

Enhanced Student Engagement through Supervised Assessment: Evidence from Kenya

Pedro Freitas
Blavatnik School of
Government, University of
Oxford
pedro.freitas@bsg.ox.ac.uk

Guillermo Romero
Blavatnik School of
Government, University of
Oxford
guillermo.romero@bsg.ox.ac.uk

Amar Lalwani
EIDU
amar.lalwani@eidu.com

Aidan Friedberg
EIDU
aidan.friedberg@eidu.com

Noam Angrist
Blavatnik School of
Government, University of
Oxford
noam.angrist@bsg.ox.ac.uk

Sharnic Djaker
Blavatnik School of
Government, University of
Oxford
sharnic.djaker@bsg.ox.ac.uk

ABSTRACT

In-app assessments are increasingly used by EdTech programs to measure students' proficiency and guide instructional decisions. Student engagement in these assessments is crucial for generating reliable learning signals; however, efforts to increase student engagement in platform-based assessments often impose additional demands on teachers. To explore an optimal assessment design, we implemented supervised assessments in which a teacher is present while students complete the assessment. We find that supervised assessments significantly increase student engagement, as evidenced by longer average assessment durations and a substantial reduction in timed-out or aborted sessions. At the same time, the number of classes participating in the assessment decreases slightly. These findings highlight a trade-off between assessment quality and participation rates, and suggest that supervision improves data quality primarily through increased student engagement.

Keywords

In-app assessment, Supervised assessment, Student proficiency, Low- and middle-income countries

1. INTRODUCTION

Assessment results constitute a key source of information for understanding the evolution of students' learning. Students and their families respond to feedback and information generated by reliable assessments [3, 4, 5].

Assessment data are also critical for teachers, as they provide information on students' learning trajectories and can inform instructional practices. A growing body of literature examines how teachers use assessment data to guide ped-

Amar Lalwani, Aidan Frieberg, Pedro Freitas, Guillermo Romero, Sharnic Djaker, and Noam Angrist. Enhanced Student Engagement through Supervised Assessment: Evidence from Kenya. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyejeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 653–656. International Educational Data Mining Society (2026).

© 2026 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

<https://doi.org/10.5281/zenodo.21040121>

agogical decisions and improve student learning outcomes [10, 9, 12, 13]. This evidence suggests that increased access to assessment information can improve instructional targeting, goal setting, and teachers' self-efficacy [6]. However, there remains limited empirical evidence on how teachers use data generated by information and communication technology (ICT) systems and how such data shape instructional practices in large-scale educational settings [7].

The usefulness of assessment data depends fundamentally on student engagement during assessment activities [8, 14]. In-platform based environments, disengagement — such as rapid guessing, non-response, or aborted sessions — can generate noisy signals about student proficiency. These issues are especially salient in unsupervised digital assessments that are completed without active teacher oversight. Proctoring offers a potential mechanism to address these challenges by introducing teacher supervision during assessments, thereby increasing student engagement and completion. However, teacher supervision also increases the effort required from teachers, which may discourage participation in more demanding assessment formats. Existing evidence on the benefits of proctoring primarily comes from higher education and focuses largely on cheating behavior in online assessments [15, 2], leaving open questions about its role in shaping engagement and data quality in large-scale primary education systems, particularly in low- and middle-income country contexts.

In this paper, we study the effects of introducing supervised assessments on a national-scale EdTech platform operating in Kenya. In September 2025, the platform tested supervised assessments in place of unsupervised assessments, with teachers proctoring students during assessment administration. This proctoring is mostly observational and similar across students in different classes and schools. Using platform data covering more than 300,000 students across approximately 10,000 schools, we examine how teacher supervision influences student engagement and participation in digital assessments.

2. INTERVENTION

EIDU is a digital education platform operating at scale in Kenya. The platform provides interactive learning content in numeracy and literacy through Android devices, with games-based content units adapted to students' individual learning pace [11, 1].

Prior to September 2025, the platform administered assessments in an unsupervised format, where assessment tasks were automatically launched on the device without active teacher oversight and learners could abort the assessments at any point of time. In September 2025, EIDU introduced and tested a supervised assessment format, in which a teacher is present and actively proctors students as they complete the assessment. During the assessment teachers observe learners without providing any feedback or scaffolding. Examples of digital assessment items of letter recognition and number identification can be found in Figures 1 and 2.

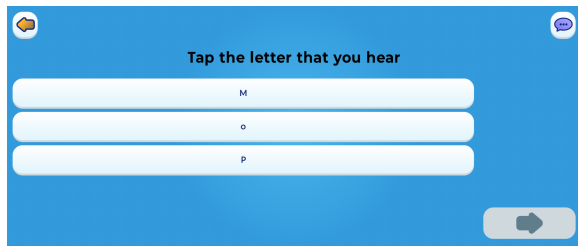


Figure 1: Letter Recognition - Digital assessment

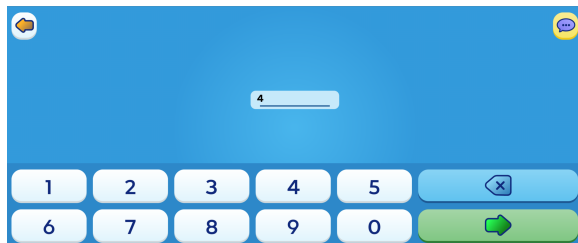


Figure 2: Number identification - Digital assessment

We examine how this introduction of teacher supervision affects student engagement and participation in digital assessments. We measure these outcomes using three school-level indicators: (i) the average time students spent on the assessment; (ii) the share of assessment sessions that were aborted or timed out; and (iii) the average number of classes per school participating in the assessment. These three indicators constitute the main outcome variables of our analysis. We additionally descriptively show how the change from unsupervised to supervised assessments changed the distribution of students' scores.

This large-scale evaluation was conducted in close coordination with county-level education authorities, who approved and monitored the implementation of the intervention. All data were anonymized, and access was restricted to researchers involved in the project.

3. DATA AND METHODS

The data covers the period from January to November 2025 and include 310,363 5-year old learners across 10,322 schools and 25,130 classes. Students are assessed in both Reading and Mathematics, completing 11 items in letter recognition and 12 number identification per assessment moment. All outcomes are aggregated at the school-level.

To examine the effects of supervised assessments, we use a pre-post research design that compares outcomes before and after the introduction of teacher supervision. The pre-intervention period is coded as $t = 0$ and the post-intervention period as $t = 1$. The post period corresponds to the introduction of supervised assessments in September 2025, when supervised assessments were introduced and tested in the platform.

For each school s and time period $t \in \{0, 1\}$, we estimate the following specification:

$$Y_{st} = \alpha + \beta S_t + \varepsilon_{st}, \quad (1)$$

where Y_{st} denotes the outcome of interest and S_t is a dummy variable equal to 1 in the post-intervention period (supervised assessments) and 0 in the pre-intervention period (unsupervised assessments). The coefficient β captures the average change in school-level outcomes following the introduction of teacher supervision. ε_{st} stands for the error term.

4. RESULTS

Tables 1–3 present the main results. In each table, Column (1) reports results for the full sample. Column (2) restricts the sample to one month before and after the introduction of supervised assessments to account for potential learning trajectory effects over time. Column (3) of Table 1 further restricts the sample to completed assessments. In addition, we report the average level for each outcome for unsupervised assessments.

Tables 1 and 2 indicate a higher level of engagement among the students who participated in the assessments. The introduction of supervised assessments is associated with increases in average assessment duration ranging from 0.09 to 0.26 minutes. This corresponds to an increase between 9% and 25% compared with unsupervised assessments. We also find a large reduction in the share of sessions that were aborted or timed out. The latter effect is particularly pronounced, with aborted or timed-out sessions declining by approximately 30 percentage points following the introduction of supervision. These patterns suggest that teacher proctoring substantially increases the likelihood that students remain engaged through assessment completion.

At the same time, supervised assessments are associated with a modest reduction in participation across classes. Table 3 shows that the average number of classes per school participating in assessments declined by approximately 0.3 classes following the introduction of supervision. This result suggests increased selectivity in class participation when assessments are administered under teacher supervision.

The distribution of assessment scores also changes following the introduction of supervision. We compute the share of correct items for each student assessment and then aggre-

Table 1: Effects on Mean duration (in minutes)

	Mean duration		
	(1)	(2)	(3)
Supervised	0.259*** (0.0371)	0.232*** (0.0372)	0.089*** (0.00925)
Av. Unsupervised	1.03		
1 Month before-after	NO	YES	NO
Only completed	NO	NO	YES
Observations	16,911	11,647	16,911
R-squared	0.004	0.003	0.007

Notes: The table presents school-level OLS regression results with mean duration as the dependent variable. The omitted category corresponds to non-supervised assessments. Standard errors clustered at the school level are reported in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2: Effects on the Share of sessions aborted

	%	
	(1)	(2)
Supervised	-0.346*** (0.00380)	-0.330*** (0.00430)
Av. Unsupervised	0.61	
1 Month before-after	NO	YES
Observations	16,911	11,647
R-squared	0.331	0.283

Notes: School-level OLS regression results with the share of aborted/timed-out sessions as the dependent variable. Standard errors clustered at the school level are in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3: Effects on Number of classes engaged

	Number of classes	
	(1)	(2)
Supervised	-0.293*** (0.00778)	-0.272*** (0.0113)
Av. Unsupervised	1.23	
1 Month before-after	NO	YES
Observations	16,911	11,647
R-squared	0.034	0.038

Notes: School-level OLS regression results with the number of classes engaged as the dependent variable. Standard errors clustered at the school level are in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

gate these measures at the school level. As shown in Figure 3, unsupervised assessments exhibit a pronounced mass of scores at zero, indicative of low student engagement and high levels of non-response. In contrast, supervised assessments display a markedly different score distribution, characterized by higher average scores and the absence of zero inflation, consistent with increased participation among participating students. As observed in Table 3, the introduction of supervised assessments led to a decline in the share of classes per school participating in the assessments. To control for this compositional effect, Figure 4 includes only classes that participated in both assessments. The figure shows a similar pattern in the score distributions for supervised and unsupervised assessments.

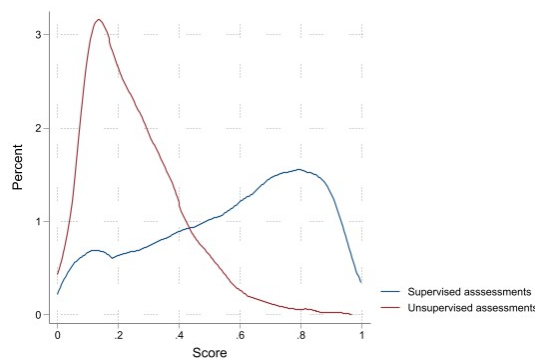


Figure 3: Score distribution, school level (scale: 0-1)

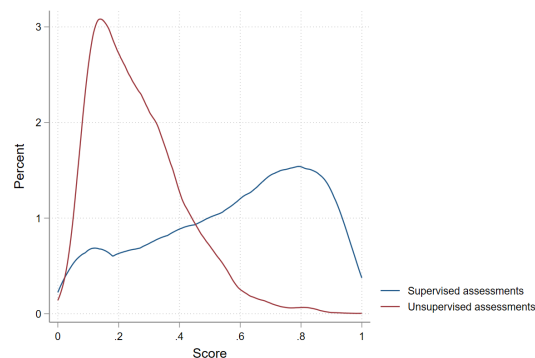


Figure 4: Score distribution for classes observed in supervised and unsupervised assessments, school level (scale: 0-1)

5. CONCLUSION

Accurate student assessment data is key to understanding students' progress and informing instructional decisions across educational contexts. The value of such data depends not only on assessment content, but also on how assessments are administered and the extent to which students are engaged during assessment activities.

As the EdTech sector continues to expand, digital assessments are increasingly embedded alongside instructional materials and used to guide personalization at scale. In this

setting, ensuring meaningful student engagement during assessment becomes critical for generating reliable learning signals. To address this challenge, EIDU introduced supervised assessments, in which teachers proctor students during assessment administration. While supervision has the potential to improve data quality, it also increases the burden on teachers.

Our results show that students engage more with supervised assessments, as measured by the increase in average time spent answering items and the strong decrease in the average share of sessions that were aborted or timed out. However, we also observe a slight decrease in the average number of classes participating in assessment, consistent with increased burden for some teachers. These findings highlight a trade-off between assessment engagement and participation when supervision is introduced at scale.

Future work should explore in-product solutions that preserve the engagement benefits of supervision while reducing participation-related burden, further enabling the collection of higher-quality digital assessment data.

References

- [1] P. Agartsupa, T. Chindanurak, and D. Suwanjinda. The effects of blended learning management in earth science on learning achievement and digital skills of grade 4 students at ekamai international school. *Human Behaviour, Development and Society*, 26, 2025.
- [2] H. M. Alessio, N. Malay, K. Maurer, A. J. Bailer, and B. Rubin. Examining the effect of proctoring on online test scores. *Online Learning Journal*, 21, 2017.
- [3] G. Azmat and N. Iriberry. The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics and Management Strategy*, 25:77–110, 2016.
- [4] M. Bobba and V. Frisancho. Learning about oneself: the effects of performance feedback on school choice. *SSRN Electronic Journal*, 2017.
- [5] R. Brade, O. Himmler, and R. Jäckle. No student left behind? relative feedback and university completion. *Journal of Economic Behavior and Organization*, 241, 2026.
- [6] R. V. Gasse, K. Vanlommel, J. Vanhoof, and P. V. Pe-tegem. The impact of collaboration on teachers' individual data use. *School Effectiveness and School Improvement*, 28, 2017.
- [7] J. Luo, M. Wang, and S. Yu. Exploring the factors influencing teachers' instructional data use with electronic data systems. *Computers and Education*, 191, 2022.
- [8] C. Penk, C. Pöhlmann, and A. Roppelt. The role of test-taking motivation for students' performance in low-stakes assessments: an investigation of school-track-specific differences. *Large-Scale Assessments in Education*, 2, 2014.
- [9] K. Schildkamp and W. Kuiper. Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 2010.
- [10] K. Schildkamp, C. Poortman, H. Luyten, and J. Ebbeler. Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, 28, 2017.
- [11] C. Sun, L. Major, R. Daltry, and N. Moustafa. Teacher-ai collaboration in content recommendation for digital personalised learning among pre-primary learners in kenya. In *L@S 2024 - Proceedings of the 11th ACM Conference on Learning @ Scale*, 2024.
- [12] P. Wohlstetter, A. Datnow, and V. Park. Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, 19, 2008.
- [13] V. M. Young. Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, 112, 2006.
- [14] G. Zamarro, C. Hitt, and I. Mendez. When students don't care: Reexamining international differences in achievement and student effort. *Journal of Human Capital*, 13, 2019.
- [15] L. Zhao, J. Peng, S. Ke, and K. Lee. Effectiveness of un-proctored vs. teacher-proctored exams in reducing students' cheating: A double-blind randomized controlled field experimental study. *Educational Psychology Review*, 36, 2024.