

The Ethiopian Anesthetist Licensing Examination and Associated Improvement in in-School Student Performance: A Retrospective Study

Yohannes Molla Asemu ^{1,2}, Tegbar Yigzaw ¹, Firew Ayalew¹, Leulayehu Akalu ^{3,4}, Fedde Scheele ^{2,5,6}, Thomas van den Akker ^{2,7}

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

Correspondence: Yohannes Molla Asemu, Jhpiego, an Affiliate of Johns Hopkins University, Ethiopia Country Office, Kirkos Sub-City, Kebele 02/03, House 693, Wollo Sefer, Addis Ababa, Ethiopia, Tel +251 911384046, Email Yohannes.Asemu@jhpiego.org

Background: Ethiopia introduced a national licensing examination (NLE) in response to growing concerns about the competence of graduates and the quality of education. This study aimed to assess the associated in-school student performance changes in anesthetist training programs following NLE implementation.

Methods: Academic records of 1493 graduate anesthetists were retrospectively obtained from eight universities before (n=932) and after (n=561) NLE implementation. Four universities were first-generation (oldest), three were second-generation, and one was third-generation (newest). We compared the yearly (Y1 to Y4) and cumulative grade point averages (GPA) to assess if there were in-school student performance differences between the two periods. The Kruskal–Wallis and Mann–Whitney *U*-tests were used to compare groups. Results are presented as a median, interquartile range, a 95% confidence interval (CI) for median differences, and Cohen's *r* effect size.

Results: Overall, there was a small to moderate improvement in student academic performance following NLE implementation. However, the statistically significant differences were limited to first-generation university students and those entering directly from high school. We found considerable positive differences in all five performance measures in first-generation university students, with Year-1 GPA and cumulative GPA measurements exhibiting large effect sizes (Cohen's *r* = 0.96 and 0.79, respectively, *p* <0.005). Those entering from high school demonstrated significant differences in four of five performance measures, with the largest positive gains on the year-1 GPA (median before [n=765] and after [n=480]: 3.11–3.30, 95% CI (0.09, 0.22), *r*=0.46, *p* <0.005). Second- and third-generation university students showed no significant differences, while nurse entrants exhibited a significant difference in their Y2GPA scores only with an actual drop in performance.

Conclusion: The Ethiopian anesthetist NLE is associated with an overall modest in-school academic performance improvement, supporting its use. The stagnant or declining performance among nurse entrants and the newest (second and third) generation university students deserve further scrutiny.

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

Introduction

In 2015, the Lancet Commission reported that an estimated five billion people worldwide desperately need essential surgical care. Access is worst in low and middle-income countries (LMICs), where nine out of ten people cannot access basic surgical care.¹ Major physician specialist surgical workforce shortages, compounded by uneven distribution of the existing workforce, disproportionately affect LMICs.^{2–4} This necessitates the production of a non-physician workforce through task-sharing in order to have cadres with shorter professional training provide services with or without supervision.⁵

In Africa, 38 out of 51 surveyed countries were previously found to implement anesthesia-related task-sharing using widely varying training standards, qualifications and scopes of practice.⁶ Ethiopia too prioritized task-sharing and increased the anesthesia workforce through the training of associate clinician anesthetists, simply referred to as ‘anesthetists’ in the country (therefore, this term will be used in the remainder of this article).^{7–9} However, simply training more health professionals may not be the answer to the vexing issue of improving access to safe essential surgical care.¹⁰ Studies conducted in Ethiopia have revealed that some graduated anesthetists may fall short of meeting competency requirements despite completing their education programs.^{11,12} In addition, an increase in perioperative medical errors attributed to anesthesia practice has been reported in recent years.¹³

Nearly a decade ago, the World Health Organization issued guidelines to transform the health workforce and improve health outcomes. One of the five key recommendations was the need to establish sustainable workforce regulation mechanisms through program accreditation and workforce licensing.¹⁰ Many countries have developed a licensing examination system to assess and monitor the inflow of newly licensed practitioners. These exams are designed and administered in different formats, assessing cognitive and psychomotor skills in single or multiple steps. The cognitive assessment formats mostly emphasize the integration of information across basic and clinical science disciplines and their application in clinical scenarios.^{14,15}

Exam validity is considered the most essential and critical criterion of the national licensing examination, without which the interpretation of results will be less meaningful and useful.^{16,17} The validation process requires the systematic synthesis of relevant evidence to support the exam’s use.¹⁸ Given that the benefits of exams should outweigh their costs and downsides, data on exam consequences, though notoriously difficult to collect, are the most fundamental source of validity evidence.^{18–20} In this regard, assessing improvements in in-school student performance (eg, academic performance differences between cohorts that took the exam and those that did not) may provide validity evidence for NLEs, if these were to demonstrate a positive impact of the NLE on the quality of pre-service education.¹⁹

Some evidence from high-income countries shows that students with better in-school academic performances perform better in NLEs, but not that NLE implementation improves students’ in-school performance.^{15,21} Published evidence assessing the associated changes in in-school student performance in LMICs following NLEs with non-physician groups is nonexistent. Previously, we qualitatively studied changes to anesthesia education due to the NLE. We found that student motivation, faculty performance, and learning and assessment practices have improved with clinically relevant written exams practiced regularly to assess higher-order cognitive thinking.^{22,23} Identifying quantitative associations may strengthen the validity evidence base and help teaching programs identify performance measures that result in good and poor changes and make necessary curricular modifications.

Ethiopia piloted a national licensing examination (NLE) from 2015–2018 using a single-step multiple-choice question (MCQ)-based format as a means to ensure that only competent graduates join the practicing workforce. Since 2019, health workforce graduates from 13 undergraduate training programs, including anesthesia, are required to pass the NLE to become licensed practitioners.²⁴ The Ethiopian NLE is currently limited to MCQs for financial, logistical, and technical challenges; psychomotor skill assessments heavily relied on assessments conducted during the university training, which employed contemporary methods such as objective structured clinical examination (OSCE) and workplace-based assessments. The aim of this study was to compare students’ in-school academic performance between those who completed their training before and after NLE implementation at multiple teaching institutions.

Materials and Methods

Study Design

We employed a cross-sectional research design by retrospectively collecting the academic records of graduate anesthetists.

Study Population and Setting

This study used data from eight of the ten public universities that have had graduate anesthetists since the start of the anesthetist NLE. Four teaching institutions were first-generation (oldest), three were second-generation, and one was third-generation (newest). Data could not be collected from the two remaining institutions due to conflict in the northern part of the country. Our study population was anesthetists who had completed their baccalaureate anesthesia training at the eight teaching

institutions between 2011 and 2021. Accordingly, academic records of 1493 eligible anesthetists were gathered. Students who did not complete the entire pre-service training or had incomplete academic records were excluded.

Training Program Description

The baccalaureate anesthesia training program accepts trainees from high school and nursing backgrounds and trains them for four and three years, respectively. A typical program for high school graduates is composed of one-year biomedical science studies, a one-year core clerkship (surgery, gynecology and obstetrics, internal medicine, and pediatrics), and two years of anesthesia specialty training (junior and senior year studies composed of theory and four practical attachments). Because of their prior qualification and experience, clinical nurses are exempted from the second-year core clerkship modules, reducing their total training duration by one year. Both groups of trainees in our study followed the same nationally standardized curriculum.

Data Source and Collection Procedure

Eight senior anesthesia instructors at the respective universities collected academic data digitally between September 20, 2021, and January 30, 2022, from university registrar offices. In-school academic performance was measured using yearly GPAs and the cumulative GPA (CGPA). GPAs are yearly unweighted composite scores of all module grades assigned for the academic year. Individual module grades are computed out of 100%, and this total score is converted to a four-point scale and recorded on student transcripts. Multiple assessment methods are employed in each module to assess all three learning domains (knowledge, skill, and attitude). Written exams in the form of context-based MCQs and structured essays, structured oral exams, viva, projects, and assignments are commonly employed knowledge assessment methods. Aside from traditional long and short exams, skill (and attitude) assessment methods for clinical-oriented courses and modules include OSCE and a variety of workplace-based assessments such as DOPS, Mini-CEX, and CBDs. The CGPA is a composite score of the yearly GPAs.

We collected the yearly GPAs (Y1 GPA, Y2 GPA, Y3 GPA, and Y4 GPA) and the CGPA with a 4-point scoring system for our study. To ensure data quality, academic records were collected by trained data collectors and cross-checked with data from respective programs, and data quality was regularly checked after submission.

Data Analysis

Academic data were cleaned and separated into two sets: before NLE (2011–2018) and after NLE implementation (2019–2021). Years 1 to 4 GPAs and CGPA were used to compare in-school student performance differences between the groups. Due to the non-normal data distribution, the Kruskal–Wallis and Mann–Whitney *U*-tests were used to compare score differences, while the Benjamini-Hochberg method²⁵ was used to adjust for multiple comparisons. The results are presented in tables as median (Md), interquartile range (IQR), and 95% confidence interval (CI) for median differences. Also, Cohen's *r* values were calculated to compare the effect size for the comparison between two medians and interpreted using Cohen (1988) criteria.²⁶ All analyses were performed using SPSS ver27 (IBM Corp.).

Results

Characteristics of Study Participants

Academic records of 1493 anesthetists, including 932 who graduated before and 561 after the NLE implementation, were analyzed, accounting for 96.8% of all enumerated anesthetists. Study participants were predominantly male (67.2%), of relatively young age (80.8% aged 17–22 years at enrollment in to their studies), had entered directly from high school (83.4%), and were enrolled in a first-generation university (67.6%).

Differences in Student Academic Performances

A Mann–Whitney *U*-test revealed a statistically significant difference in first-generation university graduates' median GPAs and CGPAs, with Y1GPA, Y2GPA, and CGPA exhibiting medium to large effect sizes ($r \geq 0.5$, $p < 0.005$). However, after adjusting for multiple comparisons, there were no statistically significant differences in the scores of second and third-generation university graduates (Table 1).

Table 1 Differences in Anesthesia Students GPAs Between the Before and After NLE Groups by University Generation, Ethiopia, 2011–2021, (n=1493)

GPAs	University Generation	Before NLE Median (IQR)	After NLE Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P-value
Y1GPA	First (n=1009)	3.09 (0.80)	3.36 (0.77)	0.22 (0.15, 0.30)	0.96	0.001**
	Second (n=392)	3.21 (0.70)	3.10 (0.98)	-0.10 (-0.22, 0.01)	0.21	0.067
	Third (n=92)	3.48 (0.63)	3.27 (0.85)	-0.18 (-0.42, -0.02)	0.13	0.036
Y2GPA	First (n=1009)	3.13 (0.66)	3.31 (0.67)	0.12 (0.06, 0.19)	0.50	0.001**
	Second (n=392)	3.11 (0.60)	3.00 (0.82)	-0.11 (-0.01, -0.22)	0.29	0.029
	Third (n=92)	3.17 (0.69)	3.12 (0.79)	-0.07 (-0.30, 0.14)	0.01	0.480
Y3GPA	First (n=1009)	3.23 (0.60)	3.35 (0.71)	0.11 (0.05, 0.17)	0.40	0.001**
	Second (n=392)	3.18 (0.59)	3.10 (0.83)	-0.08 (-0.18, 0.03)	0.17	0.146
	Third (n=92)	3.20 (0.81)	3.22 (0.87)	0.06 (-0.17, 0.29)	0.01	0.614
Y4GPA	First (n=875)	3.46 (0.60)	3.59 (0.58)	0.07 (0.01, 0.13)	0.29	0.012*
	Second (n=292)	3.22 (0.59)	3.20 (0.94)	-0.01 (-0.12, 0.14)	0.01	0.090
	Third (n=78)	3.33 (0.88)	3.41 (0.97)	0.05 (-0.30, 0.17)	0.01	0.631
CGPA	First (n=1009)	3.24 (0.54)	3.41 (0.60)	0.14 (0.09, 0.19)	0.79	0.001**
	Second (n=392)	3.21 (0.59)	3.11 (0.77)	-0.09 (-0.19, 0.002)	0.32	0.058
	Third (n=92)	3.22 (0.83)	3.26 (0.83)	0.04 (-0.25, 0.16)	0.01	0.671

Notes: * $p < 0.05$, adjusted for multiple tests; ** $p < 0.005$, adjusted for multiple tests.

Abbreviations: IQR, interquartile range; CI, confidence interval; GPA, grade point averages; CGPA, cumulative grade point averages; Y, year of study.

On the other hand, the significant performance gaps in Y1GPA before the NLE in favor of third-generation students (Md 3rd generation = 3.48; Md 1st generation = 3.09; performance gap = 0.41; 95% CI = 0.22, 0.60; $r=0.42$; $p < 0.005$) disappeared after NLE implementation, and even first-generation students significantly outperformed the second generation students in similar performance measure (Md 2nd generation = 3.10; Md 1st generation = 3.36; gap = -0.25; 95% CI = -0.35, -0.15; $r=0.75$; $p < 0.005$). Besides, after NLE implementation, first-generation students persistently outperformed the younger-generation students on the rest academic measures (Table 2).

Table 2 Differences in Anesthesia Students GPAs Among the Three Generation Students by the Before and After-NLE Status, Ethiopia, 2011–2021, (n=1493)

GPAs	Group (n)	1st Gen. Median (IQR)	2nd Gen. Median (IQR)	3rd Gen. Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P value
Y1GPA	Before NLE (n=932)	3.09 (0.80)	3.21 (0.70)	3.48 (0.63)	0.41 (0.22, 0.60) [‡]	0.42	0.001**
	After NLE (n=561)	3.36 (0.77)	3.10 (0.98)	3.27 (0.85)	-0.25 (-0.35, -0.15) [†]	0.75	0.001**
	Overall (n=1493)	3.18 (0.84)	3.15 (0.77)	3.32 (0.80)	0.20 (0.09, 0.31) [‡]	0.15	0.049

(Continued)

Table 2 (Continued).

GPA	Group (n)	1st Gen. Median (IQR)	2nd Gen. Median (IQR)	3rd Gen. Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P value
Y2GPA	Before NLE (n=932)	3.13 (0.66)	3.11 (0.60)	3.17 (0.69)	0.09 (-0.09, 0.27) [‡]	0.05	0.965
	After NLE (n=561)	3.31 (0.67)	3.00 (0.82)	3.12 (0.79)	-0.25 (-0.35, -0.16) [†]	0.81	0.001**
	Overall (n=1493)	3.14 (0.65)	3.06 (0.72)	3.15 (0.73)	-0.11 (-0.17, -0.06) [†]	0.19	0.018*
Y3GPA	Before NLE (n=932)	3.23 (0.60)	3.18 (0.59)	3.20 (0.81)	-0.07 (-0.14, -0.003) [†]	0.06	0.420
	After NLE (n=561)	3.35 (0.71)	3.10 (0.83)	3.22 (0.87)	-0.24 (-0.34, -0.15) [†]	0.76	0.001**
	Overall (n=1493)	3.29 (0.65)	3.14 (0.71)	3.21 (0.81)	-0.14 (-0.19, -0.08) [†]	0.37	0.001**
Y4GPA	Before NLE (n=765)	3.46 (0.60)	3.22 (0.59)	3.33 (0.88)	-0.21 (-0.29, -0.14) [†]	0.68	0.001**
	After NLE (n=480)	3.59 (0.58)	3.20 (0.94)	3.41 (0.97)	-0.26 (-0.37, -0.14) [†]	0.84	0.001**
	Overall (n=1245)	3.48 (0.60)	3.22 (0.72)	3.37 (0.92)	-0.22 (-0.29, -0.16) [†]	0.99	0.001**
CGPA	Before NLE (n=932)	3.24 (0.54)	3.21 (0.59)	3.22 (0.83)	-0.02 (-0.08, 0.04) [†]	0.01	0.867
	After NLE (n=561)	3.41 (0.60)	3.11 (0.77)	3.26 (0.83)	-0.24 (-0.33, -0.16) [†]	0.99	0.001**
	Overall (n=1493)	3.30 (0.56)	3.16 (0.70)	3.22 (0.81)	-0.10 (-0.15, -0.05) [†]	0.37	0.001**

Notes: All statistically significant differences are presented, while the rest non-significant ones are not. [†]Median difference between first and second-generation university students; [‡]median difference between first and third generation university students; * $p < 0.05$, adjusted for multiple tests; ** $p < 0.005$, adjusted for multiple tests.

Abbreviations: IQR, interquartile range; CI, confidence interval; GPA, grade point averages; CGPA, cumulative grade point averages; Y, year of study; gen, university generation.

A sub-group analysis by entry behavior showed that high school graduates in the NLE group showed considerable positive differences in their performance scores with small to moderate effect sizes ($r = 0.27-0.46$), except for Y4GPA. Nurse entrants, on the contrary, had a statistically significant difference in their Y2GPA scores only, but with an actual drop in performance following NLE implementation (Md before = 3.43, $n=167$; Md after = 3.14, $n=81$; median difference = -0.31 ; 95% CI = $(-0.43, -0.19)$; $r = 0.51$; $p < 0.005$) (Table 3).

Table 3 Differences in Anesthesia Students GPAs Between the Before and After NLE Groups by Entry Behavior, Ethiopia, 2011–2021, ($n=1493$)

GPA	Entry Behavior	Before NLE Median (IQR)	After NLE Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P-value
Y1GPA	High school ($n=1245$)	3.11 (0.78)	3.30 (0.89)	0.16 (0.09, 0.22)	0.46	0.001**
	Nurses ($n=248$)	3.19 (0.76)	3.02 (0.74)	-0.13 (-0.26, 0.01)	0.34	0.072
	Overall ($n=1493$)	3.14 (0.78)	3.27 (0.87)	0.11 (0.06, 0.17)	0.19	0.001**

(Continued)

Table 3 (Continued).

GPA	Entry Behavior	Before NLE Median (IQR)	After NLE Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P-value
Y2GPA	High school (n=1245)	3.09 (0.62)	3.21 (0.81)	0.10 (0.04, 0.15)	0.31	0.001**
	Nurses (n=248)	3.43 (0.69)	3.14 (0.65)	-0.31 (-0.43, -0.19)	0.51	0.001**
	Overall (n=1493)	3.14 (0.65)	3.20 (0.77)	0.03 (-0.02, 0.08)	0.07	0.248
Y3GPA	High school (n=1245)	3.22 (0.58)	3.32 (0.80)	0.07 (0.01, 0.12)	0.27	0.014*
	Nurses (n=248)	3.50 (0.58)	3.42 (0.86)	-0.09 (-0.21, 0.02)	0.05	0.098
	Overall (n=1493)	3.27 (0.61)	3.33 (0.81)	0.03 (-0.01, 0.08)	0.14	0.166
Y4GPA	High school (n=1245)	3.40 (0.60)	3.46 (0.74)	0.05 (-0.02, 0.08)	0.03	0.214
	Nurses (n=0)	NA	NA	NA	NA	NA
	Overall (n=1259)	3.40 (0.60)	3.46 (0.74)	0.03 (-0.02, 0.08)	0.03	0.214
CGPA	High school (n=1245)	3.22 (0.55)	3.33 (0.73)	0.08 (0.03, 0.13)	0.28	0.001**
	Nurses (n=248)	3.33 (0.62)	3.17 (0.66)	-0.12 (-0.23, -0.01)	0.09	0.042
	Overall (n=1493)	3.23 (0.55)	3.32 (0.72)	0.05 (0.01, 0.09)	0.14	0.029

Notes: * $p < 0.05$, adjusted for multiple tests; ** $p < 0.005$, adjusted for multiple tests.

Abbreviations: IQR, interquartile range; CI, confidence interval; GPA, grade point averages; CGPA, cumulative grade point averages; Y, year of study. NA, not applicable as nurses complete their training in 3 years.

Comparing high school and nurse entrants also revealed that all performance differences observed before NLE in favor of nurses had vanished, and even high school entrants performed better in Y1GPA after NLE (Md high school entrant = 3.30; Md nurses = 3.02; median difference = -0.20; 95% CI, -0.31, -0.07; $r = 0.24$; $p < 0.005$ (Table 4).

Table 4 Differences in Anesthesia Students' GPAs Between High School and Nurse Entrants Among the Before and After-NLE Groups, Ethiopia, 2011–2021, (n=1493)

GPA	Group (n)	High school Median (IQR)	Nursing Median (IQR)	Median Difference (95% CI)	Cohen's r	P-value
Y1GPA	Before NLE (n=932)	3.11 (0.78)	3.19 (0.76)	0.09 (0, 0.18)	0.06	0.053
	After NLE (n=561)	3.30 (0.89)	3.02 (0.74)	-0.20 (-0.31, -0.07)	0.24	0.003*
	Overall (n=1493)	3.20 (0.83)	3.12 (0.75)	-0.01 (-0.08, 0.06)	0.01	0.728
Y2GPA	Before NLE (n=932)	3.09 (0.62)	3.43 (0.69)	0.36 (0.28, 0.44)	1.49	0.001**
	After NLE (n=561)	3.21 (0.81)	3.14 (0.65)	-0.05 (-0.16, 0.07)	0.01	0.440
	Overall (n=1493)	3.13 (0.68)	3.37 (0.72)	0.22 (0.16, 0.29)	0.63	0.001**
Y3GPA	Before NLE (n=932)	3.22 (0.58)	3.50 (0.58)	0.25 (0.18, 0.32)	1.07	0.001**
	After NLE (n=561)	3.32 (0.80)	3.42 (0.86)	0.09 (-0.02, 0.19)	0.14	0.110
	Overall (n=1493)	3.26 (0.68)	3.48 (0.62)	0.19 (0.14, 0.25)	0.69	0.001**

(Continued)

Table 4 (Continued).

GPA	Group (n)	High school Median (IQR)	Nursing Median (IQR)	Median Difference (95% CI)	Cohen's r	P-value
Y4GPA	Before NLE (n=765)	3.40 (0.60)	NA	NA	NA	NA
	After NLE (n=480)	3.46 (0.74)	NA	NA	NA	NA
	Overall (n=1245)	3.43 (0.65)	NA	NA	NA	NA
CGPA	Before NLE (n=932)	3.22 (0.55)	3.33 (0.62)	0.10 (0.03, 0.16)	0.17	0.004*
	After NLE (n=561)	3.33 (0.73)	3.17 (0.66)	-0.10 (-0.20, 0.01)	0.06	0.064
	Overall (n=1493)	3.25 (0.60)	3.29 (0.64)	0.03 (-0.03, 0.09)	0.01	0.295

Notes: * $p < 0.05$, adjusted for multiple tests; ** $p < 0.005$, adjusted for multiple tests.

Abbreviations: IQR, interquartile range; CI, confidence interval; GPA, grade point averages; CGPA, cumulative grade point averages; Y, year of study. NA, not applicable as nurses complete their training in 3 years.

Although positive gender-based performance differences were found in all performance measures between the before and after NLE groups, none were statistically significant (table not shown). On the other hand, the significant gender gaps observed on Y4GPA before the NLE in favor of males (Md males = 3.47; Md females = 3.27; gender gap = 0.19; 95% CI = 0.12–0.26; $r=0.66$; $p < 0.005$) disappeared after NLE implementation (Md males = 3.51; Md females = 3.39; gap = 0.10; 95% CI = 0.01–0.18; $r=0.17$; $p > 0.05$). Although the significant gender gaps in Y1GPA, Y3GPA, and CGPA observed before NLE favoring males also narrowed, all these gaps persisted with strong large effect sizes after NLE implementation (Table 5).

Table 5 Differences in Anesthesia Students GPAs Between Male and Females Among the Before and After-NLE Groups, Ethiopia, 2011–2021

GPA	Group (n)	Female: Median (IQR)	Male: Median (IQR)	Median Estimate Difference (95% CI)	Cohen's r	P-value
Y1GPA	Before NLE (n=932)	2.88 (0.80)	3.26 (0.69)	0.31 (0.23, 0.38)	1.06	0.001**
	After NLE (n=561)	3.06 (0.92)	3.38 (0.76)	0.24 (0.14, 0.34)	0.68	0.001**
	Overall (n=1493)	2.93 (0.83)	3.29 (0.76)	0.28 (0.22, 0.34)	1.06	0.001**
Y2GPA	Before NLE (n=932)	3.05 (0.77)	3.17 (0.60)	0.12 (0.05, 0.19)	0.23	0.001**
	After NLE (n=561)	3.07 (0.81)	3.29 (0.67)	0.19 (0.11, 0.28)	0.64	0.001**
	Overall (n=1493)	3.04 (0.78)	3.20 (0.63)	0.15 (0.09, 0.20)	0.42	0.001**
Y3GPA	Before NLE (n=932)	3.00 (0.66)	3.30 (0.50)	0.21 (0.15, 0.28)	1.00	0.001**
	After NLE (n=561)	3.13 (0.89)	3.38 (0.75)	0.17 (0.08, 0.26)	0.51	0.001**
	Overall (n=1493)	3.10 (0.75)	3.35 (0.60)	0.20 (0.15, 0.25)	1.00	0.001**
Y4GPA	Before NLE (n=765)	3.27 (0.76)	3.47 (0.54)	0.19 (0.12, 0.26)	0.66	0.001**
	After NLE (n=480)	3.39 (0.84)	3.51 (0.69)	0.10 (0.01, 0.18)	0.17	0.023
	Overall (n=1245)	3.31 (0.79)	3.47 (0.61)	0.15 (0.09, 0.20)	0.59	0.001**
CGPA	Before NLE (n=932)	3.00 (0.65)	3.28 (0.44)	0.19 (0.14, 0.25)	0.88	0.001**
	After NLE (n=561)	3.14 (0.75)	3.38 (0.63)	0.15 (0.07, 0.24)	0.55	0.001**
	Overall (n=1493)	3.10 (0.69)	3.32 (0.53)	0.18 (0.13, 0.23)	0.95	0.001**

Note: ** $p < 0.005$, adjusted for multiple tests.

Abbreviations: IQR, interquartile range; CI, confidence interval; GPA, grade point averages; CGPA, cumulative grade point averages; Y, year of study.

Discussion

This study provides new consequence validity evidence on the associated improvement in in-school student performance following the implementation of NLE. Overall, modest in-school student performance changes were observed after NLE, with high school entrants and those attending their study in first-generation universities showing relevant improvements.

In this study, we found a positive shift in most student performance measures among first-generation (oldest) universities following NLE implementation. However, there was no change in student performance after NLE among the newest (second and third) generation universities, which produce more than three-quarters of the country's yearly graduate anesthetists. This could be attributed to better-quality students enrolling in first-generation universities, as well as challenges in improving educational quality in newer generation universities, such as critical shortages of educational resources, shortage of qualified faculty and lower staff profiles (oldest universities require a second degree as a mandatory requirement, while first-degree holders are still involved in teaching in newest generations), and limited student clinical exposure due to fewer clinical practice sites and preceptors.^{9,12,27}

The overall positive student performance changes following the NLE were also confined to high school entrants. In contrast, nurse students experienced a statistically significant decrease in Y2GPA, while the rest of the measurements also decreased but not significant. Besides, all performance differences observed before NLE in favor of nurses disappeared. Even high school entrants performed better in one performance parameter (Y1GPA) after the NLE. A pre-post competency assessment study in Ethiopia found similar results, with anesthetist competence scores significantly different based on entry behavior.^{9,12} Studies in other countries found no difference in the competence of graduates who followed different paths to obtain the same qualifications.²⁸ The persistent group differences reported in this study could be attributed to smaller sizes of nurse entrants in our study (17.9% of the before and 14.4% of the after) or differences in quality of education.

We found no significant gender-based differences in student performance between the before and after NLE groups. Besides, gender gaps observed before NLE in favor of males narrowed after NLE implementation. Our study findings were consistent with a government report²⁹ and the national education development roadmap,³⁰ which stated that gender gaps in the performance of higher education students had recently improved. The improvements observed in our study could be attributed to the strengthened institutional gender-responsive interventions.⁹ However, the sustained gender disparity in four of the five performance measures favoring males is alarming. A program evaluation study in Ethiopia found contradicting results, with anesthetist competence scores differences based on gender bridged after program interventions.^{9,11} The persistent group differences reported in this study could be attributed to differences in study methodology, as the current study used GPAs, whereas, the previous studies employed a one-off objective structured clinical examination (OSCE).

Another intriguing finding from this study was the decrease in in-school performance differences (between the before and after NLE groups) as the training years progressed. Given that the NLE focuses primarily on the final two years of anesthesia specialty content, it's inevitable for anesthesia schools to pay due attention to these training years to prepare their graduates for the exam.^{22,23} These senior-year-centered educational reforms, characterized primarily by repeated assessments using different methods, may have made in-school assessments more stringent, lowering senior students' performance after NLE. The rising stress and workload on senior-year students due to heightened expectations from programs to increase institutional NLE pass rates could also contribute to the decreasing performance differences. Furthermore, increased efforts by senior students to prepare for the comprehensive in-school exit exam and the subsequent NLE may contribute to lower scores by consuming more of their time and distracting their focus.

Conclusion

In conclusion, the current study will contribute to the body of validity evidence for the Ethiopian anesthetist licensing examination. The results showed that the Ethiopian anesthetist NLE is associated with modest improvements in in-school student performance. However, performance improvements were mostly positive and statistically significant only for those entering directly from high school and those attending first-generation university program. Performance stagnation and drops among nurses and newest generation university students warrant examination of education quality and systems to align institutional assessment with NLE. Furthermore, the small effect sizes and decreasing performance differences (as the

training year progressed) reinforced the need to strengthen junior-centered educational interventions to improve graduates' competence and licensing exam pass rates through the proper use of NLE data.

Recommendations

Given the rapidly increased number of anesthesia teaching institutions and newly trained anesthesiologists, a more stringent quality assurance system, including program accreditation that takes NLE pass rates into account, is needed. Younger generation universities can benefit from increased engagement in and orientation into NLE design, content (eg by sharing blueprints and sample items) and administration processes. Besides, nurse entrant training programs can benefit from proactive sharing of experiences with high school entrant programs, effective use of NLE data to inform their training, and strengthening overall systems for educational quality improvement and assurance. Collectively, we strongly believe the time has come to scale up the infantile voluntary accreditation efforts and standardize education quality through mandatory accreditation that takes NLE pass rates into account. In this regard, establishing a fully autonomous regulatory authority is a fundamental milestone.

The sustained gender disparity favoring males calls for strengthening gender-responsive institutional interventions primarily targeting junior-year students, thereby contributing to the 2030 sustainable development goals 4 and 5 on inclusive, equitable and gender-sensitive quality lifelong learning.³¹ Standardizing institutional exit exams, introducing small-scale clinically relevant summative assessments, such as year-based qualification exams, and incorporating some form of skill assessments (eg OSCE) in the NLE may help distribute the burden that caused academic performance to decline as the training year progressed and promote comprehensive learning.

Finally, given the growing concerns about the quality of patient care and safety¹³ that question the competency of graduates and the appropriateness of the NLE's decision to filter less competent graduates, further research into the relationship between licensing examination scores and in-school academic performance and patient outcomes is recommended.

Strengths and Limitations

Strengths of the Study

This study has a number of strengths. It involved all eligible anesthesiologist teaching institutions obtaining nationally representative sample. Secondly, a relatively large-scale data on student performance before and after NLE implementation was collected, allowing comparisons of changes in student performance between groups over a decade. Finally, multiple in-school performance measures were evaluated, allowing for triangulation.

Limitations of the Study

The lack of a control group was the most significant limitation of this study, limiting establishment of causal relationships. As a result, our study design may fail to distinguish between the NLE's direct effects and any other effects caused by external factors.³² However, we conducted analyses in strata of potential confounders (gender, entry behavior and university generation), which by and large did not alter our outcomes. Another important limitation is that we only looked at anesthesia programs, so our findings may not be generalizable to other programs as learning and assessment contexts may differ. Finally, data were not collected from two institutions due to security concerns, but the authors believe that their context does not differ significantly from that of the institutions involved.

Abbreviation

CGPA, cumulative grade point average; GPA, grade point average; LMICs, low and middle-income countries; NLE, National licensing examination.

Data Sharing Statement

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

Ethical Approval and Informed Consent

Ethical approval for the study was obtained from the Ethiopian Public Health Institute (EPHI-IRB-380-2021) and the Johns Hopkins Bloomberg School of Public Health Institutional Review Board (IRB # 17778). Given that we require de-identified secondary data stored in higher education institutions, individual consent was waived by the ethics review boards. However, permission to conduct the study was obtained from the Ministry of Education, and the deans of training institutions. Measures were taken to protect autonomy and data confidentiality.

Acknowledgment

The study authors would like to thank the Federal Ministry of Health, Ministry of Education, Ethiopian Association of Anesthetists and study site teaching institutions for their support. The authors would like to express special appreciation to data collectors and supervisors.

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.

Funding

This study was made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the Cooperative Agreement No. 72066320CA00008. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

Disclosure

The authors declare that they have no competing interests.

References

- Meara JG, Leather AJ, Hagander L, et al. Global surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Lancet*. 2015;386(9993):569–624. doi:10.1016/S0140-6736(15)60160-X
- Dare AJ, Ng-Kamstra JS, Patra J, et al. Deaths from acute abdominal conditions and geographical access to surgical care in India: a nationally representative spatial analysis. *Lancet Glob Health*. 2015;3(10):e646–53. doi:10.1016/S2214-109X(15)00079-0
- Epiu I, Tindimwebwa JV, Mijumbi C, et al. Challenges of anesthesia in low- and middle-income countries: a cross-sectional survey of access to safe obstetric anesthesia in East Africa. *Anesth Analg*. 2017;124(1):290–299. doi:10.1213/ANE.0000000000001690
- Kudsk-Iversen S, Shamambo N, Bould MD. Strengthening the anesthesia workforce in low-and middle-income countries. *Anesth Analg*. 2018;126(4):1291–1297. doi:10.1213/ANE.0000000000002722
- Ashengo T, Skeels A, Hurwitz EJH, Thuo E, Sanghvi H. Bridging the human resource gap in surgical and anesthesia care in low-resource countries: a review of the task sharing literature. *Hum Resour Health*. 2017;15(1):77. doi:10.1186/s12960-017-0248-6
- 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
- Ethiopian Ministry of Health. National human resources for health strategic plan 2016–2025; 2016.
- Ethiopian Ministry of Health. Health sector Transformation Plan 2015/16–2019/20 (2008–2012 EFY) 183; 2015.
- 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
- World Health Organization. *Transforming and Scaling Up Health Professionals' Education and Training: World Health Organization Guidelines 2013*. World Health Organization; 2013.
- 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
- 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
- Wamisho BL, Tiruneh MA, Teklemariam E. Surgical and medical error claims in Ethiopia: trends observed from 125 decisions made by the federal ethics committee for health professionals ethics review. *Medicoleg Bioeth*. 2019;9:23. doi:10.2147/MB.S219778
- Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med Educ*. 2016;50(1):101–114. doi:10.1111/medu.12810
- 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

16. Norcini J, Luiz T. Foundations of assessment. In: Janet G, editor. *FAIMER-Keele Master's in Health Professions Education: Accreditation and Assessment. Module 1, Unit 1*. London: FAIMER Centre for Distance Learning, CenMEDIC; 2019.
17. 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
18. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
19. 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
20. 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
21. Hu Y, Martindale JR, LeGallo RD, White CB, McGahren ED, Schroen AT. Relationships between preclinical course grades and standardized exam performance. *Adv Health Sci Educ*. 2016;21(2):389–399. doi:10.1007/s10459-015-9637-6
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):1. doi:10.1186/s12909-023-04452-5
23. Asemu YM, Yigzaw T, Desta FA, et al. Untoward consequences of the Ethiopian national licensing examination: a grounded theory study; 2022.;
24. Ethiopian Ministry of Health. Health professionals' competency assessment and licensure. Available from: <https://bit.ly/3T3YDId>. Accessed November 1, 2022.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc*. 1995;57(1):289–300.
26. Pallant J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. Routledge; 2020.
27. Salmi J, Sursock A, Olefir A Improving the performance of Ethiopian universities in science and technology: a policy note. Improving the Performance of Ethiopian Universities in Science and Technology: A Policy Note, World Bank Group; 2017.
28. Wehner L. *Comparing National Council Licensure Exam-Registered Nurse Pass Rates in Accelerated and Traditional Prelicensure BSN Programs*. Northcentral University; 2015.
29. Higher education statistics annual abstract; 2019.
30. Teferra T, Asgedom A, Oumer J, Dalelo A, Assefa B. Ethiopian education development roadmap (2018–2030). An integrated Executive Summary 101; 2018.
31. Desa U Transforming our world: the 2030 agenda for sustainable development; 2016.
32. Frye AW, Hemmer PA. Program evaluation models and related theories: AMEE guide no. 67. *Med Teach*. 2012;34(5):e288–e299. doi:10.3109/0142159X.2012.668637

Advances in Medical Education and Practice

Dovepress

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>