Open Access Full Text Article

ORIGINAL RESEARCH

Ethiopian Anesthetist Licensing Examination: A Qualitative Study of Concerns and Unintended Consequences

Yohannes Molla Asemu [1,2], Tegbar Yigzaw [1], Firew Ayalew[1], Leulayehu Akalu [3,4], Fedde Scheele [2,5,6], Thomas van den Akker [2,7]

[1]Health Workforce Improvement Program, Jhpiego, An Affiliate of Johns Hopkins University, Ethiopia Country Office, Addis Ababa, Ethiopia; [2]Athena Institute, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; [3]Ethiopian Association of Anesthetists (EAA), Addis Ababa, Ethiopia; [4]School of Anesthesia, Addis Ababa University, Addis Ababa, Ethiopia; [5]Department of Obstetrics and Gynecology, OLVG Teaching Hospital, Amsterdam, the Netherlands; [6]Department of Obstetrics and Gynecology, Amsterdam University Medical Center (AUMC), Amsterdam, the Netherlands; [7]Department of Obstetrics and Gynecology, Leiden University Medical Centre (LUMC), Leiden, the Netherlands

Correspondence: Yohannes Molla Asemu, Health Workforce Improvement Program, Jhpiego, an affiliate of Johns Hopkins University, Ethiopia Country Office, Kirkos Sub-City, Kebele 02/03, House 693, Wollo Sefer, Addis Ababa, Ethiopia, Tel +251911384046, Email Yohannes.Asemu@jhpiego.org

Background: Ethiopia increased its anesthesia workforce drastically by expanding the training of associate clinician anesthetists. Following this expansion, the Ministry of Health established an entry-level anesthesia licensing examination to ensure patient safety. However, there is limited empirical evidence on the impacts of licensing exams in low- and middle-income countries. This study aimed to explore the concerns and undesirable consequences of the anesthetist licensing examination in Ethiopia.

Methods: A qualitative design using a grounded theory approach was employed by collecting data from 10 anesthesia teaching institutions. We conducted 15 in-depth interviews with instructors and six focus groups with students and graduates who took the exam recently. Interviews and focus groups were audio-recorded, transcribed verbatim, and analyzed using Atlas.ti 23. We also extracted secondary data from the academic committee meeting minutes, curricula, faculty appraisal reports, and program quality self-review reports.

Results: Qualitative analysis revealed three central categories of concerns and untoward consequences of the anesthetist NLE: exam management, educational management, and student behavior. Exam management concerns were related to exam validity, fairness, and consistent enforcement of pass/fail decisions. The unintended consequences of the exam on education management were perceived as promoting teaching and learning for the exam, increasing faculty workload, and resulting in superficial and patchy educational reforms. Study participants also reported adverse psychosocial effects and increased cheating behaviors among students as undesirable consequences of the exam on student behavior.

Conclusion: Our study identified some concerns and unintended consequences of the Ethiopian anesthetist licensing examination. These lessons learned may contribute to improving the quality of licensing examinations in Ethiopia and beyond.

Keywords: anesthesia, anesthetist, associate clinician, non-physician, licensing examination

Introduction

For health profession regulators, it is critical to ensure that graduates master the minimum competencies to provide safe and effective care.[1] This agenda has become particularly prominent in the Ethiopian health policy context over the last decade due to burgeoning concerns about the quality and safety of healthcare resulting from a massive expansion of pre-service education to address the critical workforce shortage.[2] Ethiopia quadrupled its anesthesia workforce density from one anesthesia provider for every 333,000 people in 2012 to one for every 69,470 people in 2019[3] by expanding the training of *associate clinician anesthetists,* who in Ethiopia are simply known as anesthetists. This scale-up was made possible through a task-sharing strategy: a three-year training of nurses or a four-year training of high school graduates leading to a baccalaureate degree in anesthesia.[4–7]

1055

Anesthetists are trained to provide anesthetic care independently, aiming to improve access to safe surgical care as part of the aspiration for universal health coverage.[8] Between 2012 and 2018, the number of anesthetist teaching institutions increased ninefold, and annual graduate outputs increased more than threefold, from 98 to 340. However, concerns emerged that some health professionals might fall short of meeting competency requirements due to the rapid expansion of pre-service education.[5,7] In response, the Ministry of Health (MOH) introduced a national licensing examination in selected health professions, including anesthetists, as a requirement to receive a license to practice. A written licensing examination was piloted from 2015 to 2017 and mandated in 2019. To date, more than a thousand anesthetists have taken the national licensing examination (NLE).

High-stakes exams significantly impact stakeholders, including examinees, faculty, teaching institutions, patients, and society. Thus, regulators should regularly evaluate the effectiveness and appropriateness of such exams.[9–11] In this regard, the most crucial evidence source is consequence validity, which requires evidence supporting the claim that desirable effects outweigh undesirable ones. By viewing assessments as educational interventions, this domain of validity evidence examines the broader impact of the exam on different stakeholders.[1,12,13] However, such an impact has been sparsely studied, and the World Health Organization has flagged it as a knowledge gap in its research agenda.[14]

Some evidence shows the positive impact of licensing examinations on standards of health profession education: identifying ill-prepared graduates, providing data on competency standards of higher education institutions, driving learning, and catalyzing educational quality improvement.[15–17] NLEs hold students, faculty, and academic leaders accountable by obtaining regular performance data, thereby driving positive reforms in educational programs.[15,18–21] Our previous study has also shown that introducing the anesthetist licensing examination in Ethiopia stimulated anesthetist teaching programs, faculty, and students to improve teaching, learning, and assessment practices.[22]

However, the potential undesirable consequences of licensing examinations have received little attention in the literature. Very few studies from high-income countries mention the untoward impacts of reducing curricular innovation and discouraging diversity among teaching institutions.[17,23,24] More emphasis on NLE performance has sometimes been found to shift the focus of education to solely preparing for the exam.[9] Because of this, some argue that program accreditation and institutional self-regulation are preferable.[23,25,26] Besides, while different aspects of medical licensing examination systems have been assessed, reported, and criticized in high-income countries, published evidence on the impact of NLEs on non-physician groups, especially in low- and middle-income countries (LMICs), is lacking.[14]

In another study, we discussed the perceived positive effects of the anesthetist licensing examination.[22] On the other hand, the current study aimed to explore concerns and possible untoward consequences of the anesthetist licensing examination in Ethiopia.

Material and Methods

Study Design

We employed a qualitative study design using a grounded theory approach. The results are described according to the Consolidated criteria for reporting qualitative research (COREQ).[27] Investigators regularly refer to the 32-item checklist steps across all study phases to determine adherence to each step and ensure the study's rigor and the manuscript's comprehensiveness. We attached a completed checklist, citing where in this manuscript we address each step (Supplementary 1).

Study Population, Setting and Sampling

A total of 46 study participants were selected using purposeful sampling from all ten public teaching institutions that have had anesthetist graduates since the start of the NLE. Purposeful sampling (also known as purposive sampling) refers to the deliberate selection of study participants from different populations who have experienced the concerns and consequences of the NLE (the central study phenomenon) by the study teams.[28] In-depth interviews were conducted with 15 (two female) full-time academics who have been with their current institution since the NLE's inception. Interviewees were drawn from different departments engaged in the anesthetist training program, including anesthesia, core clerkship, biomedical sciences, and public health. Three focus groups were held with 15 (eight female) final-year students and three

more with 16 (nine female) graduates who took the exam recently. Relevant documents were also collected from eight of the ten institutions, leaving the two that were reported as unsafe for travel due to political unrest in the country's north.

Data Collection Procedure

Interview and focus group guides were developed by constructing one to two open-ended questions under each of the seven main areas highlighted in the national anesthesia education standards[29] and broadly encompassed program structure, educational input, and the learning process (Supplementarys 2 and 3). The national standards were adapted from the World Federation for Medical Education's basic medical education standards.[30] Both guides were translated into the local language, Amharic. The authors subjectively evaluated the relevance and presentation of both tools, followed by pilot testing with two volunteers.

The data set for this study was collected as a follow-up to the data gathered for another qualitative study that looked into the desirable consequences of the Ethiopian anesthetist licensing examination.[22] Data were collected by the lead author and six data collectors with practical experience performing qualitative studies and at least a master's degree in a relevant field. All data collectors attended a one-day practical training on data collection techniques, including probing, note-taking, and ethical issues.

Using contact information from their institutions, the principal investigator or a trained data collector communicated (through email, phone or in-person) with each invitee to explain the study procedures. Those who agreed to participate in the study received a study information sheet and an invitation for a face-to-face discussion.

The data collectors conducted 12 face-to-face interviews at the participants' institutions between September 20 and 24, 2021 at the participants' institution. The lead author conducted the remaining three interviews between October 5 and 10, 2021. No further interviews were carried out since saturation had been reached and no new information had emerged. On the other hand, six face-to-face focus groups were held in the capital, Addis Ababa, by creating groups with a mix of participant characteristics in terms of gender, university geographical location, and year of university establishment. Four of these focus groups were moderated by trained data collectors, with the remaining two facilitated by the lead author. Three authors supervised the data collection procedure. At the same time, the PI was actively involved in data collection and site supervision to gain a deep understanding of the context. The duration of in-depth interviews ranged between 15 and 37 minutes (mean 22 minutes), while the focus groups ran from 50 to 85 minutes (mean 65 minutes). Digital audio recording devices (Sony ICD-PX370) were used to record all conversations.

In addition, versions of curricula, nine randomly picked department academic committee minutes (2018–2020), program self-review reports (2017–2020) and NLE pass rates (2019–2021) were systematically observed and gathered from each institution.

Data Analysis

The audio data was transcribed verbatim, checked for accuracy, and prepared for analysis. The interview transcripts were shared with the respondents for verification. Six agreed, eight offered minor revisions, but one participant did not respond. Interview and focus group transcripts were separately coded and thematized inductively and iteratively by the lead author to generate a theory. The coding procedure involved initial coding (developing line-by-line tentative notions), process coding (linking actions and consequences), focused coding (distilling frequent concepts into categories), axial coding (identifying and reorganizing the most important categories), and theoretical coding (extracting the main connecting category and developing a theory). Meanwhile, memos were written and tied to the dataset, facilitating theory generation. Regular discussions were conducted between the lead author (YMA) and co-author (TY) during the study process to maximize the quality of data collection and analysis.

Subsequently, themes and concepts related to concerns and unintended consequences of anesthetist NLE were reorganized around three main categories: (a) exam management, (b) educational management, and (c) student behavior to establish interrelations and formulate a theory. The three main categories were not mutually exclusive. Analysis output and selected direct excerpts were translated from Amharic into English and presented as text. Document review findings were presented alongside transcript analysis findings as qualitative descriptions of various phenomena and proportions of variables (eg, NLE pass rates and program quality review achievements). Atlas.ti 23 was used to facilitate data analysis.

Reflexivity

A total of six investigators (all men) were involved in the study. The three co-authors (TA, FS, and TY) are physicians with a PhD and substantial expertise in qualitative research. The other three have backgrounds in anesthesia (YMA, LAG) and statistics (FAD). FAD holds a PhD and is well-versed in monitoring, evaluation, and research. YMA and TY are health professions education experts who actively supported the Ministry of Health in implementing the NLE. The first researcher (YMA) is a PhD candidate in medical education and received a fellowship and postgraduate certificates in health professions education and regulation from FAIMER and Keele University, respectively. LAG is the President of the Ethiopian Association of Anesthetists. All investigators reviewed and refined the extracted categories and the theory until they reached a consensus.

Results

Three central categories (a-c) and eight further associated sub-categories (1–8) were identified on concerns and untoward consequences of the anesthetist NLE (Table 1) and described below, supported by participant excerpts. The expanded coding tree is also available (see Supplementary 4).

Table 1 Overarching Categories and Sub-Categories Identified on Concerns and Untoward Consequences of the Anesthetist NLE

Central Category	Sub-Categories
(a) Exam management (concerns)	(1) Contested exam validity (2) Perceived exam unfairness (3) Lack of consistent consequence
(b) Educational management (untoward effects)	(4) Faculty overload and dissatisfaction (5) Focused teaching and learning to pass the exam (6) Educational reforms lack sustainability
(c) Student behavior (untoward effects)	(7) Harmful psychosocial effect (8) Growing cheating behaviors

Category: Exam Management

The three main concerns about licensing exam management were related to its validity, fairness, and consistent enforcement of pass/fail decisions, which allegedly hinder individual examinees' and programs' motivation and performance by distracting their focus.

Contested Exam Validity

Most graduates and a few instructors were concerned about the NLE's content, emphasis, and scope due to perceived misalignment in what was being taught and tested. Most graduates also indicated that the exam model, which intended to cover a range of topics in one written exam, "dragged the exam"*s focus into less relevant topics'* (GA-04), which they attributed to their poor results.

> The [licensing] exam should test our readiness in areas that we will practice in the workplace. Unfortunately, not all exam content recognized our scope or practice standard… For instance, the exam asked us how to manage anesthesia for cardiac bypass surgery and prepare a 20-year national strategic plan, which I had never heard of during my university training. Such questions lowered our exam scores. (GA-06)

Perceived Exam Unfairness

The NLE was reported as resulting in inequitable treatment of students and graduates. The disparities in learning environments among different teaching institutions and the eruption of conflict in north Ethiopia affecting some institutions raised concern for justice. A student reported:

> My institution does not have a hospital for clinical practice, and we struggled to gain adequate clinical exposure. How would it be fair to compare us with those who completed their training in a well-equipped setup? (SA-01)

At the same time, interviewees consistently reported unequal involvement of instructors from different teaching institutions in the NLE development process. Most instructors thought this contributed to differences in institutional NLE performance, as institutions with more involvement in the NLE process were reportedly better able to prepare their students by knowing the exam content and standard.

Majority of focus group participants and interviewees also see the remediation system as a source of unfairness. Some graduates who failed to pass the exam on their first attempt reported getting permission from their host institutions to continue supervised clinical practice and prepare for subsequent attempts. In contrast, other institutions denied this opportunity, "resulting in further loss of competencies" (GA 02).

Lack of Consistent Consequences

Another exam-related concern was the need for consistent consequences for examinees who took and passed, failed, or did not take the licensing examination. The disparity in enforcement of pass/fail decisions and rising unemployment rates reportedly made most focus group participants pay less attention to the exam. A graduate quoted:

> The [licensing] exam made me careless since it is not rewarding, and thus I did not understand the point of taking it. Some of my classmates who did not pass the exam are now employed and practicing anesthesia. while others who passed [the exam] did not get a job. (GA-07)

Coupled with the COVID-19 pandemic and political instability in the northern part of the country, most respondents strongly believe that the lack of consequences has contributed to the drop in overall NLE pass rates by more than nine percentage points, from 94.5% in 2019 to 85.0% in 2021, albeit determining a possible association is beyond the scope of this study.

Category: Educational Management

Increased program expectations to improve institutional exam pass rates were perceived to have resulted in a rising faculty workload. Meanwhile, inappropriate NLE data usage allegedly promoted teaching and learning for the exam. These consequences on educational management have reportedly contributed to educational reforms being less systematic, superficial, and patchy.

Faculty Overload and Dissatisfaction

The NLE-driven educational reforms and heightened accountability were reported to increase faculty workloads consuming research and personal development time and contributing to faculty dissatisfaction. A program head quoted:

> Due to the licensing exam, the department exam review committee began reviewing and rejecting low-quality in-school exams for revision. Unfortunately, some instructors became too fatigued and defensive to make the necessary revisions, grumbling about their unrecognized efforts. They started to view the reforms as burdens. (PH-01)

Focused Teaching and Learning to Pass the Exam

One of the most commonly reported unintended outcomes of the exam was the disruption it caused in learning and assessment practices. The majority of focus group participants, particularly those from lower-performing programs, reported prioritizing theoretical learning to pass the NLE.

> We were told [by instructors] the specific topics the licensing exam would focus on, so we concentrated on those areas by reading and taking many written exams. As a result, the licensing exam taught us what we needed to know to pass the exam rather than what we should know to be safe practitioners. (SA-06)

Educational Reforms Lack Sustainability

Regardless of institutional pass rates, most programs perceive that NLE pass rates accurately measure the quality of their education. Instructors from better-performing institutions claimed that there is no need for educational reforms because of their high NLE pass rate, which they see as an absolute indicator of good education quality. On the other hand, instructors from low-performing programs described themselves as desperate and attempting ill-structured educational reforms. Both programs consistently indicated a lack of institutional support to enhance and sustain reforms. An instructor quoted:

> The authorities [the university, and ministries of education and health] should evaluate whether students are learning in a way that enables them to master skills and pass the exam. Programs should first be supported to improve their education quality before the authorities try to harvest what hasn't been sown. (PI-06)

The 2017–2020 program self-review reports also indicated that anesthesia teaching institutions achieved less than 50% of basic education standards on continual quality assurance and research and development.

Category: Student Behavior

Respondents also stated that the exam had adverse psychosocial impacts on students and increased cheating behaviors, which impacted their health, productivity, and focus.

Harmful Psychosocial Effect on Students and Graduates

The majority of focus group participants reported stress and exhaustion due to NLE-driven, poorly planned multiple in-school activities that kept them busy and hampered their productivity; some began to use Khat, a stimulant leaf, to deal with the stress. As a result, low-performing students reportedly lost their confidence, while better performers were pressured to achieve higher NLE scores and raise the overall institutional mean score, a metric used to compare institutions publicly. Some students further mentioned peculiar sources of disappointment.

> Because of our seniors' low scores, we were labeled as low performers, and no one could recognize our good side. It hurts a lot to be judged solely on the performance of others. (SA-03)

Majority of the graduates described the psychosocial impact of failing the NLE as devastating due to not meeting family expectations and losing income.

Growing Cheating Behaviors

Most study participants perceived that the increasing cheating practices are hurting students' and instructors' faith in the good outcomes of the NLE. Some focus group participants (mainly those from low-performing programs) revealed that faculty repeatedly told them to do anything, even cheat, to pass the NLE. Common cheating behaviors listed by graduates and instructors included bringing phones and other unauthorized materials into exam rooms, copying answer sheets from others, attempting to take away exam papers, and f*aculty encouraging juniors to ask graduates about exam contents* (AI-02).

Discussion

Ethiopia has improved the availability of the anesthesia workforce through a task-sharing strategy by rapidly expanding anesthetist teaching institutions.[3,6] This was one of the first studies exploring the concerns and undesirable consequences of NLEs in non-physician education systems in LMICs. Because LMICs share commonalities in their health workforce landscapes, we assume that our findings from Ethiopia are helpful beyond our own country. Our findings may contribute to strengthening NLE systems across different settings.

Although NLEs are thought to focus on core areas expected of new graduates, some argue that they instead focus on what can be assessed, calling exam decisions into question. This shift in focus will affect education systems, particularly when NLE content does not match job expectations, and programs realize that what matters most is what the NLE values most.[9,10,23] The current study reported that some of the contents covered in the NLE were reportedly out of scope and less practical, contesting the exam's validity and ability to direct learning contents effectively. To maintain alignment

between the NLE content and job requirements and direct education to what is most needed, efforts must be strengthened to design the NLE using current task/job analysis reports.[12,31,32]

This study found a growing inclination toward exam-centeredness among students and faculty in a few programs. When the educational goal shifts to preparing for exams, students lose opportunities to develop additional competencies relevant to their future roles and personal development.[9,33] Addressing this exam-centeredness requires setting up an institutional assessment system capable of organizing multiple assessments to measure different competency dimensions.[34,35] In light of this, a programmatic approach of small-scale summative assessments with a holistic scope, such as year-based qualification and exit exams, might stimulate comprehensive learning. Clinical teaching could also benefit from setting minimum workplace-based assessment targets as a requirement for taking the NLE and blending the NLE with sequential OSCE, where a smaller number of stations are administered to quickly screen and exempt passing examinees from larger (sequential) OSCEs.[17]

Previously, we reported the positive impact of the NLE on driving learning by introducing quality enhancement reforms.[22] We learned in the current study that the sustainability of those reforms in some institutions relied on individual faculty commitments due to inadequate institutional support to sustain changes. Besides, due to disproportionate faculty engagement in the NLE process, only some programs with masses of informed and trained faculty reportedly made promising changes, resulting in patchy educational reforms across programs. On the other hand, programs with higher NLE pass rates demonstrated little commitment to implementing additional educational reforms because most viewed exam data as an absolute measure of educational quality. However, such exam data falls short of reflecting the complex educational system and the shortcuts some programs may take to artificially inflate exam pass rates.[9,23,36] In this regard, cultivating a continuous educational quality improvement culture through program self-review and accreditation processes will help enhance sustainable reforms.[37,38]

Given examinees' discontent with injustice in exam management, the current paper-and-pencil format exam could benefit from a phased approach to introduce computer-based testing, thereby improving result analysis and notification timeliness while portraying authentic clinical findings.[17] On the other hand, the lack of a consistent remedial system necessitates the meaningful engagement of the health and education sectors to implement an equitable and individualized examinee support system with clearly defined outcome measures in simulated and clinical settings.[39]

Cheating, in various forms, undermines the overall public protection goal of NLEs by eroding exam validity and reliability.[10,12] An integrated set of interventions is required to address this misconduct, including raising awareness about the exam among stakeholders, cultivating a healthy institutional culture, setting NLE systems that recognize clinical skills, tightening exam security, and implementing stringent policies with severe consequences for those who cheat.[40] Our findings call for a comprehensive investigation into NLE cheating behaviors, their driving forces, and practical prevention efforts.

Finally, the growing lack of interest in the exam among examinees due to inconsistent enforcement of pass/fail decisions on the one hand and unemployment on the other necessitates the effective engagement of regional regulatory entities and employers, respectively. To further limit the spread of this reluctance into training programs, we recommend utilizing NLE performance data to inform other decisions, such as institutional budget allocation and student enrollment criteria for postgraduate studies.

Strength and Limitations

We triangulated research data by collecting information from various sources (interviews, focus groups, and document reviews) and viewpoints (students, graduates, instructors, and leaders) from all eligible anesthetist training institutions nationally. Multiple measures were employed to enhance rigor, including respondent validation, team debriefing, and constant comparison method.

One potential limitation of our study is that participants may refrain from sharing their emotions openly and honestly because of the topic's sensitivity. Another limitation could be the difficulty of filtering potential biases of investigators and respondents in reporting non-existent negative impacts. To address these concerns, study participants were briefed about the study's purpose during the informed consent procedure, assuring them that the findings would not harm their

lives. Besides, we employed probing questions to explore the existing practices. Finally, though two investigators double-checked the coding and data analysis, no quantitative intercoder agreement was calculated.

Conclusion

The concerns and undesirable consequences reported in this study should be kept from overshadowing the promising positive impacts of the NLE. However, the findings should stimulate the key stakeholders to take appropriate measures. The regulatory authority must ensure that the exam is designed based on professionals' updated roles and responsibilities, with the equitable involvement of all relevant stakeholders. Furthermore, ensuring consistent enforcement of pass/fail decisions should never be overlooked. We also strongly advise the education sector to carefully interpret the exam's purpose. Programs should cultivate a comprehensive continuous quality improvement culture beyond and above NLE.

This study will improve understanding of the consequence validity evidence for an NLE in LMICs and non-physician training. We recommend conducting additional studies that quantitatively assess the exam's validity and enhance such understanding. In this regard, examining associations between NLE sores, in-school student performances, and clinical care outcomes is worthwhile.

Abbreviations

LMICs, Low- and Middle-Income Countries; MOH, Ministry of Health; NLE, National Licensing Examination.

Data Sharing Statement

The datasets generated and analyzed during the current study are not publicly available due to privacy and ethical concerns. However, they are available from the corresponding author upon reasonable request.

Ethical Approval and Informed Consent

We obtained ethical approval for the study from the Ethiopian Public Health Institute (EPHI-IRB-380-2021) and the Johns Hopkins Bloomberg School of Public Health Institutional Review Board (IRB # 17778). Permission to conduct the study was also secured from the Ministry of Health, the Ministry of Education (MOE), and teaching institutions. The school's Responsible Conduct of Research and the National Research Ethics Review Guidelines were strictly followed as guiding frameworks throughout the study. Each study participant signed informed consent prior to data collection, which included statements about publication of anonymized responses, and measures were taken to ensure privacy and confidentiality.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Acad Med*. 2016;91(6):785–795. doi:10.1097/ACM.0000000000001114
2. Ethiopian Ministry of Health. National human resources for health strategic plan 2016–2025; 2016.
3. Ethiopian Ministry of Health, National Health Workforce Update (Unpublished government report); 2019.
4. Law TJ, Bulamba F, Ochieng JP, et al. Anesthesia provider training and practice models: a survey of Africa. *Anesth Analg*. 2019;129(3):839–846. doi:10.1213/ANE.0000000000004302
5. Kibwana S, Teshome M, Molla Y, et al. Education, practice, and competency gaps of anesthetists in Ethiopia: task analysis. *J Perianesth Nurs*. 2018;33(4):426–435. doi:10.1016/j.jopan.2017.02.001
6. Asemu YM, Yigzaw T, Desta FA, Scheele F, van den Akker T. Evaluating the effect of interventions for strengthening non-physician anesthetists' education in Ethiopia: a pre- and post-evaluation study. *BMC Med Educ*. 2021;21(1):421. doi:10.1186/s12909-021-02851-0
7. Kibwana S, Woldemariam D, Misganaw A, et al. Preparing the health workforce in Ethiopia: a Cross-sectional study of competence of anesthesia graduating students. *Educ Health*. 2016;29(1):3–9. doi:10.4103/1357-6283.178931
8. Ethiopian Ministry of Health, Scope of practice for health professionals (Unpublished government report); 2021.
9. Emler TE, Zhao Y, Deng J, Yin D, Wang Y. Side effects of large-scale assessments in education. *ECNU Rev Educ*. 2019;2(3):279–296. doi:10.1177/2096531119878964
10. Norcini J, Luiz T; Foundations of assessment. *FAIMER-Keele Master's in Health Professions Education: Accreditation and Assessment. Module 1, Unit 1*. London: FAIMER Centre for Distance Learning, CenMEDIC; 2019.
11. Boulet JR. Establishing the validity of licensing examination scores. *J Grad Med Educ*. 2019;11(5):527–529. doi:10.4300/JGME-D-19-00611.1
12. AERA, AERA, APA APA, NCME NCoMiE, JCSEPT JCoSfEPT. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
13. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119 (2):166 e7–16. doi:10.1016/j.amjmed.2005.10.036
14. World Health Organization. *Transforming and Scaling Up Health Professionals' Education and Training: World Health Organization Guidelines 2013*. World Health Organization; 2013.
15. Archer J, Lynn N, Roberts M, Coombes L, Gale T, de Regand Bere S. A systematic review on the impact of licensing examinations for doctors in countries comparable to the UK. Collaboration for the Advancement of Medical Education Research; 2015.;
16. Foreman S. The accuracy of state NCLEX-RN(c) passing standards for nursing programs. *Nurse Educ Today*. 2017;52:81–86. doi:10.1016/j.nedt.2017.02.019
17. Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med Educ*. 2016;50(1):101–114. doi:10.1111/medu.12810
18. Kies S, Shultz M. Proposed changes to the United States Medical Licensing Examination: impact on curricula and libraries. *J Med Libr Assoc*. 2010;98(1):12–16. doi:10.3163/1536-5050.98.1.007
19. Gilliland WR, La Rochelle J, Hawkins R, et al. Changes in clinical skills education resulting from the introduction of the USMLE step 2 clinical skills (CS) examination. *Med Teach*. 2008;30(3):325–327. doi:10.1080/01421590801953026
20. Hauer KE, Teherani A, Kerr KM, O'Sullivan PS, Irby DM. Impact of the United States medical licensing examination step 2 clinical skills exam on medical school clinical skills assessment. *Acad Med*. 2006;81(10):S13–6. doi:10.1097/01.ACM.0000236531.32318.02
21. Archer J, Lynn N, Coombes L, et al. The impact of large scale licensing examinations in highly developed countries: a systematic review. *BMC Med Educ*. 2016;16(1):212. doi:10.1186/s12909-016-0729-7
22. Asemu YM, Yigzaw T, Desta FA, et al. Impact of anesthetist licensing examination on quality of education in Ethiopia: a qualitative study of faculty and student perceptions. *BMC Med Educ*. 2023;23(1):468. doi:10.1186/s12909-023-04452-5
23. Harden RM. Five myths and the case against a European or national licensing examination. *Med Teach*. 2009;31(3):217–220. doi:10.1080/01421590902741155
24. Gorsira M. The utility of (European) licensing examinations. AMEE Symposium, Prague 2008. *Med Teach*. 2009;31(3):221–222. doi:10.1080/01421590902741189
25. Boulet J, van Zanten M. Ensuring high-quality patient care: the role of accreditation, licensure, specialty certification and revalidation in medicine. *Med Educ*. 2014;48(1):75–86. doi:10.1111/medu.12286
26. Ranney RR. What the available evidence on clinical licensure exams shows. *J Evid Based Dent Pract*. 2006;6(1):148–154. doi:10.1016/j.jebdp.2005.12.012
27. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349–357. doi:10.1093/intqhc/mzm042
28. Creswell JW, Clark VLP. *Designing and Conducting Mixed Methods Research*. Sage publications; 2018.
29. Higher Education Relevance and Quality Agency, HERQA. National accreditation and quality improvement standards for anesthesia degree program; 2015 https://docplayer.net/6736368-Herqa-publication-series-084-higher-education-relevance-and-quality-agency.html.
30. World Federation for Medical Education WFME. *Basic Medical Education WFME Global Standards for Quality Improvement: The 2020 Revision*. 2nd ed. Ferney-Voltaire: WFME; 2020:29.
31. Raymond MR. *Job Analysis, Practice Analysis and the Content of Credentialing Examinations*. Routledge; 2016.
32. Yigzaw T, Carr C, Stekelenburg J, et al. Using task analysis to generate evidence for strengthening midwifery education, practice, and regulation in Ethiopia. *Int J Women's Health*. 2016:181–190. doi:10.2147/ijwh.s105046
33. Zhao Y. A world at risk: an imperative for a paradigm shift to cultivate 21st century learners. *Society*. 2015;52(2):129–135. doi:10.1007/s12115-015-9872-8
34. Norcini J, Anderson MB, Bollela V, et al. 2018 Consensus framework for good assessment. *Med Teach*. 2018;40(11):1102–1109. doi:10.1080/0142159X.2018.1500016
35. Konopasek L, Norcini J, Krupat E. Focusing on the formative: building an assessment system aimed at student growth and development. *Acad Med*. 2016;91(11):1492–1497. doi:10.1097/ACM.0000000000001171
36. Tymms P, Tymms P. Technical and ethical issues in indicator systems. *Educ Policy Anal Arch*. 2002;10:06. doi:10.14507/epaa.v10n6.2002

37. Van Zanten M, Norcini JJ, Boulet JR, Simon F. Overview of accreditation of undergraduate medical education programmes worldwide. *Med Edu*. 2008;42(9):930–937. doi:10.1111/j.1365-2923.2008.03092.x

38. Chhaparwal B, Christensen L, Gale R, Leinster S, Talbot M, Walters T; Foundations of assessment. *FAIMER-Keele Master's in Health Professions Education: Accreditation and Assessment. Module 2, Unit 1*. London: FAIMER Centre for Distance Learning, CenMEDIC; 2019.

39. Cheong CWS, Quah ELY, Chua KZY, et al. Post graduate remediation programs in medicine: a scoping review. *BMC Med Educ*. 2022;22(1):1–23. doi:10.1186/s12909-022-03278-x

40. General Medical Council. Misconduct procedures. Available from: https://bit.ly/3K3oxPA. Accessed March 30, 2023.