Original Article

# From a "Generalist" Medical Graduate to a "Specialty" Resident: Can an Entry-level Assessment Facilitate the Transition? Assessing the Preparedness Level of New Surgical Trainees†

Payal K Bansal,[1]*MBBS, MS (Surg)*, Vivek A Saoji,[1]*MBBS, MS (Surg)*, Larry D Gruppen,[2]*PhD*


## Abstract

**Introduction: Concerns have been raised in the literature about how well the undergraduate curriculum prepares medical students for residency. An assessment was designed and administered to entering postgraduate residents in surgery to test their preparedness vis-a-vis the competence level expected of them at the beginning of their training. This paper explores the role and place of such an assessment in the medical education continuum. Materials and Methods: Faculty members from the Department of Surgery at Bharati Vidyapeeth University Medical College (BVUMC), Pune, India and experts from the Department of Medical Education, University of Michigan Medical School, Ann Arbor designed and administered an assessment based on the multiple-choice question examination (MCQE) and objective structured clinical examination (OSCE) in June 2005 to 24 examinees from 3 different training levels at BVUMC. Results: All subsections of the MCQE showed significant correlation except the breast and endocrine section. The test showed an overall reliability of 0.8 (Cronbach's alpha). The scores and level of difficulty of the OSCE were inversely related. There was a significant difference in performance between the 3 groups and these differences were more pronounced for more complex tasks, specifically the procedural skills station, where the intern performance was particularly poor. Clinical skills reliability was 0.85. The communication skills score correlated well with the clinical skills score and also showed good reliability. Four out of the 5 new residents had below-satisfactory levels of competence for this level. Conclusion: This pilot study reveals definite educational gaps in both knowledge and skills among the residents studied. Such an intervention can be very informative, providing immense educational benefit to the learner, faculty and programme, and has an important place in the continuum of medical training.**


**Ann Acad Med Singapore 2007;36:719-24**

**Key words: Competence, Education, Postgraduate, Surgery**

## Introduction

Undergraduate medical education is considered a continuum leading into postgraduate training and ultimately medical practice.[1] To this effect, it has been suggested that measures of performance of medical students and practising doctors should be aligned and the transition should be a smooth one. However, studies show a poor correlation between medical school performance and intern (resident) performance.[2-4] The various reasons cited for this include intervening time interval,[3] the use of different measures,[2] differing contexts or deliberate measurements of different attributes.[1]

Langdale et al[5] state that although the current medical school curricula provide the educational background, the expectations of the residency programmes, especially with regard to clinical and work-related skills is much higher. The performance expected of residents may not be mere extensions of those expected (or assessed) of medical students, and the "generalist" knowledge and skill base may not be sufficient and appropriate for postgraduate training.

It has also been pointed out that medical students are often told that they would learn certain skills when they became residents; but discovered once they became

[1] Department of Surgery, Bharati Vidyapeeth University Medical College, Dhankawadi, Pune, India
[2] Department of Medical Education and Medical Education Scholar's Program, University of Michigan Medical School, Michigan, USA.
Address for Correspondence: Dr Payal K Bansal, E-5, Green Acres, Salunke Vihar Road, Wanowrie, Pune, India 411048.
Email: payal_pune2000@yahoo.com
† This paper won the Best Oral Presentation (Merit) Award at the 4th Asia-Pacific Medical Education Conference 2007.

residents, that they were expected to have learned them in medical school.[6] This raises concerns about preparedness at the entry level, which, in addition, may be different for different fields. These studies indicate that there appears to be a clear gap between the required skills and those actually possessed by new residents.

Medical schools in India follow a uniform curriculum prescribed by the Medical Council of India, both for undergraduate and postgraduate training. Admission to postgraduate courses is based on a multiple-choice question examination (MCQE) based on the undergraduate curriculum, which is designed and conducted by various universities all over the country. This system was introduced to remove ambiguity in admission procedures as well as for uniformity, and has been in place for approximately 10 years. Concerns similar to those cited above, about the inadequacy of resident preparation, have been expressed in India, where the students are expected to acquire these skills during a 1-year internship, which precedes the awarding of the medical degree. The introduction of the entrance test for postgraduate admissions has significantly diminished the quality of the internship as students spend much of their time preparing for the admission test rather than training on the wards. Educators in India have urged for reforms in this regard.[7]

Medical educators have responded to this challenge through a variety of initiatives. Some specialties have defined expectations for incoming residents for their respective fields.[8-12] A web- and simulation-based curriculum to ease the transition has been described for a surgical internship.[13] In addition, there has been renewed interest in in-training assessments to provide alignment in this continuum.[14,15]

At the entry level, Lypson et al[6] have described a postgraduate orientation assessment using the objective structured clinical examination (OSCE) format for residents from all specialties as an effective instrument for identifying gaps between medical training and residency. However, to our knowledge, no specialty-level assessment has been described that measures preparedness, in terms of desirable knowledge and skills, at the beginning of a residency.

Surgical educators agree[9,13] that the first year of residency is the most challenging and demanding time for surgical trainees and can be quite overwhelming and stressful. The new responsibility of patient care involves a multitude of tasks and requires multiple skills. These include patient evaluation, interpretation of common radiographs and the ability to perform certain procedures. Efficient time management and communication skills are important complementary skills for optimal performance. Having a baseline competency level for these tasks has important implications not only for the supervision of work and delegation of responsibility but for their own motivation and learning as well.

This becomes even more significant in the Indian context, where training in surgical disciplines is primarily through direct contact with patients, as models and simulators have yet to make way into teaching-learning programmes, primarily because of high costs. To enable a smooth transition through this challenging period, we designed, implemented and analysed an entry-level assessment, based on the specific skills and knowledge required at the beginning of a surgical residency. This article explores and discusses the possible benefits of this intervention for residents, faculty and programme, and the place of such an assessment in medical education.

## Materials and Methods

The assessment was a collaborative initiative, as a part of the FAIMER International Fellowship in Medical Education of the first author, the Department of Medical Education, University of Michigan Medical School at Ann Arbor, and experts from the Department of Surgery at Bharati Vidyapeeth University Medical College, Pune, India. The standardised test consisted of a written MCQE for testing knowledge and an OSCE for assessing clinical and communication skills.

The faculty experts reviewed the Handbook of Prerequisites for Graduate Surgical Training,[16] based on which the most common tasks that residents are expected to perform from the very outset of their training were listed by faculty consensus. We then defined the expected knowledge and skills for these tasks and based the MCQE and OSCE on this list. The experts proposed section-wise sampling of the content (Table 1) for the MCQE. It consisted of 100 one-best-response items with 6 subsections [General Surgery, Gastrointestinal (GI) and Hepatobiliary, Urology, Breast and Endocrine and two Symptomatology sections] The OSCE consisted of 6 stations (History taking – Abdominal pain; Examination – General physical examination, Abdominal examination, Thyroid swelling examination; Procedural skills – Aseptic technique, Radiograph interpretation) (Table 1). Two stations used standardised patients and 3 stations used real patients. Checklists for clinical as well as communication skills were prepared. At least 3 experts were consulted for finalising each station. The number of checklist items ranged from 25 to 40 per station. The communication skills checklists from all stations were combined to compute the total communication skills score.

In June 2005, about 1 month after residents joined the programme, a total of 24 examinees – 5 new residents, 6 senior residents and 13 interns took the examination. The MCQE was a 2-hour test. The OSCE was conducted over

Table 1. Summary of the Examination Format

| S. No | Multiple-choice question examination (MCQE) subsections | Objective structured clinical examination (OSCE) stations – content and format |
|---|---|---|
| 1 | Gastrointestinal tract and hepatobiliary | History taking for acute abdomen<br>Standardised patient + checklist |
| 2 | General surgery and paediatric surgery | General physical examination<br>Real patient + checklist |
| 3 | Urology | Abdominal examination<br>Real patient + checklist |
| 4 | Breast and endocrine | Examination of a thyroid swelling<br>Real patient + checklist |
| 5 | Symptomatology I | Radiograph interpretation<br>Radiographic films + paper/pencil test |
| 6 | Symptomatology II | Aseptic technique<br>Standardised patient + checklist |

Table 2. Score Distribution for the Three Examinee Groups

| Examinee group | MCQE percentage score (Mean ± SD) | Clinical skills percentage score (Mean ± SD) | Communication skills percentage score (Mean ± SD) | Percentage final score (Mean ± SD) |
|---|---|---|---|---|
| I | 50.30 ± 12.61 | 52.78 ± 10.02 | 54.88 ± 11.75 | 52.21 ± 10.32 |
| II | 46.20 ± 10.85 | 57.83 ± 13.57 | 65.77 ± 18.14 | 54.77 ± 12.68 |
| III | 62.66 ± 9.56 | 73.64 ± 6.80 | 80.12 ± 8.70 | 70.55 ± 7.49 |
| Total | 52.54 ± 12.71 | 59.05 ± 13.14 | 63.46 ± 16.15 | 57.32 ± 12.55 |

MCQE: multiple-choice question examination; SD: standard deviation

Table 3. Intercorrelations between Variables of Section Scores and Total Score for Multiple Choice Question Examination (MCQE)

| Section | Total | General surgery | Gastrointestinal + hepatobiliary | Urology | Breast + endocrine | Symptomatology I | Symptomatology II |
|---|---|---|---|---|---|---|---|
| Total | 1 | | | | | | |
| General surgery | 0.922**<br>0.000 | 1 | | | | | |
| Gastrointestinal + hepatobiliary | 0.891**<br>0.000 | 0.809**<br>0.000 | 1 | | | | |
| Urology | 0675**<br>0.000 | 0.491*<br>0.015 | 0.447*<br>0.029 | 1 | | | |
| Breast + endocrine | 0.458*<br>0.025 | 0.379<br>0.068 | 0.299<br>0.155 | 0.243<br>0.253 | 1 | | |
| Symptomatology I | 0.697**<br>0.000 | 0.592**<br>0.002 | 0.404<br>0.051 | 0.582**<br>0.003 | 0.249<br>0.241 | 1 | |
| Symptomatology II | 0.691**<br>0.000 | 0.593**<br>0.002 | 0.567**<br>0.004 | 0.415*<br>0.044 | 0.024<br>0.910 | 0.709**<br>0.000 | 1 |

*$P$ <0.01; **$P$ <0.001

2 days, with 12 examinees being examined each day. The length of each station was 10 minutes, with a 5-minute interval for reading instructions and completing feedback forms. All examinations were scored by the faculty members.

Data analysis was done using SPSS version 13.0 at the Department of Medical Education, University of Michigan Medical School, Ann Arbor. Student's *t*-test and one-way analysis of variance (ANOVA) were used for the comparison of mean scores. The Pearson correlation coefficient and Cronbach's alpha were used for reliability analysis.

## Results

The 24 examinees who took the test belonged to 3 different training levels. Group I comprised interns

Table 4. Intercorrelations Between Variables of the Objective Structured Clinical Examination Station – Scores and Total Score

| | Total | Station 1 | Station 2 | Station 3 | Station 4 | Station 5 | Station 6 |
|---|---|---|---|---|---|---|---|
| Total | 1 | | | | | | |
| Station 1 | 0.478*<br>0.018 | 1 | | | | | |
| Station 2 | 0.667**<br>0.000 | 0.406*<br>0.049 | 1 | | | | |
| Station 3 | 0.847**<br>0.000 | 0.141<br>0.511 | 0.534*<br>0.007 | 1 | | | |
| Station 4 | 0.833**<br>0.000 | 0.306<br>0.146 | 0.354<br>0.090 | 0.700*<br>0.000 | 1 | | |
| Station 5 | 0.844**<br>0.000 | 0.365<br>0.079 | 0.427<br>0.038 | 0.613**<br>0.001 | 0.741**<br>0.000 | 1 | |
| Station 6 | 0.799**<br>0.000 | 0.143<br>0.506 | 0.394<br>0.057 | 0.751**<br>0.000 | 0.585**<br>0.003 | 0.572**<br>0.003 | 1 |

*$P$ <0.01; **$P$ <0.001

Table 5. Score Interpretation and Grading of Performance for Competency Level

| Score | Grade | Interpretation |
|---|---|---|
| >70 | Good | • Comfortable and familiar with the tasks expected to perform |
| 60-69 | Satisfactory | • Familiar with the tasks expected to perform<br>• Needs occasional checks/appraisal. |
| 40-59 | Needs improvement | • Needs supervision for some tasks, guidance for others·<br>• Expected to improve by self-study |
| <40 | Unsatisfactory performance | • Needs to retake the test<br>• Needs remediation and close/constant supervision at work |

(n = 13), Group II new residents (n = 5) and Group III senior residents (n = 6). Table 2 shows the mean scores on the various components of the test as well as the total for the 3 groups.

For the MCQE, subjects scored highest, on an average, on the Symptomatology sections and did the poorest on the Breast and Endocrine section. The Symptomatology section scores were significantly higher in all pair-wise comparisons with other subsections ($P$ <0.05). ANOVA showed the difference in performance between the old (Group III) and new (Group II) residents in the General Surgery section to be statistically significant ($P$ <0.05). All sections showed significant intercorrelation as well as with the total score except the Breast/Endocrine section, which showed moderate correlation with the total score only (Table 3). Cronbach's coefficient alpha for the MCQE was 0.80. Subsection reliability co-efficients ranged from 0.33 to 0.78.

For the clinical skills component of the OSCE, the best average scores were seen at the history-taking station (Station 1) while the lowest scores were seen at the procedural skills station (Station 6). Paired sample $t$-test showed that the Station 1 mean was significantly higher than that of other stations except Station 5. The Station 6 mean was significantly lower than all the other stations. ANOVA showed a significant difference in the total mean clinical skills scores for the 3 groups and also for stations 3, 4, 5 and 6. Table 4 shows the correlation between the scores on the different stations, and the total score. Cronbach's coefficient alpha for the clinical skills section of the OSCE was 0.85 indicating that the examination is a reliable measure. The reliability of individual stations ranged from 0.71 to 0.90.

Communication skills were tested at all stations except the radiograph interpretation station. The communication skills score was determined for each of the 5 stations. The mean communication skills score was 63.46 ± 16. (Group I = 54.88 ± 11.76, Group II = 65.77 ± 18.14, Group III = 80.13 ± 8.70). The differences among the 3 groups were statistically significant ($P$ <0.01)

On the clinical and procedural skills stations, the faculty rated the performance of each examinee on a scale of 1 to 9 as per the guidelines provided. The numerical score was termed the global score. The global score correlated well with the clinical skills score and communications skills score, both of which were checklist-based. The final score was a composite of the MCQE score, clinical skills score and communication skills score, weighted as 4, 4 and 2 respectively. The differences among the mean final scores of the 3 groups were statistically significant ($P$ <0.01). The scoring pattern agreed upon for the new residents for an overall interpretation of their performance is shown in Table 5.

Though the OSCE format was new and everyone participated in an OSCE for the first time, the feedback indicated a high acceptance level and a good correlation between self-rating and actual performance. Almost all

participants agreed that the format should be used for testing residents regularly.

**Discussion**

This pilot study was clear in its purpose. It was meticulously planned and prepared in careful, systematic detail. Several experts were involved in its construction. The objectives were defined, blueprinting was done, a combination of standardised assessment methods were used and adequate attention was paid to the response process. By using faculty observers, we ran the risk of introducing observer bias. This was discussed prior to the implementation of the assessment and adequate time was devoted to training and preparation of each faculty scorer by the first author. Correlational analysis showed good intercorrelations among the OSCE stations, communication skills station and total OSCE score and global scores. Since there was only 1 faculty member at each station, whatever observer bias that existed was distributed equally among participating students and thus does not contribute to the variability in the scores.

The data analysis demonstrates the high reliability of the instrument. The scores compare well with the undergraduate scores and the global scores. Even though the sample size was small, this is more of a problem with "no-difference" findings. Obtaining statistically significant differences between groups in spite of small sample sizes indicate a large effect.

The overall evidence supports a well-structured comprehensive entry-level assessment with good psychometric properties, which can be used to draw inferences about the baseline competence level of these new surgical trainees.

The grading pattern segregates the residents into various levels of performance (Table 5). Four of the 5 new residents had a below satisfactory overall performance.

This assessment has been diagnostic in that the performance on skills which are required of new residents in day-to-day practice (physical examination, procedural skills, X-ray interpretation skills, etc) are below expected level. Interns' poor performance on the procedural station indicates that these skills are not being adequately learnt during internship. This is of concern and underlines the gap in skills training at the undergraduate level.

The fact that the scores correlated well with each other and behaved as one would "logically" expect is important validity evidence. Individual scores and subscores helped identify specific areas of adequate performance, as well as deficiencies. Other measures like evaluation by faculty members on the wards, retesting or further in-training assessments would help to confirm this judgment.

Further studies with larger number of examinees and similar studies across other disciplines will strengthen the case for making such assessments an integral part of the training programme.

The educational benefits that such an intervention is expected to provide to the learner, the faculty, the programme and finally, the medical educational continuum are as follows:

*1) Learner Impact*

Two important outcomes emerge for the learners. Firstly, this assessment provides critical information at the start of a new series of learning events. It defines expectations and identifies the strengths and reveals the weaknesses of the learners, at the very outset. This helps to reinforce confidence in tasks done well and provides the opportunity to reflect on areas that need improvement. Secondly, the potential of this opportunity can be harnessed to foster self-direction in the learners from the very beginning. Mentoring by faculty and a clear definition of programme objectives are complementary to the development of this learner autonomy and must be provided concurrently. The learners can individualise their learning plan. A developmental portfolio can play a useful role by guiding reflection and monitoring progress. It can provide insight to the learner regarding the utility and relevance of the assessment, which are critical to the success of the assessment in the learning cycle.[17]

*2) Faculty Benefit*

Various studies confirm the complexity of evaluating resident performance and the challenges faced by faculty members in this regard.[18-21] Useful, consistent and valid feedback is difficult to provide and the search continues for more structured and standardised formats to facilitate this process.

A baseline standardised assessment such as this serves to inform the faculty members and provides a basis for focused and meaningful feedback to learners. It makes the evaluation of progress more objective and diminishes the chances of faculty bias towards the learner. In India, each resident has a faculty dissertation advisor, who is also a mentor. This informal mentoring role can be given a more formal structure, where the responsibility of helping the learner achieve self-direction and develop autonomy in learning rests with the mentor. It would allow for closer observation of resident performance and enhance the quality of faculty evaluations. Areas with consistently poor performances can help identify topics for the formal didactic programme and remediation given when required.

*3) Implications for the Programme*

This is a formative assessment. However, since it is based on expectations at entry level, it is important for those not

meeting the expectations to do so at the earliest. How soon does the programme expect the residents to achieve at least a satisfactory performance? How will it be measured? Until such time, how does the resident continue to work – supervised or unsupervised? What will be the consequences if the desired levels are not achieved? These are challenging questions which do not have clear, definite answers. Any decision in this regard would require deliberation and discussion amongst faculty members and even then is at best tentative until further validation of the instrument or till improvement in performance to desirable/expected levels is demonstrated by the learner.

More similar assessments to increase the psychometric strength of the instrument, further in-training assessments or linking it with in-training assessments if they already exist and addition of other evaluation methods are measures that would enhance the relevance and usefulness of this assessment for the residency programme.

### *4) Role in the Medical Education Continuum*

A well defined set of expectations at the beginning of residency combined with an assessment to determine the residents' own level vis-à-vis these expectations provides information to the residents that can enable them to feel confident about their strengths and examine their areas of deficiency. A collaborative strategy between residents and faculty members to address these deficiencies, along with a clear understanding of expectations with regard to performance can enhance their work efficiency as well as learning. This, in turn, can be expected to translate into better patient care and enhanced motivation for learning. In addition, it will reduce the anxiety and stress associated with the beginning of a residency programme. An assessment at this level can therefore play a critical role in providing the missing link between the undergraduate and postgraduate levels in the medical education continuum.

***Acknowledgements***

*The authors thank the Foundation for the Advancement of International Medical Education and Research (FAIMER), Philadelphia for their support through the International Fellowship in Medical Education 2004, awarded to the first author. The study was carried out as a part of the Fellowship.*

*We gratefully acknowledge S Chandorkar, MN Ketkar, S Shinde, R Doshi, S Patankar, Priti Shah, H Lekawle and V Athawale, all of whom are faculty at the Department of Surgery, Bharati Vidyapeeth University Medical College, Pune, India for their wholehearted support and active participation in the project. We also thank Mary Oh, Consultant, Data Management and Analysis Group at the Department of Medical Education, University of Michigan Medical School, Ann Arbor, USA for her excellent help with the data analysis.*

REFERENCES

1. Wilkinson TJ, Frampton CM. Assessing performance in final year medical students. Can a postgraduate measure be used in an undergraduate setting? Med Educ 2003;37:233-40.
2. Gonnella JS, Hojat M, Erdmann JB, Veloski JJ. A case of mistaken identity: signal and noise in connecting performance assessments before and after graduation from medical school. Acad Med 1993;68(suppl): S9-S16.
3. Hojat M, Gonnella JS, Veloski JJ, Erdmann JB. Is the glass half full or half empty? A re-examination of the associations between assessment measures during medical school and clinical competence after graduation. Acad Med 1993;68(2Suppl):S69-S76.
4. Kahn MJ, Merrill WW, Anderson DS, Szerlip HM. Residency program directors evaluations do not correlate with performance on a required 4th-year objective structured clinical exam. Teach Learn Med 2001;13: 9-12.
5. Langdale LA, Schaad D, Wipf J, Marshall S, Vontver L, Scott CS. Preparing graduates for the first year of residency: are medical schools meeting the need? Acad Med 2003;78:39-44.
6. Lypson ML, Fronha JG, Gruppen LD, Wooliscraft JO. Assessing resident's competencies at the baseline: identifying the gap. Acad Med 2004;79: 564-70.
7. Bansal RK. Need for strengthening of internship (rotatory housemanship) training in India. Educ Health(Abingdon) 2004;17:332-8.
8. Bass EB Fortin AH, Morrison G, Wills S, Mumford LM, Goroll AH. National survey of clerkship directors in internal medicine on the competencies that should be addressed in the medicine core clerkship. Am J Med 1997;102:564-71.
9. Graduate Medical Education Committee. Prerequisite objectives for graduate surgical education: a study of the Graduate Medical Education Committee American College of Surgeons. J Am Coll Surg 1998;186: 50-62.
10. Martens FM, van der Vleuten CP, Grol RP, op't Root J, Crebolder H, Rethans JJ. Educational objectives and requirements of an undergraduate clerkship in general practice. The outcome of a consensus procedure. Fam Pract 1997;14:153-9.
11. Committee on Education in Obstetrics and Gynaecology and Human Reproduction. Standards of education in obstetrics and gynecology. Report of the Committee on Education in Obstetrics and Gynaecology and Human Reproduction. Int J Gynaecol Obstet 1995;50:85-95.
12. Sherokman B, Cannard K, Miller JQ. What should a graduating medical student know about neurology? The American Academy of Neurology Undergraduate Education Subcommittee. Neurology 1994;44:1170-6.
13. Meier AH, Henry J, Marine R, Murray WB. Implementation of a web- and simulation based curriculum to ease the transition from medical school to surgical internship. Am J Surg 2005;190:137-40.
14. Ringsted C, Henriksen AH, Skaarup AM, van der Vleuten CP. Educational impact of in-training assessment (ITA) in postgraduate medical education: a qualitative study of an ITA programme in actual practice. Med Educ 2004;38:767-77.
15. Hays R, Wellard R. In-training assessment in postgraduate training for general practice. Med Educ 1998;32:507-13.
16. American College of Surgeons: Committee on Resident Education: Prerequisites For Graduate Surgical Education. A Handbook for Medical Students and First-Year Surgical Residents. Available at: http://www.facs.org/education/sre/preguide.html. Accessed 7 January 2005.
17. Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? Med Educ 1999;33:276-81.
18. Herbers JE Jr, Noel GL, Cooper GS, Harvey J, Pangaro LN, Weaver MJ. How accurate are faculty evaluations of clinical competence? J Gen Intern Med 1989;4:202-8.
19. Noel G, Herbers JE Jr, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do Internal Medicine faculty members evaluate the clinical skills of residents? Ann Intern Med 1992;117:757-65.
20. Ringdahl EN, Delzell JE, Kruse RL. Evaluation of interns by senior residents and faculty: is there any difference? Med Educ 2004;38:646-51.
21. Holmboe ES, Hawkins RE. Methods for evaluating the clinical competence of residents in internal medicine: a review. Ann Intern Med 1998;129:42-8.