SN Soc Sci (2022) 2:200
https://doi.org/10.1007/s43545-022-00510-9

ORIGINAL PAPER

Transition from locally developed, faculty-written (LFW) examinations to the National Board of Medical Examiners (NBME) subject examinations in clinical medical education

Atalay Demiray[1] · Çınar Öztosun[1] · Simge Acar[1] · Sibel Sakarya[2] · Tuğba Gürsoy[3] · Barış Ata[4]

Received: 20 February 2022 / Accepted: 19 August 2022

Abstract

Background: Written examination composed of single answer multiple-choice questions (MCQs) is one of the common assessment methods in medical education. In Koç University School of Medicine (KUSOM), there has been a transition from locally developed, faculty written (LFW) MCQ exams to third-party exams such as the National Board of Medical Examiners (NBME) subject examinations. In this article, we examined this transition period in regarding scores and feedbacks. Methods: During the transition period between 2018 and 2020, KUSOM employed both exams (LFW and NBME) for clinical clerkships. Both exams were evaluated and analyzed quantitatively comparing students' scores and qualitatively using students' feedback. We compared students' scores from both exams for interrater agreement using Cohen's kappa coefficient. Seventy-five students gave written feedbacks comparing both exams in terms of content validity, language & comprehensibility, and objectivity/fairness. Results: The total of 286 pairs of scores from ten dual examinations indicated fair to moderate agreement with an unweighted kappa value of 0.41 between two exams in terms of ranking and pass/fail decision. Most students indicated a preference towards to use of NBME subject exams instead of LFW exams as a written assessment method in clinical clerkships except internal medicine clerkship. Conclusions: Third-party exams were found to be a useful tool for improving assessment of knowledge in clinical medical education. For English-taught medical schools in non-English speaking countries, like KUSOM, NBME subject examinations can replace LFW exams due to better use of English language, comprehensibility, and fairness with an acceptable agreement in decisions for students' grading.

Keywords Assessment · Multiple choice questions · Third party examination · Examination compatibility

✉ Barış Ata
barisata@ku.edu.tr

Extended author information available on the last page of the article

Published online: 19 September 2022

Abbreviations

AD	Average absolute deviation
F	Fail
KUSOM	Koç University School of Medicine
LFW	Locally developed, faculty-written
M	Median
MCQs	Multiple-choice questions
mini-CEX	Mini-clinical examinations
NBME	The National Board of Medical Examiners
OSCEs	Summative objective structured clinical examinations
P	Pass

Background

Single answer multiple-choice questions (MCQs) are commonly used as a method of summative assessment in most medical schools for the assessment of knowledge along with other methods used to assess clinical skills (Tabish 2008). MCQs aim to test the "knows" and "knows-how" levels of Miller's pyramid and can test students' higher order thinking and clinical reasoning, especially when they involve clinical scenarios or vignettes (Al-Wardy 2010). MCQ exams can be prepared either by a medical school's own faculty or by third parties. The former is commonly used in European medical schools (Custers and Cate 2018; Devine et al. 2015), while most North American medical schools employ the National Board of Medical Examiners (NBME) subject examinations (NBME 2016, 2019).

Education in Koç University School of Medicine (KUSOM) lasts for six years, the first three years of which are pre-clinical and the last three years of clinical training. The medium of education is English except for clinical patient encounters (Koç 2021). Multiple assessment methods, including formative mini-clinical examinations (mini-CEX), summative objective structured clinical examinations (OSCEs), and MCQ exams are used to evaluate students' performance. Students are required to pass each component of the assessment to successfully complete a clinical clerkship. Single answer MCQ exams were prepared by KUSOM faculty, as in all other medical schools in the country. Recently, we decided to examine whether we can replace our locally developed, faculty-written (LFW) examinations with NBME subject examinations for clinical clerkships. The motivations for such a transition and its implications are multiple and discussed later.

Herein we describe how we compared results from LFW examinations and NBME subject examinations in clinical clerkships and define our transition to foreign third-party MCQ exams in clinical clerkships.

Methods

Timeline of the transition process

The assessment committee of KUSOM continuously analyzes the quality of test items and collects feedback from students and faculty members. Assessing the reliability and validity of LFW exams has been a constant challenge due to the small number of students sitting for each exam, e.g., 20–25 students per exam. Yet, we noticed that the discriminatory indices of our MCQs were less than optimal. Therefore, reliable MCQ exams that have been validated on larger samples arose as an option. Also, the use of reliable and externally validated third-party exams would enable: (i) comparison of the output of our curriculum with that of other schools employing the same exams, (ii) monitoring students' performance across years, (iii) provision of assessment independent from the educator, (iv) improving comprehensibility of MCQs in terms of language, (v) to save time and effort of faculty members, and (vi) to have a flexible exam schedule. Thus, we decided to implement third-party examinations at KUSOM following a comparison of the results with LFW and receiving student and faculty feedback.

Given the long track record of NBME subject examinations with reproducible results across different schools over the years and their international availability, they were regarded as a reliable and valid option to replace our LFW exams in clinical clerkships (NBME 2016). In the 2017–2018 academic year, BA, both the director of the Obstetrics and Gynecology (Ob & Gyn) clerkship and the Vice Dean for educational affairs introduced the NBME subject exam in addition to the LFW exam for the first time. Twenty-five students took both the LFW and the NBME subject exam at the end of the Ob & Gyn clerkship. The results of this first trial including student feedback were presented at general faculty meetings. Following an extensive discussion of pros and cons, the Deanship decided to move forward with comparisons with LFW in all clinical clerkships, for which NBME subject examinations were available, and allow switching to NBME subject examinations when clerkship directors saw fit. The decision for employing LFW or NBME subject exams rested by clerkship directors after evaluating the validity, reliability, acceptability, and applicability of NBME subject exams.

In the following year, the Ob & Gyn clerkship decided to employ only the NBME subject exam as a written assessment method and used the NBME subject exam to evaluate the clerkship performance of students. Other clinical clerkships including internal medicine, pediatrics, surgery, neurology, and psychiatry offered NBME subject exams alongside the LFW exams at the end of clerkships to assess the agreement between scores. In the 2019–2020 academic year, pediatrics clerkship, in the 2020–2021 academic year, neurology, psychiatry, and surgery clerkships decided to employ only NBME subject exams as the only written assessment method. See Fig. 1 for the timeline of the transition process with the number of students who took both exams.

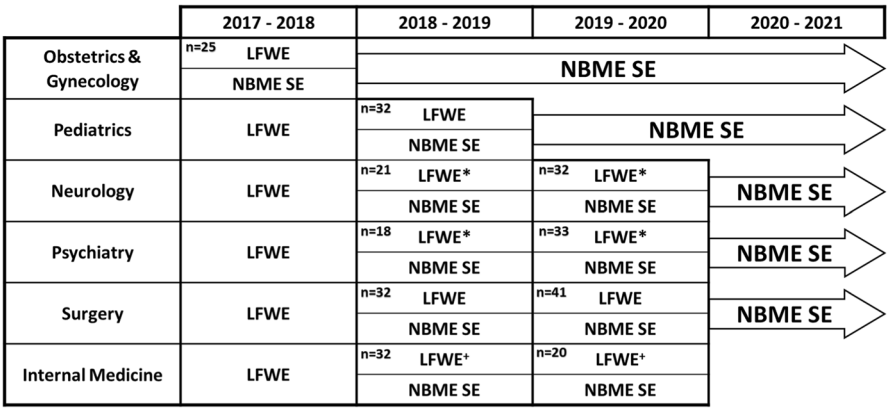

Fig. 1 Timeline of the Transition Process (n = number of students who took both MCQ exams); LFWE, locally developed, faculty written examinations; NBME SE, The National Board of Medical Examiners Subject Examinations. *Neurology and psychiatry were covered in on LFW exam. $^{+}$Internal medicine were covered in two separate LFW exams

Comparison of test scores

In the academic years 2018–2019, 2019–2020, and 2020–2021, 10 NBME subject exams were given alongside LFW exams at the end of each clerkship. There were 286 pairs of scores in ten pairs of NBME and LFW exams during this period. During this transition period, the NBME subject exams were not used for grading, and students who did not wish to take NBME subject exams in addition to the mandatory LFW exams were not included in the analyses. In the total of ten pairs of exams, only 24 students did not participate NBME subject exams indicating a 92.3% participation rate for both exams.

The results of the MCQ exam pairs (NMBE and LFW for each clerkship) were compared for ranking and pass/fail decisions. For the ranking analyses, students' scores for both MCQ exams were divided into three groups as the top 25%, the bottom 25%, and the middle 50%. Students' scores for both MCQ exams are also grouped either as pass or fail using KUSOM's grading guideline that combines properties of the norm and criterion-referenced standard setting. Briefly, the distribution of the scores was described with median (M) and average absolute deviation (AD) for each exam. Students who scored $\leq$ M–1.5 AD were given fail (F) unless their absolute score was $\geq$ 70 out of 100. For simplicity, all students who scored > M–1.5 AD or $\geq$ 70 out of 100 were accepted to have a pass (P). See Fig. 2 for sample grouping of students' scores for ranking and pass/fail decision in 2018 Ob & Gyn clerkship. The interrater agreement between the two exams with regard to ranking (top, middle, bottom) and pass/fail decision was analyzed by Cohen's weightless kappa fit test using IBM SPSS Statistics for Windows Version 26.0.

All NBME subject exams included 110 MCQs in which 100 MCQs were evaluated for the determination of the students' scores on a scale from 0 to 100 (Koç 2021). All LFW exams included 100 MCQs and students' scores on a scale

	Scores for MCQ Exams		25th - 75th Percentile		Pass/Fail Decision	
Students	LFW	NBME	LFW	NBME	LFW	NBME
STUDENT 1	92	90	top	top	pass	pass
STUDENT 2	89	70	top	middle	pass	pass
STUDENT 3	88	76	top	top	pass	pass
STUDENT 4	88	70	top	middle	pass	pass
STUDENT 5	87	81	top	top	pass	pass
STUDENT 6	87	84	top	top	pass	pass
STUDENT 7	86	60	middle	middle	pass	pass
STUDENT 8	86	83	middle	top	pass	pass
STUDENT 9	85	63	middle	middle	pass	pass
STUDENT 10	85	58	middle	middle	pass	pass
STUDENT 11	85	75	middle	top	pass	pass
STUDENT 12	85	70	middle	middle	pass	pass
STUDENT 13	84	66	middle	middle	pass	pass
STUDENT 14	84	62	middle	middle	pass	pass
STUDENT 15	81	68	middle	middle	pass	pass
STUDENT 16	79	62	middle	middle	pass	pass
STUDENT 17	79	56	middle	bottom	pass	pass
STUDENT 18	79	62	middle	middle	pass	pass
STUDENT 19	78	61	middle	middle	pass	pass
STUDENT 20	77	60	bottom	middle	pass	pass
STUDENT 21	76	56	bottom	bottom	pass	pass
STUDENT 22	75	55	bottom	bottom	pass	pass
STUDENT 23	75	57	bottom	bottom	pass	pass
STUDENT 24	64	55	bottom	bottom	fail	pass
STUDENT 25	57	21	bottom	bottom	fail	fail

OB-Gyn (n=25)		NBME Subject Exam		
		Top	Middle	Bottom
LFW Exam	Top	4	2	0
	Middle	2	10	1
	Bottom	0	1	5

OB-Gyn (n=25)		NBME SE	
		PASS	FAIL
LFW Exam	PASS	23	0
	FAIL	1	1

Fig. 2 The sample analysis of the agreement of MCQ Exams in Ob&Gyn clerkship

from 0 to 100 were determined by the number of the correct answers. Both exams were performed under the same circumstances via a computer-based system in the same computer room of the Koç University Hospital. During the transition period, students studied for the LFW exams and they did not receive any incentive

for studying NBME subject exams as only scores of LWF exams were considered for the evaluation of students' performance in clerkships.

Students' feedback for NBME exams

We collected written feedback through a questionnaire including four open-ended questions for qualitative analysis after the students sat for both NBME subject exams and LFW exams. The questionnaires were anonymously filled, and the questions regarded the content, language, and fairness of both exams. Students were also asked which one they would prefer if they were to take only one of these exams to complete their clerkship. Theme analysis was performed on the answers given to open-ended questions. See Table 1 for questions asked to students. Seventy-five students provided feedback among 286 students who took both the NBME subject exam and LFW.

Results

Comparison of test scores

The analysis of the agreement between two exams (LFW exams and NBME subject exams) in terms of pass/fail decisions and ranking of the students into three groups (top, middle, bottom) yielded similar results as shown in Table 2. Inter-rater agreement between two exams was assessed on 286 pairs of scores that were obtained from ten dual examinations. Overall, the unweighted kappa for pass/fail decision was 0.41 (95% CI: 0.25–0.57) with the observation of %87.76 ($n = 251$) agreement between two exams and the unweighted kappa for ranking the students into three groups (top, middle, bottom) was 0.41 (95% CI: 0.32–0.50) with the observation of %61.89 ($n = 177$) agreement between two exams. In the ten dual examinations, the unweighted kappas for pass/fail decisions ranged from 0.22 to 0.78 and the unweighted kappas for ranking the students into three groups ranged from 0.20 to 0.60. In the agreement analysis, we found fair to moderate agreement (unweighted kappas between 0.20 and 0.60) according to the suggested interpretation of this scale by Landis et al. (1977).

Table 1 Open-ended questions

(1) Do NBME Subject exams cover the learning objectives of the clerkship as much as LFW exams?
(2) Which exam would you prefer in terms of language and the variety of questions?
(3) Is there any topic that was assessed in the NBME subject exams but was not mentioned/taught during the clerkship?
(4) If you were to take a single exam for the assessment of the clerkship, which one would you prefer?

Table 2 The results of the agreement analysis between LFW exams and NBME exams

Agreement for Pass/Fail decisions						
N = 286			NBME Subject Exams			Observed Kappa = 0.41 Standard Error of kappa = 0.08 95% CI 0.25–0.57
			PASS		FAIL	
LFW Exams		PASS	235		20	
		FAIL	15		16	
Agreement for ranking in 3 groups (25th and 75th percentile)						
N = 286			NBME subject exams			Observed Kappa = 0.41 Standard error of kappa = 0.05 95% CI 0.32–0.50
			Top	Middle	Bottom	
LFW Exams	Top		48	22	7	
	Middle		24	81	23	
	Bottom		4	29	48	

Feedback on the two exams

Three themes emerged from the qualitative analysis: content validity, language & comprehensibility, and objectivity/fairness. In terms of content validity, the major concern was a possible difference between the coverage of NBME subject exams and the KUSOM curriculum. For instance, traumatology was mostly addressed in a clerkship named musculoskeletal diseases in the 5th year of KUSOM; however, trauma-related topics are addressed in the NBME subject exam for surgery, internal medicine and pediatrics. Additionally, one of the students stated that "*some topics are given different weight and priority between NBME subject exams and LFW exams*." Some students considered this difference as a "*challenge*" and described it as a "*better learning opportunity because we have to review more content than we normally do*.".

In terms of language and comprehension, NBME subject examinations were considered as "*more comprehensible*" among students. One of the students described this as "*It was easier to understand what the question asked*".

The objectivity/fairness of the NBME subject exam has been the most debated issue among students. Some students considered assessment by the NBME subject exams as "*unfair*", since the KUSOM curriculum did not completely overlap with the curricula of North American medical schools, and they stated that: *"NBME exams are unfair because they do not exactly cover the same topics as our curriculum."* On the other hand, some students reported that NBME subject exams were more objective and fairer than LFW and stated that: "*we do not have to focus on specific interests or priorities of our individual faculty members, and it is fairer to be evaluated for relevant knowledge."* Most students rated both exams as fair. Among the students who rated the use of NBME exams as unfair, the most common reason was differences in clerkship curriculum and NBME subject exam coverage. Indeed, several students pointed out a need for curriculum modifications to harmonize contents with NBME subject exam coverage.

The analysis of the preference of students for MCQ exams revealed that most of the students preferred NBME subject exams for Surgery, Pediatrics, Ob&Gyn,

Neurology, and Psychiatry clerkships, however, they preferred the LFW exam for internal medicine clerkship. The main reason behind this preference was concerns regarding the above-mentioned difference in content since internal medicine clerkship covered a wide range, and even though the content is similar across the world, relative importance and priorities can vary to some extent between different locations and communities. Another reason why most students preferred the LFW exam for internal medicine can be the division of internal medicine into two separate clerkships in the KUSOM curriculum. This separation might have influenced the students' perception of the content of NBME subject examinations.

We did not get written feedback on NBME subject exam experiences from the faculty members, but they were presented and discussed during the faculty meetings. Faculty members raised two main concerns for NBME subject examinations: the epidemiology of diseases is different between U.S. and Turkey and NBME questions not being shared with faculty members. As faculty members could not see what was asked to their students, they expressed their concerns as "*feeling less in control as they cannot assess what they have taught to students*". On the contrary, some faculty members reported their positive impressions such as increased student engagement and participation during lectures rather than obsessively trying to take notes and focusing on PowerPoint presentations, when they knew that they would be assessed by a third-party exam.

Although we did not collect quantitative data on how much time it took to prepare a LFW exam for faculty and clerkship directors, since NBME subject exams are ordered by clerkship coordinators, the latter clearly takes much less time to prepare.

Current status

As of the academic year 2020–2021, all clerkships for which there is an available NBME subject exam, i.e., Ob&Gyn, Pediatrics, Surgery, Psychiatry, and Neurology decided to apply only NBME subject exams as a summative assessment method. The internal medicine curriculum is being revised to improve content agreement with NBME subject exams for transition.

Discussion

We presented our experience of transition from LFW to third-party examinations by comparing the agreement between exam results and evaluating the feedback from students and faculty. The agreement analysis for ranking and pass/fail decision between the results of LFW exams and NBME subject exams indicates fair to moderate agreement, which was regarded acceptable to employ NBME subject exams instead of LFW exams in clinical clerkships.

There can be several reasons contributing to the lack of good to perfect agreement between two exams. Students could have been less motivated to study for NBME subject exams, which were usually given a few days before LFW exams since the former would not contribute to their pass/fail status. Some differences in the content

coverage of exams could have also played a role. Yet, a perfect agreement is rarely achieved in such comparisons. Indeed, we were aware of differences between our curriculum and NBME subject exam coverage before we started the transition, and a less than perfect or good agreement was anticipated. We were looking forward to identifying areas that were not adequately covered by KUSOM curriculum through detecting areas that our students were underperforming compared with other test takers. One example was the expansion of adolescent care in Pediatrics upon noticing our students were underperforming in questions regarding adolescent care.

Our additional motivations for using third-party MCQ exams are consistent with those mentioned in the relevant literature. First and foremost, a major drawback of LFW exams is the lack of sufficient time or expertise on part of most faculty members to prepare clear, high-quality MCQs that match the learning objectives (Jozefowicz et al. 2002; AlMahmoud et al. 2015). Third-party MCQ exams help to save time and effort of faculty members, which could be allocated for clinical education with more clinic hours for interaction between faculty and students in the clinical settings.

Second, poorly formulated MCQs by clinicians, all of whom may not have comprehensive formation in education, may simply assess recall capabilities or test-taking skills instead of assessing higher-order thinking skills or clinical reasoning (Baig et al. 2014). In schools where the educational language is English, non-native English speaker faculty members may have difficulty in preparing clear, understandable MCQs. Third, some faculty members tend to prepare questions focused on their individual research or clinical interests rather than prespecified learning objectives for undergraduate students, which leads to a mismatch between the assessment and the learning objectives of the curriculum (Kelly et al. 2012; Downing 2002). Indeed despite multiple rounds of training sessions and the availability of educational videos on MCQ preparation for faculty members, matching MCQs and learning objectives as well as generating MCQs with acceptable discriminatory indices remained a challenge at KUSOM. Furthermore, medical students taking LFW exams may anticipate these mismatches between MCQs and the learning objectives, and study what they anticipate being asked in LFW exams rather than focusing on learning objectives. Therefore, NBME subject exams motivated students to focus on prespecified learning objectives of the curriculum rather than specific interests of individual faculty members.

Lack of an independent committee for the preparation of MCQs or peer review process for the review of MCQs before the examination makes these downsides of LFW exams inevitable (Jozefowicz et al. 2002). Despite the presence of an assessment committee at KUSOM, and the efforts to implement such a system for several years now, it has not been possible to effectively implement a proper peer review process for the last ten years in our experience. Another reason that makes the use of third-party exams appealing is that they are readily available to use at any time to provide a flexible exam schedule without causing a time constraint or cost of energy for the faculty members (NBME 2019).

Finally, third-party MCQ exams can provide externally validated results and an opportunity to compare the curriculum's output both in terms of its potential graduates between different batches and students' performance compared to their peers

internationally (Al Ojaimi et al. 2020). Since the LFW exams are, unfortunately, not standardized or externally validated in several aspects such as content and facility, they do not enable the evaluation of the performance of the school, educators, and students compared to other institutions. On the other hand, besides circumventing above mentioned pitfalls of LFW exams, NBME subject exams can also be used to evaluate the current educational program and the effects of curricular changes, if performed, between different years (Williams 1993; Wright and Baston 2017; Hoffman 1993).

It was expected that some students would rate the use of NBME subject examinations as unfair, since curriculum, in its status at the time, was designed differently and differences in coverage was inevitable. This was one of the main reasons for not switching to NBME examinations right away, but through a transition period during which content coverage would be assessed and improved. We could have better phrased the question as "If you were to take a single exam for the assessment of the clerkship, assuming similar content coverage which one would you prefer based on exam quality?" As explained above, we regarded switching to third party examinations as an opportunity to rethink our curriculum, both in terms of coverage and organization, and wanted to use the observed differences between coverage of our clerkship content and NBME subject examinations as a guide in this process. The observed differences led to changes in curriculum and gradually every clerkship voluntarily switched to NBME examinations.

While the cost of NBME subject examination tests, which are purchased at a price per examinee basis, can be regarded as a disadvantage at first, it saves the faculty's time that could be spared for education, research, or patient care, and eventually create additional economic value. In KUSOM's case, the annual cost of NBME subject examinations is around 10.000 USD per year. Such expenditure can be a burden for medical schools with a limited budget and a large number of students. Second, in some clerkship programs learning objectives may not perfectly match with the topics that are covered in third-party exams. The substantial discrepancy would render the use of third-party exams implausible due to low content validity. Third, like in our case, there can be concerns about the differences in the epidemiology of health issues in different regions as this may affect the contents of the exam and distribution of the questions. Although this may seem like an important concern at first, in practice, the number of such questions can generally be negligible for pass/fail decisions. Moreover, results of NBME subject exams are not reported as pass/fails, but as scores for individual students with suggested thresholds for pass/fail decisions. Thus, we had the liberty to assess the distribution of scores within their examinees and apply our own hybrid norm and criterion-based system for picking our own threshold for a particular exam. It should be noted that given our students' increasing performances on NBME subject examinations, our thresholds are not lower than the thresholds suggested by NBME. Besides, other assessment methods like mini clinical exams, OSCEs, quizzes can be used for the assessment of competency in areas that are locally more relevant but are not given weight in third-party examinations (Kelly et al. 2012). Last but not least, locally monitoring of the quality of third-party examinations can be challenging if the exam questions are not disclosed to faculty members due to copyright concerns (NBME 2021).

Conclusions

Herein we described our transition process including our motivations for seeking third-party examinations, rationale for selecting a particular third-party exam provider, methods we used for assessing agreement between two exams. We concluded that third-party MCQ exams, i.e., NBME Subject Examinations in our case, had an acceptable agreement with LFW exams for students' grading decisions while taking less time and effort from the faculty for the preparation of MCQs. Besides fair agreement for grading, third-party exams such as NBME subject exams provided additional benefits including evaluation of the curriculum, comparison with other schools using the same exams, tracking of students' performance across years, better-formatted MCQs in terms of comprehensibility and language. Cost and adaptation issues are administrative hurdles that can be managed. Some discrepancy between the curriculum and exam coverage can be regarded as an opportunity rather than a limitation since reliable high quality third party examinations provide educational leadership with valuable feedback to rethink the curriculum and assessment methods. In our case we decided that the advantages of the transition outweighed disadvantages. Our experience can be useful for other schools considering the use of third-party examinations.

Acknowledgements We thank all faculty and students of Koç University School of Medicine during the study period for their efforts to help us to assess this unique transition period which the article is based on.

Author contributions BA designed the article and acquired the data regarding students' grade and written feedbacks. AD, SS, TG made quantitative analysis and interpreted students' scores for the agreement analysis between two examinations. BA, ÇÖ, SA made qualitative analysis and interpreted students' feedback regarding the change in the assessment methods. AD, ÇÖ, SA drafted the article and BA, SS, TG revised it critically for important intellectual content. All authors read and approved the final manuscript.

Funding Not applicable.

Data availability The datasets generated and/or analyzed during the current study are not publicly available because Koç University School of Medicine has a policy of non-disclosure of students' grades publicly but are available from the corresponding author on reasonable request.

Declarations

conflict of interest The authors declare that they have no competing interests.

Ethical approval The need for consent was waived by the Koç University Committee on Human Research (2021.166.IRB3.075).

Consent for publication Not applicable.

References

Al Ojaimi M, Khairallah M, Younes R, Salloum S, Zgheib G (2020) National board of medical examiners and curriculum change: what do scores tell us? a case study at the University of Balamand Medical School. J Med Educ Curric Dev 7:2382120520925062

AlMahmoud T, Elzubeir MA, Shaban S, Branicki F (2015) An Enhancement-focused framework for developing high quality single best answer multiple choice questions. Educ Health (abingdon) 28(3):194–200
Al-Wardy NM (2010) Assessment methods in undergraduate medical education. Sultan Qaboos Univ Med J 10(2):203–209
Baig M, Ali SK, Ali S, Huda N (2014) Evaluation of multiple choice and short essay question items in basic medical sciences. Pak J Med Sci 30(1):3–6
Custers E, Cate OT (2018) The history of medical education in Europe and the United States, with respect to time and proficiency. Acad Med 93:S49–S54
Devine OP, Harborne AC, McManus IC (2015) Assessment at UK medical schools varies substantially in volume, type and intensity and correlates with postgraduate attainment. BMC Med Educ 15:146
Downing SM (2002) Threats to the validity of locally developed multiple-choice tests in medical education: construct-irrelevant variance and construct underrepresentation. Adv Health Sci Educ Theory Pract 7(3):235–241
Hoffman KI (1993) The USMLE, the NBME subject examinations, and assessment of individual academic achievement. Acad Med 68(10):740–747
Jozefowicz RF, Koeppen BM, Case S, Galbraith R, Swanson D, Glew RH (2002) The quality of in-house medical school examinations. Acad Med 77(2):156–161
Kelly WF, Papp KK, Torre D, Hemmer PA (2012) How and why internal medicine clerkship directors use locally developed, faculty-written examinations: results of a national survey. Acad Med 87(7):924–930
KUSOM. Koç University School of Medicine 2021. https://medicine.ku.edu.tr/en/education/undergraduate-medical-education/. Accessed January 2021
Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174
NBME ® Subject Examinations Program Guide. National Board of Medical Examiners 2021. https://www.nbme.org/sites/default/files/2021-04/NBME_Subject_Exam_Program_Guide_2021.pdf. Accessed February 2021.
NBME clinical clerkship subject examination survey: Summary of results 2016. https://www.nbme.org/PDF/SubjectExams/Clerkship_Survey_Summary.pdf. Accessed February 2019.
NBME. NBME Annual Report 2019 National Board of Medical Examiners 2019. https://www.nbme.org/sites/default/files/2020-03/NBME_Annual_Report_2019.pdf. Accessed February 2021.
Tabish SA (2008) Assessment methods in medical education. Int J Health Sci (qassim) 2(2):3–7
Williams RG (1993) Use of NBME and USMLE examinations to evaluate medical education programs. Acad Med 68(10):748–752
Wright WS, Baston K (2017) Use of the national board of medical examiners comprehensive basic science exam: survey results of US medical schools. Adv Med Educ Pract 8:377–383

Authors and Affiliations

Atalay Demiray[1] · Çınar Öztosun[1] · Simge Acar[1] · Sibel Sakarya[2] · Tuğba Gürsoy[3] · Barış Ata[4]

[1] Students at 4th, 5th and 6th Years, Koc University School of Medicine, Istanbul, Turkey

[2] Department of Public Health, Koc University School of Medicine, Istanbul, Turkey

[3] Department of Pediatrics, Koc University School of Medicine, Istanbul, Turkey

[4] Department of Obstetrics and Gynecology, Koc University School of Medicine, Istanbul, Turkey