The 41st Annual Conference of The Association for Educational Assessment in Africa (AEAA)

Assessment of the CBT System in Ethiopia's Secondary School Leaving Certificate Examination (2016 E.C./2024 G.C.): Practices, Challenges, and Future Prospects

(Abitew Asmare, Fikre Sahile, Marefu Lerebo and Wondwosen Eyesuswork)

Sub theme: Leveraging Technology in Educational Assessment

Presented by: Wondwosen Eyesuswork

August 25/2025, Addis Ababa, Ethiopia

41st AEAA Annual Conference

Presentation Overview

August 2025
EAES, Addis Ababa, Ethiopia

- **Introduction**
  - **Purpose of the Study**
  - **Theoretical Framework of the Study**
  - **Global Experiences for Inspiration**
- **Methodology**
  - **Research Methodology**
  - **Participant of the study**
- **Result and Discussion**
  - **Key Practices in CBT Implementation**
  - **Major Challenges and Future Prospects**
  - **CBT system's usability, compatibility, and data management**
  - **Performance Patterns: CBT vs. PPBT**
- **Conclusion, recommendations and implication of the research**

Introduction

August 2025
EAES, Addis Ababa, Ethiopia

- **Ethiopia's national public examinations** have historically been conducted using traditional **Paper-and-Pencil Based Testing (PPBT)** methods. **Key exams include:**
  - **ESLCE** (Grade 12, 1950–2003)
  - **EGSECE** (Grade 10, 2001–2019)
  - **EHEECE** (2004–2010)
  - **EUEE** (2011–2021)
  - **ESSLCE** (2021–present)
- **National exams**, overseen by **EAES**, are crucial for students' **academic and career paths** (Mamaru, Getachew, & Tafese, 2023). However, the traditional **PPBT system** faces increasing challenges:
  - Logistical delays and slow result processing
  - Growing risks of malpractice and exam leakage (Dessalegn & Mekonen, 2024)
  - High administrative costs

Introduction…

August 2025
EAES, Addis Ababa, Ethiopia

- Ethiopia is transitioning to a hybrid exam system, retaining PPBT in most regions while advancing CBT for its benefits in security, rapid scoring, real-time monitoring, and reduced logistics (Bati & Workneh, 2021).
- Despite these benefits, the **rollout of CBT presents substantial challenges**, especially in Ethiopia's varied educational and technological landscape:
  - **Infrastructure disparities** between urban and rural areas limit equitable access (Bati & Workneh, 2021; Welesilassie & Gerencheal, 2025).
  - **Low digital literacy** among students and educators slows adoption (Welesilassie & Gerencheal, 2025).
  - **Technical and administrative complexities** remain unresolved (Welesilassie & Gerencheal, 2025).
  - **Digital divides between private and public** schools risk widening inequalities (Woldehanna & Araya, 2016)
- **Research gap –** Current studies rarely address **the practical and operational realities of CBT in Ethiopia's high-stakes exams.** While Zebediwos Zekarias (JEAE, 3[1]) offers insights into **educators' experiences during early ESSLCE CBT rounds**, most research still centers on ICT readiness, overlooking key pedagogical, managerial, and policy issues (Dessalegn & Mekonen, 2024).

Purpose of the Study

August 2025
EAES, Addis Ababa, Ethiopia

- This study seeks to critically explore how CBT can be **effectively, equitably, and securely integrated** into Ethiopia's high-stakes examination system. Specifically, it aims to answer the following research questions:
  - What are the **current practices** in implementing CBT within Ethiopia's hybrid testing model, and which exam center type—**university-level** or **school-level**—demonstrates greater effectiveness?
  - What are the **future prospects and challenges** of scaling CBT in Ethiopia's Secondary School Leaving Certificate Examination (ESSLCE), and which exam center type is more affected?
  - How user-friendly is the current **CBT platform** for students and administrators? What improvements are necessary, and which exam center type offers a better user experience?
  - How does **student performance** in CBT differ across school types and regions? What equity gaps emerge from these differences?
  - What steps can be taken to improve the **accessibility**, **security**, and **scalability** of CBT in Ethiopia, in line with international best practices?

Theoretical Framework of the Study

August 2025
EAES, Addis Ababa, Ethiopia

- This study is anchored in a multifaceted framework that ensures the **validity, fairness, reliability, and inclusivity** of Ethiopia's hybrid assessment system.
  - **Psychometric Theories** such as *Classical Test Theory (CTT)* and *Item Response Theory (IRT)* provide statistical grounding for test reliability and adaptive design (Crocker & Algina, 1986; Embretson & Reise, 2013).
  - **Cognitive Load Theory** (Sweller, 1988) emphasizes reducing unnecessary mental effort during digital testing, especially critical for students with low digital exposure.
  - The **TPACK framework** (Mishra & Koehler, 2006) promotes meaningful integration of technology into assessment, beyond digitizing existing formats.
  - **Constructivist and Sociocultural theories** (Piaget, 1970; Vygotsky, 1978) support context-aware, scaffolded assessments that reflect learners' backgrounds and interactions.
  - **Validity and fairness principles** (Messick, 1989; Kane, 2006; Popham, 2010) ensure that exams are not only psychometrically sound but also equitable and accessible.

41st AEAA Annual Conference

Global Experiences for Inspiration

August 2025
EAES, Addis Ababa, Ethiopia

- International experiences offer practical models for Ethiopia's journey. These cases highlight the importance of **system readiness**, **gradual scaling**, and **inclusive planning** for Ethiopia's successful hybrid exam reform.
  - **Nigeria (JAMB):** Implemented biometric authentication to enhance exam security and conducted nationwide mock CBTs to build candidate familiarity, improve trust in the system, and address access-related challenges (Ajeyalemi et al., 2021; Udoh & Essien, 2024).
  - **Finland:** Successfully transitioned to a fully digital examination system by integrating secure, centralized platforms with comprehensive, nationwide teacher training programs, ensuring both technical readiness and pedagogical alignment (Savolainen, 2017).
  - **Georgia:** Demonstrated that Computer Adaptive Testing (CAT) can be effectively scaled at the national level when supported by strategic investment in technology, infrastructure, and human resource capacity (Sadeghi & Khonbi, 2017).
- ***Lesson*:** Invest in infrastructure + training + gradual scaling = equitable, robust CBT system

Methodology

August 2025
EAES, Addis Ababa, Ethiopia

- This study adopted a **mixed-methods approach** to gain both **quantitative rigor** and **qualitative depth** in evaluating Ethiopia's **Computer Based Testing System (CBT)** across diverse exam settings.
- A **descriptive design** was used to map HTS rollout, challenges, and impacts through **surveys** (227 ICT experts, 2,126 students, and exam administrators), **FGDs** (ICT staff and exam leaders), and **document analysis** (reports, policies, and international case studies), enabling both quantitative trends and qualitative insights while benchmarking Ethiopia's HTS against global standards.
- The study used **purposive sampling** to select zones and exam centers (universities and schools) with prior CBT deployment, ensuring relevance to hybrid testing.
- Within these, **stratified random sampling** captured diverse school types (public/private/boarding) and urban–rural settings, representing the 237,938 candidates across CBT and PPBT formats.

Methodology…

August 2025
EAES, Addis Ababa, Ethiopia

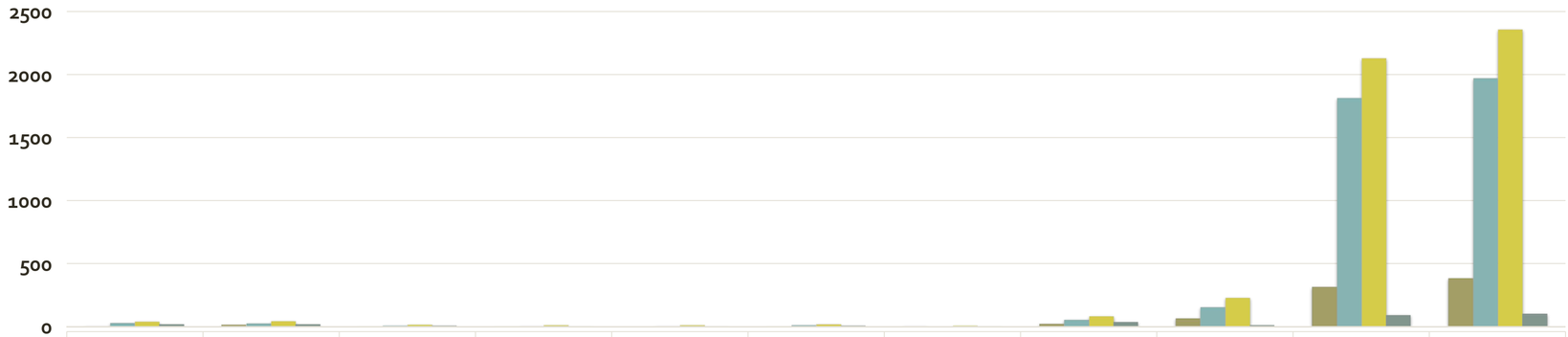
- The data were analyzed using SPSS v25 with **descriptive statistics, chi-square tests, and regression analysis** to summarize **patterns, examine group differences, and identify key performance** predictors.
- **Descriptive statistics**: Summarized performance patterns and perception ratings. To identify general trends and distributions
- **Chi-square tests**: Tested differences in perceptions across school types and locations. To verify whether digital exam experience varied based on context
- **Regression analysis**: Predicted student performance using variables like school type, location, and exam mode. To reveal **what factors most influenced performance** and explain score variability (CBT scores showed higher variability).
- Ethical standards were upheld by obtaining informed consent, ensuring anonymity and confidentiality, and securely handling data. A rapport note accompanied the instruments to build trust and encourage honest, unbiased responses.

41st AEAA Annual Conference

Participant of the Study

August 2025
EAES, Addis Ababa, Ethiopia

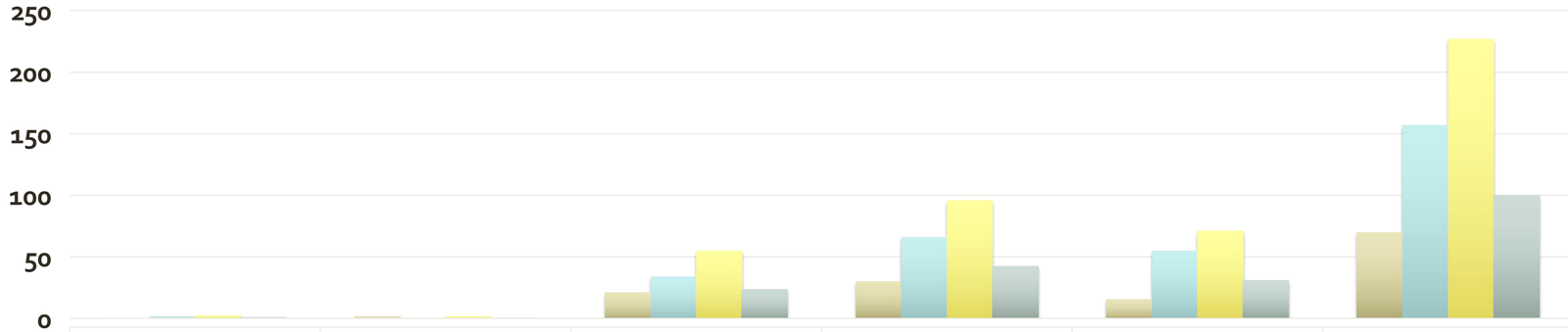
Number Participant in the study



	Proctors	Supervisors	Chief	Cluster Coordinators	Cluster Head	Zone Exam Admin	Region Exam Admin	ICT Expert	Administrator Total	Candidates Total	Grand Total
School	9	17	4	3	2	3	6	26	70	313	383
University	32	28	10	7	8	14	2	56	157	1813	1970
Total	41	45	14	10	10	17	8	82	227	2126	2353
%	18.06	19.82	6.17	4.41	4.41	7.49	3.52	36.12	9.65	90.35	100.00

School University Total %

CBT Exam Administrators Familiarity with Digital Tools



	Very Low	Low	Moderate	High	Very High	Total
School	1	2	21	30	16	70
University	2	0	34	66	55	157
Total	3	2	55	96	71	227
%	1.32	0.88	24.23	42.29	31.28	100.00

School University Total %

41st AEAA Annual Conference

Key CBT Practices Observed

August 2025
EAES, Addis Ababa, Ethiopia

Thematic Area	Key Findings	Discussion & Implications
Overall Satisfaction	87.7% (Universities) vs. 77.5% (Schools) rated “Good/Very Good”; no significant difference (Exam Administrators’ perception)	Aligning with Bello and Abdullah (2021), quality factors significantly impact CBT user satisfaction
Standards & Support	Positive ratings in exam standards, instructions, and examiner support (Exam Administrators’ perception)	Consistent with Fehintola (2018) – system perception not center-dependent.
Candidate Preparedness	Schools gave more “Fair” ratings; variation noted but not statistically significant (Exam Administrators’ perception)	Candidate preparedness is influenced by digital infrastructure disparities (Dejene & Tilahun, 2024).
Communication & Clarity	High satisfaction in both center types; no major differences observed (Exam Administrators’ perception)	Aligning with Bello and Abdullah (2021), who emphasized the role of clear communication in user satisfaction
Qualitative Insights	Rural school candidates reported infrastructure challenges: poor internet, limited equipment, weak support.	Infrastructure gaps influence delivery success (Fehintola, 2018).
Statistical Analysis	Crosstab shows no significant difference in CBT perception by center type	Emphasizes parity in core delivery, despite setting.

41st AEAA Annual Conference

Key CBT Practices Observed…

August 2025
EAES, Addis Ababa, Ethiopia

Thematic Area	Key Findings	Discussion & Implications
Coordination & Supervision	Candidate Perception: 96% (Universities) vs. 75% (Schools) rated Good/Very Good	Effective CBT implementation relies on strong coordination and supervision, as emphasized by Alaeddinoglu (2022) in ensuring quality assessment delivery.
Candidate-Proctor Ratio	Positive Perception: 97% (Universities) vs. 90% (Schools); schools gained from smaller group sizes	May be linked to center size and scale (Chukwuma-Nosike & Chukwuma, 2023).
Anti-Cheating Measures	High Ratings: 94% (Universities) vs. 96% (Schools); university responses were more consistent. Some cases of side-cheating (e.g., scrolling) reported.	AI-based proctoring, and behavioral analytics to detect suspicious activity, all supported by robust cybersecurity protocols (Bello & Abdullah, 2021; Wang et al., 2019; Meeks et al., 2020).
CBT Relevance	Positive Ratings: 80% (Universities) vs. 89% (Schools); universities favored on qualitative grounds	Digital readiness affects perception (Khan et al., 2019).
Digital Access & Internet	No major statistical difference; qualitative reports highlight rural school struggles	Reflects ongoing digital divide (Ogechukwu, 2019; Chukwuma-Nosike & Chukwuma, 2023).

⚠ Challenges Encountered

August 2025
EAES, Addis Ababa, Ethiopia

Thematic Area	Key Findings	Discussion & Implications
Technical Performance	Rated Good/Very Good by ~60% in both schools & universities.	Aligns with digital exam reliability findings (Fluck et al., 2009; Omanio et al., 2024).
Infrastructure Adequacy	Only 28% (schools) and 44% (universities) rated positively.	Highlights digital divide (Chen et al., 2023; Zhao, 2024; Adjei et al., 2013).
Fairness & Equity	Low ratings: 38% (schools), 44% (universities).	Reflects regional disparities and policy gaps (SAYO, 2024).
Academic Dishonesty	Positive control: 62% (schools), 67% (universities).	Stronger systems in universities (Afridon et al., 2024); school vulnerability noted (Moyo et al., 2024).
Technical Support	Effective support noted: 78% (universities), 65% (schools).	Tied to trained staff and faster resolution (Marina, 2015).
Candidate Awareness	Higher in schools (77%) than universities (67%).	May reflect targeted orientation or varied implementation strategies.
Time Management	Manageable time reported by 63% in both settings.	Supports digital flow efficiency (Omanio et al., 2024).

Future Prospects

August 2025
EAES, Addis Ababa, Ethiopia

Thematic Area	Key Findings	Discussion & Implications
Optimism for Future CBT Use	89% (universities) and 82% (schools) expect CBT improvement in 5 years.	Bennett (2009); OECD (2015) — Digital assessments improve efficiency and transparency.
Feasibility of Full Transition	84% (universities) vs 75% (schools) believe CBT could replace paper-based exams.	Fehintola (2018) — Institutional readiness, esp. digital maturity, boosts CBT success.
Equity for Special Groups	Only 54% (universities) and 50% (schools) feel CBT adequately serves special-needs and rural candidates.	Nguyen et al. (2023); Graham et al. (2017) — Equity must be embedded in digital exam design from inception.
Impact of Tech Upgrades	77% (universities) and 76% (schools) optimistic about tech's positive role.	Mohamed Nafrees et al. (2023) — Tech (e.g., AI, cloud) boosts testing if backed by governance.
Stakeholder Expectations Met	79% (universities) and 80% (schools) feel stakeholders' expectations were met.	Dede et al. (2018) — Stakeholder alignment builds accountability and trust.
Collaboration Support	Higher in universities (75%) than schools (72%), though difference is not significant.	Emphasizes slightly stronger institutional cohesion at university level.

41st AEAA Annual Conference

CBT system's usability, compatibility, and data management

August 2025
EAES, Addis Ababa, Ethiopia

Aspect	Findings	Discussion & Implications
System Usability	Majority find CBT user-friendly: Schools 83.3%, Universities 86.8%.	Supports Fagbola et al. (2013) on user-friendliness critical for adoption.
	Reliability rated positively: Schools 80%, Universities 86.8%.	No significant difference by center type (Abass et al., 2017), but schools report navigation/support challenges due to digital literacy gaps (Romawanti et al., 2024).
	Interface intuitiveness: Schools 73.3%, Universities 84.2%.	Rural school users struggle with login, data display, and lack real-time support (Osei & Boateng, 2023; Eko et al., 2022).
System Compatibility	Positive compatibility ratings: Schools 73.3%, Universities 75%.	Consistent with Mutula (2021) on infrastructure gaps in Sub-Saharan Africa; Baghdady & Zaki (2019) on governance challenges.
	Both sectors optimistic about scalability and future upgrades	Moodle-based exams face security risks, poor coordination, and weaknesses of biometric authentication systems (Ally, 2022; Sabbah, 2017; Kumar et al., 2024).
Data Management & Security	High confidence in data security: Over 80% rate personal data protection and accuracy as Good in both settings.	Recurring backend failures demand stronger infrastructure, monitoring, and security—encryption, backups, and firewall/DNS safeguards (Moustafa et al., 2021; Sharma et al., 2024; Luecht, 2016).

41st AEAA Annual Conference

Performance Patterns: CBT vs. PPBT

August 2025
EAES, Addis Ababa, Ethiopia

Key Insight	Findings	Discussion & Implications
Higher CBT Scores	CBT candidates had higher average scores (Mean = 39.72) than PPBT (Mean = 31.10), but also showed higher variability (CV 44.7% vs. 34.4%).	Mhlongo (2023) linked better CBT scores to positive digital use, while Alwazzan & Alyousef (2019) warned of higher EFL anxiety highlighting the need to boost experience and reduce anxiety.
School Type Advantage	Scores vary widely by school type boarding (95.2%), community (72.7%), private (46.8%), public (3.9%) and CBT candidates (26.3%) outperform PPBT (7.4%).	Gaps reflect resource disparities (Abdulkareem, 2023) and CBT engagement benefits (Nguyen et al., 2017), highlighting need for equity and infrastructure.
Test Equivalence Concerns	Greater score inconsistencies between CBT and PPBT raise validity issues, possibly due to unfamiliarity with digital formats.	Ebrahimi et al. (2019) showed CBT-PBT score differences tied to computer familiarity and attitudes, highlighting the need to consider these for fair CBT testing.
Need for Test Validation	Validity varied by subject, with STEM weaker and humanities stronger, underscoring the need for advanced psychometric methods.	CBT validity gaps in STEM vs. humanities call for advanced methods—CTT, IRT, DIF, equating—to ensure fairness (Teka et al., 2024; Mutluer & Çakan, 2023; Alordiah, 2015).

📌Conclusion and Recommendations

August 2025
EAES, Addis Ababa, Ethiopia

- In conclusion, Ethiopia's move to **Computer-Based Testing** in the Secondary School Leaving Certificate Examination is a bold and transformative step.
- The 2024 CBT-based ESSLCE improved security and efficiency but faced school-level challenges in infrastructure and digital readiness, underscoring the need for a national strategy.
- To strengthen Ethiopia's CBT system for the ESSLE exam (2016 E.C./2024 G.C.), the study proposes:
  - ✓**Expand ICT Infrastructure** – Improve internet, power, and computer access, especially in schools and rural centers.
  - ✓**Build Digital Capacity** – Provide hands-on CBT training for students, teachers, and administrators.
  - ✓**Standardize CBT Protocols** – Develop national guidelines on authentication, security, and administration.
  - ✓**Customize and Automate the System** – Align the platform with Ethiopia's curriculum and workflows, integrating automation for scheduling, scoring, and reporting. Ensure inclusivity through accessible design, and enhance security using biometric data and remote proctoring.
  - ✓**Promote Collaboration** – Engage EAES, MoE, regions, telecoms, and donors in coordinated CBT development.
  - ✓**Monitor and Evaluate** – Establish a feedback system to track CBT performance and guide continuous improvement.

Implications of the Research

- **CBT's Benefits & Gaps:** Computer-Based Testing streamlines grading, reduces malpractice, enables secure item banking, and supports test equivalence for fair exam versions. Yet, its potential is limited by ICT, connectivity, and technical disparities—especially in rural areas—hindering equitable access.
- **Key Requirements:** Sustainable adoption requires targeted **ICT investments, nationwide training, standardized procedures, robust item bank development, psychometric safeguards for test equivalence, and ongoing system development and customization** to fit Ethiopia's context. Reforms must address technological, administrative, and pedagogical readiness.
- **Strategic Roadmap:** Findings guide policymakers and exam boards toward scalable, inclusive CBT systems that integrate global best practices with local needs, modernize assessments, and reduce educational inequality.

⚠ Limitation of the Research

August 2025
EAES, Addis Ababa, Ethiopia

This study faced several limitations that may influence the broader applicability of its findings :

- Rural exam centers lacked stable internet and electricity, limiting nationwide CBT scalability.
- Students, teachers, and administrators especially in schools struggled with CBT tools due to limited exposure.
- The study did not fully assess accessibility for marginalized groups (e.g., students with disabilities, low-income backgrounds).
- No real-time system monitoring or **backend analysis was included; findings relied mostly on self-reported data.**
- CBT and PPBT formats **were not compared for item equivalence**, missing key psychometric insights.

Thank you for your attention and engagement !

Questions

Comment

Suggestion