

**DIGITAL LITERACY IN EDUCATION: UTILISATION OF
ARTIFICIAL INTELLIGENCE TOOLS IN TEST, MEASUREMENT,
MONITORING AND EVALUATION AMONG PUBLIC SENIOR
SECONDARY SCHOOL TEACHERS IN RIVERS STATE, NIGERIA**

Nwankwo, Emmanuel Chijioke Obinna

Dept. of Educational Psychology Guidance and Counseling,
Ignatius Ajuru University of Education, Port Harcourt
08128106945 nwankwoobinna13@gmail.com

Abstract

This study “examined the extent to which teachers in public senior secondary schools in Rivers State apply artificial intelligence tools to test development, test administration and scoring, and the monitoring and evaluation of learning. A descriptive survey design was adopted. The population was 5,833 teachers from 320 public Senior Secondary Schools in Rivers State and the sample was 375 drawn through random sampling technique. Three research questions and two hypotheses guided the study. A structured questionnaire titled Artificial Intelligence in Assessment Practices Questionnaire (AIAPQ) was validated by two professionals in measurement and evaluation. The instrument was pilot tested and Cronbach Alpha was used to establish the reliability. The instrument was administered and the 356 completed copies were collected and analyzed using mean, grand mean and standard deviation. Independent samples t-test were used to test the hypotheses”. Utilization was low across all three assessment domains, while infrastructural and training constraints were rated high. Gender produced no significant difference, but school location did. Structured retraining and a state policy on artificial intelligence in assessment are recommended.

Keywords: Digital literacy, artificial intelligence, educational measurement, continuous assessment, Rivers State.

Introduction

Digital “literacy has moved from a peripheral skill to a working requirement of the teaching profession, and nowhere is this shift sharper than in assessment. Decisions about promotion, certification and the allocation of remedial attention all rest on evidence generated through testing, and those decisions can be no better than the evidence behind them. Black and Wiliam (2018) argue that classroom assessment earns its value not from the act of grading but from the instructional decisions it makes possible. The National Policy on Education (Federal Republic of Nigeria, 2013) reflects this position by placing continuous assessment at the centre of practice at every level of the system.

In practice the assessment cycle is demanding. A teacher determines the purpose of a test, prepares a table of specification, writes and moderates items, administers the test, marks the responses, analyses item performance, interprets the results and reports them. Each stage takes time. In a Rivers State senior secondary school where one teacher may handle four or five arms of fifty students, the later stages are usually the first to be abandoned. Item analysis is seldom carried out, feedback reaches the student long after it can shape learning, and the record of progress amounts to little more than a column of scores in a mark book.

Artificial intelligence has begun to alter this picture. Owan et al. (2023) mapped the applications of large language models onto the full testing cycle, from purpose determination and blueprint preparation through item generation, assembly, administration, scoring, interpretation and reporting. Automated scoring systems now handle objective and constructed responses at scale, adaptive platforms adjust item difficulty in response to examinee performance, and analytics tools summarise class-level patterns that would take a teacher several hours to compile by hand (Bulut et al., 2024; Zawacki-Richter et al., 2019). The monitoring function is where the change is sharpest. Progress can be tracked across successive assessments, students at risk of failure can be flagged early in a term, and the evaluation of teaching effectiveness can draw on evidence rather than impression (Mpolomoka, 2025).

These gains carry costs. Automated scoring can reproduce the biases present in the data used to build it, and the sources of that bias are not

always visible to the teacher who acts on the mark (Johnson & McCaffrey, 2023; Mehrabi et al., 2021). Systematic reviews of fairness, accountability and transparency in educational artificial intelligence report that questions of due process and redress remain largely unsettled (Memarian & Doleck, 2023), while work on explainability shows that many systems in classroom use cannot account for the judgements they produce (Khosravi et al., 2022). Student assessment data is also sensitive, and the platforms that process it are typically hosted outside the schools whose students they describe. UNESCO accordingly framed its competency expectations for teachers around human agency and accountability rather than technical fluency alone (Miao & Cukurova, 2024; Miao & Holmes, 2023), and Perkins et al. (2024) proposed a graded scale for declaring how much artificial intelligence is permitted in a given assessment task.

Nigeria has moved quickly on the delivery side of assessment. The Joint Admissions and Matriculation Board administers computer-based testing to more than two million candidates each year, and in 2025 the Federal Ministry of Education directed the West African Examinations Council and the National Examinations Council to migrate their examinations to the same mode, beginning with objective papers. The National Artificial Intelligence Strategy places skills development and sectoral adoption among its stated priorities (Federal Ministry of Communications, Innovation and Digital Economy, 2024), and Bali et al. (2024) trace a steady rise in artificial intelligence applications across Nigerian education. Computerising delivery, however, is not the same as bringing artificial intelligence into measurement. A test can be delivered on a screen and still be built, marked and interpreted by exactly the methods used forty years ago. Senior secondary teachers in Rivers State are therefore approaching a computerized examination environment without any assurance that their own assessment practice has changed.

Rivers State presents a useful setting for examining this question. The state comprises twenty-three local government areas grouped into three senatorial districts, with 320 public senior secondary schools staffed by 5,833 teachers (Aderiye & Wagbara, 2024). Its schools divide sharply between the Port Harcourt and Obio-Akpor urban corridor, where power and connectivity are comparatively stable, and the riverine communities of the Niger Delta, where some schools are reached only by boat and where

mains electricity and mobile data are unreliable and expensive. Any account of digital literacy in assessment within the state has to reckon with that division.

The available Nigerian evidence is thin and points in a consistent direction. Thomas et al. (2024) found moderate awareness of artificial intelligence among lecturers in a federal university but noted that awareness had not converted into practice. Eze and Onah (2024) reported that lecturers in Enugu State relied mainly on conversational tools such as ChatGPT, with little use of anything more specialised. Ibrahim et al. (2024) found perception running well ahead of utilisation in educational assessment. Ofem and Chukwujama (2024) drew comparative lessons from Estonia, China, the United States and Australia and concluded that infrastructure and policy, not enthusiasm, were the binding constraints for Nigeria, a conclusion echoed for the Estonian case by Granström and Oppi (2025). Ofem et al. (2025), working with a large national sample, showed that preparedness to use artificial intelligence in classroom assessment depended more on technological readiness and pedagogical belief than on attitude alone.

Within Rivers State itself, Aderiye and Wagbara (2024) established a strong positive relationship between artificial intelligence usage in assessment and grading and teacher performance in public senior secondary schools. That study treated usage as a single correlate, so it could not say which parts of the assessment cycle teachers actually reach for such tools and which they leave untouched, and it did not examine whether the urban and riverine parts of the state differ. The present study addresses both gaps by disaggregating the assessment cycle into three functional domains and by testing school location as a source of difference.

Research Questions

1. What is the extent of AI tools utilisation by teachers in test development and item construction in public senior secondary schools in Rivers State?
2. To what extent is AI tools utilisation by teachers in test administration and scoring of male and female students' in public senior secondary schools in Rivers State?

3. To what extent is AI tools utilisation by teachers in Urban and Rural public senior secondary schools' in Rivers State?

Hypotheses

H₀₁: There is no significant difference in AI tools utilisation by teachers in test administration and scoring male and female students' in public senior secondary schools in Rivers State.

H₀₂: There is no significant difference in AI tools utilisation by teachers in Urban and Rural public senior secondary schools in Rivers State.

Artificial Intelligence (AI)

Artificial "Intelligence (AI) is a technological tool which performs human tasks in solving complex problems with human intelligence through Computer Controlled Machine. Artificial Intelligence (AI) is the ability of a digital computer or Computer Controlled Machine to perform task that typically require human intelligence such as reasoning, learning from experience, solving complex problems (Copeland, 2026)

Method

The study adopted a descriptive survey design, which suits an enquiry seeking to describe existing practice in a defined population without manipulating any variable. The area of study was Rivers State, a Niger Delta state of twenty-three local government areas.

The population consisted of 5,833 teachers in the 320 public senior secondary schools in Rivers State and the sample was 375 drawn through simple random sampling technique.

The instrument titled Artificial Intelligence in Assessment Practices Questionnaire (AIAPQ) was validated by two professional who majored in Educational Measurement and Evaluation and the instrument was rated on a four-point scale of Very High Extent (VHE), High Extent (HE), Low Extent (LE) and Very Low Extent (VLE). The instrument was pilot tested with 30 teachers in public senior secondary schools in Bayelsa State who are not within the study area. Cronbach Alpha was used to establish the reliability.

375 instruments were administered by the researcher with the help of three trained research assistants and 356 were retrieved on the spot and the data collected was analyzed using mean, grand mean and standard deviation to answer the research questions. The independent samples T-test was used to test the hypotheses at 0.05 level of significant.

Results

Table 1: Demographic characteristics of respondents (n = 356)

Variable	Category	Frequency	Percentage
Gender	Male	159	44.7
	Female	197	55.3
Academic qualification	NCE	62	17.4
	B.Ed / B.Sc (Ed)	232	65.2
	M.Ed and above	62	17.4
School location	Urban	201	56.5
	Rural / riverine	155	43.5
Teaching experience	10 years and below	190	53.4
	Above 10 years	166	46.6

Table 2: Mean ratings on the use of artificial intelligence tools in test development and item construction

S/N	Item	Mean	SD	Decision
1	Generating draft objective test items from a scheme of work or topic outline	2.51	0.88	High Extent
2	Preparing a table of specification for term examinations	2.21	0.84	Low Extent
3	Rephrasing ambiguous or poorly worded test items	2.75	0.81	High Extent
4	Matching test items to stated behavioural objectives	2.46	0.83	Low Extent
5	Generating plausible options for multiple-choice items	2.42	0.85	Low Extent
6	Drafting marking schemes and rubrics for essay questions	2.27	0.88	Low Extent
Grand mean		2.44	0.55	Low Extent

Table 2 shows that only two of the six practices in this cluster attracted mean ratings above the 2.50 criterion. Teachers were most willing to use artificial intelligence for the editorial task of rewording unclear items and least willing to use it for the structural task of preparing a table of specification. The cluster mean of 2.44 falls within the low extent” band.

Table 3: Mean ratings on the utilisation of AI tools in test administration and scoring of male and female students.

S/N	Item	Mean	SD	Decision
7	Administering class tests or continuous assessment on AI digital platforms	2.25	0.80	Low Extent
8	Automated marking of objective test responses using AI	2.56	0.85	High Extent
9	Assisted marking of essay and theory responses with AI	2.12	0.84	Low Extent
10	Producing individualised written comments on students' scripts using AI	2.23	0.83	Low Extent
11	Detecting unusual answer patterns suggesting examination malpractice	2.09	0.80	Low Extent
12	Checking students' assignments for originality and undeclared AI use	2.47	0.83	Low Extent
Grand mean		2.29	0.55	Low Extent

Automated marking of objective responses was the single practice in Table 3 to reach the criterion. Assisted marking of essay and theory responses recorded 2.12 and the detection of irregular answer patterns 2.09, the two lowest in the cluster. The cluster mean of 2.29 indicates use to a low extent.

Table 4: Mean ratings on utilisation of AI tools in Urban and Rural secondary schools.

S/N	Item	Mean	SD	Decision
19	Irregular electricity supply in the school	3.38	0.68	High Extent
20	Weak or costly internet connectivity, especially in riverine communities	3.31	0.70	High Extent
21	Lack of training on AI applications in test and measurement	3.21	0.73	High Extent
22	Cost of data and subscription to AI platforms	3.19	0.72	High Extent
23	Absence of a state or school policy on AI use in assessment	3.04	0.76	High Extent
24	Concern about the privacy of students' assessment records	2.87	0.81	High Extent
25	Doubts about the accuracy and fairness of AI-generated marks	2.86	0.75	High Extent
Grand mean		3.12	0.40	High Extent

Every item in Table 5 exceeded the criterion. Electricity supply and internet connectivity attracted the highest ratings, followed by lack of training and the cost of data and subscriptions. Concerns about privacy and about the fairness of machine-generated marks recorded the lowest means in the cluster.

Table 5: Independent samples t-test of mean utilisation ratings by gender and school location

Variable	Group	N	Mean	SD	df	t	p	Decision
Gender	Male	159	2.27	0.48	354	-1.30	.194	Not significant
	Female	197	2.34	0.52				
School location	Urban	201	2.41	0.50	354	4.43	<.001	Significant
	Rural / riverine	155	2.17	0.48				

Note. Significance tested at the .05 level.

Table 6 shows a t-value of -1.30 for gender at 354 degrees of freedom with a probability value of .194, which is greater than .05. The first null hypothesis was therefore retained. For school location the t-value was 4.43 with a probability value below .001, so the second null hypothesis was rejected. Teachers in urban schools recorded a higher mean utilisation rating than their colleagues in rural and riverine” schools.

Discussion

The “picture that emerges is one of narrow, shallow adoption. Teachers reach for artificial intelligence where the task is linguistic and the output can be checked at a glance, such as rewording an ambiguous item or marking an objective test, and avoid it where the task is technical and the output would have to be trusted. This is consistent with Thomas et al. (2024), who found awareness among Nigerian educators running ahead of practice, and with Eze and Onah (2024), whose respondents relied on general conversational tools rather than anything built for assessment. Owan et al. (2023) set out the full range of what these tools can do across the testing cycle; the present findings suggest that Rivers State teachers are working in a small corner of that range.

The significant difference by school location is the most consequential finding of the study. Urban teachers returned a mean of 2.41 against 2.17

for their rural and riverine colleagues, and although both figures fall within the low extent band, the gap is real and it runs in the direction the constraint ratings would predict. Power and connectivity were rated the two heaviest constraints overall, and both bear far more heavily on schools in the creeks than on schools along the Port Harcourt corridor. The concern is that the migration of school certificate examinations to computer-based testing will meet a teaching workforce whose familiarity with digital assessment is already unevenly distributed within the same state. A reform designed to raise the integrity of examinations may, without deliberate compensation, widen an existing divide.

Concerns about fairness and privacy were rated lower than infrastructural concerns, though still above the criterion. This ordering is worth watching. The measurement literature treats algorithmic bias and the opacity of automated judgement as first-order threats to validity (Johnson & McCaffrey, 2023; Khosravi et al., 2022; Memarian & Doleck, 2023), yet respondents here ranked them below the cost of data. The likeliest explanation is that a constraint becomes visible only through use, and use is still too limited for these problems to have been met. If adoption accelerates without a corresponding rise in critical awareness, teachers may come to accept machine-generated marks with less scrutiny than they would apply to a colleague's marking.

The absence of a gender difference agrees with Eze and Onah (2024) and Thomas et al. (2024), both of which reported no significant difference between male and female respondents. Taken together with the location finding, this suggests that what separates users from non-users in Rivers State is not who the teacher is but where the teacher works. That reading is consistent with Ofem et al. (2025), who found technological readiness a stronger predictor of preparedness than attitude, and with Ertmer and Ottenbreit-Leftwich's (2010) argument that the conditions surrounding practice matter as much as the beliefs of the practitioner. It also aligns with Aderiye and Wagbara (2024), whose Rivers State respondents showed that where such tools are used they relate positively to teacher performance. The problem is not that the tools fail to help. It is that most teachers in the state are not yet in a position to use them.

Conclusion

Teachers in “public senior secondary schools in Rivers State use artificial intelligence tools in assessment to a low extent across test development, test administration and scoring, and the monitoring and evaluation of learning, with the monitoring function least developed of the three. Constraints of electricity, connectivity, training, cost and policy were all rated high. Gender made no significant difference to utilisation, but school location did, with urban teachers reporting greater use than those in rural and riverine schools. As the national examination bodies move towards computer-based testing, the gap between what these tools can do for educational measurement and what is being done with them in Rivers State classrooms remains wide, and it is a gap of preparation, infrastructure and policy rather than of willingness.

Recommendations

1. The Rivers State Senior Secondary Schools Board should mount structured retraining for teachers that goes beyond general familiarity with artificial intelligence to the specific measurement tasks reported here as least practised, particularly item analysis, progress tracking and the interpretation of performance data. Such training should be delivered by measurement and evaluation specialists rather than by information technology units alone.
2. Retraining and infrastructural support should be weighted deliberately towards schools in the rural and riverine local government areas, since a uniform intervention applied across the state would preserve rather than close the gap this study has identified.
3. The State Ministry of Education should issue a written policy stating which assessment functions may be delegated to artificial intelligence, what moderation is required for automatically generated items, how the use of such tools is to be declared, and what recourse a student has where a mark has been produced with algorithmic assistance.
4. Schools should be supported to digitise their continuous assessment records, since analytics and monitoring tools cannot operate on data that exists only in paper mark books. A modest, well-maintained school assessment database would yield more than the purchase of any single platform.

5. Faculties and institutes of education in the state should incorporate artificial intelligence literacy for assessment into their teacher preparation programmes, so that newly posted teachers enter the school system able to use and to question these tools.
6. Further research should extend this enquiry to junior secondary and primary levels, and should complement self-report with observation of actual assessment practice, since respondents may report intention more accurately than behaviour.

References

- Aderiye, P. C., & Wagbara, C. D. (2024). Artificial intelligence usage and teacher's performance in public senior secondary schools in Rivers State, Nigeria. *Journal of Association of Educational Management and Policy Practitioners*, 5(1), 252–260.
- Bali, B., Garba, E. J., Ahmadu, A. S., Takwate, K. T., & Malgwi, Y. M. (2024). Analysis of emerging trends in artificial intelligence for education in Nigeria. *Discover Artificial Intelligence*, 4(1), 1–17.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551 <https://doi.org/10.1080/0969594X.2018.1441807>
- Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilo va, P. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges.
- Copeland, B. J. (2026). Machine learning from theory to algorithm an overview. *Journal of Physics Conference Series*.
- Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255–284.
- Eze, A. O., & Onah, B. I. (2024). Analysis of educational artificial intelligence tools utilized by vocational education lecturers in teaching and learning in public universities in Enugu State, Nigeria. *E-Learning and Digital Media*. Advance online publication. <https://doi.org/10.1177/20427530241297289>

- Federal Ministry of Communications, Innovation and Digital Economy. (2024). National Federal Republic of Nigeria. (2013). National policy on education (6th ed.). NERDC Press.
- Granström, M., & Oppi, P. (2025). Assessing teachers' readiness and perceived usefulness of AI in education: An Estonian perspective. *Frontiers in Education*, 10, Article 1622240. <https://doi.org/10.3389/feduc.2025.1622240>
- Ibrahim, A.-W., Taura, A. A., Iliyasu, A., Shogbesan, Y. O., & Lukman, S. A. (2024). Artificial intelligence (AI): Perception and utilization of AI technologies in educational assessment in Nigerian universities. *Edukasiana: Jurnal Inovasi Pendidikan*, 3(3), 367–380.
- Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, Article 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 35. <https://doi.org/10.1145/3457607>
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 5, Article 100152. <https://doi.org/10.1016/j.caeai.2023.100152>
- Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO.
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- Mpolomoka, D. L. (2025). Utilizing artificial intelligence for assessment in higher education. *Pedagogical Research*, 10(3), em0243.
- Ofem, U. J., & Chukwujama, G. (2024). Sustainable artificial intelligence-driven classroom assessment in higher institutions: Lessons from Estonia, China, the USA, and Australia for Nigeria. *European*

Journal of Interactive Multimedia and Education, 5(2), e02403.
<https://doi.org/10.30935/ejimed/15265>

- Ofem, U. J., Orim, F. S., Edam-Agbor, I. B., Amanso, E. O. I., Eni, E., Ukatu, J. O., Ovat, S. V., Osang, A. W., Dien, C., & Abuo, C. B. (2025). Teachers' preparedness for the utilization of artificial intelligence in classroom assessment: The contributory effects of attitude toward technology, technological readiness, and pedagogical beliefs with perceived ease of use and perceived usefulness as mediators. *Frontiers in Education*, 10, Article 1568306. <https://doi.org/10.3389/feduc.2025.1568306>
- Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(8), em2307. <https://doi.org/10.29333/ejmste/13428>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6), 49–66.
- Thomas, G., Gambari, A. I., Yusuf, H. T., Abanikannda, M. O., & Daramola, F. O. (2024). Assessment of lecturers' awareness of artificial intelligence for education in a Nigerian university. *Educational Technology Quarterly*, (4), 378–389.
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s4123>