MEDICAL EDUCATION

Performance of students in the final examination in paediatrics: importance of the "short cases"

Z Hijazi, I G Premadasa, M A A Moussa

Arch Dis Child 2002;**86**:57–58

See end of article for authors' affiliations

Correspondence to:
Dr Z Hijazi, Associate
Professor in Paediatrics,
Faculty of Medicine,
Kuwait University, PO Box
24923, Kuwait 13110;
zeinat@hsc.kuniv.edu.kw

Accepted for publication
15 October 2001

Aims: To determine which component of the final examination in paediatrics at Kuwait University best predicted the final outcome.

Methods: The performance of 356 medical students in the short cases, the long case, multiple choice questions (MCQs), and the essay components of the final examination in paediatrics at Kuwait University was correlated with the final grade, and the mean difference between each component and the final score calculated.

Results: The correlation was highest for the short cases, followed by MCQs, the long case, and essays. The mean difference between the final score and that of short cases was not significant, but was highly significant for the other components.

Conclusion: Results show that performance in the short cases component of the final examination in paediatrics is a better discriminator of competence than that in the long case.

A wide variety of examination types is in use in the assessment of students' achievements at medical school. They include written tests such as the essay type, short answer questions, and multiple choice questions (MCQs), which constitute the standard form of knowledge testing; and the "long case" and the "short cases" which assess clinical skills. Some examiners view the long case, the short cases, and the viva as being more useful than the other techniques.¹ Recently, the objective structured clinical examination (OSCE), has received wide acceptance.^{2–4}

The Department of Paediatrics at the Faculty of Medicine, Kuwait University organises training in paediatrics for medical students during their second clinical year. One group of students follow a clerkship of three months duration in paediatrics, while others enter comparable groups in obstetrics and gynaecology, psychiatry, or community medicine. On completion, students take an "end of block" examination, which consists of written (50 MCQs of "best answer" and "single true/false" types with no negative marking) and clinical components. The two sections contribute 12 and 18 marks, respectively, to the final examination which is administered to all groups at the end of the year. In the final examination the written section offers candidates 100 MCQs (14 marks) and four out of five essay questions (14 marks). The long case in the clinical section (21 marks) is of 60 minutes duration, and requires history taking and physical examination of a patient. The candidate is then examined on the case for 30 minutes. The two short cases (14 marks), where the candidate is asked to show a particular clinical skill, such as examination of the abdomen or heart, and data interpretation (7 marks) are of 10 minutes duration each. Unlike in the long case, the examiner closely observes the candidate's performance at all stages in the short cases. The clinical components (both block and final) constitute 60% of the final mark. Final grades are expressed as F, C, C+, B–, B, B+, A–, and A, corresponding to score ranges of <60, 60–69, 70–79, 80–83, 84–86, 87–89, 90–94, and 95–100, respectively. Candidates who score 85% and above, or just short of 60% (F) are given an oral examination (viva), for a possible distinction pass (A+), or a passing grade, respectively.

A previous investigation showed a high positive correlation between candidates' performances in the different components of the end of block examination and the final examinations.⁵ Some of the local examiners feel that the short

cases discriminate candidates more effectively than the long case, and that this aspect needs to be reflected in the weighting of the different examination components.

The objectives of this investigation were to: (1) determine which component of the final examination most effectively identifies the competent student; and (2) assess whether the performance of the candidates in the examination is gender related.

METHODS

The records of candidates' performances during the period 1991 to 1999 were reviewed.

The individual scores that each student obtained in the different components of the final examination were correlated with the aggregate score. The mean difference between the scores for each component and the aggregate scores was computed.

The candidates were divided into three groups, based on the aggregate score: $\geq 80\%$ (group 1), 70–79% (group 2), and $< 70\%$ (group 3). The deviations of the individual scores from the aggregate score in each component of the final examination of those in group 1 and group 3 (the relatively high ability and the relatively low ability candidates, respectively) were analysed.

The grade distribution was analysed on the basis of gender.

Pearson's correlation (r) was calculated to assess the correlation between scores of the individual components and the aggregate score of the final examination; the χ^2 test was used to analyse the significance of differences; χ^2 for trend was used to investigate the trend in proportions between males and females; and Student's t test was used to analyse differences between means.

RESULTS

Of 356 students (99.7% valid cases), 160 (44.9%) were males and 196 (55.1%) females.

Abbreviations: MCQ, multiple choice question; OSCE, objective structured clinical examination

Table 1 Mean differences between the aggregate score (75.65) and the score for each component in the final examination

	Essay	MCQs	Long case	Short cases
All students (n = 356)	2.02* (6.53)	4.76* (7.13)	-3.32* (8.67)	-0.43§ (7.87)
Group 1, high ability (n = 114)	+4.66* (5.88)	5.74* (5.47)	-3.64* (6.62)	-1.92† (6.59)
Group 3, low ability (n = 75)	-1.45‡ (6.09)	4.13* (9.77)	-3.53* (8.74)	3.61* (9.21)

Results expressed as mean (SD).

* p < 0.001; † p < 0.002; ‡ p < 0.04; § not significant.

Table 2 Grade distribution in the sample of candidates (n = 356), comprising 160 males and 196 females, at the final examination

Grade	F	C	C+	B-	B	B+	A-	A	Total
Male	8 (5)	31 (19.4)	73 (45.6)	22 (13.8)	17 (10.6)	4 (2.5)	4 (2.5)	1 (0.6)	160 (100)
Female	3 (1.5)	40 (20.4)	86 (43.9)	22 (11.2)	16 (8.2)	15 (7.7)	8 (4)	6 (3)	196 (100)

Percentages of row total in parentheses.

 χ^2 for trend in proportions = 4.339, p = 0.037.

Performance in the components of the final examination and the aggregate score

The correlation (*r*) between the aggregate score and the short cases, the MCQs, the long case, and the essay were 0.77, 0.71, 0.68, and 0.67 respectively (significant at the 0.01 level).

The mean score deviation of -0.43 ± 7.87 between the scores in the short cases and the aggregate scores showed no significant difference, while the corresponding mean scores of each of the other components of the examination were highly significant (table 1).

The mean score deviation between the aggregate score and the short cases score was the lowest in the high ability group (group 1). In the low ability group (group 3), the lowest mean deviation score was observed between the aggregate score and the score for the essay section.

Performance and gender

Table 2 shows the distribution of grades achieved by the male and female students. The highest proportion of students (males as well as females) fell within the C+ grade. However, there was a small but significant difference (χ^2 for trend in proportions = 4.339, p = 0.037) in favour of the females.

DISCUSSION

In general, candidates who performed well in the individual components of the final examination received higher scores for the overall performance. This suggests that the different sections of the examination assessed similar or related competencies. In addition, the substantially high correlation coefficient between the scores for the short cases and the overall score indicates that the performance in the short cases is a better determinant of the overall rank than the scores in the other sections.

A review of the aggregate score and the scores obtained in the individual components of the final examination on the basis of ability levels revealed that students who are competent on an overall basis had acquired satisfactory clinical skills. However, the less competent candidates showed poor clinical skills, although their level of knowledge may have been satisfactory.

All except one of the seven candidates who failed the final examination in paediatrics had failed in the short cases section. Three of them, however, passed in the long case component. This shows that the short cases would have identified nearly all the poor achievers more effectively than the long case.

The format of the short cases includes close supervision and provides for assessing the candidate in more than one subject area. This would contribute to increasing the validity and the reliability of the assessment, as was reflected in the higher positive correlation between the short cases score and the aggregate score.

This investigation confirms the observation of a previous report that short cases are more useful than the long case in assessing clinical competence.¹

According to this study, female candidates achieved a better level of overall performance than the male students, reinforcing the perception that is held by some that male students, in general, may not be performing as satisfactorily as the females.

Conclusion

This study shows that the short cases component of the final examination in paediatrics is more effective than other components at discriminating among the candidates, with respect to the assessment of clinical skills. To reflect this feature, the relative weighting of the two groupings used at present may need revision. The short cases may also be transformed to an OSCE format, which would enable the assessment of a wider range of individual subskills, thus improving the content sampling of the examination. The long case, which provides the opportunity of assessing the competencies of analysis, synthesis, and evaluation, should, however, be retained within any modifications that may be contemplated.

Authors' affiliations

Z Hijazi, Department of Paediatrics, Faculty of Medicine, Kuwait University, Kuwait

I G Premadasa, Kuwait Institute for Medical Specialization, Kuwait

M A A Moussa, Department of Community Medicine & Behavioural Sciences, Faculty of Medicine, Kuwait University, Kuwait

REFERENCES

- 1 Wakeford R. Examiners' beliefs about the utility and the appropriateness of the different examination techniques. *Med Educ* 1985;19:505-6.
- 2 Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination. *Med Educ* 1979;13:41-54.
- 3 Newble DI. Eight years' experience with a structured clinical examination. *Med Educ* 1988;22:200-4.
- 4 Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length requirements for reliable performance-based examination scores. *Acad Med* 1993;68:224-9.
- 5 Alsaied K, Premadasa IG, Alsaied M, et al. Evaluation of undergraduate students in paediatrics at Kuwait University. *Med Teach* 1995;3:289-96.