

Validity And Reliability Evaluation of an Automated Results Management System for Public Secondary Schools in Cross River State, Nigeria

Maurice Udie Angib¹, Sunday Okechukwu Abonyi², Julius Angib Angib³

^{1&2}Department of Science Education (Computer Science), Ebonyi State University, Abakaliki, Nigeria

³Department of Computer Science, Federal College of Education Obudu, Cross River State, Nigeria

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150900086>

Received: 22 September 2026; Accepted: 27 September 2026; Published: 15 October 2026

ABSTRACT

The increasing volume of students' academic records and the continued reliance on manual or semi-manual result-processing procedures in many public secondary schools create substantial challenges relating to accuracy, consistency, timeliness, retrieval and preservation of academic information. This study empirically evaluated the validity and reliability of an Automated Results Management System (ARMS) developed for public secondary schools in Cross River State, Nigeria. The evaluation adopted a quantitative software-validation approach involving criterion-related concurrent validation and repeated-processing reliability testing. Existing academic records of 300 Senior Secondary Two (SS II) students drawn from 10 purposively selected public secondary schools constituted the principal empirical dataset. Students' Continuous Assessment scores and examination scores were entered into the ARMS and processed to generate subject totals, grades, overall totals, averages, class positions and broadsheet entries. For validity assessment, system-generated outputs were compared item-by-item with corresponding manually processed school records used as the reference criterion. Reliability was assessed by processing the same 300-record dataset on three separate occasions without modification of the input data. Results showed that all 300 ARMS-generated records agreed with the reference records across the six evaluated result components, producing an overall agreement rate of 100% and zero recorded computational mismatch. The manual and automated procedures also produced identical aggregate scores ($\Sigma X = 18,735$) and mean scores ($M = 62.45$). Across the three reliability trials, the ARMS repeatedly produced an aggregate score of 18,735 and a mean of 62.45, indicating complete output reproducibility. The study therefore provides empirical evidence that the developed ARMS possesses strong computational validity and deterministic reliability for the processing of secondary-school academic results under the evaluated conditions. The findings demonstrate the potential of appropriately validated educational information systems to improve accuracy, consistency and integrity in school result management. Large-scale deployment should nevertheless be accompanied by further testing of usability, security, scalability, fault tolerance and performance under concurrent-user conditions.

Keywords: Automated Results Management System, Educational Information System, Software Validation, Concurrent Validity, Reliability.

INTRODUCTION

Accurate management of students' academic records is a fundamental responsibility of educational institutions. Examination and continuous-assessment results provide evidence of learners' academic achievement and support decisions relating to promotion, placement, certification, counselling, admission and educational planning. Consequently, errors in the computation, storage or reporting of students' results can have consequences extending beyond administrative inconvenience to the educational progression and opportunities of individual learners.

Despite rapid advances in information and communication technologies, result processing in many developing-country school systems continues to depend substantially on manual records, isolated spreadsheets and other semi-computerized procedures. Such approaches require teachers and examination officers to perform repetitive calculations involving continuous-assessment scores, examination marks, totals, averages, grades and class positions. Manual procedures can increase workload and create opportunities for arithmetic error, transcription error, duplication, inconsistent grading and delayed publication of results.

Within the public secondary-school system of Cross River State, Nigeria, these challenges are particularly important because academic records must be managed across numerous schools, classes, subjects and academic sessions. The underlying doctoral study from which the present article was derived identified continued problems involving delayed result processing, computational inaccuracies, poor storage and retrieval of records, vulnerability of records to unauthorized access and difficulties in obtaining historical academic information. An Automated Results Management System was therefore developed to automate major activities involved in academic-result processing.

The existence of a functioning software application does not, however, establish that its outputs are accurate or dependable. Software intended for use in educational decision-making must be subjected to systematic evaluation before routine institutional deployment. Software-quality evaluation increasingly emphasizes empirical verification of whether information systems produce intended outputs consistently and meet defined quality requirements. International software quality frameworks similarly recognize functional suitability and reliability as fundamental dimensions of software quality.

Noël, Taramasco, and Márquez (2022), in a systematic review of information-system quality assessment, observed that although formal software quality standards contain numerous quality attributes, empirical procedures for operationalizing those attributes remain comparatively limited. They consequently emphasized the importance of measurable evaluation mechanisms rather than relying only on descriptive assertions of system quality. Similarly, Souza-Pereira, Pombo, and Ouhbi (2022) demonstrated that systematic measurement of software quality characteristics provides empirical evidence of whether a system meets stakeholder expectations. More recent work by Santokhee, Augusto, and Brodie (2024) also emphasizes linking measurable software quality characteristics directly to functional requirements and evaluating them through empirical evidence.

For an academic result management application, two particularly important qualities are validity and reliability. Validity concerns whether the application produces results corresponding to the correct or established values for the same input. Reliability concerns whether the system produces stable and reproducible outputs when identical operations are repeated under the same conditions. These properties are crucial because even an attractive or user-friendly system would be unsuitable for academic result management if it produced incorrect grades, averages or student positions.

An automated result system should therefore not merely calculate faster than the traditional process; it must demonstrate that its calculations are correct and reproducible. This requirement distinguishes the present study from development studies that principally describe system design, interfaces and functionality without subjecting system outputs to substantive empirical verification.

Problem Statement

Manual and semi-manual academic-result processing exposes educational institutions to computational errors, delayed result preparation, inconsistent records and difficulties in retrieval. Automated systems offer a potential solution, but their usefulness depends on the accuracy and reproducibility of the outputs they generate. Introducing inadequately validated software into an academic environment may simply replace one source of error with another and may create errors on a larger scale because automated calculations are replicated across many records.

Although various school result applications have been developed in Nigeria and similar contexts, much of the available development literature concentrates on system architecture and functionality. For example, previous Nigerian work on online result processing has demonstrated the feasibility of centralized digital processing for secondary school results (Ezema & Ogbuokiri, 2022). However, there remains a need for empirical evidence showing whether a newly developed application reproduces established academic results accurately and whether such outputs remain stable across repeated processing.

The central problem addressed in the present study was therefore the need to determine empirically whether the developed ARMS generated academic results that agreed with established school records and whether the system consistently reproduced the same outputs when identical datasets were processed repeatedly.

Purpose of the Study

The study evaluated the validity and reliability of the developed Automated Results Management System for processing students' academic records in public secondary schools in Cross River State, Nigeria.

Specifically, the study sought to:

- determine the level of agreement between academic results generated by the ARMS and corresponding results obtained through the established school result processing procedure; and
- determine the consistency of ARMS-generated academic results across repeated processing of the same dataset.

Research Questions

The following research questions guided the study:

- What level of validity does the developed Automated Results Management System demonstrate when its outputs are compared with corresponding established academic records?
- To what extent does the developed Automated Results Management System produce consistent results when the same students' academic dataset is processed repeatedly?

LITERATURE REVIEW

Automated Academic-Result Management

An Automated Results Management System is an information system designed to capture, process, store, retrieve and report students' academic records with minimal manual computation. Its principal operations may include student registration, subject registration, score entry, addition of continuous-assessment and examination scores, automatic grading, computation of averages, ranking, generation of broadsheets and preparation of individual report cards.

Automation is particularly relevant to result processing because many academic calculations are repetitive and rules-based. Once approved grading and computational rules have been correctly encoded, software can apply them consistently across large numbers of student records. Digital systems can additionally improve storage and retrieval by maintaining records within structured databases instead of paper registers and independent files.

Ezema and Ogbuokiri (2022) developed an online result processing system for secondary schools and observed that digital processing could provide centralized storage and facilitate result management and dissemination. This demonstrates the relevance of automation to secondary school administration.

However, the benefits of automation presuppose that the implemented computational logic produces correct outcomes.

Software Validity and Output Accuracy

Validation asks whether a software product satisfactorily fulfils its intended purpose. Within the present context, computational validity refers specifically to whether the result management application correctly translates input scores into the expected totals, grades, averages, positions and result reports.

Criterion-related concurrent validation is appropriate where output from a new method can be directly compared with an established reference for the same cases. The process provides more direct evidence than simply asking users whether they believe a system is accurate. Each value generated by the system can be compared against the corresponding reference value, and disagreements can be identified quantitatively.

Software quality research supports the need for measurable approaches to system evaluation. Noël et al. (2022) found that internationally recognized quality frameworks provide useful structures for evaluation but argued that quality attributes must ultimately be translated into practical measures. Souza-Pereira et al. (2022) similarly showed that measurable software evaluation enables investigators to identify whether a system actually satisfies expected quality conditions rather than assuming quality on the basis of its design.

In academic result processing, output validation is particularly appropriate because expected computational outputs can be independently established. If identical raw scores and grading rules are processed through two methods, a properly implemented deterministic algorithm should produce equivalent results.

Software Reliability and Repeatability

Reliability refers to the capacity of a software system to perform required functions consistently under specified conditions. For deterministic result management operations, one basic dimension of reliability is repeatability: identical input data processed under identical rules should yield identical outputs.

The contemporary software quality literature conceptualizes reliability as broader than repeatability alone. It can include availability, fault tolerance and recoverability. Nevertheless, output repeatability remains indispensable. An information system cannot reasonably be regarded as reliable if repeated processing of identical academic records yields different grades, totals or rankings.

Santokhee et al. (2024) emphasize that software-quality characteristics should be linked to specific measurable requirements. Applied to academic result processing, repeatability can be operationalized by running an identical dataset on multiple occasions and comparing the outputs. Complete agreement across repeated trials provides evidence of computational stability, although it does not by itself establish resilience to infrastructure failure, high transaction load or cybersecurity threats.

Research Gap

A considerable proportion of educational result management research focuses on designing databases, creating web interfaces and demonstrating system functions. Fewer studies present structured empirical comparisons between automated results and established records followed by repeated processing tests of the same data.

The present study addresses this gap by subjecting an operational ARMS to criterion-based output validation using actual school records and by examining reproducibility across repeated processing trials. It therefore contributes empirical evidence concerning the computational quality of educational administrative software within a public secondary-school environment.

MATERIALS AND METHODS

Research Design

The study employed a quantitative software-evaluation design. Two complementary procedures were used:

criterion-related concurrent validation for determining accuracy and repeated-processing testing for determining computational reliability.

The approach formed part of a larger Research and Development study guided by the ADDIE framework in which the ARMS was analyzed, designed, developed, implemented and evaluated. The present paper is restricted specifically to the empirical validity and reliability evaluation.

Study Context

The investigation was conducted within public secondary schools in Cross River State, Nigeria. The operational environment was selected because public secondary schools routinely generate continuous-assessment and examination data and maintain formal academic-result records suitable for comparison with outputs from the developed ARMS.

Dataset and Sampling

The principal evaluation dataset consisted of the existing academic records of 300 Senior Secondary Two students drawn from 10 purposively selected public secondary schools.

Senior Secondary Two records were suitable because the students remained within the school system and possessed continuing academic records that could be processed without relying on hypothetical or artificially generated observations.

For each student, relevant biodata and academic scores were obtained from existing school records. The score inputs included Continuous Assessment components and examination scores. Nine school subjects were represented in the result processing exercise. No personally identifying information is presented in this article.

Automated Results Management System

The ARMS was developed to automate the principal academic result processing functions required in the study environment. The application accepted students' biodata and assessment scores and generated academic outputs including:

- subject totals;
- grades;
- overall totals;
- averages;
- class positions;
- report information; and
- broadsheet entries.

The system incorporated a centralized database to support storage and retrieval of students' academic records.

Validity Evaluation Procedure

Criterion-related concurrent validation was used to evaluate output accuracy. Existing manually prepared academic records obtained from the participating schools served as the reference criterion for comparison

because these represented the officially maintained results against which the newly developed system was being evaluated.

The following procedure was applied:

First, students' Continuous Assessment and examination scores were extracted from the participating schools' existing records. Second, the same input data were entered into the ARMS. Third, the ARMS automatically processed the records and generated the relevant results. Fourth, ARMS-generated outputs were compared item-by-item with the corresponding reference records.

Comparison focused on:

- subject totals;
- grades;
- overall totals;
- averages;
- class positions; and
- broadsheet entries.

For each component, the number of matches and mismatches was determined. Percentage agreement was calculated as:

$$\text{Agreement (\%)} = (\text{Number of matching records} / \text{Total number of compared records}) \times 100$$

A high proportion of agreement was interpreted as evidence of concurrent computational validity.

Reliability Evaluation Procedure

Repeated processing testing was used to determine computational reliability. The same dataset containing the 300 students' academic records was processed using the ARMS on three separate occasions without changing the input values or computation rules.

Aggregate results and mean scores were compared across Trial 1, Trial 2 and Trial 3. Output consistency was assessed by determining whether repeated processing reproduced identical results.

The original thesis additionally reports Pearson Product Moment Correlation as a reliability statistic. Because identical records were reproduced across trials, the relationship between repeated outputs was perfect. For the present article, exact reproducibility is treated as the principal reliability evidence because it directly tests the deterministic expectation that unchanged inputs should produce unchanged outputs.

Data Analysis

Data were analyzed using frequencies, percentages, sums and means. Percentage agreement was used for criterion-related validity. Reliability was assessed through comparison of repeated processing outputs.

The larger study also reported paired-sample comparison and correlation procedures. However, where every paired difference equals zero, the variance of the difference distribution is zero. Under such conditions, percentage agreement and exact discrepancy analysis are more informative for software validation than conventional significance testing because the substantive finding is exact equivalence rather than failure to detect a difference.

This reporting approach is adopted to provide a statistically transparent interpretation suitable for an international empirical audience.

RESULTS

Criterion-Related Validity of the ARMS

The first analysis examined agreement between the established manually processed records and outputs produced by the ARMS.

Table1 Agreement Between Reference Records and ARMS-Generated Results (N = 300)

Result component	Records compared	Matches	Mismatches	Agreement
Subject totals	300	300	0	100.00%
Grades	300	300	0	100.00%
Overall totals	300	300	0	100.00%
Averages	300	300	0	100.00%
Positions	300	300	0	100.00%
Broadsheet entries	300	300	0	100.00%

The results indicate complete agreement between ARMS-generated outcomes and the comparison records. No mismatch was recorded for any of the six evaluated result components. The observed agreement rate was therefore 100%.

This means that, for the 300 records evaluated under the specified grading and processing rules, the ARMS reproduced the expected academic outcomes without recorded computational discrepancy.

Comparison of Aggregate Result Statistics

Aggregate statistics from the reference procedure and ARMS procedure were also compared.

Table2 Aggregate Comparison of Reference and ARMS Results

Measure	Reference records	ARMS outputs
Number of student records	300	300
Sum of scores	18,735	18,735
Mean score	62.45	62.45
Mean difference	0.00	—

The sum of students' scores obtained from both processing procedures was identical at 18,735. The mean was also identical at 62.45, producing a mean difference of zero.

These results reinforce the record-level analysis and demonstrate that agreement was maintained at both

individual output and aggregate levels.

Repeated-Processing Reliability

The second research question investigated whether the ARMS would reproduce the same output when an unchanged dataset was processed repeatedly.

Table3 Reproducibility of ARMS Outputs Across Three Processing Trials

Trial	Number of records	Sum of scores	Mean score
Trial 1	300	18,735	62.45
Trial 2	300	18,735	62.45
Trial 3	300	18,735	62.45

All three processing trials generated the same aggregate score of 18,735 and the same mean score of 62.45. No change in output was observed when the dataset was repeatedly processed without modification.

The reproducibility rate was therefore 100% under the evaluated conditions. The results demonstrate deterministic computational stability: identical input values subjected to identical processing rules yielded identical outputs.

DISCUSSION

The principal finding of this study is that the developed Automated Results Management System demonstrated complete computational agreement with the reference academic records across the evaluated dataset. All 300 student records matched across subject totals, grades, overall totals, averages, class positions and broadsheet entries.

This result is important because accuracy is central to the acceptability of academic administrative systems. Errors in academic results may have direct implications for promotion, certification, placement and other educational decisions. Consequently, the value of an automated result-processing application rests not merely on reducing the time required for computation but on correctly executing the underlying academic rules.

The observed result supports the broader position in software quality research that information systems should be evaluated through measurable attributes rather than solely through descriptions of system functions. Noël et al. (2022) found limited evidence of practical mechanisms for operationalizing information-system quality standards and called for measurable assessment methods. The item-by-item criterion comparison applied in the present study constitutes such an operational measure because it provides a direct quantitative indicator of computational accuracy.

The results are also consistent with the logic of quality-evaluation approaches discussed by Souza-Pereira et al. (2022), who emphasized that software quality must be assessed against defined characteristics and measurable expectations. In the current study, the expected requirement was unambiguous: the ARMS should reproduce correctly determined academic outcomes from the same input scores. The absence of discrepancies across 300 cases provides strong evidence that the underlying computational functions operated correctly within the tested dataset.

The reliability analysis produced similarly strong findings. Across three processing occasions, the application generated the same total score and mean score without variation. Such reproducibility is

particularly important for deterministic educational systems because an identical academic record should not yield different grades or positions merely because processing occurs on another occasion.

The result also illustrates an important distinction between computational reliability and the broader concept of software reliability. The present evaluation demonstrates computational repeatability: the same inputs produced the same outputs. International software-quality perspectives, however, conceptualize reliability more broadly to include dimensions such as availability, maturity, recoverability and fault tolerance. Therefore, the finding should not be interpreted to mean that every dimension of system reliability has been established.

This distinction improves the interpretation of the 100% result. The evidence supports the conclusion that the application's computational procedures were stable and reproducible within the evaluation conditions. It does not automatically establish how the system would behave under prolonged server failure, large-scale concurrent access, malicious interference, unstable internet connectivity or hardware malfunction.

The study additionally contributes to educational information-system research in the Nigerian context. Previous work, including Ezema and Ogbuokiri (2022), has demonstrated the practical value of online result processing in secondary schools. The present investigation extends that development-oriented literature by providing explicit output validation and repeatability evidence based on actual academic records.

Another important methodological implication concerns the interpretation of perfect agreement. Conventional null-hypothesis tests are not always the most informative tools for assessing deterministic software outputs. When every pair of observations is identical, the paired differences contain no variability. The important information is therefore not that a statistical test failed to detect a difference, but that the observed discrepancy count was zero. For software validation of deterministic calculations, agreement measures and error rates can consequently provide clearer evidence than significance tests.

Practical Implications

The findings have implications for school administrators, examination officers, teachers and educational policymakers. A computationally valid automated result system can reduce repeated arithmetic operations and improve consistency in the application of result-processing rules.

Automated computation may also reduce opportunities for accidental arithmetic errors when totals, averages and grades are generated. Centralized processing could furthermore support timely preparation of broadsheets and report cards.

For education authorities, system-level automation may support standardization of result-processing procedures across participating schools. However, deployment should be accompanied by clear data-governance rules and appropriate access controls because academic records constitute sensitive institutional information.

Contribution to Knowledge

The study makes three main contributions.

First, it provides empirical evidence based on actual secondary-school academic records rather than solely demonstrating application interfaces or system architecture.

Second, it operationalizes concurrent software validation through item-by-item comparison between new-system outputs and an established reference dataset.

Third, it demonstrates repeatability through repeated processing of an unchanged 300-record dataset, thereby providing a transparent measure of computational reliability.

Collectively, these contributions strengthen evidence for the use of empirical software evaluation in educational administration.

LIMITATIONS

The findings should be interpreted within the boundaries of the study.

First, the validation involved 300 SS II student records from 10 public secondary schools in one Nigerian state. Although sufficient for the intended evaluation, further studies involving larger datasets and additional schools would strengthen evidence for general deployment.

Second, the study assessed computational accuracy and repeatability but did not comprehensively evaluate every software-quality characteristic. Performance efficiency under very large datasets, security penetration resistance, availability, fault tolerance and disaster recovery require specialized testing.

Third, reliability was examined under controlled repeated-processing conditions in which the input dataset remained unchanged. This is appropriate for evaluating computational reproducibility but does not constitute a complete operational-reliability assessment.

Fourth, the reference criterion consisted of existing school records. Where manual records contain undetected human errors, direct agreement with such records may not necessarily represent mathematical correctness. Future validation should therefore combine comparison with school records and independently recomputed reference values.

Fifth, user-centred evaluation of usability, satisfaction, perceived usefulness and ease of use was outside the specific scope of this paper. Such dimensions should be investigated before large-scale institutional adoption.

CONCLUSION

This study empirically evaluated the validity and computational reliability of an Automated Results Management System developed for public secondary schools in Cross River State, Nigeria.

Using academic records of 300 SS II students from 10 public secondary schools, ARMS-generated outputs were compared with corresponding established school records across subject totals, grades, overall totals, averages, class positions and broadsheet entries. Complete agreement was observed, with 300 of 300 records matching and no recorded computational mismatch.

The system also demonstrated complete repeatability when the same dataset was processed on three separate occasions. Each trial produced an aggregate score of 18,735 and a mean score of 62.45.

The results therefore provide empirical evidence that the ARMS possesses strong criterion-related computational validity and deterministic reliability under the conditions in which it was evaluated. The findings support the potential adoption of validated automated systems for improving the accuracy and consistency of academic-result processing.

Nevertheless, claims of total software quality should extend beyond computational validity. Before state-wide deployment, additional evaluation should examine security, scalability, usability, performance efficiency, availability, fault tolerance and disaster-recovery capability.

RECOMMENDATIONS

Based on the empirical findings, public secondary schools and relevant education authorities may consider phased implementation of the developed ARMS after appropriate institutional and technical review.

Before large-scale deployment, the system should undergo independent security assessment, concurrent-user/load testing, backup and recovery testing, and structured usability evaluation involving teachers, examination officers and school administrators.

A state-level implementation should include standardized grading rules, data-governance protocols, role-based access privileges, secure authentication, audit logging, scheduled backups and user training.

Future empirical studies should evaluate the application with larger multi-school datasets and examine performance under realistic concurrent-use conditions.

Further research should also assess user acceptance, usability, security and long-term operational reliability so that computational accuracy can be integrated with broader internationally recognized software-quality requirements.

REFERENCES

1. Ezema, H. K., & Ogbuokiri, B. (2022). Online result processing system for secondary schools. *International Journal of Computer Sciences and Engineering*, 10(9), 1–5. <https://doi.org/10.26438/ijcse/v10i9.15>
2. International Organization for Standardization. (2018). ISO 9241-11:2018 Ergonomics of human-system interaction—Part 11: Usability: Definitions and concepts. ISO.
3. Laudon, K. C., & Laudon, J. P. (2020). *Management information systems: Managing the digital firm* (16th ed.). Pearson.
4. Noël, R., Taramasco, C., & Márquez, G. (2022). Standards, processes, and tools used to evaluate the quality of health information systems: Systematic literature review. *Journal of Medical Internet Research*, 24(3), e26577. <https://doi.org/10.2196/26577>
5. Pressman, R. S., & Maxim, B. R. (2020). *Software engineering: A practitioner's approach* (9th ed.). McGraw-Hill.
6. Santokhee, A., Augusto, J. C., & Brodie, L. (2024). Towards engineering higher quality intelligent environments: A multi case study approach. *Software Quality Journal*, 32, 1075–1135. <https://doi.org/10.1007/s11219-024-09678-0>
7. Sommerville, I. (2016). *Software engineering* (10th ed.). Pearson.
8. Souza-Pereira, L., Pombo, N., & Ouhbi, S. (2022). Software quality: Application of a process model for quality-in-use assessment. *Journal of King Saud University – Computer and Information Sciences*, 34(7), 4626–4634. <https://doi.org/10.1016/j.jksuci.2022.03.031>
9. Taherdoost, H. (2016). Validity and reliability of the research instrument: How to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5(3), 28–36.