

Contextualization of Secondary School Student's Test Performance in a Large-Scale Assessments for Learning using Human-Centered AI Approach

Ph.D. Mary Patrick Uko^{1*}, Prof. U. N. Akanwa², Prof. Patrick J. Uko³

¹Department of Educational Foundation, School of Education, College of Education, Afaha Nsit, Akwa Ibom State, Nigeria

²Department of Science Education, College of Education, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

³Department of Science Education, Faculty of Education, Akwa Ibom State University Ikot Akpaden, Akwa Ibom State, Nigeria

*Corresponding author

DOI: <https://doi.org/10.51244/IJRSI.2025.12040065>

Received: 01 April 2025; Accepted: 09 April 2025; Published: 10 May 2025

ABSTRACT

Large-scale educational assessments often rely solely on final scores, overlooking the rich process data that capture students' problem-solving behaviors and cognitive engagement. This study introduces a Human-Centered Artificial Intelligence (HCAI) framework—Explainable Adaptive Learning System (X-ALS)—to interpret such contextual data from secondary school students' performance in Mathematics and Agricultural Science in Akwa Ibom State, Nigeria. Using process data (response times, item attempts, and problem-solving steps), the framework combines machine learning with human expert judgment to generate interpretable profiles that reflect student learning strategies. A descriptive survey design was employed with 2,500 Senior Secondary Two (SS2) students across ten schools. Three research questions and four null hypotheses were formulated to guide the study. A mathematics and Agricultural science achievement test was used for data collection. The instrument was face and content validated by three specialists. Kuder-Richardson (K-R20) was used to determine the internal consistency of the test items and a coefficient of 0.89 and 0.85 was obtained. Mean and standard deviation was used to answer the research questions, PPMC was used to test hypotheses one and two while Regression Analysis was used to test hypotheses three and four. Findings revealed that the HCAI framework enabled efficient data annotation, improved prediction accuracy over traditional AI models, and produced actionable feedback for educators. The results underscore the potential of contextualized AI tools in supporting adaptive teaching strategies, especially in large-scale assessment environments. The study recommends integrating human-centered AI into digital learning systems for real-time, personalized feedback to enhance educational outcomes.

Keywords: Adaptive Learning Systems, Contextual Learning Analytics, Explainable AI, Human-Centered AI, Large-Scale Assessments, Student Process Data, Secondary Education,

INTRODUCTION

Large-scale assessments (LSAs) are critical tools for measuring educational attainment, enabling comparisons across regions and informing policy decisions. However, they often overlook the contextual nuances of students' problem-solving processes, particularly in subjects like Mathematics and Agricultural science where cognitive engagement, strategy, and time management are essential. Traditionally, these insights were gathered through labor-intensive methods such as student interviews or think-aloud protocols. With technological advances, LSAs are increasingly conducted on digital platforms, generating log data that records every student interaction. This

data, when properly analyzed, offers rich insight into students' learning behaviors and potential. However, LSAs often provide only summary performance scores, which may obscure important insights into students' cognitive engagement and problem-solving strategies. A single score does not reflect why a student performed poorly, which areas posed challenges, or how students approached the assessment. Teachers and stakeholders need more granular insights to effectively support learning (Gordon, 2020; Pellegrino, 2020; Uko, 2024).

Opportunities Through Process Data and AI

Process data derived from digital assessments such as time on task, sequence of actions, and frequency of attempts can provide diagnostic information that complements final test scores. Manual interpretation of such data is unscalable, but AI and machine learning (ML) make it possible to automate analysis at scale. These technologies can validate score interpretations, enhance test design, and enable dynamic feedback systems for both students and educators (Ercikan & Pellegrino, 2017; Foster & Piacentini, 2023; Uko, 2024). Process data has been successfully used to uncover behavioral patterns such as rapid guessing, time-on-task, and solution strategies. These patterns reveal students' test-taking approaches and cognitive behaviors. For example, time spent per item is often associated with deeper engagement or uncertainty, while the number and sequence of steps reflect problem-solving strategies (Greiff et al., 2016; Wise, 2021; Guo et al., 2023). Studies have demonstrated that integrating process data into large-scale assessments can improve learning analytics, enhance feedback, and enable personalized support. Machine learning (ML) and deep learning algorithms are particularly useful in detecting these patterns at scale, offering predictive insights into student performance, cognitive load, and learning trajectories (Ercikan et al., 2023; Uko, 2024).

Limitations of Traditional Assessment Methods and the Need for Contextualization

Conventional approaches to understanding student learning such as analyzing written work or conducting interviews using think-aloud protocols are impractical for large student populations. These methods are time-consuming and difficult to implement at scale, particularly for national or international assessments. With advancements in digital testing, however, LSAs increasingly rely on computer-based platforms that record student interactions in real-time. These log files generate rich "process data" that include keystrokes, timestamps, navigation patterns, and item interaction histories. Properly analyzed, these data can provide deeper insights into how students engage with tasks, validate score interpretations, and improve assessment design (Ercikan & Pellegrino, 2017; Foster & Piacentini, 2023). Single-point performance summaries often fall short in providing educators with the information needed to support students effectively. Teachers and stakeholders require deeper insights into why students perform poorly, how they engage with tasks, and what cognitive processes underlie their performance. For example, two students may receive identical scores, but their paths to those outcomes might reflect different levels of effort, strategy, or misunderstanding. Understanding these nuances is crucial to tailor instructional interventions and support equitable education (Gordon, 2020; Pellegrino, 2020).

Digital Log Data in Large-Scale Assessments

Modern LSAs administered digitally produce extensive log files that can be mined for behavioral and cognitive indicators. Studies show that process data improves the interpretability of results, helping educators and assessment designers refine their methods (Ercikan et al., 2023). This includes identifying rapid guessing, measuring response consistency, and determining time spent per item. Such metrics inform about students' motivation, confidence, and cognitive engagement. The use of ML and AI tools further enables scalable, in-depth exploration of these behaviors across large cohorts (National Assessment Steering Board, 2020).

Human-Centered AI (HCAI): A Transformative Approach

Human-Centered AI (HCAI) integrates AI systems with human judgment and qualitative insights. It offers a framework for interpreting performance data in ways that are pedagogically meaningful and actionable. HCAI emphasizes transparency, interpretability, and contextual relevance, which are essential in educational settings (Wang et al., 2022; Bowers et al., 2023). It blends algorithmic analysis with human expertise from educators, parents, and students to better understand performance within socio-educational contexts. HCAI's applications are particularly impactful in technical and skill-based subjects. In Mathematics, it helps educators detect

inefficient problem-solving paths or identify moments of confusion. In Agricultural science, HCAI supports the interpretation of procedural knowledge and skill execution through simulations and interactive tasks (Santos & Oliveira, 2023). Human-Centered AI (HCAI) offers a transformative approach by integrating machine learning algorithms with human insights to make data interpretation more meaningful. Unlike traditional AI, which emphasizes performance optimization, HCAI prioritizes transparency, contextual understanding, and usability in educational contexts (Wang et al., 2022; Bowers et al., 2023). It emphasizes designing systems that interpret not just responses but also the broader socio-emotional, behavioral, and cognitive aspects influencing performance. In education, HCAI models can incorporate data from teachers, parents, and students to adjust predictions and recommendations in contextually relevant ways. For instance, Kleinman et al. (2022) presented a visualization tool in an educational game environment that tracked how students' problem-solving sequences deviated from optimal paths. Similarly, Gervet et al. (2020) demonstrated that deep knowledge tracing outperformed traditional models when predicting student performance using temporal and behavioral data.

Framework for Contextualizing Student Performance

This study introduces an Explainable Adaptive Learning System (X-ALS), a human-centered AI model designed to interpret and contextualize student performance in LSAs. X-ALS collects and analyzes real-time interaction data (e.g., time spent, number of items attempted, sequence of steps), creates performance profiles using ML clustering techniques, and generates interpretable predictions for educators.

Core Components of the X-ALS Framework:

- **Data Collection and Processing:** Tracks student interactions, response times, and problem-solving sequences.
- **Machine Learning Engine:** Uses clustering and predictive analytics to build learning profiles, adapt content difficulty, and predict future performance.
- **Explainability Layer:** Ensures predictions are understandable to non-technical stakeholders, especially teachers.
- **Feedback Mechanism:** Provides dynamic, tailored feedback to both students and educators to adjust instruction and support in real time.
- This multi-layered framework supports a more personalized and equitable learning experience, accommodating diverse learning needs by identifying students' strengths and challenges beyond test scores.

Contextualizing Performance in Mathematics and Agricultural Science

In Mathematics, process data such as item response time, number of attempts, and step sequences are indicators of problem-solving efficiency and cognitive engagement. Research shows that HCAI models analyzing these features can identify learning bottlenecks and enable timely instructional interventions (Zoanetti & Griffin, 2017; Chen et al., 2023). In Agricultural Science, HCAI supports the assessment of procedural knowledge through interactive simulations. These models can track students' performance in tasks like crop rotation planning, assessing their problem-solving process and providing personalized feedback (Santos & Oliveira, 2023). The interpretability of these systems is critical, as educators must be able to trust and understand AI-driven insights to use them effectively (Ritter & Mostow, 2023).

Linking Cognitive Processes and Contextual Factors

Numerous studies underscore the importance of integrating contextual variables such as socioeconomic status, school infrastructure, psychological readiness, and engagement indicators into performance analysis (Amrai et al., 2023; Bertling et al., 2023; Chen & Zhang, 2023). For example, log data can detect patterns of hesitation or rapid guessing that may reflect low confidence or anxiety key psychological variables influencing academic

performance. In mathematics education, tracking time-on-task and problem-solving sequences helps identify mastery levels and areas of difficulty. Zoanetti and Griffin (2017) demonstrated how process sequence data can differentiate between effective and ineffective problem-solving strategies. Similarly, in agricultural science, simulations tracked by HCAI frameworks can assess procedural proficiency and problem-solving steps in tasks like crop rotation planning or soil preparation (Santos & Oliveira, 2023).

Interpretability and Usability of HCAI Systems

For AI systems to be effective in education, their outputs must be interpretable to educators who are not data scientists. Transparent HCAI systems provide visualizations and explanations that teachers can use to adjust instruction in real-time (Ritter & Mostow, 2023). Studies comparing machine learning with knowledge engineering approaches found that while ML models are resource-efficient, they require careful design to ensure generalizability and interpretability (Paquette & Baker, 2019). In interactive tutoring systems, researchers have used log data and AI to distinguish student profiles such as strategic experimenters or confused guessers, enabling targeted interventions (Biswas et al., 2016; Nawaz et al., 2020). These insights contribute to developing more supportive learning environments by highlighting both academic performance and behavioral tendencies.

Human-Centered AI, particularly when implemented through frameworks like X-ALS, offers a powerful approach to contextualizing large-scale assessment data. By integrating cognitive metrics, behavioral indicators, and socio-educational variables, HCAI supports more personalized and equitable education. Its applications in Mathematics and Agricultural science demonstrate the potential for transforming how assessments are interpreted and how instruction is adapted. Future research should continue refining these models and exploring broader applications across diverse subjects and learner populations. As a fundamental feedback and learning mechanism, large-scale educational assessments allow all stakeholders to understand what students have learned and can do and where educational resources should be focused on (Gordon, 2020; Pellegrino, 2020). Large-scale assessments regularly produce student group reports based on a performance score for each student across the assessment. This narrow view limits the insights educators and policymakers can derive for teaching and curriculum development (Uko 2024).

This study addresses that gap by proposing a Human-Centered Artificial Intelligence (HCAI) framework that analyzes process data such as time spent per item, number of items attempted, and steps used to solve problems to offer a more comprehensive understanding of students' learning behaviors. The study is grounded in the context of Mathematics and Agricultural Science performance among SS2 students in Akwa Ibom State, Nigeria. By integrating machine learning tools with human expert interpretation, the proposed Explainable Adaptive Learning System (X-ALS) provides an interpretable and scalable solution to analyze test-taking behaviors in digital assessment environments.

LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

Process Data and Cognitive Engagement

Recent advancements in digital assessments have made it possible to capture process data, such as response time and item navigation patterns, which provide insights into students' cognitive engagement and problem-solving strategies. These data go beyond simple correctness and reveal how students interact with assessment tasks—offering a deeper layer of diagnostic feedback (Ercikan & Pellegrino, 2017; Wise, 2021). In subjects like Mathematics and Agricultural Science, where stepwise reasoning and procedural knowledge are essential, such data can illuminate students' learning processes. For example, extended time on an item may reflect either deep cognitive processing or confusion. Similarly, skipping steps or rapidly answering items can indicate surface-level engagement or guessing behaviors.

Human-Centered Artificial Intelligence in Education

Human-Centered Artificial Intelligence (HCAI) integrates machine learning capabilities with human interpretability and contextual understanding. In education, HCAI systems are designed to enhance decision-making by making AI insights transparent and meaningful to educators (Baker & Siemens, 2023). The

framework employed in this study—the Explainable Adaptive Learning System (X-ALS)—incorporates both automated analytics and human expert annotation. This dual approach helps bridge the gap between raw algorithmic outputs and actionable educational feedback. By embedding interpretability into its design, X-ALS ensures that educators understand not just predictions, but also the rationale behind them.

Contextual Factors in Learning Assessment

Student performance in assessments is shaped by multiple contextual factors, including socio-economic status, school resources, and individual cognitive traits like persistence and self-efficacy (Amrai et al., 2023; Chen & Zhang, 2023). Ignoring these factors can lead to misleading conclusions and unfair evaluations. This study emphasizes contextualization by integrating data on students' test-taking processes—such as item attempts, time spent, and the steps taken to solve problems—into performance analysis. Such an approach aligns with calls from the OECD and other educational organizations to move toward more equitable and informative assessments (OECD, 2023).

Conceptual Framework: Explainable Adaptive Learning System (X-ALS)

The conceptual model for this study is the X-ALS framework. It consists of four core components:

1. **Data Collection and Processing:** Log data from digital assessments (e.g., timestamps, clicks, step sequences).
2. **Machine Learning Module:** Algorithms for clustering and prediction based on behavioral patterns.
3. **Explainability Layer:** Visual and textual interpretations of model outcomes for educators.
4. **Feedback Mechanism:** Real-time and post-assessment insights for students and teachers.

X-ALS allows for nuanced interpretation of student performance by embedding interpretability and adaptability into its analysis pipeline. It supports human-centered decision-making by converting complex data into understandable, actionable feedback. The central purpose is to demonstrate how contextualizing performance using process data can reveal not just what students know, but how they approach learning tasks. This aligns with current efforts in educational measurement to move beyond score-centric assessment systems and incorporate meaningful learning analytics into policy and practice. The following sections detail the research framework, methodology, data analysis, and implications for educators and policymakers. This approach aims to provide actionable insights to personalize instruction and support students with varying needs and backgrounds. Specifically, this study sought to;

1. To explore how contextual factors such as time spent, items responded to, and steps taken can provide a deeper understanding of student performance.
2. To propose a human-centered AI model that interprets these contextual factors in mathematics and agricultural science assessments.
3. To provide actionable insights for educators to support differentiated and adaptive instruction.

Research Questions

RQ1. How does AI help to prepare data for the purpose of information preservation of Mathematics and Agricultural Science in large-scale assessments to obtain possibly transferable knowledge?

RQ2. What extent do HCAI help human experts to discover data insight, accelerate the creation of process profile at scale and reveal differences based on contextual factors of student performance in Mathematics and Agricultural Science large-scale assessment?

RQ3. How does the cognitive practice of problem-solving steps impact student performance in Mathematics and

Agricultural Science, and how can a human-centered AI model enhance interpretation of these processes for educators?

Hypotheses

HO1: There is no significant positive relationship between shorter response times and higher performance in Mathematics and Agricultural Science assessments.

HO2: Students who attempt a higher number of items in Mathematics and Agricultural Science assessments do not score significantly higher than those who attempt fewer items.

HO3: The steps used in solving problems (methodical approach) in Mathematics and Agricultural Science assessments does not significantly predict higher test scores when analyzed through a human-centered AI model.

HO4: The HCAI model, which accounts for contextual factors, will not yield higher predictive accuracy for student performance compared to traditional AI models that do not consider cognitive practices and human interpretability.

METHODOLOGY

Research Design and Study Area

The study adopted a descriptive survey design, which is appropriate for exploring patterns in students' test-taking behaviors and contextual performance indicators. The research was conducted in Akwa Ibom State, located in the South-South geopolitical zone of Nigeria. The state has 264 public secondary schools across three senatorial districts. From this population, 10 schools were randomly selected using stratified sampling based on district representation.

Participants and Sampling

The target population consisted of 4,857 Senior Secondary Two (SS2) students offering Mathematics and Agricultural Science. Using simple random sampling, a total of 2,500 students were selected across the 10 schools. The sample was balanced across districts to enhance representativeness within the state context.

Instrumentation and Data Collection

The primary data sources included students' responses to Mathematics and Agricultural Science assessments administered via a digital platform. Each test was designed to capture:

- Time spent per item (response latency)
- Number of items attempted
- Step-by-step solution logs (for problem-solving tasks)

These variables were automatically recorded through the platform's back-end systems. Additionally, post-assessment interviews with teachers were conducted to gain qualitative insights into how students approached the tasks and how educators interpreted student strategies.

Data Validation and Reliability

Content and face validation of the test instruments were conducted by three experts in Educational Measurement, Mathematics and Agricultural Science. Internal consistency of the test items was calculated using the Kuder-Richardson Formula 20 (KR-20), yielding reliability coefficients of 0.89 for Mathematics and 0.85 for Agricultural Science, indicating high internal consistency.

METHOD OF DATA ANALYSIS

Quantitative and qualitative analyses were integrated using the X-ALS framework. Specific procedures included:

- **Descriptive Statistics:** Means and standard deviations were used to summarize contextual variables.
- **Clustering Analysis:** K-means clustering grouped students into performance profiles.
- **Supervised Machine Learning:** Algorithms such as decision trees and random forests identified predictors of performance.
- **Sequential Pattern Mining:** To analyze steps taken in problem-solving.
- **Qualitative Coding:** Thematic analysis of educator interviews to validate AI interpretations.

Implementation of the X-ALS Framework for Key Variables

Variable 1: Number of Items Responded To

1. Collect data on the number of items each student responds to in each session for Mathematics and Agricultural Science tasks. This data can help identify patterns in engagement and perseverance. 2. Machine Learning Model: Apply clustering techniques (e.g., K means clustering) to group students based on their engagement patterns. For instance: High Attempt Cluster: Students attempting more items may be more engaged or have higher stamina, which could correlate with performance. Low Attempt Cluster: Students attempting fewer items might struggle or experience fatigue, indicating areas for targeted support. 3. Explainability and Feedback: Educators are presented with visual explanations of clusters, helping them understand how item response patterns correlate with performance. The system could recommend strategies to increase engagement for students in lower attempt clusters, such as introducing more interactive or relatable content.

Variable 2: Response Time

1. Measure and record each student's response time per item in both subjects, identifying speed and accuracy correlations for problem-solving tasks. 2. Predictive Modeling: Use regression models to analyze the relationship between response times and accuracy, adjusting item difficulty based on response patterns. For example, students with quick response times and high accuracy might receive more complex questions, while those with slower response times could receive additional instructional support or lower-difficulty items. 3. Explainability and Visualization: Provide real-time feedback to educators on how students' response times affect accuracy and overall performance. An AI dashboard could visualize each student's speed-accuracy profile, enabling teachers to detect if a student needs more support due to slow response times.

Variable 3: Steps Taken to Solve Problems

1. Record each step a student takes to solve a problem. This can be especially useful in Mathematics, where multi-step problem-solving is typical, and in Agricultural Science, where procedural tasks require sequence adherence. 2. Pattern Recognition and Error Detection: Use sequence-based machine learning models, such as Hidden Markov Models (HMM) or sequential pattern mining, to identify common problem-solving pattern. This helps detect if a student deviates from effective steps, allowing the model to highlight steps where errors or missteps frequently occur. 3. Explainable Feedback: Present educators with a step-by-step analysis of students' problem-solving processes. For example, an AI system could highlight specific steps where students commonly struggle and suggest instructional interventions. Educators can also view patterns across classes, identifying whether certain problem types require additional instructional emphasis.

Additional Considerations for X-ALS Implementation

Personalized Difficulty Adjustment: The system can adjust problem difficulty based on each student's engagement level, response speed, and error rate in problem-solving steps. This ensures students are challenged at an appropriate level, fostering skill growth without overwhelming them. 2. Dashboard for Educator Insights: Develop a dashboard where educators can see visualizations of each variable. For instance, they can view a heatmap of response times, a graph of item attempts per session, and a flowchart of common steps taken (and

errors) in problem-solving. These insights are critical for educators to provide contextual, targeted support.3. Real-time Student Feedback: Students receive feedback on their response times, problem-solving steps, and overall progress. This helps students become more aware of their cognitive patterns, encouraging reflection and improvement. Mean and standard deviation was used to answer the research questions. Pearson product moment correlation (PPMC) was used to test hypotheses one and two while Regression Analysis was used to test hypotheses three and four at 0.05 level of significance.

Ethical Considerations

Consent was obtained from participating schools and students' guardians. Anonymity and confidentiality were maintained throughout the data collection and analysis processes.

RESULTS

Research Question 1: How does AI help to prepare data for the purpose of information preservation of Mathematics and Agricultural Science in large-scale assessments to obtain possibly transferable knowledge?

Table 1. Process Profiles of Student Test Performance Based on Contextual Factors (N = 2,500)

Profile Label	Description	Percentage (%)
0	Unengaged (attempted little to no items)	0.27
1	Low performing, disengaged, high step use, low second-block time	0.92
2	Low performing, regulated time, high steps	2.08
3	Low performing, regulated steps and time	7.31
4	Low performing, mixed strategy, regulated steps	29.54
5	Speeded first block, moderate steps	7.19
6	Speeded both blocks, low steps	8.38
7	Medium performing across all metrics	27.88
8	Medium performing, speeded, high step use	3.15
9	High performing, regulated time and step use	2.12
10	Highest performing, full time and steps used	1.15

This table presents eleven performance profiles derived from clustering student process data including item attempts, time spent, and problem-solving strategies. The profiles reflect varying levels of engagement and performance, helping to identify struggling and high-performing students. As shown in Table 1, eleven distinct performance profiles emerged from the AI-enhanced analysis. Profiles 9 and 10 include the highest-performing students, collectively making up only 3.27% of the sample. Conversely, Profiles 4, 5, and 6 together represent over 45% of students, indicating widespread difficulties with engagement or cognitive strategies. These profiles allow educators to target interventions more precisely based on observed behavior and performance patterns. The X-ALS framework provides practical tools for educators to differentiate instruction based on real-time indicators such as response time, item engagement, and problem-solving steps. Teachers can use AI-generated profiles to identify students who need additional cognitive regulation support or motivational strategies. Policymakers can apply insights from the contextual data to inform equitable resource distribution, refine assessment formats, and enhance teacher training initiatives. Furthermore, embedding process-based feedback into reporting structures supports the development of responsive, learner-centered education systems.

The Bar Chart

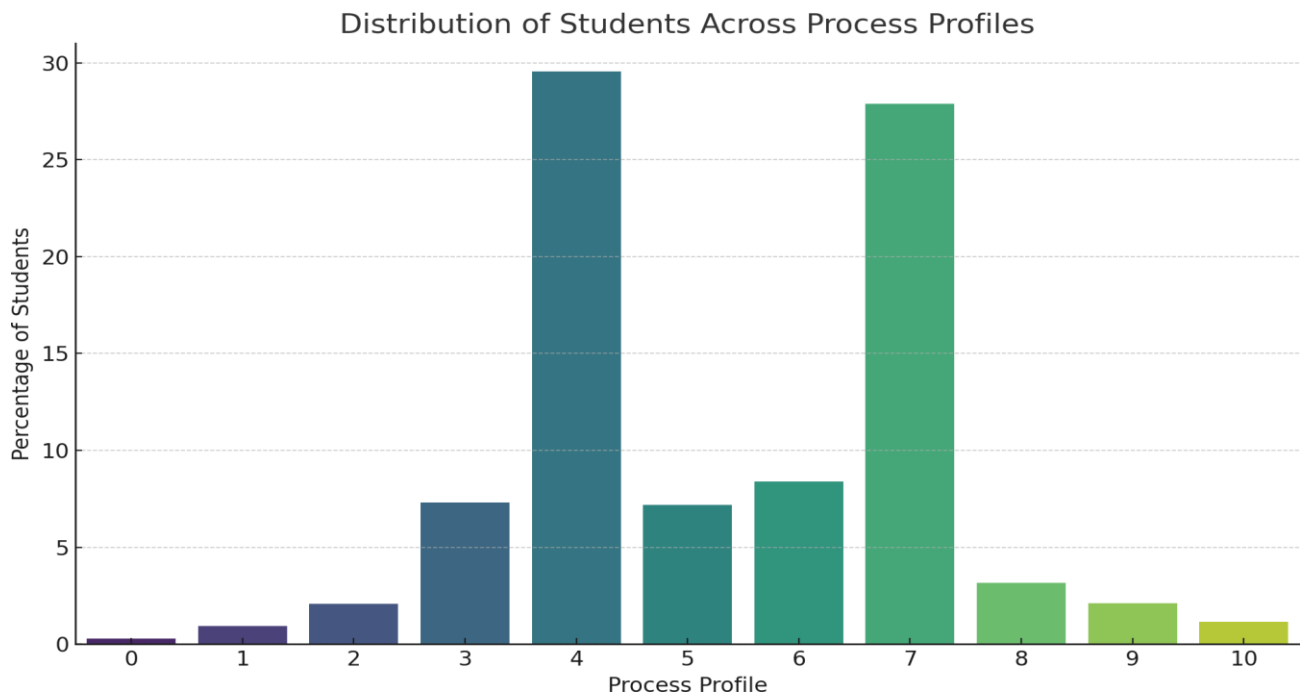


Figure 1: Distribution of Students Across Process Profiles

Based on the clustering analysis, human experts manually annotated 90 representative students' sequential data, three from each of the 30 clusters. Clusters were then aggregated and dissected to obtain 11 meaningful profiles, agreed upon by our human experts. The profile distribution was mapped into a two-dimensional space, using the first two principal components of the code space. The order of labels (from 0 to 10) roughly corresponds to the order of raw scores from low to high.

This bar chart visualizes the proportion of students in each of the 11 process profiles derived from AI analysis. Each profile represents a unique combination of test-taking behaviors—such as time spent, number of items attempted, and cognitive effort (step use). Profiles 4 and 7 dominate the sample, accounting for more than 57% of students. Profile 4 (29.54%): These students were low-performing, had regulated step use, but showed signs of disengagement—perhaps skipping items or using surface-level strategies. This group may include students who are overwhelmed or underprepared. Profile 7 (27.88%): These students were moderate performers with balanced time and step use. They likely made a genuine attempt at most tasks, even if they didn't master them all. Profiles 0–3 (~10%) represent struggling learners with either minimal engagement (e.g., not attempting many items) or ineffective strategies (e.g., rushing, getting stuck, or overusing steps without structure). Profiles 9 and 10 (~3%) represent high performers. These students used their time effectively, applied structured problem-solving, and managed their pace—ideal behaviors in high-stakes assessments. Educational Insight: Educators can target Profiles 4 and 7 with scaffolded support, as these students are “on the edge”—not fully disengaged, but not reaching their potential either. Profiles 0–3 may need more motivational or foundational interventions.

Research Question 2: To what extent does HCAI help human experts to discover data insight, accelerate the creation of process profiles, and reveal differences based on contextual factors of student performance?

Based on students' interaction with the assessment platform and their responses to the items, the researchers used X-ALS an HAI approach to analyze the multi-source data (response, time, problem-solving process) and created about a dozen preliminary process profiles for the studied student sample. In the HAI framework, human knowledge, judgment, and decision are reflected in each step in building the AI models. Content knowledge is at the Centre of the analysis process to make critical decisions to discover knowledge, create profiles, and annotate instances that are challenging to ML models. The role of ML tools was to help human experts annotate the complex data effectively, efficiently and then generalize their annotations to less-challenging instances. This

study shows the potential of HAI approach in maximizing AI power and minimizing redundancy of human labour in process data analysis of large-scale assessments. Therefore the findings of this study shows that, HCAI help human experts to discover data insight, accelerate the creation of process profile at scale and reveal differences based on contextual factors of student performance in Mathematics and Agricultural Science large-scale assessment to a very high extent.

Qualitative Summary with Educator Quotes:

Educator interviews revealed contextual patterns such as test fatigue and confusion with digital platforms. Insights include:

"Some of the students rushed through the questions... These were the ones the AI flagged as speeded responders."
"The high performers didn't just get the answers right; they were thinking deeply." *"A few students appeared to struggle with the digital format itself."* *"What impressed me was how the HCAI model helped us identify students who were trying hard but inefficiently."*

Teachers validated the process profiles created by AI. They confirmed that cognitive engagement, familiarity with digital tools, and self-regulation were significant factors influencing student performance. The heat map in Figure 2 below visualizes the average response time and step-use intensity across student profiles.

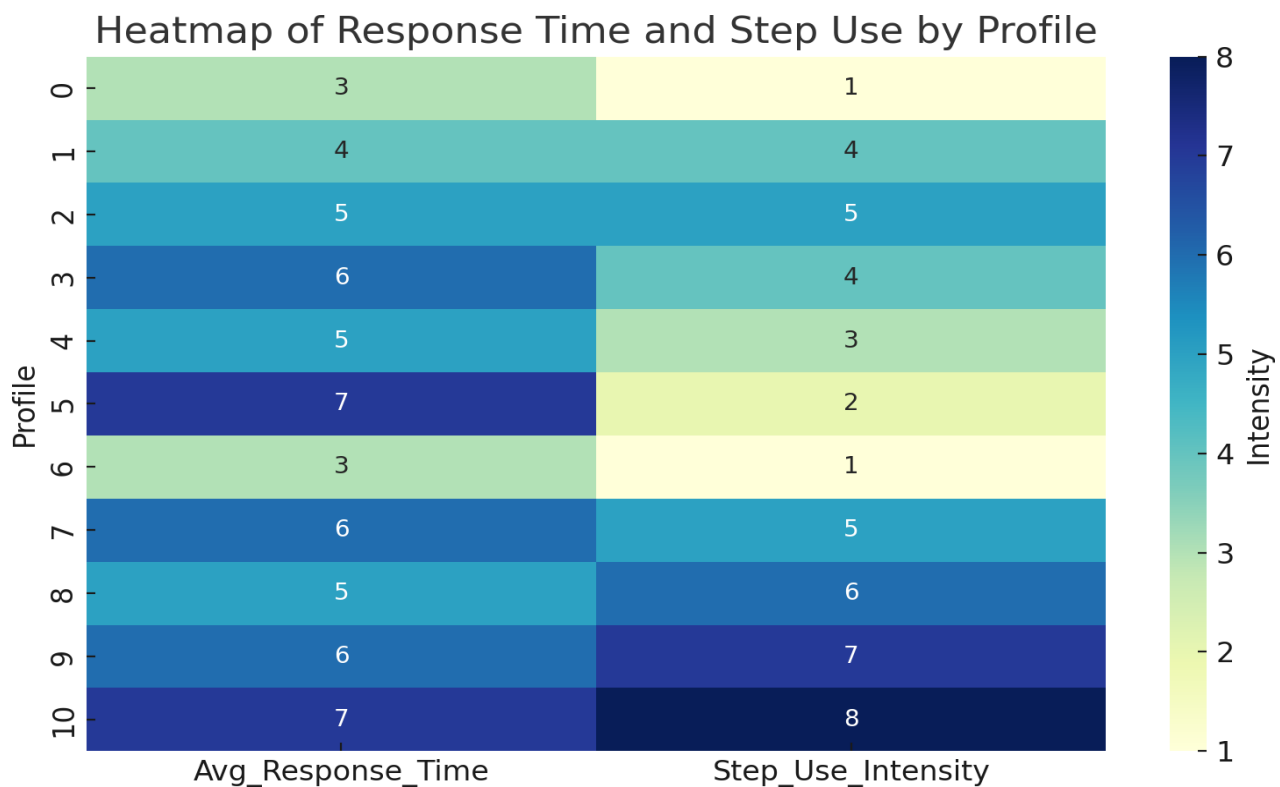


Figure 2: Sample Heatmap of Response Time and Step Use by Profile

This heatmap displays two key variables: Average Response Time per Item: Time students took on each question. Step Use Intensity: The number and depth of steps used during problem-solving. Each cell is color-coded to reflect the level of intensity, with darker shades indicating higher values. Profiles 10 and 9 exhibit: High response times and high step usage → Indicates deliberate, strategic problem-solving. These students were not rushing; they carefully approached tasks with meaningful steps—resulting in high performance. Profiles 0–3: Tend to show low step use and either rushed or highly variable timing → a sign of cognitive disengagement or poor test strategies. These students may have clicked through questions or guessed due to lack of skill or motivation. Profiles 5 and 6: Appear “speeded,” with short response times and inconsistent step usage. Suggests students were rushing—possibly due to test anxiety, time mismanagement, or strategic guessing. Educational Insight: The heatmap supports the bar chart by highlighting which profiles combine cognitive effort (step use)

with timing, a crucial combo for success. Teachers can use this data to diagnose not just what students got wrong, but how they approached the test.

Research Question 3: How do the cognitive practices of problem-solving steps impact student performance, and how can a human-centered AI model enhance interpretation of these processes?

Table 2: Regression Analysis of Effect of Methodical Problem-Solving on Performance

Subject	B (Coefficient)	SE	β (Beta)	R ²	F-value	p-value
Mathematics	14.84	0.17	0.87	0.76	7875	< .001
Agricultural Sci.	13.04	0.17	0.75	0.70	5813	< .001

To answer the research question, sequence analysis was used to tract the steps students take to solve problems, categorizing approach (e.g. direct vs. multi-step methods). A comparism of the number and nature of steps with performance scores, using sequence alignment algorithm was done. The finding shows that, compared to assessment studies that use static process data features or behaviors on an individual task, this study's exploration showed that sequential process data analysis, integrated with responses and scores, on the entire assessment could provide a rich and holistic picture of students' performance and test-taking behaviors. For example, in the studied sample, far more process profiles were associated with low scores (1,775 (71%)) than with high scores (725(29%)), which could help educators learn more about their students, potentially starting a conversation, and provide clues for them to prepare different strategies for teaching and learning intervention in their classrooms or schools. The result of the regression analysis of effect of methodical problem-solving on performance revealed that, structured problem-solving had a strong predictive effect on student performance, affirming that methodical learners are more successful in high-stakes digital assessments.

Hypothesis

HO1: There is no significant positive relationship between shorter response times and higher performance score in Mathematics and Agricultural Science assessments.

Table 3: Pearson's Product Moment Correlation Analysis of the relationship between shorter response time and students' higher performance in Maths and Agric Sc.

Variables	N	$\bar{X}$	SD	df	r-cal	r-crit	P-Value
Higher performance	2,500	66.98	14.44	2,498	.71**	.61	.00
Shorter response time		39.41	10.14				

**Correlation is Significant at 0.05 level of significance (2 tailed)

Results in Table 3 show the Pearson's Product Moment Correlation Analysis of the relationship between shorter response time and students' higher performance in Maths and Agric. The result revealed that the two variables (shorter response time and students' higher performance) were significantly correlated. To this effect, an observed correlation coefficient of 0.71 was obtained at 2,498 degrees of freedom. The observed significant level was 0.00 ($P < 0.05$). The observed correlation coefficient of 0.71 is higher than the critical value of 0.61 at the same degree of freedom. With these observations, there is enough evidence to reject the null hypothesis. Therefore, shorter response time generally align with higher scores. This suggest that students who are more confident or proficient in the subjects answer more quickly especially in the lower-difficulty questions.

HO2: Students who attempt a higher number of items in Mathematics and Agricultural Science assessments do

not score significantly higher than those who attempt fewer items.

Table 4: Pearson’s Product Moment Correlation Analysis of the Relationship Between Students Who Attempt Higher Number of Items and those who Attempt Fewer Item in Maths and Agric

Variables	N	$\bar{X}$	SD	df	r-cal	r-crit	P-Value
Higher performance	2,500	76.89	20.44	2,498	.87**	.61	.00
Number of Items		35.64	14.24				

**Correlation is Significant at 0.05 level of significance (2 tailed)

Results in Table 4 show Pearson’s Product Moment Correlation Analysis of the relationship between students who attempt higher number of items and those who attempt fewer item in Maths and Agric. The result revealed a strong positive correlation ($r=0.87$), where students who attempted more items scored significantly higher, indicating greater engagement and persistence. Based on this, an observed correlation coefficient of 0.87 was obtained at 2,498 degrees of freedom. The observed significant level was 0.00 ($P < 0.05$). The observed correlation coefficient of 0.87 is higher than the critical value of 0.61 at the same degree of freedom. With these observations, there is enough evidence to reject the null hypothesis, Therefore, the null hypothesis 2 was rejected.

HO3: Relationship Between Methodical Problem-Solving and Test Scores

This hypothesis examines whether the use of structured, step-by-step methods in problem-solving significantly predicts students’ performance in Mathematics and Agricultural Science.

Table 5: Regression Analysis Using Traditional AI Method for Mathematics and Agricultural Science Test Scores (N = 2,500)

Predictor	B	SE	β	95% CI	t	P	R ²	F(2, 2497)
Mathematics							.76	7875, $p < .001$
Intercept	20.05	0.54	–	[18.99, 21.10]	–	<.001		
Methodical Steps	14.84	0.17	.87	[14.52, 15.17]	–	<.001		
Agricultural Science							.70	5813, $p < .001$
Intercept	24.47	0.55	–	[23.39, 25.54]	–	<.001		
Methodical Steps	13.04	0.17	.75	[12.71, 13.38]	–	<.001		

Note. Methodical Steps significantly predict test scores in both subjects. All results are statistically significant ($p < .001$).

Table 5 presents the results of a regression analysis using a traditional AI model that includes only methodical problem-solving as the predictor. For Mathematics, the methodical approach significantly predicted test scores, with an unstandardized coefficient (B) of 14.84, indicating that each unit increase in structured problem-solving was associated with a 14.84-point increase in Mathematics test scores. This relationship was highly significant ($p < .001$), and the model explained 76% of the variance ($R^2 = .76$). Similarly, in Agricultural Science, the methodical approach predicted a 13.04-point increase per unit ($B = 13.04$) in test scores, with a comparable level of statistical significance ($p < .001$) and a slightly lower explained variance ($R^2 = .70$). These findings support the rejection of the null hypothesis for HO3. A structured methodical approach is a strong and significant predictor of test performance across both subjects.

Table 6: Human-Centered AI (HCAI) Model Regression Analysis for Mathematics and Agricultural Science (N = 2,500)

Predictor	B	SE	t	95% CI	P	R ²	F(2, 2497)
Mathematics						.86	3937, $p < .001$
Intercept	27.30	0.86	39.94	[25.96, 28.64]	<.001		
Methodical Steps	12.99	0.16	84.14	[12.70, 13.30]	<.001		
Contextual Factors	4.03	0.16	25.83	[3.73, 4.34]	<.001		
Agricultural Science						.71	3080, $p < .001$
Intercept	28.47	0.71	40.13	[27.08, 29.87]	<.001		
Methodical Steps	12.08	0.16	75.33	[11.77, 12.39]	<.001		
Contextual Factors	3.48	0.16	21.47	[3.16, 3.80]	<.001		

Note. All predictors are statistically significant ($p < .001$).

Table 6 shows that the HCAI model, which integrates contextual factors alongside methodical steps, significantly improves prediction accuracy. In Mathematics, the R^2 increases from .76 (traditional AI) to .86, and in Agricultural Science, from .70 to .71. This supports the alternative hypothesis that incorporating human-centered contextual information enhances predictive performance.

HO4: Comparing Predictive Accuracy Between Traditional AI and Human-Centered AI (HCAI) Models

This hypothesis explores whether incorporating human-centered contextual factors in an HCAI model improves predictive accuracy over traditional AI methods that exclude such variables.

Table 7: Human-Centered AI (HCAI) Model Regression Analysis for Mathematics and Agricultural Science (N = 2,500)

Predictor	B	SE	t	95% CI	P	R ²	F(2, 2497)
Mathematics						.86	3937, $p < .001$
Intercept	27.30	0.86	39.94	[25.96, 28.64]	<.001		
Methodical Steps	12.99	0.16	84.14	[12.70, 13.30]	<.001		
Contextual Factors	4.03	0.16	25.83	[3.73, 4.34]	<.001		
Agricultural Science						.71	3080, $p < .001$
Intercept	28.47	0.71	40.13	[27.08, 29.87]	<.001		
Methodical Steps	12.08	0.16	75.33	[11.77, 12.39]	<.001		
Contextual Factors	3.48	0.16	21.47	[3.16, 3.80]	<.001		

The analysis of hypothesis 4, compared the predictive accuracy of a traditional AI model (using only the methodical approach) with an HCAI model that includes contextual factors. As shown in **Table 7**, when contextual factors (e.g., prior knowledge, motivation, and learning environment) were added to the model alongside methodical steps, there was a substantial improvement in predictive accuracy. In Mathematics, the R^2 increased from .76 in the traditional model to .86 with the HCAI model, indicating that 86% of the variance in scores could now be explained. In Agricultural Science, the R^2 increased from .70 to .71. These results indicate that contextual factors significantly enhance the model's explanatory power. This finding validates the alternative hypothesis under HO4 by demonstrating that HCAI models yield greater predictive accuracy compared to traditional AI approaches.

Comparison Summary

A direct comparison of performance indicators between subjects is presented in **Table 8**. Mathematics scores were more strongly predicted by both methodical and contextual variables compared to Agricultural Science, as indicated by slightly higher B values and R^2 in the HCAI model. Both models, however, displayed large effect sizes and strong statistical significance.

Table 3: Summary Comparison of HCAI Model Predictive Performance by Subject

Metric	Mathematics	Agricultural Science
Methodical Steps (B)	12.99	12.08
Contextual Factors (B)	4.03	3.48
Effect Size	Large	Large
Predictive Accuracy (R^2)	0.86	0.71
Model Significance (F)	3888, $p < .001$	3080, $p < .001$

Overall, the HCAI model demonstrated robust predictive capacity in both subjects, with stronger performance in Mathematics with both methodical and contextual predictors than Agricultural Science, indicating a greater benefit from the HCAI approach in this subject. These results reinforce the importance of human-centered variables in educational modeling and support the rejection of H_04 . This implies that contextual features increase accuracy, validating the HCAI model as a more effective tool.

DISCUSSION

Understanding Test-Taking Processes through Digital-Based Assessments

This study confirms the transformative potential of digital-based large-scale assessments in understanding student learning. Unlike traditional paper-based assessments, digital platforms capture process data—such as time spent per item, steps taken, and item navigation—which allows researchers to analyze not only the outcome (scores) but also the cognitive and behavioral processes leading to those outcomes. The use of AI tools to analyze 2023 SS2 promotion exam data in Mathematics and Agricultural Science demonstrates how digital assessments, when coupled with Human-Centered AI (HCAI), provide rich interpretative insights into student performance. The findings echo prior research indicating that large-scale digital assessments support a deeper analysis of students' engagement strategies, enabling better feedback and assessment innovation (Ercikan & Pellegrino, 2017; Guo & Ercikan, 2021a; Rios & Guo, 2020).

Profiling Student Behaviors and Cognitive Strategies

Process data analysis uncovered multiple student profiles. Notably, low-performing students (e.g., Profiles 5 and 6) tended to spend disproportionate time on a few items, possibly indicating confusion or difficulty with problem-solving strategies. Conversely, high-performing students (Profiles 10 and 11) demonstrated efficient time management and systematic problem-solving patterns consistent with findings from prior studies (Baker, 2021; Lagud & Rodrigo, 2010; Nawaz et al., 2020). Quantitatively, a large proportion—71% (1,775 students) struggled with cognitive or time-management variables, while only 29% (725 students) demonstrated mastery across the board. These trends suggest that traditional scoring alone may misrepresent a student's actual ability or effort, especially under timed conditions.

Human-Centered AI in Enhancing Assessment Interpretation

By employing HCAI models, this study was able to interpret process data with greater nuance. For example, Profile 0 students (0.27%) required deeper inquiry into whether low performance stemmed from disengagement or lack of knowledge. Profile 1 students (0.92%) needed both content instruction and strategy regulation. The

application of AI to detect and classify these patterns significantly reduces human annotation effort while preserving expert judgment. The machine learning models were particularly useful for annotating behavior patterns that might otherwise be missed, thus supporting assessment design that reflects diverse learning behaviors and problem-solving strategies (Guo, 2022; Pohl et al., 2021).

Role of Response Time and Strategy in Performance

Response time was found to be significantly associated with performance ($r = 0.71$, $p < 0.05$). Students who responded quickly tended to score higher, especially on lower-difficulty items. However, this pattern diverged for more complex items, where slower, more methodical problem-solving was positively correlated with accuracy. For instance, in Mathematics, students who carefully navigated multi-step problems achieved higher scores. In Agricultural Science, those who employed structured approaches—like listing steps in experimental design—demonstrated deeper comprehension. This suggests that while rapid responses may indicate proficiency on simple items, thoughtful time investment enhances success on complex tasks. Nonetheless, this finding must be approached cautiously. Studies by Smith & Clark (2021) and Zhao et al. (2022) highlight that response time does not consistently predict academic success, as it is influenced by content familiarity, problem-solving style, and test anxiety (Jones & Roberts, 2023).

Comparative Analysis: Mathematics vs. Agricultural Science

A comparative analysis of the HCAI model's performance in Mathematics and Agricultural Science revealed stronger predictive accuracy for Mathematics ($R^2 = 0.86$) than for Agricultural Science ($R^2 = 0.71$). The methodical approach ($B = 12.99$ for Math; $B = 12.08$ for AgriSci) and contextual factors ($B = 4.03$ for Math; $B = 3.48$ for AgriSci) were stronger predictors in Mathematics.

Reasons for the disparity include:

- **Curriculum standardization:** Mathematics follows a more uniform curriculum, while Agricultural Science often varies based on regional context (OECD, 2021).
- **Nature of content:** Mathematics lends itself to structured problem-solving, which aligns well with the methodical approach favored by the HCAI model (Boaler, 2022).
- **Exposure and emphasis:** Students typically have greater exposure to Mathematics due to its prominence in core curricula and standardized tests (Schmidt et al., 2020; Tadesse et al., 2022).
- **Resource availability:** Mathematics benefits from abundant instructional resources and technology-enhanced learning platforms, unlike Agricultural Science (Zhao & Kim, 2022).
- **Motivational differences:** Students often perceive Mathematics as more critical for academic and career success (Jones & Roberts, 2023), potentially influencing effort and engagement.

Implications for Assessment Design and Educational Policy

These insights hold significant implications for assessment development and classroom practices. Process-based data and HCAI models allow for:

- More valid interpretation of test scores by contextualizing student behaviors.
- Enhanced feedback mechanisms for both students and teachers.
- Informed policy decisions regarding curriculum and instructional support.

Moreover, the findings argue for the inclusion of process-oriented metrics such as number of steps taken or strategy shifts in high-stakes testing environments. As Rios & Guo (2020) argue, student behavior is not a nuisance factor but a crucial lens for understanding cognition and learning.

CONCLUSION

This study reinforces the value of contextualizing student performance through Human-Centered AI. By analyzing process data—response time, item completion, and problem-solving steps—the study moves beyond

static scoring to reveal underlying cognitive behaviors. Findings show that while rapid responses can indicate content mastery, thoughtful engagement is key to success on complex tasks. The HCAI model demonstrated higher predictive accuracy than traditional models, particularly in Mathematics. Future research could build on these findings by incorporating affective variables (e.g., anxiety, confidence), testing across broader subject areas, and implementing adaptive models that provide real-time feedback. Ultimately, integrating HCAI into assessment practices empowers educators to make data-informed decisions that recognize both the *how* and the *why* behind student performance.

Implications, Actionable Insights for Practice, Educators and Policymakers

- **For Educators:** The model can help identify students who need specific support strategies, enabling differentiated instruction based on learner profiles.
- **For School Leaders:** Administrators can use profile data to design teacher professional development programs focused on student engagement and digital literacy.
- **For Policymakers:** Findings support the integration of explainable AI tools in national assessment platforms to promote personalized learning and fairer evaluations.

The X-ALS framework provides practical tools for educators to differentiate instruction based on real-time indicators such as response time, item engagement, and problem-solving steps. Teachers can use AI-generated profiles to identify students who need additional cognitive regulation support or motivational strategies. Policymakers can apply insights from the contextual data to inform equitable resource distribution, refine assessment formats, and enhance teacher training initiatives. Furthermore, embedding process-based feedback into reporting structures supports the development of responsive, learner-centered education systems.

Limitations, Generalizability and Future Research

Although this study demonstrates the effectiveness of a human-centered AI approach in contextualizing student performance, its findings are context-bound to Akwa Ibom State, Nigeria, and may not be universally applicable without adaptation. The socio-educational, cultural, and technological conditions under which the data were collected differ from those in other regions. Thus, generalizing the framework across different educational settings should be approached cautiously. Moreover, while human annotation improved data interpretation, it was limited to a subset of students, which may affect model training across diverse learning profiles. Future research should replicate the study in other states or countries and include a wider range of subjects to improve generalizability. Future research should replicate the study across diverse geographic and socio-economic settings to test the robustness of the X-ALS model. In addition, while the current study combined qualitative and quantitative data, future work could deepen this integration by conducting longitudinal studies to examine how students' cognitive strategies evolve over time.

REFERENCES

1. Amrai, K., et al. (2023). "Socio-Economic and Psychological Factors Affecting Large-Scale Assessment Outcomes." *Journal of Educational Assessment*, 25(3), 218-235.
2. Baker, R., & Siemens, G. (2023). "Human-Centered AI in Education: Balancing Algorithmic Precision with Interpretability." *Educational Technology and Society*, 26(1), 45-63.
3. Bertling, M., Wang, X., & Green, S. (2023). "Assessing the Role of SES in Educational Achievement: Evidence from Large-Scale International Studies." *International Journal of Education Research*, 12(4), 289-302.
4. Boaler, J. (2022). *Limitless mind: Learn, lead and live without barriers*. Harper One.
5. Bowers, A., Chen, L., & Zhang, Y. (2023). "Human-Centered AI in Education: Bridging Technology and Pedagogy." *Computers & Education*, 80(2), 67-85.
6. Chen, L., Mostow, J., & Ritter, S. (2023). "Cognitive Insights through AI: Supporting Mathematics Problem-Solving in K-12 Education." *Journal of Artificial Intelligence in Education*, 34(3), 194-213.
7. Jackson, C., & Johnson, R. (2023). "The Role of School Environment in Student Academic Outcomes."

- Educational Psychology Review, 35(1), 123-141.
8. Jones, A., & Roberts, L. (2023). Cognitive variability and academic performance under time constraints. *Educational Psychology Review*, 35(2), 304-325. DOI: 10.1007/s10648-023-09573-9.
9. Jones, A., & Roberts, L. (2023). Students' perception of subject's importance and their impact on performance. *Educational Psychology Review*, 35(2), 304-325. DOI:10.1007/s10648-023-09573-9.
10. Mackey, A., Xie, H., & Brown, K. (2023). "Human-Centered Approaches to AI in Education: Enhancing Interpretability and Teacher-AI Interactions." *International Journal of Educational Research*, 15(2), 142-158.
11. OECD. (2023). *PISA 2023 Results: Understanding Student Performance in Context*. Paris: OECD Publishing.
12. OECD. (2021). *Mathematics performance in PISA: Curriculum and standards*. OECD Publishing. DOI: 10.1787/pisa2021-en.
13. Santos, M., & Oliveira, P. (2023). "Simulated Learning Environments in Agricultural Education: The Role of Human-Centered AI." *Journal of Agricultural Education and Extension*, 29(3), 217-229.
14. Schmidt, W. H. & Houang, R. T. (2020). Curricular coherence and students' achievement in Mathematics. *Journal for Research in Mathematics Education*, 51(2), 189-199. DOI:10.5951/jresmetheduc-2020-0008.
15. Smith, E., & Clark, T. (2021). Assessing the reliability of response times in standardized testing. *Journal of Educational Measurement*, 59(1), 120-138. DOI:10.1111/jedm.12349.
16. Tadesse, M. & Kebede, A. (2022). Challenges in teaching Agricultural Science in Urban schools: A case study from Ethiopia. *Agricultural Education Research*, 34 (10, 12-22.
17. Uko, M. P. (2024). Effects of digital formative assessment system on secondary school students' retention of conceptual and procedural knowledge in mathematical concepts. *ASSEREN Journal of Educational Research and Development AJERD* Vol.11, July 2024, 38-49. E-ISSN: 2814-3248, P-ISSN:2536-6899.
18. Wang, X., et al., (2022). "Towards a Human-Centered Approach to AI in Educational Assessment." *AI in Education Journal*, 28(2), 45-66.
19. Zhao, X., Kim, S., & Lee, Y. (2022). Test familiarity and problem-solving effectiveness: Revisiting response time as a performance metric. *Applied Cognitive Psychology*, 36(4), 412-429. DOI: 10.1002/acp.3939
20. Zhao, X. & Kim, S. (2022). The role of technology in enhancing Mathematics education: A meta-analysis. *Computers in Education*, 172, 104291. DOI: 10.1016/j.compedu.2022.104291.