https://doi.org/10.32815/jitika.v19i2.1189

Volume 19, Issue 2, 2025

Jurnal Ilmiah Teknologi Informasi Asia

Journal Homepage: https://jurnal.asia.ac.id/index.php/jitika

Article

Development of a TOEFL test website for the "Plug-in" Training Institute

Arya Bima Wijaya *, Yohana Tri Widayati, Satrio Agung Prakoso

Teknik Informatika, Fakultas Teknik dan Informatika, Universitas AKI Semarang, Semarang, 50139, Indonesia

***Abstract*—The TOEFL test is an international standard for measuring English language ability. This research developed a web-based TOEFL examination system at the "Plug-in" Training Institute using the Scrum method to address the shortcomings of the old system, such as audio control, score accuracy, and insufficient detail in recording results. The data were collected through observation, interviews, and literature study. Scrum was applied in three sprints that focused on the listening section navigation, score calculation, and answer logging. Features for single audio playback, a timer, automatic score calculation, and a detailed results display were successfully implemented. Testing showed the system to be stable and accurate according to TOEFL standards. This system improves efficiency, fairness, and user experience in digital TOEFL examinations. Suggestions for development include broader testing, UI/UX optimization, a data export feature, and more detailed answer logging for analysis.**

***Keywords*—audio playback control; score calculation; scrum method; TOEFL test system; web-based examination.**

1. Introduction

The Test of English as a Foreign Language (TOEFL) is one of the international standard methods used to measure English language proficiency. With the increasing need for this test among students and professionals, a digital system is required to support flexible and efficient test administration.

In the current digital era, information technology has changed many aspects of life, including the field of education. According to Febriansyah & Voutama (2024), information technology has influenced many areas of life, including education, in an expanding digital age. This advancement affects the education sector with impacts such as easier access and reduced paper usage, which are relevant solutions for the current situation.

In line with this, Khazanah & Purnama (2024) also state that traditional examination processes often face several obstacles, such as difficulties in administration, creation of varied questions, and a complicated and time-consuming assessment process.

Various previous studies have discussed the importance of digitalization in examination administration. Sanjaya et al. (2025) state that time-consuming traditional examination systems can be replaced by more modern digital platforms, emphasizing the need for an online examination system that increases efficiency, transparency, and accountability (Pauzi & Kharisma, 2025). Meanwhile, Abdi et al. (2024) show the importance of security and fairness in web-based TOEFL examinations through the implementation of question randomization algorithms, such as Fisher-Yates, to prevent cheating and improve fairness among participants.

However, although various online examination systems have been developed, gaps still exist between user needs and the performance of available systems. The old system used by the "Plug-in" Training Institute—the institution where this research was conducted—for its web-based TOEFL examination showed a number of significant constraints (Maulana & Lubis, 2022). One of the main problems was the system's inability to

* Corresponding author.
E-mail Address: aryabimaw13@gmail.com (A.B. Wijaya)

Author E-mail(s): ABW (aryabimaw13@gmail.com), YTW (yohana.tri@unaki.ac.id), SAP (satrio.agung@unaki.ac.id)
Digital Object Identifier 10.32815/jitika.v19i2.1189
Manuscript submitted 14 Agustus 2025; revised 15 Agustus 2025; accepted 15 Agustus 2025.
ISSN: 2580-8397(O), 0852-730X(P).

process participant answers after the audio in the listening section had played, preventing participants from continuing the test properly (Marbun et al., 2016). Additionally, the system incorrectly displayed a score of 0 for participants, whereas in an official TOEFL exam, it is highly unlikely for a participant to receive a zero score. Another issue that arose was the identical final scores among different participants, which raised doubts about the accuracy and fairness of the assessment process.

Furthermore, the system did not provide a recovery mechanism for participants who were disconnected from the test session due to signal interference or internet connection problems (Oktapianti, 2024). As a result, participants who could not re-access the exam questions had to repeat the process from the beginning or even lost the opportunity to complete the test.

These problems indicate that the old system was unable to provide a reliable, valid, and standards-compliant examination experience (Lubis & Irmayana, 2019). Therefore, the development of a new TOEFL examination website is necessary as a solution to these limitations (Diraja, 2018). The new system will also be designed to handle technical disruptions such as lost connections, with features for saving progress and flexible re-access. With this new system, the administration of the TOEFL examination at the "Plug-in" Training Institute is expected to become more efficient, effective, and trustworthy.

2. Method

2.1. Subject and object of the study

The subject of this research is the website-based TOEFL examination system developed for the "Plug-in" Training Institute. The object of the research includes the features and data of the examination, such as questions, participant answers, scores, completion times, and participant data integrated into the system.

2.2. Data Collection Method

This study aimed to obtain data related to the needs and development of a web-based TOEFL examination system using three data collection methods: indirect observation, interviews, and literature study. Observation was conducted through the analysis of documentation and existing TOEFL systems to understand the examination flow, from registration, listening, structure, and reading sections to the presentation of scores. Interviews were conducted with 3 TOEFL instructors, 1 administrative staff member, and 25 test participants to explore needs, constraints, and expectations for a more efficient system. A literature study was performed by reviewing literature related to information system development, TOEFL examination methods, and web-based testing technologies from various sources, such as books, journals, and previous research.

2.3. Data Sources and Types

This research used data sources obtained from respondents and documents, both statistical and non-statistical, in accordance with the subject and objectives of the research. With a qualitative approach from an emic perspective, the researcher prioritized the views of informants obtained through interviews or observation. The data sources consist of two types: (1) Primary data, obtained directly from respondents through interviews to determine the functional and non-functional requirements of the web-based TOEFL examination system; and (2) Secondary data, derived from journals, books, articles, and documentation of similar systems, serving as a theoretical foundation, for comparison, and as a reference for selecting technology and development methods.

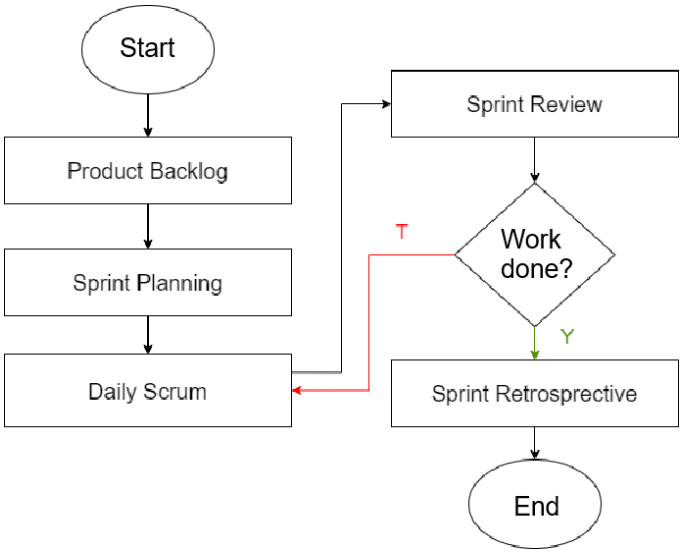

Fig. 1. Scrum Method

2.4. System Development Method

A system development methodology is a structured approach used in designing and building information systems. In this research, the development of the TOEFL examination website was carried out using the Scrum method (Wijaya et al., 2024), an Agile Development framework that is iterative, adaptive, and responsive to changing user needs (Noveandini et al., 2023).

The selection of the Scrum method was based on the need for a system development process that is complex, dynamic, and requires a quick response to user feedback. The TOEFL examination system has a number of important features, such as question randomization, a timer, an audio-based listening question display, and automatic scoring. These features require regular testing and evaluation to function as expected and be easy to use for both participants and administrators (Noveandini et al., 2023).

Scrum is a framework well-suited for developing web-based and mobile systems due to its adaptive and collaborative nature, and its ability to produce a system gradually and measurably through specific sprints.

In this study, system requirements were formulated as user stories, which are brief descriptions from a user's perspective that explain a desired feature or function. The format used follows the pattern: "As a [*usertype*], I want [*goal*], so that [*reason*]." For example: "As a test participant, I want to hear the listening audio only once so that the test atmosphere resembles an official TOEFL exam." This approach helps the team understand user expectations, prioritize development, and ensure that every feature created has a direct use for the end user.

As shown in Fig. 1, the scrum method applied in this study was used to manage the development flow of the TOEFL examination system in a phased and structured manner (Nurmasani et al., 2024). Using Scrum, the development of the TOEFL examination system was carried out in the following stages:

a. Product backlog
   All system requirements were summarized in a product backlog list based on the analysis of problems from the previous TOEFL system used at the "Plug-in" Training Institute. These requirements included input for listening and reading questions, automatic test results, and score reports.
b. Sprint planning
   This stage was used to plan the features to be developed within one sprint period. For example, development began with the audio listening function, then continued to the score report feature in the next sprint.
c. Sprint (development)
   The development process was carried out within a predetermined time frame (e.g., two weeks per sprint). In this stage, system features were built using the Laravel Framework and MySQL as the database.
d. Daily scrum
   Evaluations were conducted regularly to monitor daily work progress, address obstacles encountered, such as audio playback errors or answer validation issues, and ensure each part proceeded according to plan.
e. Sprint review
   After a sprint was completed, a review of the developed features was conducted. An example is testing the listening question feature, which can only be played once and automatically moves to the next question after the time expires.
f. Sprint retrospective
   An evaluation of the previous development process was conducted to identify shortcomings or challenges that arose and to formulate improvement steps for the next sprint.

With the use of user stories in Scrum, system development becomes more focused on real user needs, minimizing the risk of irrelevant features and ensuring system acceptance when implemented at the "Plug-in" Training Institute. The use of Scrum in this project allowed the TOEFL examination system to be developed gradually and measurably. This is important, as a TOEFL examination system requires high validity and reliability in presenting questions, managing time, and calculating participant scores. With the scrum method, each feature could be thoroughly tested and adjusted based on user feedback from the "Plug-in" Training Institute before the system was fully launched (Nurmasani et al., 2024).

3. Results

3.1. System analysis

System analysis is a key stage in the development of information systems. At this stage, user needs are identified, and the functional and technical specifications of the system to be developed are defined. The results of this analysis serve as a reference for the system design and implementation process.

The system requirements in this study were obtained through indirect observation, by analyzing documentation and existing systems, and through interviews with related parties, such as administrators and managers. The analysis process was also conducted with reference to the Scrum methodology approach.

3.1.1. Problem Identification

Before creating or developing a system, it is necessary to identify problems. This aims to help in developing a system by determining the strategy and design that align with the desires and needs of the system's users.

The following are some of the problems encountered in the TOEFL examination system at the "Plug-in" Training Institute:

a. After the audio in the listening section had played, it was not possible to continue the examination.
b. The system incorrectly displayed a score of 0 for participants.
c. The final scores of different participants were identical.

3.1.2. Problem Analysis

In the development of a website-based TOEFL examination system, the identification and analysis of problems in the old system are important to ensure that the new system can resolve these obstacles. Based on observations and testing, several main problems were found that disrupted the smooth administration of the exam and the accuracy of the evaluation results. The following is an analysis of each problem:

1. After the audio in the listening section played, it was not possible to continue the examination.
   This problem indicates an error in the question flow management or the audio playback function in the listening section. The system failed to provide an automatic transition to the next question after the audio finished playing. This could be caused by:
   a. Mismatch in logic between audio control and question navigation.
   b. Absence of an event listener or validation function to handle playback status.
   c. Lack of thorough testing on the listening section.
2. The system incorrectly displayed a score of 0 for participants.
   A score of 0 appearing even though the participant had completed the exam indicates a failure in the score calculation process or the storage of answer data. Possible causes include:
   a. Participant's answer data was not read correctly by the system due to an output error.
   b. Score calculation was performed before the complete data was received by the system.
   c. The database storage schema failed to capture the results correctly (e.g., empty fields, backend processing errors).
3. The final scores of different participants were identical.
   The recurrence of identical scores among participants suggests a potential failure in the question

Table 1. Product backlog

No	Backlog Name	Description	Priority
1	Navigation and Listening Control.	Improve the flow of listening questions so participants can proceed after the audio finishes.	High
2	Score Calculation Logic.	Improve the evaluation system so that participant scores are calculated based on the number of correct answers.	High
3	Question Completion Log.	Add a feature to record the number of correct answers for audit and score validation purposes.	High

Table 2. Sprint planning

Sprint	Backlog Name	Task Description	Estimated Time
Sprint 1	Navigation and Listening Control.	Redesign the flow of listening questions after audio playback & Limit audio to a single play.	1 week
Sprint 2	Score Calculation Logic.	Read participant answers from the database accurately & Develop score calculation logic based on the number of correct answers.	1 week
Sprint 3	Question Completion Log.	Record the number of correct questions for each participant in each section.	1 week

randomization system or the scoring function. Some possible causes are:

a. The scoring function used a default value or a logical loop error that produced a fixed value.
b. An error in saving the scores to the database caused data to be overwritten or copied between participants.

Through this problem analysis, the new system being developed is expected to resolve these three core problems by:

a. Improving the navigation flow and controls in the listening section.
b. Ensuring the accuracy of the score calculation logic.
c. Adding a feature to log the results of question completion.

3.2. Scrum Implementation

At this point, the scrum method was used to develop the web-based TOEFL examination system. This method was chosen because it is one of the most widely used and organized for iterative software development. In this study, the scrum implementation included documenting the scrum work carried out sequentially, from the initial scrum stage through three sprint processes used to refine the features (Tumbade et al., 2024).

3.3. Product Backlog

At this stage, the researcher conducted interviews to determine system needs. Subsequently, the researcher reviewed literature relevant to this study. This included studying literature on previous research, UML diagrams, ERDs, databases, and the Scrum method. The results of the interviews with sources were then analyzed and summarized as a list of system requirements, which are presented in Table 1. This table shows the required product backlog along with the description and priority of each item, making it easier for the development team to plan the development stages.

3.3.1. Sprint Planning

After the product backlog was obtained, the researcher planned the sprints to determine which product backlog items would be worked on during each sprint. The complete sprint plan, including the backlog name, task description, and estimated completion time, is shown in Table 2. This table helps the development team prioritize tasks and monitor the progress of system development in a structured manner.

3.3.2. Daily Scrum

A daily scrum is a short, scheduled meeting held every day during a sprint. The main purpose of this activity is to monitor development progress, identify obstacles, and align the daily work plan to ensure the sprint is completed on time.

Although the development of this TOEFL examination system was done individually, the principles of the daily scrum were still applied to maintain discipline and transparency regarding the designed schedule. Self-evaluation was conducted through daily records of feature progress, technical issues encountered, and the plan for completing tasks for the next day.

3.3.2.1. Sprint 1

The first sprint focused on improving the system for the listening section, particularly in terms of audio playback control, navigation between parts, and timer integration. Although the previous system was usable, a redevelopment was undertaken to enhance control and align the exam flow with TOEFL standards. One of the main goals was to ensure that the audio could only be played once, and participants could only proceed to the next part after finishing the audio.

Table 3 shows the daily log during sprint 1, detailing the activities and development results for each day. The daily scrum was conducted every day during the one-week sprint. Daily progress was used to implement the single-play audio feature, structure the listening page by parts, and refine the transition between question parts. A timer was also installed to ensure participants complete the exam within the specified time.

Audio playback was controlled using the JavaScript onended

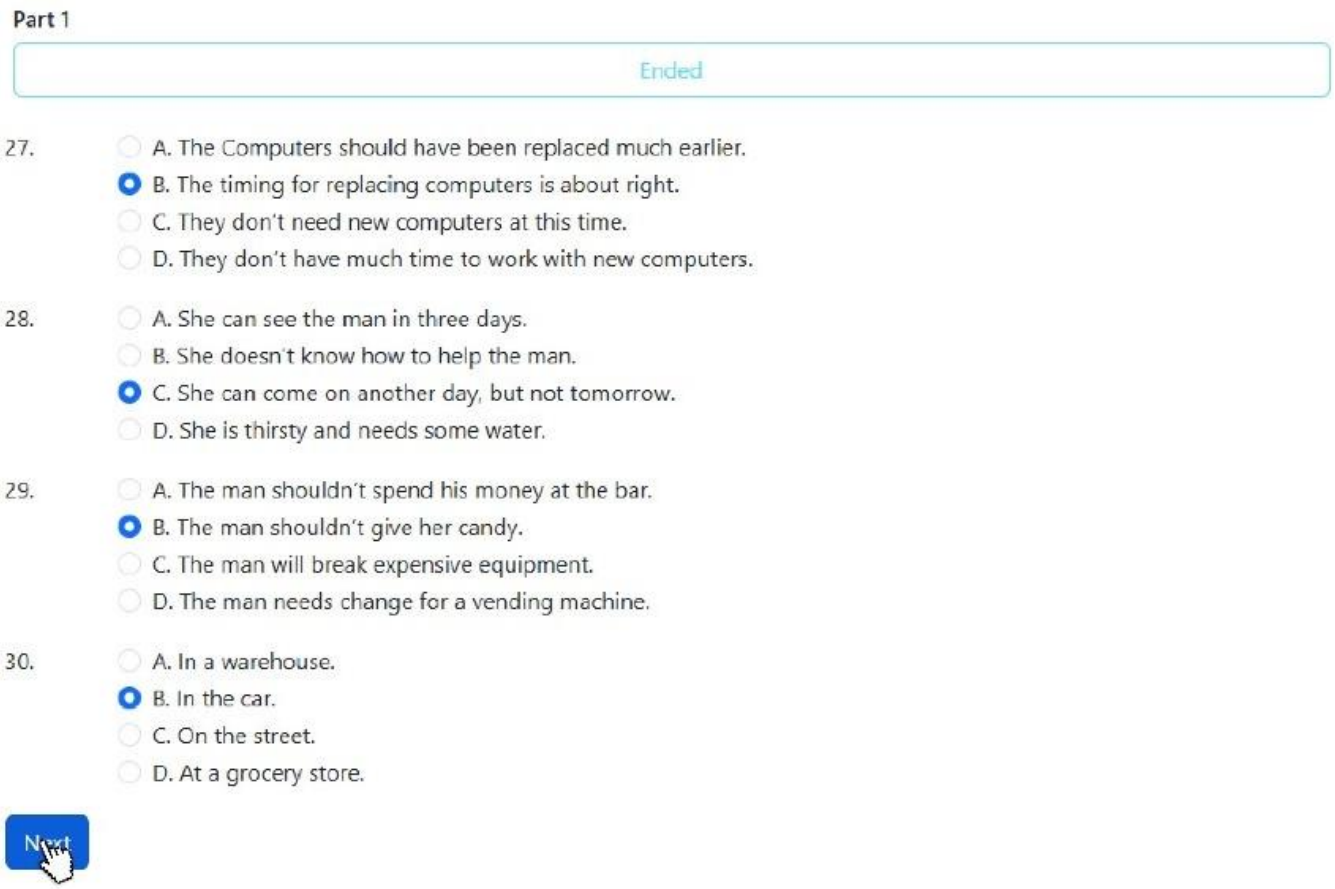

Fig. 2. Listening section

event listener, a function that runs automatically when the audio finishes playing. With this logic, the "Next" or "Finish" button, which was previously unclickable, becomes clickable when the audio playback is complete, allowing the participant to proceed to the next question. This was done to ensure that the audio is fully heard before the participant answers or moves to the next part.

Additionally, participants cannot replay the audio, nor can they click the navigation button while the audio is playing. This is done to maintain the integrity of the exam and provide an experience that approximates a real TOEFL test.

As shown in Fig. 2, the audio playback control mechanism ensures that the "Next" button is only activated after the audio has finished playing, consistent with the official TOEFL exam mechanism. Audio playback is controlled by the JavaScript onended event listener (Khoirurrizal et al., 2024), so the "Next" or "Finish" button is only active after the audio is complete, preventing participants from replaying it or using the navigation buttons while the audio is running.

3.3.2.2. Sprint 2

The second sprint focused on validating and strengthening the TOEFL score calculation system and presenting the test results for each category (Listening, Structure, and Reading). Although the scoring system had been functioning well previously, this sprint was conducted to ensure that the entire process of score calculation, value conversion, and result display fully complied with the expected standards.

Table 3. Daily log sprint 1

Day	Activity	Result
1	Redesign of listening part page.	Questions separated by part, navigation buttons prepared.
2	Integration of audio and Play button control.	Audio can only be played once, button disabled after play.
3	onended JavaScript logic, revealing navigation button.	Navigation button is only clickable after audio finishes playing.
4	Browser testing.	Compatible across browsers.

The main features reviewed in this sprint included the convertScore() function, used to convert the number of correct answers to a TOEFL score based on the category, and the finalScore() function, which calculates the final score based on the three categories using the TOEFL Paper-Based Test (PBT) scale formula. The system uses a specific conversion table for each category and produces a final score with the formula:

$$\text{FinalScore} = \frac{(\text{Listening} + \text{Structure} + \text{Reading}) \times 10}{\text{CategoryCount}}.$$

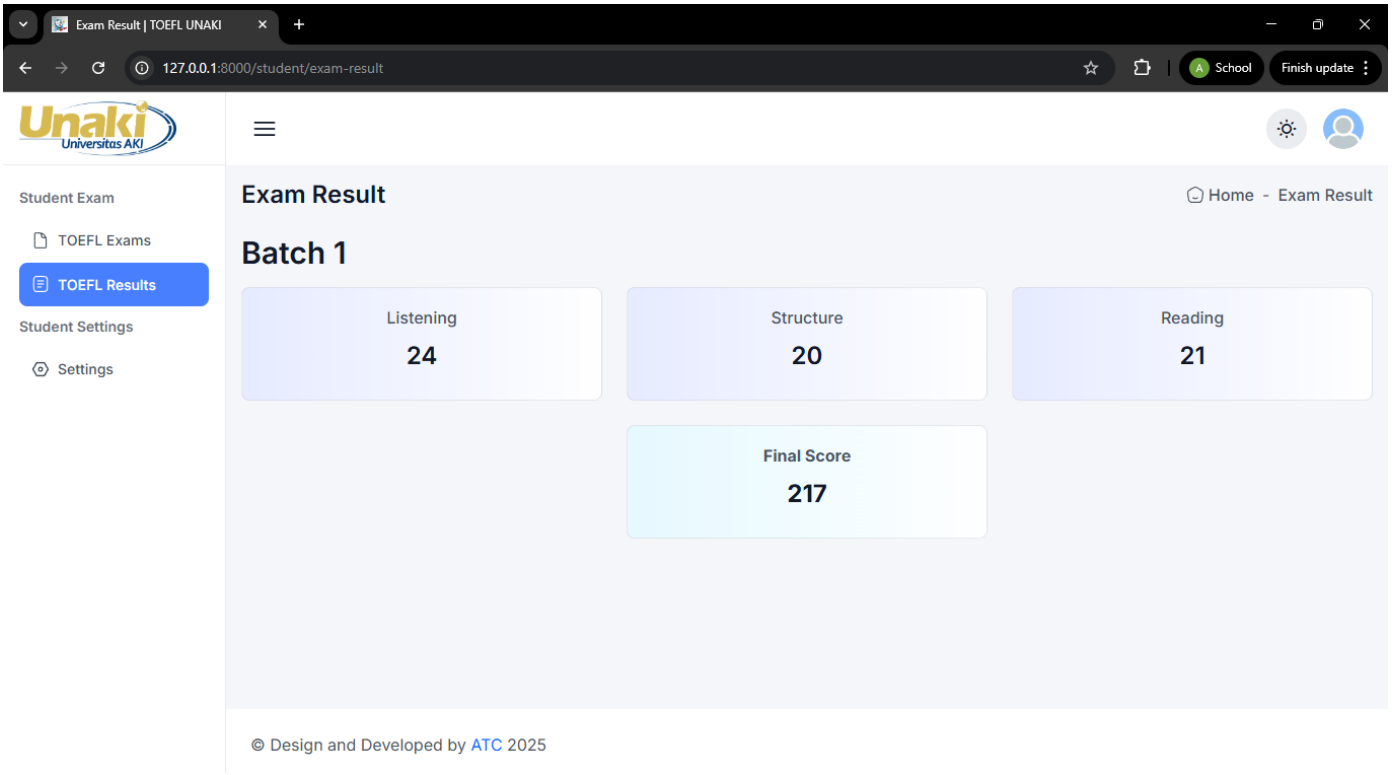

Fig. 3. Display of score calculation results

Through this process, the system can generate a final score in the realistic and accurate range of 310-677, consistent with common TOEFL scoring practices.

The daily log for sprint 2 is displayed in Table 4, which details the testing activities and results for each day. The table outlines the validation and testing steps, including testing the score calculation functions, simulating participant answers, and adjusting the test result display.

To ensure the accuracy of the scoring system's logic, a test was conducted where none of the answers were correct. The purpose of this test was to see how the system handles the minimum value and to ensure that the resulting score remains consistent with the TOEFL PBT conversion standards.

Upper-bound testing with all correct answers yielded a final score of 677, corresponding to the maximum standard score for the TOEFL PBT.

In addition, random tests were conducted with varying combinations of correct answers in each category (e.g., listening 35, structure 25, reading 40). The scores from the system were compared with manual calculations in all trials, and all results showed a 100% match, with no score difference. The simulation results and final score display can be seen in Fig. 3.

This simulation demonstrates that even if a participant provides no correct answers, the system still provides a final score of 217, in line with the minimum score standard for the TOEFL Paper-Based Test (PBT). This indicates that the system has used the correct conversion table and does not return a value of 0, which is unrealistic in the context of TOEFL scoring.

Table 5 shows the reference standard for score conversion for each category. For instance, a participant with zero correct answers in all sections will have their score converted to 24 for listening, 20 for structure, and 21 for reading. This table is used by the system to ensure that extreme cases still produce a valid score according to TOEFL standards.

Table 4. Daily log sprint 2

Day	Activity	Result
1	Review convertScore() and finalScore() functions.	Verified that the system uses the standard TOEFL PBT conversion table.
2	Simulate answer submission and score calculation.	Scores per category and final score were as expected and showed no anomalies.
3	Validate potential for 0 scores or identical scores between participants.	System worked well, scores differed according to input.
4	Cross-check system results with manual calculations.	No score difference; calculation formula proven accurate. Cross-checking was performed on 10 sets of participant answer data, covering variations in the number of correct answers from 0 to 50 in each category. Testing included lower bound (0 correct), upper bound (50 correct), and several random combinations. All manual calculations were identical to the system's results in every scenario.
5-7	Adjust test result display and multi-batch testing.	Scores can be displayed per batch, per category, and the final score is clear.

Based on the table, if a participant gets zero correct answers

Table 5. Standard for score conversion

Correct Answers	Listening	Structure	Reading
0	24	20	21
1	25	20	22
2	26	21	23
3	27	22	23
4	28	23	24
5	29	25	25
6	30	26	26
7	31	27	27
8	32	29	28
9	32	31	28
10	33	33	29
11	35	35	30
12	37	36	31
13	38	37	32
14	39	38	34
15	41	40	35
16	41	40	36
17	42	41	37
18	43	42	38
19	44	43	39
20	45	44	40
21	45	45	41
22	46	46	42
23	47	47	43
24	47	48	43
25	48	49	44
26	48	50	45
27	49	51	46
28	49	52	46
29	50	53	47
30	51	54	48
31	51	55	48
32	52	56	49
33	52	57	50
34	53	58	51
35	54	60	52
36	54	61	52
37	55	63	53
38	56	65	54
39	57	67	54
40	57	68	55
41	58	-	56
42	59	-	57
43	60	-	58
44	61	-	59
45	62	-	60
46	63	-	61
47	65	-	63
48	66	-	65
49	67	-	66
50	68	-	67

in all sections (listening, structure, and reading), the number of correct answers (zero) is converted to a score of 24 for listening, 20 for structure, and 21 for reading.

The system has successfully handled extreme cases (all wrong answers) and still produces a valid final score according to the TOEFL scoring logic. Referring to Table 4, Table 5, and Fig. 3, it can be concluded that the conversion and score calculation system works in a stable, measurable, and standard-compliant manner.

Table 6. Daily log sprint 3

Day	Activity
1	Men Identify the need to record the number of correct questions per section per participant.
2	Trace the relational structure between models (Exam, ExamSession, UserAnswer, etc.).
3	Develop backend logic to calculate the number of correct answers for each section.
4	Store and format the results into a structured array (per category/section).
5	Insert this data into the exam_result view to be displayed to the user.

3.3.2.3. Sprint 3

In the third sprint, the main development focus was on recording and displaying the number of questions answered correctly by each TOEFL test participant in each section: listening, structure, and reading. This feature aims to provide participants with more detailed information about their performance in each part of the test. In addition to displaying the converted final score, the system now also shows the metric of the number of correct answers and the total number of questions for each category. The daily development log for sprint 3 can be seen in Table 6, which details the development activities each day.

This sprint aimed to calculate and record the number of correct answers obtained by the participant for each section of the TOEFL exam. Besides increasing result transparency, this feature also provides educational value by allowing participants to know which areas need improvement.

This information enriches the test result report and provides deeper insight for both participants and instructors to conduct more targeted evaluation and learning. The implementation of the correct answer count feature can be seen in Fig. 4, which shows the system's user interface when displaying the final score along with the number of correct answers per section.

3.3.3. Sprint Review

A sprint review was conducted after the completion of the three main sprints in the development of the web-based TOEFL examination system. This review aimed to evaluate the development work, ensure the features function as expected, and identify opportunities for further development.

In sprint 1, the system successfully completed the Listening Section, which included automatic audio playback, timer settings, and locking the navigation button until the audio was finished. This feature was tested by users and functioned well. The addition of audio.onended logic proved effective in maintaining test integrity.

In sprint 2, the system successfully calculated TOEFL scores based on the conversion of correct answers to standard values. The scores from each section—listening, structure, and reading—were converted using a predetermined table, and then

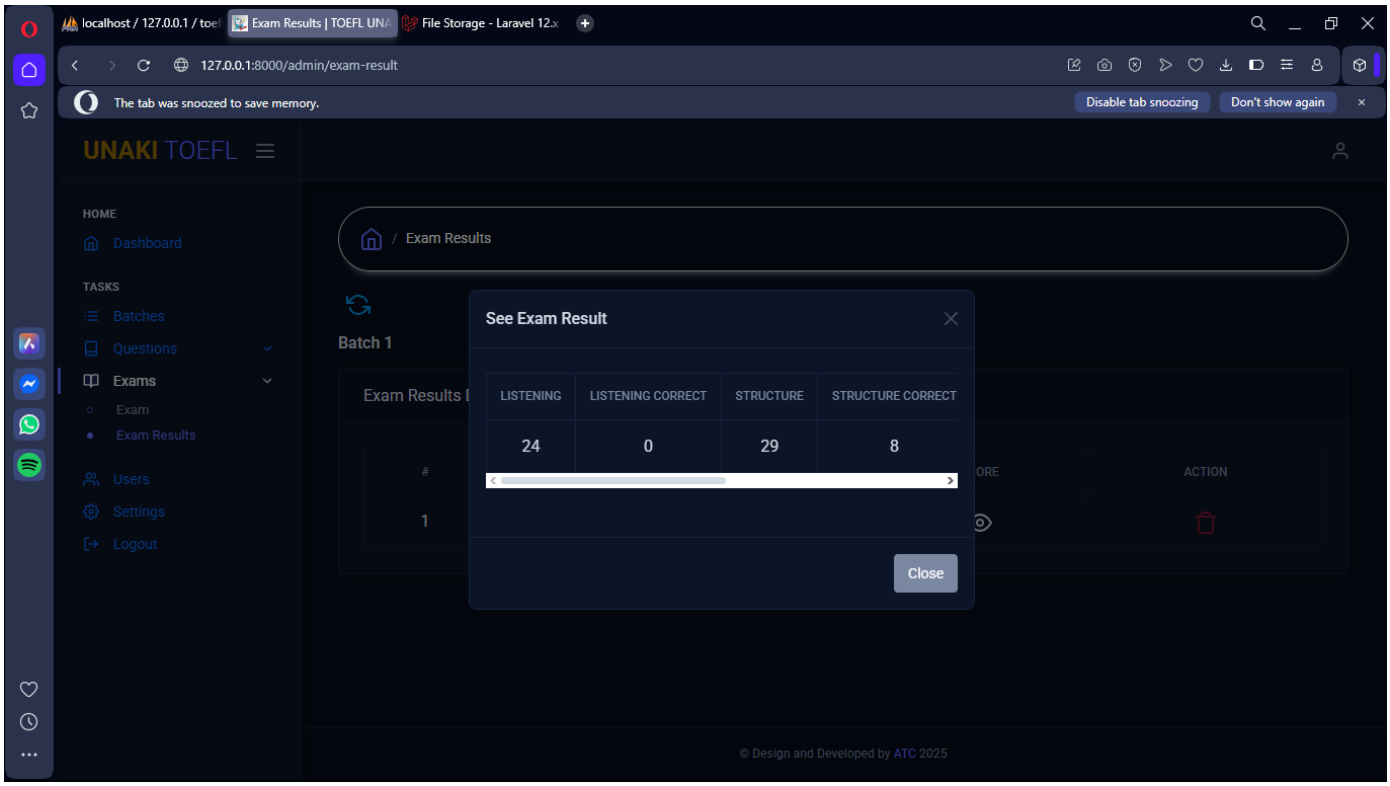

Fig. 4. Display of correct answer results

the final score was calculated using the standard TOEFL formula. The system also displayed the final score properly on the test results page.

Next, sprint 3 focused on recording the number of questions answered correctly by participants in each section. The results of this calculation are displayed on the test results page and provide more comprehensive information for participants. Users can see not only their score but also their actual performance in terms of the number of correct answers out of the total questions in each part.

A summary of the development and feature implementation results from the three sprints can be seen in Table 7. During the sprints, team communication was conducted regularly through daily scrums to monitor daily progress. Any technical obstacles found were promptly discussed and addressed. Testing was performed after each sprint to ensure that the features were stable and aligned with the agreed-upon user stories.

Overall, all planned core features were successfully implemented. The system operates stably and can be used by participants to take the TOEFL exam online. Feedback from users during the review indicated satisfaction with the completeness of the features and ease of use.

3.3.4. Sprint Retrospective

In the sprint retrospective stage, the team reflected on the work done during the three sprints. Some positive achievements included:

a. Sprint 1: The audio, timer, and navigation button features functioned well, but documentation and initial testing needed improvement.
b. Sprint 2: TOEFL score conversion was accurate, but testing for extreme cases and documentation of the score table were needed.
c. Sprint 3: The addition of correct/incorrect question data clarified results for participants, although the controller code needed to be tidied up and an admin UI added to the backlog.

Improvement steps include creating a testing checklist, documenting changes, adding extreme case testing, refactoring code, and planning for future features.

The results of this reflection are structured in Table 8, which summarizes the feature requirements, implementation notes, and evaluation during the sprint retrospective.

4. Discussion

The development of the web-based TOEFL examination system at the “Plug-in” Training Institute shows that the application of the Scrum method was effective in resolving the problems present in the previous system. The iterative nature of the development process made it easier for the team to make improvements based on feedback obtained in each sprint, leading to a gradual increase in system quality.

The implementation of audio control with a “once play” feature through the onended event listener successfully prevents participants from replaying the audio without authorization, thereby maintaining the integrity of the examination. Additionally, the score calculation logic, which refers to the TOEFL PBT conversion table, produces more

Table 7. Sprint review summary

Feature Requirement	Sprint	Estimate	Notes
Redesign listening flow & limit audio to single play.	Sprint 1	1 Week	Feature works as expected, tested successfully, UX functions well.
Read answers and develop score calculation logic.	Sprint 2	1 Week	core conversion is standard-compliant, final score appears correctly on the results page.
Record number of correct questions per section.	Sprint 3	1 Week	More detailed result information.

Table 8. Sprint retrospective summary

No	Feature Requirement	Description	Sprint Retrospective
1	Listening audio navigation and control.	Audio must be automatic, "Next" button active after audio finishes.	Already working well, but needs technical documentation and additional validation.
2	Convert score to TOEFL scale.	High accuracy required.	Already tested with extreme values (all wrong answers) resulting in a minimum score of 217, consistent with TOEFL PBT standards. Minimum to maximum value conversion table prepared for documentation.
3	Display correct and unanswered results.	Admin sees number of correct questions.	Implemented, controller needs refinement.
4	Provide final score and category details.	Participant sees score for each category and final score.	Functions well, suggested to add a visual graph for UI/UX improvement.

accurate assessments and avoids errors such as zero scores or identical scores among participants. The detailed score reporting feature also simplifies evaluation by instructors for each section, such as Listening, Structure, and Reading.

These findings are consistent with previous research indicating that online examination systems built with an iterative approach tend to be more stable and user-friendly. In practice, this system can serve as a reference for other developers aiming to build web-based examination platforms, particularly in maintaining the validity and reliability of the results.

Although the system has undergone functional testing internally through daily scrums, score calculation cross-checks, and extreme value testing (minimum and maximum scores), User Acceptance Testing (UAT) could not be conducted at this stage due to time and logistical constraints. The results of UAT are expected to provide direct evaluation from participants and institution managers regarding the system's feasibility, ease of use, and feature suitability for real-world needs. UAT is recommended for future development to allow for more comprehensive testing before widespread use.

Other limitations of the research include testing being conducted at only one course institution, so the results may not be representative of conditions in other locations with different devices and network quality. Furthermore, the user experience aspect was not measured quantitatively, so the evaluation of user satisfaction remains qualitative.

5. Conclusion

Based on the results of developing the TOEFL examination website system using the scrum method at the "Plug-in" Training Institute, it can be concluded that the application of the scrum method provides an effective and structured approach by dividing the work stages into measurable sprints. Through the three sprints conducted, the system was successfully developed in stages with consistent testing and evaluation. This system resolves various problems found in the previous system, such as imperfect navigation in the listening section, inaccurate score calculation, and the absence of detailed recording of participant performance. Important features like a timer, audio control, automatic score conversion, and recording the number of correct answers have been successfully implemented according to requirements. Overall, the developed system has been running well and is ready for use in digital TOEFL simulations or training.

Suggestions for further development of this system are as follows:

a. System testing should be conducted more broadly and continuously to find bugs or deficiencies from various usage perspectives.
b. The user interface (UI/UX) needs to be optimized to be more user-friendly and responsive on various devices, especially mobile devices.
c. A feature for exporting test result data to PDF or Excel format could be added to facilitate documentation and reporting.
d. The system could be enhanced by adding a feature to record participant answers in more detail, such as noting the choice selected for each question, for further analysis.

Data Availability

All author data has been presented in the article.

Declaration of Conflicting Interests

The authors declare that they have no competing financial

interests or personal connections that could be considered to influence the research presented in this article.

Author Contributions

All authors designed the article, contributed to writing the content, and revised the manuscript. All authors read and approved the final version of the manuscript.

References

Abdi, M., Amru, A., Rafiqa, S. F., & Safriadi, S. (2024). Implementasi algoritma fisher-yates untuk pengacakan dinamis soal TOEFL dalam ujian online berbasis web. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *7*(5), 1286–1291. https://doi.org/10.32672/jnkti.v7i5.8086

Diraja, Z. (2018). *Pembangunan aplikasi simulasi tes TOEFL berbasis web* [Universitas Esa Unggul]. https://digilib.esaunggul.ac.id/public/UEU-Undergraduate-10082-jurnal.Image.Marked.pdf

Febriansyah, M. R., & Voutama, A. (2024). Rancang bangun sistem ujian online berbasis website menggunakan metode waterfall. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(3), 2640–2647. https://doi.org/10.36040/jati.v8i3.9563

Khazanah, D. P. N., & Purnama, G. (2024). Perancangan sistem ujian online menggunakan metode pengembangan Waterfall berbasis web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 2332–2339. https://doi.org/10.36040/jati.v8i2.9475

Khoirurrizal, M. F., Hidayat, C. R., & Ruuhwan, R. (2024). Analisis perbandingan framework front-end Javascript Solidjs dan Vuejs pada pengembangan website interaktif. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4106

Lubis, L. R., & Irmayana, A. (2019). Analisis kesulitan mahasiswa IPTS dalam menyelesaikan soal-soal TOEFL. *Jurnal Education and Development*, *7*(3), 118. https://doi.org/10.37081/ed.v7i3.1202

Marbun, Y. Y., Isnanto, R. R., & Martono, K. T. (2016). Pembuatan aplikasi TOEFL sebagai media pelatihan Bahasa Inggris berbasis web. *Jurnal Teknologi Dan Sistem Komputer*, *4*(1), 83. https://doi.org/10.14710/jtsiskom.4.1.2016.83-92

Maulana, C., & Lubis, I. A. (2022). Students' problem of taking TOEFL test at STMIK Royal Academic year 2020/2021. *Journal of Science and Social Research* , *5*(1), 5. https://doi.org/10.54314/jssr.v5i1.811

Noveandini, R., Wulandari, M. S., & Hakim, A. (2023). Penerapan Metode Scrum pada rancang bangun sistem informasi penjualan Toko Sepatu Rabbani Shoes. *Seminar Nasional Teknologi Informasi Dan Komunikasi STI&K (SeNTIK)*, 192–198. https://ejournal.jak-stik.ac.id/index.php/sentik/article/view/3454

Nurmasani, A., Kurniawan, F. D., Hartanto, A. D., & Fajri, I. N. (2024). Penerapan Metode Scrum pada pengembangan sistem informasi pencatatan magang. *Information System Journal*, *7*(01), 34–44. https://doi.org/10.24076/infosjournal.2024v7i01.1616

Oktapianti, S. (2024). TOEFL score of Tadris Chemistry students at UIN Antasari in achievement of CEFR standard. *Tarbiyah: Jurnal Ilmiah Kependidikan*, *13*(1), 55–63. https://doi.org/10.18592/tarbiyah.v13i1.11417

Pauzi, Y., & Kharisma, I. L. (2025). Pengembangan website ujian online pada SMP Putra Perintis . *MAJU: Indonesian Journal of Community Empowerment*, *2*(1), 153–165. https://doi.org/10.62335/cqex4k29

Sanjaya, H. R., Situmorang, L., M Syahrul, Kalmany, L., & Pamungkas, R. W. P. (2025). Pengembangan sistem ujian online berbasis mobile dengan Agile-Scrum untuk meningkatkan efisiensi dan pengalaman pengguna. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *4*(1), 60–69. https://doi.org/10.69916/jkbti.v4i1.209

Tumbade, M. O., Hartomo, K. D., & Purnomo, H. D. (2024). Pengembangan perangkat lunak berbasis website menggunakan kombinasi Metode Scrum dan V-Model. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(3), 447–460. https://doi.org/10.25126/jtiik.938456

Wijaya, R. I., Usman, M. L. L., Wibowo, F. M., & Gustalika, M. A. (2024). Rancang bangun sistem infromasi tracer study berbasis website menggunakan Metode Scrum (studi kasus: Institut Tekologi Telkom Purwokerto). *Journal of Software Engineering and Multimedia (JASMED)*, *2*(1), 10–20. https://doi.org/10.20895/jasmed.v2i1.1235

Arya Bima Wijaya is currently pursuing a Bachelor of Computer Science degree at the Faculty of Engineering and Informatics, Universitas AKI (Abadi Karya Indonesia) Semarang. His current research focus is on the development of information systems and web-based applications, particularly on the development of TOEFL examination websites.

Yohana Tri Widayati is currently an active lecturer at the Faculty of Engineering and Informatics, Universitas AKI Semarang.

Satrio Agung Prakoso is currently a lecturer at the Faculty of Engineering and Informatics, Universitas AKI Semarang. He is currently serving as the head of the Informatics Engineering study program at Universitas AKI Semarang.