Assessing the Impact of Increasing Exam Burden on Final Year Medical Students - an observational study

AUTHOR(S):
- Heather Keenan, East and North Herts Trust, UK (Presenter)
- Jack Haywood, East and North Herts Trust, UK

ABSTRACT

Background: Final-year medical students in the UK face a growing burden of examination. In addition to written and clinical exams, they must also take the situational judgement test (SJT) and prescribing skills assessment (PSA). In 2022, the Medical Licensing Assessment (MLA) will begin. Students are therefore never far away from the next exam during this crucial year. The aims of this study were: 1. To evaluate whether proximity to their next exam impacts student perception of teaching quality, and; 2. To assess how proximity to exams impacts student self-reported learning priorities.

Summary of Work: We designed four teaching sessions to deliver to approximately 50 final-year students, delivered six times by the same facilitator. Students completed a questionnaire after each session, rating their current priority (preparing for exams versus preparing to become a doctor), the relevance of the teaching session to this priority, and the overall quality of the teaching on a five-point Likert-type scale. These ratings were correlated against number of days until the next examination, of which the SJT, PSA and first day of finals were considered.

Summary of Results: Students were on average 37.4 days away from their next exam when teaching sessions were delivered (min 3 days, max 92 days, SD 26.8). Data collection is ongoing. Early results indicate: 1. The closer students are to an exam, the more they prioritise passing exams above learning to become a doctor. 2. Students rate teaching sessions that they consider relevant to their current needs more highly. 3. Student rating of teaching session quality decreases closer to exams.

Discussion and Conclusions: As we move towards introducing more assessments for medical students, in addition to an already significant burden of other exams, it is important to consider the impact this has on the ultimate goal of producing effective doctors. Our results indicate that if there is always a major exam looming on the horizon, we run the risk of entirely removing their focus from learning how to be an effective doctor.

Take-home Messages: We should be wary of adding yet more examinations to the already crowded schedule of final-year medical students.
Licensing examinations - a retrospective: (Un)intended emphases in Germany's national licensing examinations?

AUTHOR(S):
- Nikolas Psathakis, German Medical Students' Association (bvmd e.V.), Germany (Presenter)
- Aurica Ritter, German Medical Students Association (bvmd e.V.), Germany
- Lucas Thieme, German Medical Students Association (bvmd e.V.), Germany
- Birgitta Kütting, The German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy, Germany
- Jana Jünger, The German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy, Germany

ABSTRACT

Background: By 2021, the German National Institute for state examinations in Medicine, Pharmacy and Psychotherapy (IMPP) will have devised a new content outline for Germany's national licensing examinations (NLE) with the intention of shifting to a more competency- as well as role-focused assessment based on Germany's non-binding national curriculum ('NKLM', 2015) and the CanMEDS framework. A joint research group consisting of IMPP researchers, medical students from various German medical schools coordinated by the German medical students' association (bvmd) conducted a 'post-examination' review of prior medical licensing examinations as an initial benchmark on the representation of the different professional roles and specialties in medicine under the currently effective content outline from 2013, ('GK2', 4th revision)

Summary of Work: Throughout 2018, the IMPP and bvmd reviewed 16 NLEs. (Spring 2015 - Fall 2018) The examinations analysed consist of 8 similar to the USMLE Step 1 as well as 8 similar to the USMLE Step 2CK resulting in 5160 items analysed. For data reliability, standardised procedures in conjunction with a specialised online platform have been used.

Summary of Results: Our preliminary results of the item review indicate an imbalance in terms of which professional roles and competencies in terms of past exams are focused on. The final data will include the spread of disciplines, roles and competencies plus secondary indicators e.g. in- or outpatient setting etc. - separately for both Step 1/2CK equivalents.

Discussion and Conclusions: As Germany's current content outline for its NLEs is focusing mainly on traditionally-organised knowledge, its balancing of professional roles and competencies struggles to fully comply with coming standards and expectations. A standardised and systematic review as conducted can yield a more comprehensive picture of medical examinations' strengths and weaknesses.

Take-home Messages: Our results show that reviewing passed examinations by students is a helpful tool for the conceptual design process of future national licensing examinations. Joint working groups including both fully-licensed physicians and students might prove beneficial in including all relevant perspectives.
Development of national licensing exam

AUTHOR(S):

- Zalim Balkizov, ASMOK, Russia (Presenter)
- Tatiana Semenova, Ministry of Health of Russia, Russia
- Zhanna Sizova, Federal Center for Accreditation of the Ministry of Health of Russia, Russia

ABSTRACT

Background: Medical education system in Russia requires 6 years of formal undergraduate training in medical school and 2 years of training in residency programmes. Access to the medical profession based on diploma and formal certification after residency programs. Following problems was identified: Certification by providers of education leads to conflict of interests and low quality of assessment. Health practitioners complains on low level of skills in graduates. Procedure of certification is very formal and compromised.

Summary of Work: New regulation (based on the Federal law) implements 3 types of accreditation - primary after graduation from the medical school, primary specialized after residency programs, and reaccreditation every 5 years of practice. Both primary and primary specialized accreditation includes 3 steps - MCQ, OSCE and oral examination. New institution was founded to run the system for whole Russia - Federal Center for Accreditation. Center develops all assessment materials including MCQs, OSCE stations and tasks for oral exams.

Summary of Results: Primary accreditation was implemented in 2017 for medical school graduates and independent assessment of clinical skills became strong motivation for both - students and faculty members to improve training in clinical skills centers.

Discussion and Conclusions: Implementation of OSCE is very limited - only 5 stations was adopted nationwide for the accreditation step 2. However, independent assessment of clinical skills became strong motivation for both - students and faculty members to improve training in clinical skills centers. Oral exams demonstrates low validity and going to be replaced by long cases. Selection and development of independent examiners turned out to be biggest challenge in organization of exams.

Take-home Messages: Implementation of new methods of assessment requires support from the regulators and from the professional society.
The impact on Angoff standard setting of changing to open-resource assessment

AUTHOR(S):
- Mike Tweed, Otago Medical School, New Zealand (Presenter)
- Jen Desrosiers, Otago Medical School, New Zealand
- Tim Wilkinson, Otago Medical School, New Zealand

ABSTRACT

Background: Assessment of healthcare students should reflect eventual practice. Healthcare professional practice encourages use of a wide range of resources, looking up information when needed, which is preferable to guessing. However, implementing resource availability in student examinations ('open-book') might alter the expected performance. We explored the potential impact on standard setting of an open-resource examination.

Summary of Work: Two pools (A and B) of 40 MCQs were created. Students sat two exams, each of 60 minutes, made up of each pool in randomised order, with access to resource also randomised. The closed instructions were standard. The open instructions allowed any resource to be used, but no communication. Prior to the students sitting the exams, a panel of five standard setters was convened. Following discussion of the format of the examination, student ability and the instructions to the students, the panelists undertook a process of judging proportion correct for minimally competent students. For 10 practice questions, this was done considering closed instructions and open instructions, with discussion of judgments after each question. Following this, the panelists made independent judgments for the 80 questions without discussion.

Summary of Results: The overall pass standard was different for pool A (59.4% for open resource instructions v 47.9% for closed, p<0.001) and pool B (58.8% for open v 46.5% for closed, p<0.001). There was no difference in variance of the pass-mark between open and closed instructions for both pool A and B. Of the 80 individual questions, 78 were deemed to have a higher pass-mark under open conditions, 2 a higher pass-mark under closed conditions, but no differences were statistically significant. There was no difference in variance of the pass-mark between open and closed instructions for any question.

Discussion and Conclusions: The panelists judged that exams would be easier under open resource conditions, and adjusted their decisions accordingly. In reality, there was no difference between in correct scores by instruction type. This study was underpowered for question level analysis. Areas for exploration include how resource utility affects judgments and performance.

Take-home Messages: Standard setters expected, potentially erroneously, better student performance due to availability of resources.
Setting defensible standards in small cohorts: understanding when borderline regression works

AUTHOR(S):
- Matt Homer, Leeds Institute of Medical Education, UK (Presenter)
- Richard Fuller, LIME, UK
- Godfrey Pell, LIME, UK
- Jennifer Hallam, LI ME, UK

ABSTRACT

Background: Standard setting, particularly in OSCEs, is always challenging, especially in examinations with small cohorts (e.g. n<50). Examinee-centred approaches (e.g. borderline regression - BRM) are generally thought to be problematic in such context, and institutions often prefer to rely on test-centred approaches such as Angoff, which can be time-consuming and are not particularly reliable in themselves.

Summary of Work: In three quite different small-cohort OSCE contexts, we investigate station- and test-level BRM metrics, and the visual relationship between global grades vs. total scores, to better understand where BRM can ‘work’: 1. PLAB2 - The examination of international medical graduates wanting to practice in the UK (n=30-35 per administration, 18 stations per exam, 198 administrations, station level data). 2. Senior sequential examination of undergraduate students in a large medical school - only a small proportion of candidates have to sit the second sequence (n=20 to 50, 10-12 stations with 25-26 stations in full sequence, four administrations, candidate level data). 3. Physician associates (n=25, 16 stations, two administrations, candidate level data) - successful candidates will practice as part of a team alongside fully qualified doctors.

Summary of Results: We find that BRM works well in the vast majority of stations (the percentage of problematic stations are 5%, 18% and 9% respectively across the three contexts). Where problems occur, this is generally due to an insufficiently strong relationship between global grades and checklist scores to be confident in the standard set by BRM for these stations (i.e. low R squared), often as the result of insufficient spread in scores in these stations.

Discussion and Conclusions: From a practical point of view, extant cut-scores, preferably based on previous satisfactory station performance in a large cohort, should be available for all stations in small cohort exams so that when problems with BRM are diagnosed (e.g. a lack of spread of scores) these cut-scores can substitute.

Take-home Messages: We have shown that where there is sufficient spread of ability in a small cohort, BRM will generally also provide defensible standards but, as ever, care needs to be taken in the overall design of station level scoring instruments.
Standard Setting Using a Modified Post-Examination Borderline Group Method for Simulation-Based Performance Assessment

AUTHOR(S):

- Na Jin Kim, The Catholic University of Korea, College of Medicine, Seoul, South Korea (Presenter)
- Young-Min Kim, The Catholic University of Korea, College of Medicine, Seoul, South Korea
- Sung A Lee, The Catholic University of Korea, College of Medicine, Seoul, South Korea
- Su Young Kim, The Catholic University of Korea, College of Medicine, Seoul, South Korea

ABSTRACT

Background: We have evaluated various standard setting methods to find an optimal standard setting method for simulation-based performance assessment (SBPA) and remediation in our medical college. In 2018, we have implemented a modified post-examination borderline group method (PBGM) for the SBPA to compensate the rater qualification problem in examinee-centered standard setting during the examination.

Summary of Work: We implemented a SBPA comprised of 6 clinical performance examination and 6 OSCE stations for procedural skills at the end of 3rd year clinical clerkship. The SBPA was conducted for 2 days with 2 different sets of 12 stations. During the examination, content experts and experienced standardized patients assessed the students' performance using checklists and a global rating scale (outstanding-clear pass-borderline-clear fail). After examination, 13 experienced standard panelists have decided the cut scores through a modified PBGM with 3-round setting procedure. We compared the cut scores and failure rates calculated from the 4 methods - PBGM, BGM and borderline regression method (BRM) during examination, and a relative method [mean minus one standard deviation (SD)].

Summary of Results: During 3 rounds of the setting procedure, SD of the setters' scores were decreased. Overall cut score (sum of the cut scores) calculated from the PBGM were similar to the BRM but higher than those of the BGM and relative method. Failure rates by two decision rules including overall cut score and minimum passing station numbers were different among the methods (15.1% in BRM, 12.8% in PBGM, 10.5% in BGM and 5.8% in the relative method). In the post-setting survey, the panelists were responded that the modified PBGM is easier than modified Angoff method which was implemented in 2017 and is more appropriate for standard setting for the SBPA in our college. In an on-line survey, 89% of students responded the SBPA and remediation were helpful for their competency improvement.

Discussion and Conclusions: A modified PBGM was feasible and practical for standard setting in a SBPA and remediation in our medical college.

Take-home Messages: A modified PBGM can provide reasonable and defensible approach to standard setting for relatively low-stakes, criterion-referenced SBPA in a medical school or college.