

Rating errors in the assessment of clinical competence of newly graduated doctors

Short Report

Medical Education Study Group, KIMS

Objectives Clinicians and academicians attached to health care institutions, who share the responsibility for training medical graduates, need to guard against common errors of measurement when assessing trainees' competencies. This study aims to identify the rating errors that may occur during the assessment of the clinical competence of interns.

Methods Trainers assigned responsibility for the internship program assessed the competencies of 45 internship trainees in the March 2002 cohort in six domains (case presentation, clinical skills, diagnosis, therapy, handling of emergencies and professional behavior) using a structured format provided. Competencies related to different clinical problems were rated under 1 = *Unsatisfactory*, 2 = *Below expected standard*, 3 = *At expected standard*, 4 = *Above expected standard* or 5 = *Excellent*.

Results 4868 assessments were made of the performances of 45 interns who completed rotations in Surgery, Medicine, Obstetrics & Gynecology, and Pediatrics. 66% of all ratings fell in Excellent category (mean = 4, mode = 5). The rotations in Surgery and

Pediatrics elicited more Excellent ratings (37.6% and 31.8% respectively) than other specialties. Of all trainers, Registrars allocated the highest proportion (46%) of Excellent ratings. When considering the number of trainers in each category, Excellent ratings had been assigned to a high extent by Assistant Registrars (72%), and Registrars (70%).

Conclusion A high proportion of the ratings given by the trainers fell at the upper end of the Rating Scale. This could be a result of the assessments being affected by errors of leniency, compromising their usefulness. It is desirable that the Heads of the respective Clinical Departments take steps to organize appropriate training programs so that the trainers improve their knowledge in the concepts underlying the assessment of clinical competencies and skills and methods of the use of the instruments.

Key words: clinical skills, internship, measurement, outcome measurement errors

Bull Kuwait Inst Med Spec 2005;4:32-34

Introduction

Before being granted a license to practice, medical graduates follow a one-year internship that requires the trainee to complete rotations in the specialties of Medicine, Surgery, Obstetrics & Gynecology and Pediatrics. Some internship training programs include additional rotations in fields such as Family Practice.

Internship training has been the focus of considerable attention from authorities such as the General Medical Council (GMC) in the UK during the recent past,¹ and as a result a number of measures has been introduced that is aimed at improving the training program. Clinicians serving at health care institutions and academic staff attached to university

medical faculties share the responsibility for training medical graduates. An area that demands scrutiny is the assessment of clinical skills, and assessors and examiners evaluating the competencies of trainees need to guard against rating errors that could compromise the usefulness of the results. Among the common errors observed are the following: *error of leniency* (being generous), *error of stringency* (being unduly strict), *error of central tendency* (avoiding the extremes of a score range) and the *halo effect* (prior knowledge about the candidate affecting the score).^{2,3} This study aims to determine the rating errors that may occur when trainers involved in internship training assess the clinical competencies of intern medical graduates.

Report

The internship trainees in the March 2002 cohort had their competencies in six domains

Medical Education Study Group, Kuwait Institute for Medical Specialization, Kuwait.

Correspondence: Medical Education Study Group, KIMS, P.O. Box 1793, Kuwait 13018. Tel: 2410027; Fax: 2410028; e-mail: cmecenter@kims.org.kw

assessed by the assigned trainers. These domains were case presentation, clinical skills, diagnosis, therapy, handling of emergencies and professional behavior. A set of *Interaction Cards*, which provided a format for the trainer to evaluate the intern's level of performance in the six listed domains, was made available. Table 1 shows the distribution of the domains among the encounters observed. The trainers rated the competencies using a five-point rating scale (1 = *Unsatisfactory*, 2 = *Below expected standard*, 3 = *At expected standard*, 4 = *Above expected standard* or 5 = *Excellent*). The assessments were repeated in relation to different clinical problems, and the ratings at the individual assessments were averaged and recorded under Overall Performance at end of rotation.

4868 assessments were made of the performances of 45 interns who completed rotations

Table 1. Focus of interaction in the clinical performance domains assessed

Clinical Performance Domain	n	%
Clinical skills		
History		
Directly observed	2893	59.4
Not observed	130	2.7
Case sheet		
Directly observed	2728	56.0
Not observed	157	3.2
Physical exam		
Directly observed	2572	52.8
Not observed	198	4.1
Professional Behavior		
Directly observed	2508	51.5
Not observed	195	4.0
Case presentation		
Written	1843	37.9
Verbal	1589	32.6
Diagnosis (clinical judgment)	1214	24.9
Therapy	1079	22.2
Handling of emergencies	618	12.7

in Surgery, Medicine, Obstetrics & Gynecology, and Pediatrics (Table 2). The ratings for 1 = *Unsatisfactory* were included under 2 = *Below expected standard* for analysis. Data were analyzed using counts of the ratings and their percentages in each category of responses. The ratings assigned and their distribution were as follows: nearly all the ratings given at the

Table 2. Distribution of ratings (% in parentheses) in each category among the designations and specialties of evaluators

Designations & Specialties	Rating			
	Below Expected Standard	At Expected Standard	Above Expected Standard	Excellent
Consultant				
Medicine	1 (0.85)	18 (15.72)	32 (28.24)	62
Obs & Gyne	0	0	0	8
Pediatrics	0	31 (16.23)	73 (38.21)	87
Surgery	0	12 (12.12)	32 (32.32)	55
Faculty Member				
Medicine	0	5	3	
Obs & Gyne	0	1	2	1
Pediatrics	0	1	0	4
Surgery	0	0	4	2
Asst. Registrar				
Medicine	0	2	4	
Obs & Gyne	0	0	2	18
Pediatrics	0	0	2	0
Surgery	0	0	1	10
Registrar	0	15 (28.30)	19 (35.84)	37
Obs. & Gyne	0	1 (0.57)	19 (35.84)	155
Pediatrics	0	4 (2.76)	27 (18.62)	114
Surgery	0	4 (2.88)	22 (15.83)	109
Senior Registrar	0	31 (29.25)	48 (45.28)	
Obs & Gyne	0	0	22 (34.92)	41
Pediatrics	0	3 (2.16)	32 (23.02)	104
Surgery	0	3 (1.10)	81 (29.67)	189
Total	1 (0.06)	131 (8.02)	461 (28.21)	1041 (63.71)

assessments were positive, with a high proportion (65.9%) falling within the *Excellent* category (mean = 4, mode = 5, $p < 0.001$). The rotations in Surgery (37.6%) and Pediatrics (31.8%) elicited more *Excellent* ratings than other specialties. Obstetrics & Gynecology, Pediatrics, and Surgery had high proportions of *Excellent* assessments. Of all trainers, Registrars had allocated the highest proportion (46%) of *Excellent* ratings. Within as well as across the four specialties, the ratings given for the performances were consistently close to the upper end of 5 in the Rating Scale. When considering the number in each category, Assistant Registrars (72%), and Registrars (70%) had assigned *Excellent* ratings to a high extent. Evaluators holding the designation of Consultant appeared less lenient than those of other ranks in using the different scale points.

Discussion

ASSESSMENT OF PERFORMANCES

An unusually large proportion of the performances had been rated as Excellent. When the normal distribution of attributes within a population is considered, it would be expected that the majority of the subjects would fall into a group that may be designated as average or satisfactory (64% of the population coming within ± 1 SD of the mean). Individuals who qualify to be considered as excellent should constitute only a small minority, their abilities falling near the +3 SD of the mean in a distribution. Thus, the results obtained suggest that the evaluators' expertise in the field of assessment or the manner in which the instrument is used should be reviewed.

EXPERTISE OF EVALUATOR

A topic of major interest to the administrators of internship training is the expertise of the evaluators in the area of educational measurement. The majority of the tutors who are given the responsibility for training the interns are clinicians who are primarily concerned with patient care. It may be assumed that their expertise in concepts and techniques of educational measurement is limited. Thus, the results observed emphasize the need for conducting appropriate training programs that deal with the themes of educational measurement and clinical assessment.

USE OF ASSESSMENT TOOL

Some of the tutors may view that assigning low or negative ratings to the trainees may contribute to the Training Unit itself being perceived somewhat negatively by the trainees and the administrators. Additionally, such a situation may be considered as leading to trainees not being posted for rotations in future, thus reducing the personnel available for sharing the clinical duties workload.

The excessively high proportion of Excellent ratings may also be related to the tendency on

the part of the evaluator to please the trainee. The reluctance to assign negative (or low positive) ratings may be associated with the local cultural background, where negative aspects are not usually emphasized. The observation is in sharp contrast to the reported 2.8:1 ratio of positive to negative comments reported in the evaluation of postgraduate trainees.⁴ Training the trainers in assessment methods needs to be considered as a remedial measure so that the effects of the error of leniency are eliminated, or at least minimized.

Heads of Departments of the concerned hospitals need to take steps to ensure that the clinical tutors who are given responsibility for training the interns are adequately competent in the underlying concepts and the techniques of assessing clinical skills. When it is observed that trainers are deficient in such competencies, measures should be implemented so that training in clinical skills and their assessment could be undertaken in a satisfactory manner.

References

1. General Medical Council. *The New Doctor*. London: GMC; 2003.
2. Ebel, Robert L. *Essentials of educational measurement*. 3rd ed. N.J: Englewood Cliffs; 1979.
3. Thorndike, Robert L, editor. *Educational measurement*. 2nd ed. American Council on Education; 1971.
4. Rhoton MF. A new method to evaluate clinical performances and critical incidents in anesthesia: quantification of daily comments by teachers. *Med Educ* 1990;24:280-9.

Members of Medical Education Study Group of the Kuwait Institute for Medical Specialization:

Nabila Abdella
Khaled F. Al-Jarallah
Mohamed A.A. Moussa
Nawaf Al-Mutairi
I.G. Premadasa
Diaa Shehab