

Medical Students' Training For a Final Examination in Surgery – The Technion Experience

Nahum Rosenberg MD, Rachel Nave M.Sc, Zeev Gazit MA, Shmuel Reis MD,
Michael Soudry MD , Ahud Sternberg MD

The Ruth & Bruce Rappaport Faculty of Medicine
Technion- Israel Institute of Technology
Haifa, Israel

Abstract: There is a controversy on the best educational method for improving medical graduates' achievements in final examinations. We hypothesize that final “tuning” of knowledge by a mock examination is effective for success in the final examination.

In the years 2003-2005, graduates of the Faculty of Medicine in Technion, Israel Institute of Technology were exposed to mock examinations in general surgery and surgical sub-specialties. The 2004 and 2005 examinations were validated. The standardized (according to the annual national average) scores were compared for graduates in 2003-2005 and 2002.

There was no significant change in the results among Technion graduates for the year 2003 national examination compared to the year 2002, but in 2004 and 2005 the Technion graduates significantly improved their standardized scores compared to the 2002 graduates ($p=0.01$, $p<0.001$ respectively). There was no additional educational intervention in the curriculum for the years 2003-2005 in comparison to the 2002 class. Therefore we suggest that a validated mock examination, designed according to the final examination format combined with a debriefing session, improves the medical graduate's achievements in the final certification examination.

According to the Israeli Higher Education authority the requirements for the MD certification of medical school graduates include a successful pass of a national examination in general surgery and the surgical subspecialties at the minimal score of 60%. The performance in this examination should reflect the level of knowledge of the graduates in different aspects of surgery at a level sufficient for their future general practice, providing that the test is validated at the reliability level of at least 0.8 (Novik MR and Lewis C 1967, Cronbach LJ 1951). In the Ruth & Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, the students are instructed in general surgery and the surgical subspecialties during focused clerkships of 7 weeks during the fifth year of their studies. An additional 15 weeks are dedicated to the surgical sub-specialties' clerkships during the sixth year of studies. There are nine general surgery, four orthopaedics, three intensive care, four urologic, two plastic surgery and two vascular surgery wards that are affiliated with the faculty of medicine and involved in clinical teaching of the students. Although the surgical teaching curriculum is uniform for all the teaching wards, some deviations in instruction and emphasis on the important subjects can occur according to the personal preferences of the lecturers and tutors. Therefore, a final controlled “tuning” of the provided knowledge might be important in order to reach a desired uniform level of knowledge that is essential for the students' success in the final examination. Because

the degree of insufficient understanding of different topics varies among the students, especially because they are instructed in different teaching environments, it is important that all the curriculum topics will be reviewed with the whole forum of the students who are eligible to sit the final examination. In order to achieve this goal of a final verification of essential knowledge, we chose to expose the students to a preparation mock examination (PE) in general surgery and surgical subspecialties that is structurally similar to the final examination. The PE was followed by didactic explanation of the unclear or controversial topics before the full forum of students immediately after analyzing the PE's results and prior to the final national examination. We hypothesized that the PE, combined with systematic review of the problematic or unclear topics revealed by this examination, will improve the performance of the students in the final examination. We hereby present the three years' experience and evaluation of the implementation of this method in our institution.

Materials and Methods

During three years, from 2003 to 2005, the medical school graduates took a preparation examination (PE) two weeks prior to the national final examination in general surgery and surgical subspecialties.

The test comprised 150-160 multiple choice questions (MCQ) based on the national final examination blueprint.

The PE was designed according to the following guidelines:

- The test comprised 150-160 multiple choice questions (MCQ).
- The number of questions in each surgical subspecialty was determined according to the proportion of every subject in the previous year's final national examination.
- In the year 2003 examination, the questions were drawn from the database of the previous internal faculty tests. The 2004 and 2005 PEs were freshly written by representatives of each of the affiliated surgical and surgical subspecialty wards. These representatives were instructed by an item-writing workshop. The items were quality-controlled by a dedicated team. Care was taken not to involve members who were active in preparing the national test.
- In preparation for the 2005 test, an additional validation of the PE by a group of interns was implemented. Five surgical interns, who were 6-12 months after their final certification examination in surgery, took the PE and submitted suggestions for insufficiently formulated MCQ items. These items were either corrected or entirely omitted from the final PE version.
- In the years 2004 and 2005, following the PE, an instructional session that aimed to help students figure out topic areas where they needed more preparation took place before the whole forum of graduates.

The Technion graduates' scores on the national ex-

aminations were standardized by dividing the individual scores by the overall average results of all graduates from other three Israeli medical schools in each year between 2002-2005. We discuss the standardized values rather than raw scores in order to eliminate the effect of variations in test difficulty.

As already mentioned, the Technion graduates in 2002 were not exposed to the PE. Therefore the standardized results of the Technion graduates in national examinations in the years 2003-2005 were compared to the standardized national examination results in 2002 by the t-test after verifying the normal distribution; p values below 0.05 established a statistically significant difference between the groups. The reliability values of the PE and the national examination were estimated by Cronbach's alpha.^{1,2} We considered alpha values above 0.8 as evidence of adequate reliability.

The numbers of Technion graduates who took the final national surgery examination between 2002 – 2005 were 77, 67, 68 and 64, respectively, for each year.

Results

The reliability coefficients of the PEs for 2003-2005 were 0.66, 0.54, and 0.87, respectively. The reliability coefficients of the national surgery examinations between 2002-2005 were 0.8, 0.8, 0.82, and 0.8, respectively. There was no significant change in the standardized results among Technion graduates in the year 2003 national examination comparing to their standardized results in the year 2002, but in the years 2004 and 2005 the Technion graduates significantly improved their standardized

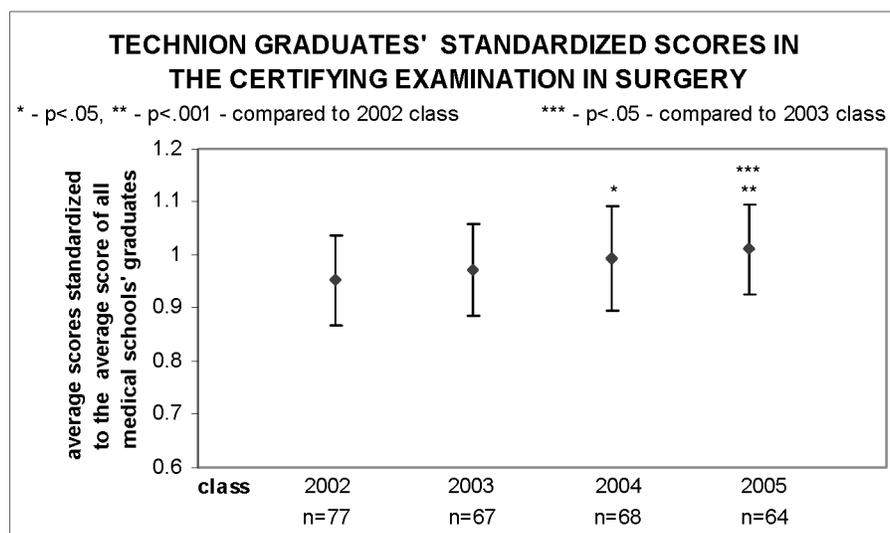


Figure 1: Vertical bars represent standard deviation from the mean values.
n = number of students who took the test

results in the national examinations as compared to the 2002 class ($p=0.01$ and $p<0.001$, Figure 1). The class of 2005 had significantly higher standardized scores on the national examination compared to the year 2003 graduates ($p=.008$) but no significant change compared to the 2004 class. The standardized results of the 2004 class were not different from the 2003 one (Figure 1).

Discussion

There is a controversy about the best educational method that is effective for improving medical graduates' achievements in final examinations.^{3,4} There is some previous evidence, from a study on dentistry students, that coaching students by mock examinations can improve their performance on the final written examination by revealing the weak areas in their knowledge and by exposing the students to the previously unfamiliar examination format.⁵ The authors are not aware of a similar study on medical students about preparation for a final written examination in surgery. There are several reports of a positive impact of mock examinations for preparation for an oral board examination in surgery,^{6,7} but because of the different basic structures of oral and written MCQ-type examinations, it may be difficult to deduce the effectiveness of the written PE from the reported experience with a oral mock examination. Therefore, the presented data on the effectiveness of a written PE might be of value for its implementation as an effective educational method for preparation for a final written certification examination in surgery.

We present here a PE intervention, prior to the final national examination, that was designed to improve the final examination performance. Since we were not aware of other didactic evidence for proven tools for constructing an effective written PE, we approached the design of the PE empirically by building on our previous experience.

The PE in the year 2003, that was composed of questions from the pool of previous internal examinations in surgery, didn't cause any improvement in the students' subsequent national examination standardized results, and its reliability was low (at 0.66).

In the following two years, when the newly designed PE was implemented, the graduates significantly improved their national examination scores (the 2004 and 2005 examinations). Although a newly designed year-2004 PE, with the addition of past-exam debriefing, presented a parallel significant improvement in the national standardized results comparing to the national 2002 test, its reliability was still lower than desired at 0.54. The 2004

PE didn't appear to contribute to a significant improvement in the national examination scores compared to the year 2003 examination, probably because both the 2003 and 2004 PEs had low reliability values; i.e, the addition of the post-PE review process in 2004 came with parallel improvement in the national test results comparing to the 2002 but was not sufficient to improve the results comparing to the 2003 national test when the PE was initially implemented. At this stage the identified weak point of the PE was addressed in order to improve its reliability coefficient. We aimed to achieve the desired 0.8 value of the reliability coefficient, the value consistently achieved in the national examination. Therefore, in the year 2005 the PE design was supplemented by re-validating the test with the surgical interns, who had taken their national final tests in the previous year. Following this process, the reliability value of the 2005 PE increased to 0.87. Subsequently, after the exposure to the PE process the year 2005 graduates achieved significantly higher standardized results in the national examinations as compared to the years 2002 (prior to the PE implementation) and to 2003 when the students were exposed to the initial PE. But the year 2005 results were not significantly different from the year 2004 scores. Therefore further improvement tools are needed for the future design of a better PE process. Because there were no recent radical changes in the teaching curriculum of general surgery and its subspecialties in the Faculty of Medicine in the Technion, we can attribute these changes, at least partially, to the PE design, being also aware of the multi-factorial nature of the graduates' achievements in the final examinations.⁵

We are aware of the limitations of this study design: this is a description of an educational intervention and a retrospective analysis of its results, and as such it is inferior to a prospective experimental design. Nevertheless, it addresses the essential attributes of an educational intervention: reliability, validity, feasibility, acceptability and educational impact. The nature of faculty life makes a prospective experimental design in this area practically impossible.

In conclusion we have demonstrated that a PE that is equivalent to the national examination blueprint and supported by validation processes and a focused review session played a role in improving of medical graduates' achievements in the final national examinations in surgery.

References

1. Novik, MR, Lewis C. Coefficient alpha and the reliability of composite measurements. *Psychometrika* 32:1-13,1967.

2. Cronbach L J. Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297-334, 1951.
3. Werner LS, Bull BS. The effect of three commercial coaching courses on Step One USMLE performance. *Med. Ed.* 37: 527-531, 2003.
4. Baerheim A, Meland E. Medical students proposing questions for their own written final examination: evaluation of an educational project. *Med. Ed.* 37: 734-738, 2003.
5. Dadian T, Guerink K, Olney C, Littlefield J. The effectiveness of a mock board experience in coaching students for the Dental Hygiene National Board Examination. *J. Dental. Ed.* 66: 643-648, 2002.
6. Stawski WS. Evolution of a mock oral board examination program in surgery. *Am. Surg.* 60:603-605,1994.
7. Sako EY, Petrusa ER Jr, Paukert JL. Factors influencing outcome of the American Board of Surgery certifying examination: an observational study. *J. Surg. Res.* 105: 75-80, 2002.

Correspondence

Nahum Rosenberg MD
Department of Orthopaedic Surgery "A"
Rambam Medical Center
POB 9602
Haifa 31096, Israel

Tel: +972 4 8542527
Fax: +972 4 8542022
Email: nahumrosenberg@hotmail.com