Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Innovative test item analysis using optical mark recognition technology: An evaluation

Ruth G. Luciano *

College of Information and Communications Technology, Nueva Ecija University of Science and Technology, Cabanatuan, Philippines

ARTICLE INFO

Article history:
Received 9 September 2024
Received in revised form
6 January 2025
Accepted 2 April 2025

Keywords:
Optical mark recognition
Test item analysis
Automated assessment
Software evaluation
Educational tools

ABSTRACT

This study focuses on the need for effective tools to improve assessment processes in education by developing and evaluating a software application that analyzes test items using Optical Mark Recognition (OMR) technology. Traditional methods of test item analysis are often slow and unreliable due to manual handling and limited statistical insights. The proposed software aims to automate the creation, analysis, and management of test items, making the process more efficient for educators. The study follows a mixed-method approach, using qualitative methods for software design and quantitative evaluation based on ISO/IEC 25010 software quality standards. Developmental research principles guide the continuous improvement of the system to meet educational goals and user needs. Initial assessments by IT experts and users confirm the system's functionality and ease of use. Recent advancements in automated assessment systems highlight the potential of OMR-based technology to make test item analysis faster and more accurate. The evaluation phase uses quantitative measures to assess the system's reliability, efficiency, and user satisfaction. Findings from related studies on question difficulty prediction further support improvements to the software, ensuring it meets the demands of modern educational assessment.

1. Introduction

In the 21st-century educational system, assessment plays a crucial role in ensuring the quality of learning provided to students. According to de Guzman and Adamos (2020), effective assessment mechanisms offer accurate feedback and are central to the learning process, gauging and enhancing student learning. They also articulated that assessment is used to identify students' learning needs, monitor their progress, and evaluate their performance against established learning outcomes. It is crucial for students, teachers, and stakeholders to comprehend the purpose and significance of assessments and their connection to overall student learning progress.

A high-quality test is essential in evaluating students' achievement in line with subject objectives, allowing students to self-assess their understanding at the end of each term (Yan et al., 2020). In addition, they mentioned that assessments help identify individual students' weaknesses, providing valuable feedback for improvement.

However, generating examinations is a labor-intensive and challenging task for teachers. Often, the same test items are reused without proper statistical item analysis, leading to deficiencies in teaching and unreliable assessments. Currently, teachers in the Department of Education (DepEd), particularly in the research locale, use an Item Analysis System based on MS Excel, which requires manual input of students' test scores, adding to the teachers' workload, especially for SHS teachers who handle multiple subjects.

Traditionally, teachers used spreadsheet software to compute and generate Test Item Analysis results, requiring manual input of students' test results. For example, a teacher with 50 students must select the top and bottom 27% scorers, 14 students each, and manually input their test answers into the spreadsheet. This manual process fails to guarantee test quality, as ambiguous or misleading test items are not addressed, and teachers often use the same test questions repeatedly. Test Item Analysis is valuable for enhancing teachers' skills in test construction and identifying areas of the

* Corresponding Author.
Email Address: rcgluciano@gmail.com
https://doi.org/10.21833/ijaas.2025.04.001
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-8532-6971

curriculum needing more emphasis or clarity (Olipas and Luciano, 2024). To address these issues, this research focuses on developing an Automated Test Item Analysis System utilizing Optical Mark Recognition (OMR) and the assessment made by the IT experts and end-users on the initial implementation of the said system. The said system is designed to create, analyze, and collect high-quality test items, helping teachers develop standardized quarter tests for a more efficient and reliable learning process. Recent studies have demonstrated the benefits of automated systems in test item analysis. For example, Kurdi et al. (2020) compared various methods for predicting multiple-choice question difficulty, finding that data-driven approaches can significantly enhance the quality and fairness of assessments. Similarly, AlKhuzaey et al. (2024) highlighted the importance of textual complexity as a predictor of question difficulty, which can be integrated into automated systems to improve item analysis. By leveraging OMR and recent advancements in automated test item analysis, the innovative system for test item analysis aims to reduce teachers' workload, improve test quality, and provide more reliable and actionable feedback for students.

2. Related works

Recent research on test item analysis and assessment systems has focused on developing more efficient and accurate methods for evaluating student performance and improving educational outcomes. A key area of study is the use of automated test item analysis systems, which leverage technologies such as Optical Mark Recognition (OMR), machine learning (ML), and artificial intelligence (AI) to streamline the process of analyzing test results. These systems provide immediate feedback on student performance, enabling educators to make data-driven decisions. For instance, an AI-based test item analysis system was shown to enhance the reliability and validity of assessments by identifying patterns in student responses and flagging poorly designed or biased questions (Khosravi et al., 2022).

Optical Mark Recognition (OMR) has gained prominence for its ability to automate the grading of multiple-choice tests. Studies such as that by Das et al. (2022) demonstrated the efficiency of OMR systems in handling large volumes of answer sheets, reducing the time and human error associated with manual grading processes. OMR-based systems can also generate detailed item analysis reports that help instructors identify test questions that are too difficult, too easy, or unclear, thus informing test design improvements (Akhtar and Kovacs, 2024). Recent developments in OMR technology as with the study of Tinh and Minh (2024) have integrated AI algorithms that improve accuracy in reading answer sheets because it uses a fast object detection model to detect markers effectively and then segment the answer sheet into smaller regions for the mark reader model to recognize the user's selections accurately.

Furthermore, web-based assessment systems are increasingly being adopted in educational institutions, allowing for more flexible and scalable testing environments. These platforms enable students to take assessments remotely while offering real-time feedback and performance analytics. Research by Ngqondi et al. (2021) highlighted the advantages of such systems, particularly their ability to accommodate large numbers of students while maintaining high levels of test security through features like randomized question banks and proctoring mechanisms. Additionally, test portals equipped with assessment analytics capabilities can provide insights into student learning patterns, which are instrumental in curriculum design and instructional strategies (Vashishth et al., 2024).

Another growing area of interest is the application of big data and learning analytics in assessment systems. These technologies allow educators to analyze vast amounts of assessment data to uncover trends in student performance, predict future outcomes, and identify at-risk students early (Alalawi et al., 2024; Ikegwu et al., 2024). According to recent studies, integrating learning analytics into test item analysis can help educators pinpoint specific areas where students struggle, enabling personalized learning interventions that cater to individual student needs (Shoaib et al., 2024). Such systems also contribute to adaptive testing mechanisms, where test difficulty adjusts in real-time based on the student's performance, providing a more accurate measure of their abilities (Raj and Renumol, 2024).

Data security and privacy are crucial considerations in developing and deploying online assessment systems. Studies have emphasized the need for robust encryption methods and secure data storage solutions to protect sensitive student data from unauthorized access (Afridon et al., 2024). As test item analysis systems evolve, it is essential to ensure that data privacy regulations, such as the General Data Protection Regulation (GDPR), are adhered to, particularly when storing assessment results and student profiles on cloud-based platforms.

As a whole, recent literature on test item analysis and assessment systems highlights significant advancements in the use of OMR, AI, learning analytics, and cloud-based platforms. These innovations are enhancing the accuracy and efficiency of educational assessments while also providing valuable insights into student learning behaviors. However, ongoing attention to data privacy and system security is critical as these technologies continue to evolve and become more widespread in educational settings.

3. Methodology

This study adopts a sequential explanatory mixed methods design to develop and evaluate the App for

innovative test item analysis using Optical Mark Recognition (OMR). Initially, quantitative data is collected to assess the app's performance against ISO software quality assurance standards (ISO/IEC 25010), focusing on attributes like functionality, reliability, usability, and efficiency. This phase, according to Attia et al. (2024), provides objective metrics regarding the app's effectiveness. Subsequently, qualitative methods are employed to delve into user experiences and preferences, emphasizing customer interaction, feedback, and iterative adjustments throughout the software development process.

Qualitative data, gathered through interviews or surveys, explores user interactions and informs further app refinement based on iterative feedback loops. Tkalich et al. (2025) mentioned that thematic analysis integrates these qualitative insights with quantitative results, offering a comprehensive understanding of the app's development and users. By sequentially integrating quantitative and qualitative approaches, this mixed methods design ensures a robust evaluation framework for enhancing the innovative capabilities of the test item analysis app.

As regards, assessment of system performance based on ISO/IEC 25010 standards and user acceptability. Data will be gathered on attributes such as functionality, reliability, availability, efficiency, security, maintainability, and portability. The collected data will be organized in a database or spreadsheet for analysis.

Descriptive statistics such as frequency count, percentage, and weighted mean will be used to summarize the results. The frequency count will show how often each rating was given, the percentage will show the distribution of responses, and the weighted mean will show the overall average score that reflects how participants perceive the quality of the system.

The results will be interpreted using a 4-point Likert scale. A score between 3.35 and 4.00 means the system is highly functional, reliable, available, efficient, secure, maintainable, portable, and effective. A score between 2.25 and 3.24 means the system is functional, reliable, available, efficient, secure, maintainable, portable, and effective. A score between 1.75 and 2.24 means the system is slightly functional, reliable, available, efficient, secure, maintainable, portable, and effective. Lastly, a score between 1.00 and 1.74 means the system is not functional, reliable, available, efficient, secure, maintainable, portable, or effective.

This scale offers a clear framework for evaluating the system's strengths and weaknesses across various attributes. Higher scores indicate strong performance, while lower scores highlight areas for improvement. This analysis provides a comprehensive evaluation based on international standards and user feedback.

4. Results and discussion

This section presents the findings of the study and offers a comprehensive analysis of their implications, elucidating their contribution to the existing body of literature and addressing the research questions articulated at the outset.

4.1. Design and development process undertaken to create the innovative test analysis system using OMR

The app for innovative test item analysis with OMR was successfully completed by following the different phases of the Agile Model. Each phase is described and explained below. Defining requirements: This initial phase involved planning the components for the upcoming work in collaboration with various stakeholders. Priorities and specifications were established during this stage. The researcher and her former thesis advisee conducted on-site meetings and interviews with the Senior High School (SHS) faculty and school head who are employed at the locale of the study. As part of the agreement, a letter of request and a series of questions were sent to the clients to gather their perspectives on the system's functionality. This process helped identify the issues and challenges encountered during the quarterly assessment and evaluation of SHS learners. A thorough review of current assessment policies, guidelines, rules, regulations, forms, and reports mandated by the Department of Education (DepEd) was also conducted. Table 1 lists the problems and difficulties identified in the current process, along with the proposed solutions.

Table 1: Problems encountered in the current process and proposed solutions

Transaction workflow	Problems and difficulties encountered	Solutions
Examination process	• Time-consuming manual checking of test papers • Susceptibility to human errors • Compromised accuracy of test scores due to student cheating	Use of bubble sheets and optical mark recognition (OMR) to automate test paper checking
Generation of test item analysis	• Manual input of test results in MS Excel • Reuse of test questions despite the generated analysis • Rejected test items are not automatically deleted, compromising test quality	Development of an automated test item analysis system using optical mark recognition to expedite test item analysis
Test bank	• Test questions stored in MS Word without a manageable database	Integration of a comprehensive test item banking system to easily manage and locate test items

Following the interviews, it was evident that the manual checking of test papers and the subsequent test item analysis were time-consuming tasks, creating a significant burden for SHS teachers, especially those handling multiple subjects outside their specialization. The use of manual checking

methods, sometimes involving students, increased the risk of inaccuracies and cheating. Several studies have documented these challenges, highlighting how traditional methods of assessment, which often rely on manual processes, introduce risks such as miscalculations, delayed feedback, and even opportunities for student dishonesty (Braun et al., 2023). Additionally, Fagbohun et al. (2024) in their research mentioned that by eliminating manual steps in test scoring, educators can focus on more critical teaching tasks and ensure that the evaluation process remains objective and transparent.

System design: In this phase, design specifications were determined for the development of the interface, software, and database. The system design was based on approved guidelines and agreed-upon software and hardware specifications between the researchers and stakeholders. This stage included the creation of flow diagrams to demonstrate the system's functional design. These diagrams are effective in illustrating the logical flow of data and operations within the system, providing stakeholders with a clear understanding of how the system will function upon implementation. These visual representations, according to Masegosa et al. (2024), also aid in identifying potential design issues early in the development process, allowing for adjustments before coding begins.

Technical specifications: The system is comprised of several components, including software, hardware, and end-users.

a) Software Specification: The App for innovative test item analysis employs a range of web-based software components and frameworks to facilitate efficient user interaction with the system. These include HTML 5 for structuring web pages, PHP for server-side scripting, MySQL for database management, jQuery 1.10.2 for simplifying JavaScript programming, Bootstrap 3.1.1 for responsive web design, and JavaScript for client-side scripting. According to Igbokwe (2023), the system can ensure compatibility across various web browsers such as Google Chrome, Mozilla Firefox, and Microsoft Edge to enhance accessibility and usability for users.
b) Hardware Specification: The hardware requirements of the software app for innovative test item analysis encompass essential components necessary for input, processing, output, and storage functionalities. These include Personal Computers (PCs) and laptops for accessing and using the system, printers and scanners for document processing and data input, and a Local Area Network (LAN) for connecting multiple devices within a restricted area. The system operates efficiently with a minimum specification of a Pentium IV or higher processor, at least 2GB of memory, and 5GB of free storage space. It is compatible with major operating systems such as Windows, Linux, and Mac OS, ensuring flexibility and accessibility across different computing platforms (Swiecki et al., 2022).

Table 2 presents the minimum hardware and software specifications required for personal computers and laptops.

Table 2: System requirements for PC and laptop

Hardware	Minimum specification	Software	Minimum specification
Processor	Pentium IV or higher	Operating system	Windows, Linux, Mac OS
Memory	At least 2GB or higher	Web browser	Google Chrome, Mozilla Firefox, Microsoft Edge
Storage	5GB of free space	MS office	2007 or higher

End-users' description of the roles and functionalities accessible to authorized users can be summarized as follows:

- Administrator. Can perform various system functionalities after authentication, such as updating student and school information, generating test item analysis and examination, and managing user accounts.
- Teacher. Can access functionalities like generating examinations, creating test questions, promoting students, and importing test scores.

The Administrator, after authentication, has full control over the system, including updating school and student information, managing user accounts, generating test item analyses, and overseeing examination processes. On the other hand, the teacher's role focuses on classroom-related functionalities, such as creating and generating examinations, formulating test questions, promoting students, and importing test scores. This role-based access ensures that each user interacts with the system according to their responsibilities, promoting security and operational efficiency in managing assessments.

Through the login process, administrators and teachers can access the system and perform their respective functions using authenticated credentials. The design ensures that only authorized users can access the application.

Procedural design: Procedural design in the app for innovative test item analysis translates structural elements into detailed procedural explanations, as depicted in various diagrams below:

- OMR form scanner: The OMR Form Scanner process involves several steps depicted in Fig. 1. Users begin by accessing the OMR Form Scanner and creating a template within the system. They then select images and scan them, with scanned data automatically saved as an Excel file. To analyze the scanned OMR data for item difficulty

and metrics, users copy the Excel file into a designated bubble sheet Excel file.

- Users module: Fig. 1 outlines the functionalities accessible through the user or teachers' module. Upon logging in, teachers can view profiles, access students' test results, generate test item analysis reports, manage the test bank, create examinations, import reports and scores, update contact information, and modify passwords. The session ends upon logging out.
- Admin module: Fig. 1 also illustrates the flow of processes on the Admin's Module. Admins begin by logging in and verifying their account. They can access and update site details, manage teacher and student data, assign subjects, view student lists, manage test banks, analyze test items, and perform other administrative functions (Delgado et al., 2015). The system terminates upon logging out.

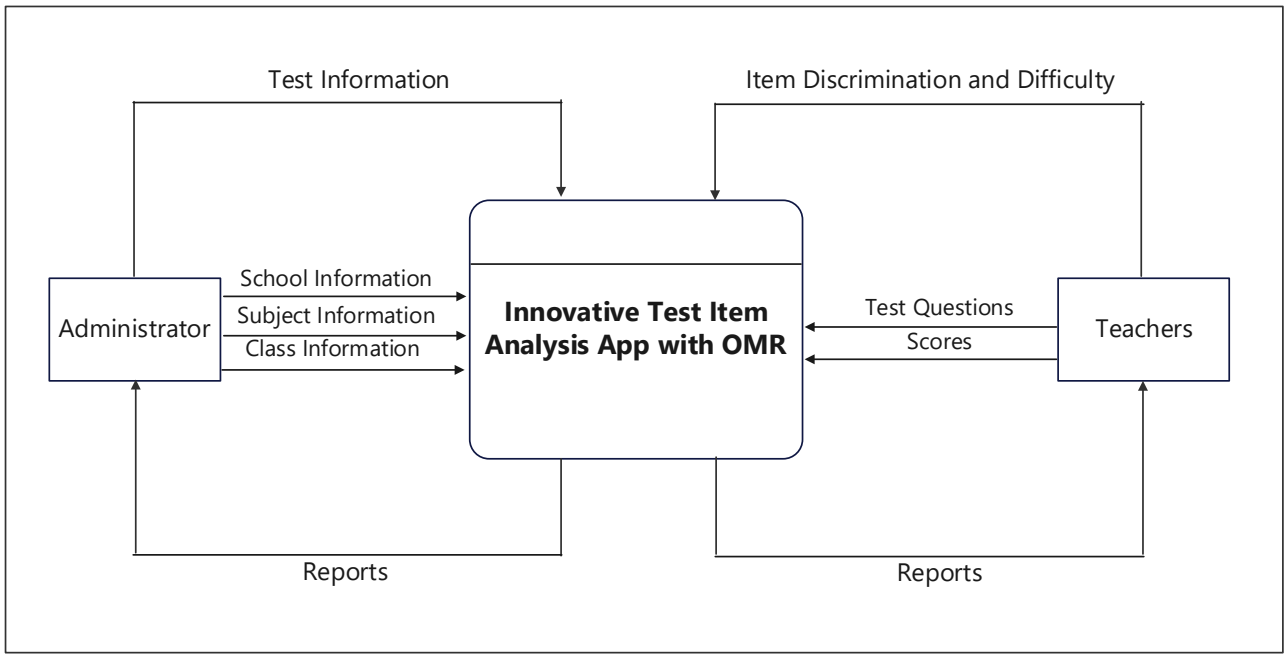

Fig. 1: Process flow on the admin module

Software design: Software design transforms user requirements into coding and implementation strategies. Using Data Flow Diagrams (DFD), the system's functionalities and data flow are graphically represented.

Interface design: Interface design focuses on the aesthetic and functional aspects of the graphical user interface (GUI), ensuring ease of use. Elements such as buttons, navigation keys, menus, and inputs facilitate user interaction. The interface for the app for innovative test item analysis was developed using Bootstrap, HTML 5, CSS, and jQuery, ensuring responsiveness across devices (Pushpakumar et al., 2023). The interface design prioritizes both aesthetic appeal and functionality, ensuring that users can easily navigate the system. It incorporates elements like buttons, menus, and input fields that facilitate seamless user interaction. By using the above-mentioned tools, the interface ensures responsiveness across various devices, offering a consistent user experience regardless of the platform.

Development: The development phase involved coding the system using PHP for scripting, JavaScript and JQuery for programming, and Bootstrap, CSS, and HTML 5 for the front-end interface. MySQL managed the database, ensuring cross-platform compatibility and responsiveness in a web-based environment (Vassallo et al., 2019). During this phase, the system was coded using multiple programming tools, ensuring the interface was modern, responsive, and easy to use, while MySQL handled the database management to ensure cross-platform compatibility in a web-based environment.

User Acceptance Testing: User acceptance testing involved end-users accessing the system via LAN, logging in, and evaluating its features. Detected bugs and errors were resolved, and user feedback was incorporated before system release (Davis and Venkatesh, 2004). In this phase, end-users, primarily teachers and administrators, accessed the system via a local area network (LAN) to test its features. Any bugs or issues encountered were addressed, and feedback from users was incorporated into the final adjustments to ensure the system met user expectations before its official release.

Releasing: The final stage included system deployment in the SHS Department, orientation and training for end-users, and evaluation based on ISO 25010 Software Product Quality Standards (Gupta and Gayathri, 2022). A user manual was provided to guide users, ensuring seamless integration and usability. The final stage of system deployment involved rolling out the software to the Senior High School Department, conducting orientation and training sessions for end-users, and evaluating the system based on ISO 25010 Software Product Quality Standards. A user manual was also provided

to ensure that users could operate the system effectively and integrate it smoothly into their workflow.

Generally speaking, this system stands out from other available applications and previous studies due to its advanced integration of OMR technology, automated test item analysis, and comprehensive test item management. Unlike traditional systems that rely on manual data entry or scattered files, it streamlines the entire process by automating the scanning, analysis, and reporting of test results, significantly reducing time and errors. Its role-based access ensures that teachers and administrators have distinct functionalities tailored to their specific needs, enhancing both security and usability.

Additionally, the system's ability to export OMR data into Excel files for further analysis provides users with greater flexibility, allowing for more in-depth manipulation of the data as needed. These features, combined with their user-friendly interface and scalable design, offer a more efficient and comprehensive solution for managing assessments and improving the quality of test items.

4.2. Evaluation of the app for innovative test item analysis using OMR based on ISO 25010/IEC 25010 criteria by IT experts

Alem (2020) inferred that data interpretation involves the process through which data are reviewed to reach an informed conclusion. It assigns meaning to the information analyzed and determines its significance and implications. The interpretation and analysis of data for the evaluation of the App for innovative test item analysis using Optical Mark Recognition (OMR) by IT experts aim to make sense of the numerical data collected, analyzed, and presented.

Functional suitability: Functional suitability measures the degree to which a system provides the functions that meet stated and implied needs when used under specified conditions. This characteristic includes functional completeness, functional correctness, and functional appropriateness.

The three sub-categories under functional suitability received an overall weighted mean rating of 3.80 (very functional). Functional completeness is rated 3.80, indicating that the system meets all specified goals and objectives. Functional correctness rated 3.60, confirming the system's ability to provide accurate results. Functional appropriateness received a 4.00 rating, showing that the system facilitates the accomplishment of specified tasks and objectives. The results imply that the system provides all necessary features for the operation of the SHS Department for test item analysis, successfully automating the test item analysis process and meeting stated and implied needs under specified conditions. According to ISO/IEC 25010:2011, functional suitability is crucial for determining the overall effectiveness of a software system in meeting user requirements. This also means that the was evaluated as highly functional by IT experts, demonstrating strong performance across functional completeness, correctness, and appropriateness. It successfully meets all specified goals, provides accurate results, and facilitates the accomplishment of tasks, particularly in automating the test item analysis process. This indicates that the system effectively addresses the needs of the SHS Department and operates efficiently under the given conditions.

Performance efficiency: Performance efficiency represents the performance of the app for innovative test item analysis using OMR relative to the number of resources used under stated conditions. This characteristic includes time behavior, resource utilization, and capacity. The performance efficiency obtained a weighted mean of 3.93 (Very Efficient). Time behavior rated 3.80, indicating the system's ability to respond and process user requests efficiently. Resource utilization and capacity both rated 4.00, demonstrating that the system provides accurate results with minimal resources and can handle user requests without difficulty.

The overall rating of 3.93 indicates that the system is very efficient, capable of processing information simultaneously without delays or technical issues. Efficient software performance is essential for enhancing user satisfaction and productivity. The system's high efficiency rating reflects its ability to process tasks quickly and handle simultaneous operations without experiencing slowdowns or technical issues. This level of performance is crucial for maintaining smooth workflows, reducing downtime, and ultimately increasing both user satisfaction and productivity in managing test item analysis.

Compatibility: Compatibility is the degree to which the app for innovative test item analysis using OMR can exchange information with other products, systems, or components while sharing the same hardware or software environment. This characteristic includes co-existence and interoperability. A quantitative rating of 3.80 for co-existence suggests that the application functions effectively alongside other applications in a shared environment. Interoperability rated 4.00, signifying that the system's resources can be used by other applications simultaneously without affecting functionality. The overall mean of 3.90 implies high compatibility'.

The high compatibility rating indicates that the system can be used simultaneously with other applications and share the same hardware or software environment without performance issues. This aligns with findings that emphasize the importance of compatibility in ensuring seamless integration and operation across different systems (van Berkel et al., 2020). The system's high compatibility rating means it can function alongside other applications without causing conflicts or performance degradation, ensuring smooth integration into existing workflows. This capability is essential for institutions that rely on multiple software solutions, as it enables efficient

multitasking and interoperability, enhancing overall system effectiveness.

Usability: Usability measures the degree to which specified users can achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. This characteristic includes appropriateness, recognizability, learnability, operability, user error protection, user interface, and accessibility. The usability was rated 3.76 (Very Usable). Appropriateness, recognizability, learnability, operability, user interface, and accessibility all rated 3.80, while user error protection rated 3.40.

The overall usability rating of 3.76 signifies that the system is user-friendly, easy to learn, and effective in achieving specified goals. Usability is a critical factor in software adoption and user satisfaction. This high level of usability is essential for promoting software adoption and enhancing user satisfaction, as it ensures that users can achieve their goals efficiently while minimizing frustration and errors during use.

Reliability: Reliability measures the degree to which the app for innovative test item analysis using OMR performs specified functions under specified conditions for a specified period. This characteristic includes maturity, availability, fault tolerance, and recoverability. The reliability was rated 3.75 (Very Reliable). Maturity and availability were rated 4.00, fault tolerance rated 3.40, and recoverability rated 3.60.

The high reliability rating indicates that the system can be trusted to perform consistently over time, even in the presence of faults. Reliable software systems are essential for maintaining operational stability and users' trust. With high scores in maturity and availability, the system demonstrates robust performance and readiness for use in real-world conditions. This level of reliability is crucial for ensuring operational stability and fostering user trust, as it reassures users that they can depend on the system for accurate and timely test item analysis.

Security: Security measures the degree to which the app for innovative test item analysis using OMR protects information and data so that access is appropriate to authorization levels. This characteristic includes confidentiality, integrity, non-repudiation, accountability, and authenticity. The security was rated 3.64 (Very Secure). Confidentiality rated 4.00, integrity rated 3.80, non-repudiation and accountability both rated 3.40, and authenticity rated 3.60.

The overall security rating of 3.64 indicates that the system effectively protects data and maintains integrity, preventing unauthorized access and modifications. Security is paramount in protecting sensitive information and ensuring compliance with data protection regulations (Saarela et al., 2017). With strong scores in confidentiality and integrity, the system demonstrates a robust framework for protecting sensitive information while ensuring that access aligns with user authorization levels. This focus on security is essential for protecting user data and complying with data protection regulations, thereby reinforcing user confidence in the system's reliability and safety.

Maintainability: Maintainability measures the degree of effectiveness and efficiency with which the app for innovative test item analysis using OMR can be modified to improve, correct, or adapt to changes in the environment and requirements. This characteristic includes modularity, reusability, analyzability, modifiability, and testability. The maintainability was rated 3.64 (Very Maintainable). Modularity rated 3.87, reusability rated 3.8, analyzability rated 3.40, and both modifiability and testability rated 3.60.

The overall maintainability rating of 3.64 indicates that the system can be easily updated and modified, ensuring continued performance and adaptability. Maintainable systems are crucial for long-term sustainability and cost-effectiveness. High scores in modularity and reusability indicate that the system's components can be efficiently repurposed or adjusted, facilitating ongoing improvements. This level of maintainability is vital for ensuring the system's long-term sustainability and cost-effectiveness, as it reduces the need for extensive overhauls and supports continuous enhancement in response to user needs.

Portability: Portability measures the degree of effectiveness and efficiency with which the app for innovative test item analysis using OMR can be transferred from one environment to another. This characteristic includes adaptability, installability, and replaceability. The portability was rated 3.80 (very portable). Adaptability, installability, and replaceability all rated 3.80.

The high portability rating indicates that the system can be used in various environments and is adaptable to hardware and software upgrades. Portability enhances the system's versatility and usability across different platforms (Nayebi et al., 2012). With equally high scores in adaptability, installability, and replaceability, the system can be easily integrated into new settings and updated as necessary. This level of portability enhances the system's versatility, ensuring that it remains functional and user-friendly across diverse platforms and technology landscapes, ultimately improving user accessibility and experience.

Fig. 2 summarizes the evaluation of the IT experts on the app for innovative test item analysis based on ISO/IEC 25010 Software Product Quality Standards. The results indicate that the system complies with various software quality criteria, suggesting its high suitability for implementation in the DepEd Senior High School, particularly at the research locale.

The results highlight the system's alignment with essential software quality criteria, such as functional suitability, usability, reliability, security, maintainability, and portability. This strong compliance indicates that the system not only meets the necessary performance and operational standards but also demonstrates a high degree of suitability for implementation within the

Department of Education's Senior High School, particularly in the specific research locale. Overall, these findings suggest that the system is well-equipped to enhance the efficiency and effectiveness of test item analysis processes in educational settings.

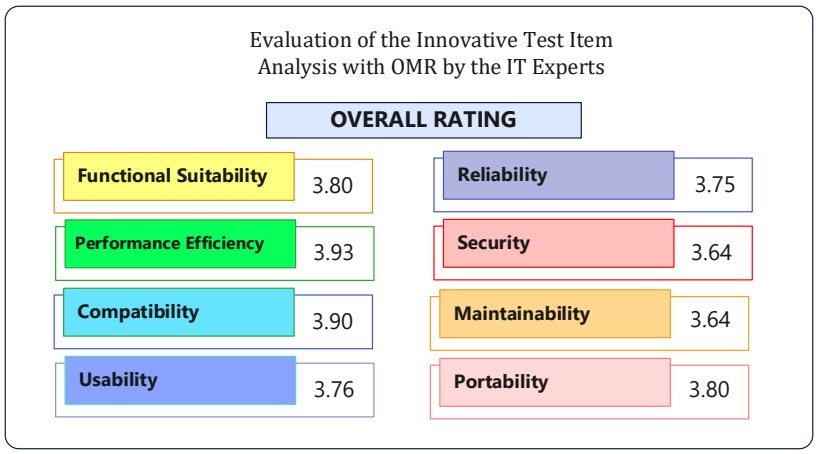

Fig. 2: Summary of evaluation made by the IT experts

4.3. Users' evaluation of innovative test analysis system using OMR based on ISO 25010 software product quality standards

End-users, primarily teachers, evaluated the app for innovative test item analysis using OMR based on functional suitability, performance efficiency, and usability.

Functional suitability: The functional suitability received an overall weighted mean rating of 3.80 (very functional), indicating the system meets all specified user requirements.

The high rating suggests that the system is functionally suitable for end-users' needs, effectively automating the manual process of evaluating students. User satisfaction with functional suitability is crucial for successful system adoption. End-users, primarily teachers, evaluated the app for innovative test item analysis using OMR across key dimensions of functional suitability, performance efficiency, and usability. The system achieved an overall weighted mean rating of 3.80 (very functional) for functional suitability, indicating that it successfully meets all specified user requirements. This high rating reflects the system's effectiveness in automating the previously manual process of student evaluation, enhancing the efficiency of test analysis. Such user satisfaction regarding functional suitability is essential for fostering successful adoption and sustained use of the system within educational settings.

Performance efficiency: The performance efficiency received an overall rating of 3.67 (Very Efficient), indicating efficient resource utilization and capability to handle user requests.

Efficient performance reduces the time required for tasks, enhancing user productivity and satisfaction. Performance efficiency is a key determinant of user acceptance and system success (Shneiderman et al., 2016).

Usability: The usability received an overall rating of 3.73 (Very Usable), indicating the system is user-friendly and easy to learn.

High usability ensures that end-users can effectively and efficiently use the system, leading to higher satisfaction and productivity. Usability is a critical factor in the overall user experience and system success (Krug, 2013). The app for innovative test item analysis using OMR received an overall rating of 3.67 (Very Efficient) for performance efficiency. This rating indicates that the system effectively utilizes resources and can efficiently handle user requests, which contributes to a smoother operational experience. Enhanced performance efficiency minimizes the time required to complete tasks, thereby increasing user productivity and satisfaction, making it a crucial factor for user acceptance and the overall success of the system. Summary of users' evaluation: Fig. 3 summarizes the evaluation by end-users, indicating high acceptance, adaptability, and integration readiness. The results suggest that system is very functional, efficient, and usable, making it highly recommended for use in DepEd Senior High School, particularly in one national high school in the schools division of Nueva Ecija.

These positive assessments indicate strong acceptance and adaptability of the system, along with its readiness for integration into existing educational frameworks. Consequently, the results suggest that the app is highly recommended for implementation in DepEd Senior High Schools, particularly within a national high school in the schools division of Nueva Ecija, highlighting its potential to enhance educational processes.

4.4. Level of effectiveness of system implementation

Implementation involves executing a plan or design to ensure proper operation in its environment. Effective implementation is critical for ensuring that systems perform as intended and meet user requirements (Haleem et al., 2022). According to Delone and McLean (2003), user satisfaction is a key indicator of successful system implementation

and overall effectiveness. Table 3 below shows the data on the evaluation of the system's implementation effectiveness.

The system received an overall rating of 3.86 (highly effective), indicating successful implementation and user satisfaction. This rating reflects that the system meets user requirements and provides significant advantages for test item analysis, improving the overall evaluation process. These findings are consistent with previous research highlighting that well-implemented systems not only fulfill user needs but also enhance operational efficiency (Venkatesh et al., 2003).

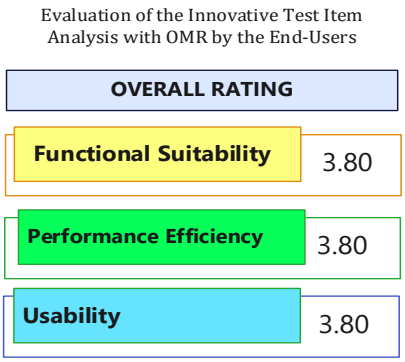

Fig. 3: Summary of the end-users' evaluation

Table 3: Evaluation of the effectiveness of the system's implementation

Item statements	Weighted mean	Verbal description
1. The system is easy to use	4.00	Highly effective
2. The system has recognizable icons	3.80	Highly effective
3. The system can accurately deliver user commands	3.80	Highly effective
4. The system provides easy access to reports and generates reports easily	3.80	Highly effective
5. Username and password are used as access passes for authorized users	3.80	Highly effective
6. The application can be accessed using any device	4.00	Highly effective
7. Mistakes in using the system can be easily recovered	3.60	Effective
8. The organization of the information in the application interface screens is clear	3.80	Highly effective
9. The system reduces the time consumed in checking examinations and generating test item analysis	4.00	Highly effective
10. The system has all the necessary functions it should have for test item analysis	4.00	Highly effective
Mean	3.86	Highly effective

The findings also indicate that the system successfully meets user requirements and enhances the evaluation process, as reflected in the high ratings for various functionalities, including ease of use, recognizable icons, accurate command delivery, and efficient report generation. Notably, the ability to access the application on any device and its effectiveness in reducing the time required for checking examinations further contribute to user satisfaction.

5. Conclusion

The development and implementation of an innovative test item analysis system utilizing Optical Mark Recognition (OMR) technology has shown significant promise in enhancing the efficiency and accuracy of educational assessments. This study highlights the critical need for automated assessment tools in modern educational settings, addressing the inefficiencies and limitations of traditional manual methods. The new system successfully automates the creation, analysis, and management of test items, significantly reducing teachers' workload and improving the quality of assessments. Evaluations based on ISO/IEC 25010 software quality standards by IT experts confirm the system's high performance across various criteria, including functional suitability, performance efficiency, compatibility, usability, reliability, and security. The system demonstrates robust functionality, efficient performance, and high compatibility with existing educational tools, ensuring seamless integration and user satisfaction. Furthermore, the innovative system's ability to provide accurate and reliable assessments contributes to better educational outcomes by offering actionable feedback to students and educators.

To maximize the benefits, it is recommended that the system be widely implemented across educational institutions with comprehensive training programs to ensure effective use. Continuous improvement and regular updates should be prioritized to enhance functionality and incorporate new features. Integration with curriculum development will ensure alignment with learning objectives, creating assessments that accurately measure student learning. Future development should expand the system to accommodate diverse assessment types, ensuring a comprehensive solution for educational needs. Robust data privacy and security measures must be implemented to protect sensitive information, and collaborative efforts between educational institutions, developers, and policymakers should be encouraged to refine and expand the system.

5.1. Further studies

In light of the promising results from the evaluation of the app for innovative test item analysis using OMR, further studies could explore several avenues to enhance its impact and utility. First, future research could investigate the long-term effects of implementing the system on teachers' efficiency and student outcomes, particularly in various educational contexts and diverse subjects.

Additionally, studies could focus on the integration of advanced analytics, such as machine learning algorithms, to provide deeper insights into student performance and item effectiveness.

Exploring user experience in different educational settings could also yield valuable feedback for refining the user interface and functionality, ensuring that the system remains intuitive and accessible to all educators.

Finally, assessing the scalability of the application in larger educational institutions or across multiple schools could provide insights into its adaptability and potential for broader implementation, supporting a more widespread shift toward innovative assessment practices in education.

Compliance with ethical standards

Ethical considerations

The study involved human participants in the form of interviews, surveys, and user testing. Participation was voluntary, and informed consent was obtained from all participants. All data collected were treated with strict confidentiality and were used solely for research purposes. No personally identifiable information was recorded or disclosed. The study followed ethical standards in accordance with the institutional research guidelines of Nueva Ecija University of Science and Technology.

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Afridon M, Tedyyana A, Ratnawati F, Julianto A, and Faizi MN (2024). Optimizing data security in computer-assisted test applications through the advanced encryption standard 256-bit cipher block chaining. International Journal of Advanced Computer Science and Applications, 15(8): 163-170. https://doi.org/10.14569/IJACSA.2024.0150817

Akhtar H and Kovacs K (2024). Measurement precision and user experience with adaptive versus non-adaptive psychometric tests. Personality and Individual Differences, 225: 112675. https://doi.org/10.1016/j.paid.2024.112675

Alalawi K, Athauda R, Chiong R, and Renner I (2024). Evaluating the student performance prediction and action framework through a learning analytics intervention study. Education and Information Technologies, 30: 2887–2916. https://doi.org/10.1007/s10639-024-12923-5

Alem DD (2020). An overview of data analysis and interpretations in research. International Journal of Academic Research in Education and Review, 8(1): 1-27.

AlKhuzaey S, Grasso F, Payne TR, and Tamma V (2024). Text-based question difficulty prediction: A systematic review of automatic approaches. International Journal of Artificial Intelligence in Education, 34: 862–914. https://doi.org/10.1007/s40593-023-00362-1

Attia S, El-Degwy A, and Attia M (2024). Usability and fitness testing for building performance simulation tools. Journal of Building Performance Simulation, 17(4): 460-479. https://doi.org/10.1080/19401493.2024.2341089

Braun J, Schläfli R, Schmocker D, and Wilding B (2023). Experiences and critical reflection on online-assessment with Excel case studies: Review on a successful online-assessment practice as well as the adaptation to a remote setting due to the COVID-19 pandemic. In: Hummel S and Donner MT (Eds.), Student assessment in digital and hybrid learning environments: 9-36. Springer Fachmedien Wiesbaden, Wiesbaden, Germany. https://doi.org/10.1007/978-3-658-42253-0_2

Das B, Majumder M, Sekh AA, and Phadikar S (2022). Automatic question generation and answer assessment for subjective examination. Cognitive Systems Research, 72: 14-22. https://doi.org/10.1016/j.cogsys.2021.11.002

Davis FD and Venkatesh V (2004). Toward preprototype user acceptance testing of new information systems: Implications for software project management. IEEE Transactions on Engineering Management, 51(1): 31-46. https://doi.org/10.1109/TEM.2003.822468

de Guzman AB and Adamos JL (2020). Like the layers of an onion: Curricular noticing as a lens to understand the epistemological features of the Philippine K to 12 secondary mathematics curriculum materials. Educational Research for Policy and Practice, 19(3): 389-409. https://doi.org/10.1007/s10671-020-09264-8

Delgado AJ, Wardlow L, McKnight K, and O'Malley K (2015). Educational technology: A review of the integration, resources, and effectiveness of technology in K-12 classrooms. Journal of Information Technology Education: Research, 14: 397-416. https://doi.org/10.28945/2298

DeLone WH and McLean ER (2003). The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems, 19(4): 9-30. https://doi.org/10.1080/07421222.2003.11045748

Fagbohun O, Iduwe NP, Abdullahi M, Ifaturoti A, and Nwanna OM (2024). Beyond traditional assessment: Exploring the impact of large language models on grading practices. Journal of Artificial Intelligence and Machine Learning and Data Science, 2(1): 1-8. https://doi.org/10.51219/JAIMLD/oluwole-fagbohun/19

Gupta S and Gayathri N (2022). Study of the software development life cycle and the function of testing. In the International Interdisciplinary Humanitarian Conference for Sustainability, IEEE, Bengaluru, India: 1270-1275. https://doi.org/10.1109/IIHC55949.2022.10060231 **PMid:35048294 PMCid:PMC8768444**

Haleem A, Javaid M, Qadri MA, and Suman R (2022). Understanding the role of digital technologies in education: A review. Sustainable Operations and Computers, 3: 275-285. https://doi.org/10.1016/j.susoc.2022.05.004

Igbokwe IC (2023). Application of artificial intelligence (AI) in educational management. International Journal of Scientific and Research Publications, 13(3): 300-307. https://doi.org/10.29322/IJSRP.13.03.2023.p13536

Ikegwu AC, Nweke HF, and Anikwe CV (2024). Recent trends in computational intelligence for educational big data analysis. Iran Journal of Computer Science, 7(1): 103-129. https://doi.org/10.1007/s42044-023-00158-5

Khosravi H, Shum SB, Chen G, Conati C, Tsai YS, Kay J, Knight S, Martinez-Maldonado R, Sadiq S, and Gašević D (2022). Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 3: 100074. https://doi.org/10.1016/j.caeai.2022.100074

Krug S (2013). Don't make me think, revisited: A common-sense approach to web usability. 3rd Edition, New Riders, Indianapolis, USA.

Kurdi G, Leo J, Parsia B, Sattler U, and Al-Emari S (2020). A systematic review of automatic question generation for educational purposes. International Journal of Artificial Intelligence in Education, 30: 121-204. https://doi.org/10.1007/s40593-019-00186-y

Masegosa AR, Cabañas R, Maldonado AD, and Morales M (2024). Learning styles impact students' perceptions on active learning methodologies: A case study on the use of live coding

and short programming exercises. Education Sciences, 14(3): 250. https://doi.org/10.3390/educsci14030250

Nayebi F, Desharnais JM, and Abran A (2012). The state of the art of mobile application usability evaluation. In the 25th IEEE Canadian Conference on Electrical and Computer Engineering, IEEE, Montreal, Canada: 1-4. https://doi.org/10.1109/CCECE.2012.6334930

Ngqondi T, Maoneke PB, and Mauwa H (2021). A secure online exams conceptual framework for South African universities. Social Sciences and Humanities Open, 3(1): 100132. https://doi.org/10.1016/j.ssaho.2021.100132

Olipas CNP and Luciano RG (2024). Analyzing test performance of BSIT students and question quality: A study on item difficulty index and item discrimination index for test question improvement. International Journal of Information Technology and Computer Science, 16(3): 1-11. https://doi.org/10.5815/ijitcs.2024.03.01

Pushpakumar R, Sanjaya K, Rathika S, Alawadi AH, Makhzuna K, Venkatesh S, and Rajalakshmi B (2023). Human-computer interaction: Enhancing user experience in interactive systems. In the International Conference on Newer Engineering Concepts and Technology: E3S Web of Conferences, EDP Sciences, Paris, France, 399: 04037. https://doi.org/10.1051/e3sconf/202339904037

Raj NS and Renumol VG (2024). An improved adaptive learning path recommendation model driven by real-time learning analytics. Journal of Computers in Education, 11(1): 121-148. https://doi.org/10.1007/s40692-022-00250-y **PMCid:PMC9748379**

Saarela M, Hosseinzadeh S, Hyrynsalmi S, and Leppänen V (2017). Measuring software security from the design of software. In the 18th International Conference on Computer Systems and Technologies, Association for Computing Machinery, Ruse, Bulgaria: 179-186. https://doi.org/10.1145/3134302.3134334

Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, and Diakopoulos N (2016). Designing the user interface: Strategies for effective human-computer interaction. 6th Edition, Pearson, London, UK.

Shoaib M, Sayed N, Singh J, Shafi J, Khan S, and Ali F (2024). AI student success predictor: Enhancing personalized learning in campus management systems. Computers in Human Behavior, 158: 108301. https://doi.org/10.1016/j.chb.2024.108301

Swiecki Z, Khosravi H, Chen G, Martinez-Maldonado R, Lodge JM, Milligan S, Selwyn N, and Gašević D (2022). Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3: 100075. https://doi.org/10.1016/j.caeai.2022.100075

Tinh PD and Minh TQ (2024). Automated paper-based multiple choice scoring framework using fast object detection algorithm. International Journal of Advanced Computer Science and Applications, 15(1): 1174-1181. https://doi.org/10.14569/IJACSA.2024.01501115

Tkalich A, Klotins E, Sporsem T, Stray V, Moe NB, and Barbala A (2025). User feedback in continuous software engineering: revealing the state-of-practice. Empirical Software Engineering, 30: 79. https://doi.org/10.1007/s10664-024-10557-2

van Berkel N, Goncalves J, Wac K, Hosio S, and Cox AL (2020). Human accuracy in mobile data collection. International Journal of Human-Computer Studies, 137: 102396. https://doi.org/10.1016/j.ijhcs.2020.102396

Vashishth TK, Sharma V, Sharma KK, Kumar B, Panwar R, and Chaudhary S (2024). AI-driven learning analytics for personalized feedback and assessment in higher education. In: Nguyen TV and Vo N (Eds.), Using traditional design methods to enhance AI-driven decision making: 206-230. IGI Global, Pennsylvania, USA. https://doi.org/10.4018/979-8-3693-0639-0.ch009

Vassallo K, Garg L, Prakash V, and Ramesh K (2019). Contemporary technologies and methods for cross-platform application development. Journal of Computational and Theoretical Nanoscience, 16(9): 3854-3859. https://doi.org/10.1166/jctn.2019.8261

Venkatesh V, Morris MG, Davis GB, and Davis FD (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3): 425-478. https://doi.org/10.2307/30036540

Yan Z, Brown GT, Lee JCK, and Qiu XL (2020). Student self-assessment: Why do they do it? Educational Psychology, 40(4): 509-532. https://doi.org/10.1080/01443410.2019.1672038