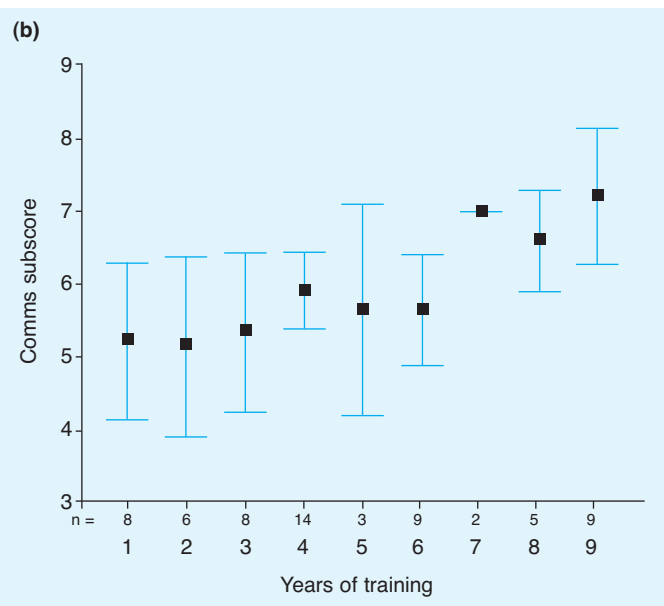


Fig 2b. Communication subscores (mean 95% confidence interval) by year of training. n = evaluations/years of training.

becomes more critical in the reliable documentation of performance in a low scoring trainee.

Assessors found in general that it was easier to evaluate, especially with more senior trainees, if they were seeing a patient under the care of another consultant. This meant that neither the patient nor the trainee had an expectation of input to the clinical management and this enhanced the observer role of the assessor and left the trainee with unambiguous clinical responsibility for the encounter. Verbal feedback was generally possible

even in a busy clinic or ward round situation although completing the forms was sometimes more appropriate afterwards. The giving of a score of  $\leq 3$  for any element (but in particular for overall competency) may necessitate extra time for discussion and more detailed constructive feedback.

The mean time spent on individual evaluations (about 25 minutes including feedback) was similar to previously reported figures (eg 25.6 minutes)<sup>2</sup> and to the RCP pilot. It should be noted, however, that this is 25 minutes from the

trainee perspective. From the perspective of the assessor, the total time taken finding the trainee, preparing, finding or awaiting the patient and going to the next assessment was substantially longer and in a given clinical session only three or four Mini-CEX/DOPS evaluations were achieved (ie approximately an hour/assessment). Variability of trainee numbers and available assessors, disposition across different hospital sites and nature of specialty may all mean that the above system of session allocation is not suitable for all and maximum flexibility of approach is likely to be necessary for effective implementation. Furthermore, caution is required in specifying generic time allocations for assessors in different specialties in relation to contracts of work.

The current forms indicate that assessment should be scored to 'reflect the performance...against that reasonably expected at their stage of training and level of experience' implying that better performance by a more senior trainee should not necessarily attract a higher score than lesser performance by a more junior trainee. In this study, average overall competence scores and communication subscores did increase, however, with year of training. Given that in the neurology training system the PRHOs and SHOs do not necessarily become the SpRs in that specialty this may simply reflect a sampling issue. On the other hand, it is difficult not to give a higher mark when an observed behaviour is clearly better even making allowance for the guidance above. We suspect that the increasing scores with longer training do actually reflect improved performance ie they provide some evidence of validity. In the US study<sup>2</sup> performed on 421 residents in 21 programmes mean (sd) overall competence scores were 6.7 (0.9) and counselling/communication 6.8 (0.9) compared to 5.95 (0.95) and 5.92 (1.21) in the present study although differences in trainee level, assessor and healthcare system make direct comparisons difficult.

Several assessors and trainees found that the process added some stress to a clinical encounter particularly in the early stages: this largely reduced with familiarity. Training and preparation of assessors will probably be important to minimise this. Trainees are increasingly familiar with observed short assess-

ments from the early stages of medical school: the great majority of our trainees welcomed the opportunity to be observed and to receive constructive feedback on their performance.

In conclusion, we found that it was feasible to undertake Mini-CEX and DOPS evaluations on our trainees, that the evaluations were constructively undertaken and were perceived as useful. Further work is needed to explore the handling of low scores, the contribution of scores to the overall evaluation process, the generalisability of scores across syllabus elements and the standardisation of performance for specialists<sup>4</sup> bearing in mind increasingly well-recognised principles of assessment.<sup>5</sup> Consultant assessors need appropriate time to undertake these evaluations with a guide of three to four per session being achievable in the context of neurology.

### Acknowledgments

Winnie Wade and James Wilkinson (Education Department, RCP) for helpful advice and information about the RCP pilot study. Cardiff Neurology Department trainees for their support for the process and positive cooperation.

### References

- 1 Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* 1995;123:795-9.
- 2 Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003;138:476-81.
- 3 Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. *Acad Med* 2003;78(10 Suppl):S33-5.
- 4 Southgate L, Hays RB, Norcini J *et al*. Setting performance standards for medical practice: a theoretical framework. *Med Educ* 2001;35:474-81.
- 5 Epstein RM. Assessment in medical education: review article. *N Engl J Med* 2007;356:387-96.