I.J. Information Technology and Computer Science, **2024, 3, 1-11**
Published Online on June 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2024.03.01

Analyzing Test Performance of BSIT Students and Question Quality: A Study on Item Difficulty Index and Item Discrimination Index for Test Question Improvement

Cris Norman P. Olipas*
College of Information and Communications Technology, Nueva Ecija University of Science and Technology, Cabanatuan City, Philippines
E-mail: olipas.cris@gmail.com
ORCID iD: https://orcid.org/0000-0002-2856-9448
*Corresponding author

Ruth G. Luciano
College of Information and Communications Technology, Nueva Ecija University of Science and Technology, Cabanatuan City, Philippines
E-mail: rcgluciano@gmail.com
ORCID iD: https://orcid.org/0000-0001-8532-6971

Received: 06 January 2024; Revised: 17 March 2024; Accepted: 12 April 2024; Published: 08 June 2024

Abstract: This study presents a comprehensive assessment of the test performance of Bachelor of Science in Information Technology (BSIT) students in the System Integration and Architecture (SIA) course, coupled with a meticulous examination of the quality of test questions, aiming to lay the groundwork for enhancing the assessment tool. Employing a cross-sectional research design, the study involved 200 fourth-year students enrolled in the course. The results illuminated a significant discrepancy in scores between upper and lower student cohorts, highlighting the necessity for targeted interventions, curriculum enhancements, and assessment refinements, particularly for those in the lower-performing group. Further examination of the item difficulty index of the assessment tool unveiled the need to fine-tune certain items to better suit a broader spectrum of students. Nevertheless, the majority of items were deemed adequately aligned with their respective difficulty levels. Additionally, an analysis of the item discrimination index identified 25 items suitable for retention, while 27 items warranted revision, and 3 items were suitable for removal, as per the analysis outcomes. These insights provide a valuable foundation for improving the assessment tool, thereby optimizing its capacity to evaluate students' acquired knowledge effectively. The study's novel contribution lies in its integration of both student performance assessment and evaluation of assessment tool quality within the BSIT program, offering actionable insights for improving educational outcomes. By identifying challenges faced by BSIT students and proposing targeted interventions, curriculum enhancements, and assessment refinements, the research advances our understanding of effective assessment practices. Furthermore, the detailed analysis of item difficulty and discrimination indices offers practical guidance for enhancing the reliability and validity of assessment tools in the BSIT program. Overall, this research contributes to the existing body of knowledge by providing empirical evidence and actionable recommendations tailored to the needs of BSIT students, promoting educational quality and student success in Information Technology.

Index Terms: Item Difficulty Index, Item Discrimination Index, Learning Outcomes, Question Enhancement, Systems Integration and Architecture, Test Performance.

1. Introduction

Assessment holds significant value in the realm of higher education on multiple grounds. Firstly, it offers valuable feedback to both educators and learners, thereby facilitating the refinement of their teaching methodologies and learning approaches [1]. Secondly, assessment propels learning by directing students toward effective learning practices and gauging their degree of accomplishment [2]. Thirdly, it assists in the assessment item’s quality evaluation and the creation

of assessment tools [3]. Moreover, assessment serves as a pivotal factor in analyzing the overall well-being and viability of the higher education system [4]. Lastly, in the context of information technology projects, evaluation and assessment procedures contribute towards the continual enhancement of quality and efficient project management [5]. By directly linking the proposed assessment methodologies to anticipated outcomes such as enhanced learning efficacy and curriculum alignment, it becomes evident that assessment plays a pivotal role in shaping the educational landscape. Through effective assessment strategies, educators can refine teaching methodologies, guide students towards effective learning practices, and ensure the alignment of curricula with educational objectives. This comprehensive approach not only enhances pedagogy and scholarship but also fosters the cultivation of students' aptitudes, contributing to the overall improvement of higher education institutions. Additionally, the methodologies and objectives of evaluating academic performance, such as the appraisal of acquired knowledge and skills, are instrumental to the progression of tertiary education. The implementation of dual assessment methodologies, which contrast student-generated scores with those calculated by instructors, can likewise prove advantageous in nurturing a more proficient student body. All in all, assessment constitutes a critical element of higher education, as it facilitates the enhancement of pedagogy and scholarship, gauges institutional efficacy, ensures the integrity of educational projects, and fosters the cultivation of students' aptitudes.

In the realm of educational assessment, item analysis pertains to the systematic approach of acquiring and scrutinizing information derived from students' responses to evaluate the merit of test items. This entails the assessment of the difficulty level, discrimination power, and distractor efficiency of multiple-choice questions (MCQs) utilized in examinations. The objective is to recognize items that necessitate refinement or elimination from the question bank, as well as to establish a dependable and valid question bank. Item analysis can be executed through conventional techniques such as classical test theory (CTT), item response theory (IRT) or by integrating exam logs and exam-taker behavior documented by electronic assessment platforms (EAPs) [6, 7]. In CTT, item difficulty and discrimination are calculated to determine the item quality [8]. IRT, on the other hand, uses model parameters to evaluate item characteristics, with estimates obtained from sample data [9]. Both approaches aim to create tests with a minimal number of items that yield the desired level of reliability [10]. The theoretical framework underpinning item analysis methodology is essential for understanding the statistical models and methods used in constructing tests and generating scores for individuals [11]. It provides a basis for interpreting item responses and estimating latent traits, such as ability, which underlie the observed item responses [12].

The analysis facilitates the determination of the effectiveness of each item, the dependability of the entire examination, and provides insights that exceed the scope of conventional item analyses [13]. Furthermore, it enables the identification of underdeveloped MCQs and the enhancement of their quality [14, 15].

The assessment of item difficulty is a vital component in evaluating the quality of assessment. It is a valuable tool that provides information on the level of difficulty of each item in a test, enabling educators to identify the strengths and weaknesses of students in specific topics or subjects [16-18]. Through the analysis of difficulty index, educators can make necessary modifications and improvements to test items to ensure that students' abilities are accurately measured and reliable measurement results are provided [19]. Additionally, the item difficulty index is of utmost importance in computerized adaptive testing, where item difficulty is dynamically adjusted based on the tester's present ability [20]. This aids in the selection of items that are consistent with the test taker's ability and optimizes the adaptive learning process. All in all, the item difficulty index plays a significant role in enhancing the validity and effectiveness of assessments by ensuring that test items accurately measure students' abilities, thereby providing valuable insights for instructional improvement.

The utilization of the item discrimination index is a crucial factor in the field of educational and psychological research, as it serves to appraise an item's capacity to differentiate between individuals with high and low levels of learning outcomes. This index furnishes valuable insights into an item's ability to discriminate between varying levels of performance. Its application is particularly useful in evaluating the quality of diagnostic items, and it is also an integral component in the construction of tests and item selection algorithms. By identifying item difficulty and discrimination levels, the index can facilitate the improvement of test items and the design of reinforcement and enrichment activities for educators. Additionally, the item discrimination index is pertinent to the validation testing of instruments, as it gauges an item's ability to be correctly answered by individuals with high total scores and incorrect by those with low total scores [16, 21-23].

Prior investigations into item analysis have concentrated on various aspects of this technique. For example, one such study delved into the difficulty index, discrimination index, and distractor efficiency of multiple-choice questions on a block examination [24]. In addition, a different study developed and assessed a user-friendly electronic item analysis software named SuperSETIA [25]. Moreover, research has been conducted on the psychometric properties of the Authority Acceptance scale from the Teacher Observation of Classroom Adaptation-Revised (AA-TOCA-R), which is implemented to identify childhood conduct problems [26]. Another study examined teachers' competency in performing item analysis and their utilization of item analysis software [27]. Finally, a novel measure known as angsta was introduced in a study to appraise the distribution of responses to distractions in multiple choice questions [28].

However, despite the extensive body of research on item analysis, there remains a notable gap in the literature. While previous studies have explored item difficulty and discrimination indices in various educational contexts, limited research has investigated how these indices specifically relate to the test performance of Bachelor of Science in Information Technology (BSIT) students. Thus, this study seeks to address this significant research gap by examining the impact of item difficulty and discrimination indices on the test performance of BSIT students. This endeavor

contributes to the growing body of literature and unveils new insights, thereby enhancing our understanding of test question construction, validation, and improvement. This research not only has the potential to benefit BSIT education but also provides valuable insights for educators and researchers working in the broader field of educational assessment.

Statement of the Problem

In general, this study sought to analyze the test performance of the BSIT students and the question quality in the Systems Integration and Architecture course to provide a basis for the improvement of the test question. Specifically, it aimed to describe the following:

- How may the test performance of the students in the Systems Integration and Architecture (SIA) course be described when grouped accordingly in upper and lower groups?
- How may the test items for the test questions in SIA course be described and analyzed in terms of their:
    - item difficulty index; and
    - index of discrimination?
- What improvement may be made on the midterm test based on the result of the study?

2. Methodology

This study utilized a cross-sectional research design. This study presents a comprehensive assessment of the test performance of Bachelor of Science in Information Technology (BSIT) students in the System Integration and Architecture (SIA) course, coupled with a meticulous examination of the quality of test questions, aiming to lay the groundwork for enhancing the assessment tool. Employing a cross-sectional research design, the researchers collected data from 200 senior BSIT students enrolled in the System Integration and Architecture course during the first semester of the Academic Year 2022-2023 from a higher education institution in Nueva Ecija, Philippines. Using the purposive sampling technique, the researchers identified the respondents, as it was necessary that the respondents for this study were the students who took the course SIA and were able to use the test questions in a major assessment scheduled for the course.

While the purposive sampling technique allowed for the selection of respondents directly relevant to the study's objectives, it may introduce potential biases. For instance, students who were absent during the assessment or who performed poorly may have been less likely to participate, leading to a skewed representation of student performance. Additionally, limiting the study to a single institution may restrict the generalizability of the findings to other BSIT programs or educational contexts. Furthermore, the reliance on data from a single assessment may not capture the full range of student abilities or accurately reflect the effectiveness of the assessment tool over time.

Despite these potential limitations, the study provides valuable insights into the test performance of BSIT students and the quality of assessment tools used in the SIA course. Future research could adopt longitudinal designs or include a more diverse sample of institutions to enhance the generalizability of findings and mitigate potential biases. Additionally, incorporating multiple assessments throughout the academic year would provide a more comprehensive understanding of student performance and the effectiveness of assessment practices within the BSIT program.

The test question which was subjected to quality assessment was self-made by one of the researchers handling the course of that specific semester. The test was composed of 100-item questions administered for one and a half hours to the respondents in different schedules. It was a multiple-choice type of assessment. The items included in the test questions were anchored and aligned to the learning outcomes of the course. The researchers ensured that the contents of the test questions were valid.

3. Results and Discussion

The graphs and textual discussion presented below were generated from the scores of the SIA students.

3.1. The Test Performance of the BSIT Students in the Systems Integration and Architecture Course

The table presents the test performance of BSIT students enrolled in the Systems Integration and Architecture course, divided into two distinct groups: the "Upper Group" and the "Lower Group." The "Upper Group" comprises several students, and their corresponding raw scores, equivalent scores, and verbal descriptions are listed. Likewise, the "Lower Group" consists of other students, and their respective raw scores, equivalent scores, and verbal descriptions are also provided. The equivalent scores appear to be presented as percentages, indicating the level of achievement in the test. Verbal descriptions assigned to the performance of students in both groups offer insights into their overall performance levels.

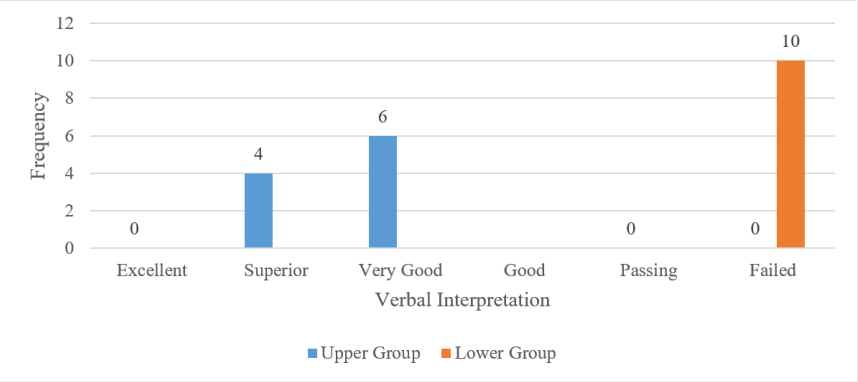

Fig.1. Data visualization of the test performance in the upper and lower groups

Table 1. Test performance of the BSIT students in the upper and lower groups

Upper Group				Lower Group			
Student	**Raw Score**	**Equivalent**	**Verbal Description**	**Student**	**Raw Score**	**Equivalent**	**Verbal Description**
1	43	89.09	Very Good	1	21	69.09	Failed
2	44	90.00	Superior	2	25	72.73	Failed
3	41	87.27	Very Good	3	25	72.73	Failed
4	46	91.82	Superior	4	24	71.82	Failed
5	46	91.82	Superior	5	24	71.82	Failed
6	39	85.45	Very Good	6	23	70.91	Failed
7	45	90.91	Superior	7	22	70.00	Failed
8	40	86.36	Very Good	8	21	69.09	Failed
9	40	86.36	Very Good	9	21	69.09	Failed
10	39	85.45	Very Good	10	17	65.45	Failed
Mean Score		**88.45**	**Very Good**	Mean Score		**70.27**	**Failed**

A close analysis of the data highlights a noticeable discrepancy in test performance between the two student segments. Raw scores of Upper Group students were distributed between 39 and 46, with equivalent scores spanning 85.45% to 91.82%. These results demonstrate exceptional performance, according to the descriptive labels "Very Good" and "Superior." In contrast, the "Lower Group" demonstrated notably lower performance levels. Students in this group obtained raw scores between 17 and 25, corresponding to equivalent scores ranging from 65.45% to 72.73%. These scores are associated with the verbal description "Failed," indicating that students in the "Lower Group" did not meet the required level of proficiency in the test. The Upper Group's impressive performance standard stands out when compared to the Lower Group's average score of 70.27%, demonstrating a significant disparity in overall performance.

Significant implications arise from the table's findings for various stakeholders in the education institution. At first glance, the considerable discrepancy in performance between the two groups underscores the existence of understanding disparities within their knowledge of course content. Customized teaching or supplementary support must be provided to address the disparities and elevate the performance of students in the "Lower Group." This has been supported by [29], explaining that when designing instructional materials, it is necessary to understand the learning styles and needs of the students to improve the quality of their performance.

The data also stresses the importance of reviewing and improving the curriculum to cater better to the distinct learning requirements of students. Curriculum designers ought to customize teaching techniques, material, and evaluations to accommodate diverse learning preferences and skill levels. Reviewing and enhancing the curriculum is of the utmost importance in order to effectively cater to the unique learning requirements of students. This practice guarantees that the curriculum remains up-to-date and progressive, guiding students toward achieving their intended objectives and equipping them with the necessary skills to become successful members of the workforce [30]. Continual evaluation of the curriculum facilitates the identification of potential challenges and enables the implementation of essential strategies to counteract them, such as managing metadata, reducing staff workload, and conducting evaluations [31]. Through the alignment of the curriculum with learning objectives, the possibility of misalignment is greatly minimized, resulting in more efficient and effective learning outcomes [32]. Moreover, the integration of requirements elicitation concepts and learning assessments into the curriculum serves to add value to those students pursuing a degree in information systems and information technology, as it enhances their capacity to evaluate new information system needs in a business environment [33]. All in all, the evaluation and enhancement of the curriculum ensures that it is tailored to meet the specific needs of students, enhances their overall learning experience, and prepares them for future success.

It can be noted that data can be leveraged by educators to modify their teaching approaches and offer bespoke instruction suited to the distinct needs of learners. The Lower Group students' underperformance demands timely and effective interventions to provide the necessary assistance, thus facilitating their recovery and boosting their academic achievement. Additionally, the analysis of Item Difficulty Index and Item Discrimination Index can aid educators in refining test questions and assessments. By improving the quality and clarity of test items, assessments can become more accurate indicators of student learning and mastery of the subject matter.

The implications of the results point towards the importance of using data-driven approaches to enhance the learning experience, promote academic success, and foster a more inclusive educational environment for all students in the program. Through targeted interventions, curriculum enhancements, and assessment improvements, educators and institutions can work together to elevate the academic achievements of all students.

A. Item Difficulty Index of the Test Question in Systems Integration and Architecture Course

Figure 2 shows the data visualization of the item difficulty index (DIF I) for the lower and upper groups. Presenting the data in visual format makes it easier to grasp and acquire immediate insights of the results collected.

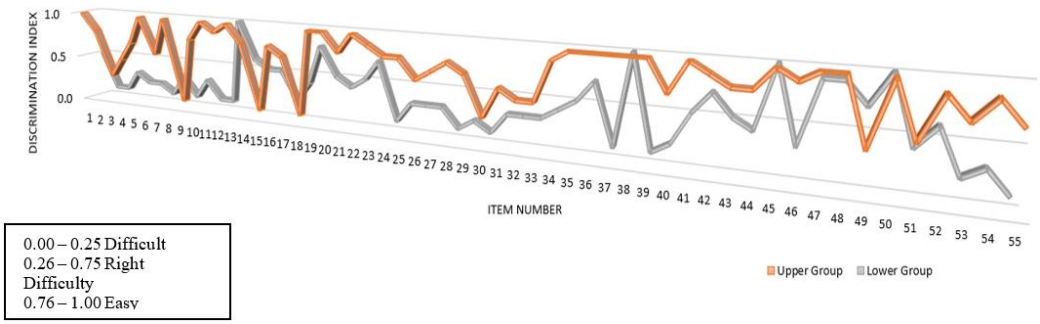

Fig.2. Data visualization of the item difficulty index for the lower and upper groups

Meanwhile, in Table 2, a comprehensive presentation of the data is shown to provide details of each item analysis with respect to its DIF I.

The table presents an analysis of test questions based on the Item Difficulty Index (DIF I) for two distinct groups: the "Upper Group" and the "Lower Group." The IDI is utilized to assess the difficulty level of each question, and the interpretations provide insights into how well the questions align with the abilities of students in both groups. The interpretations are categorized into "Easy," "Right Difficulty," or "Difficult," indicating whether the questions were appropriately challenging or too easy/difficult for the students.

For the "Upper Group," several questions scored an DIF I of 1.0, including Questions 1, 6, 8, 11, 13, 14, 16, 20, 22, 25, 26, 27, and 30. This indicates that most students in the Upper Group answered these questions correctly with relative ease. Additionally, Questions 3, 4, 5, 7, 10, 12, 15, 17, 18, 21, 23, and 28 had DIF I score ranging from 0.5 to 0.9, placing them in the "Right Difficulty" category, suggesting an appropriate level of challenge for students, which they effectively addressed. Finally, Question 9 had an DIF I of 0.1, indicating that most students found it challenging to answer, classifying it as "Difficult."

In the "Lower Group," Question 31 achieved an DIF I of 0.6, placing it under the "Right Difficulty" category, indicating an appropriate level of challenge for the Lower Group students. Questions 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, and 54 all attained DIF I score between 0.1 to 0.9, making them "Right Difficulty" questions, appropriately aligned with the abilities of the Lower Group. On the other hand, Question 29 and Question 55 received DIF I score of 0.2 and 0.1, respectively, indicating they were considered "Difficult," presenting challenges for the Lower Group students.

The analysis of the Item Difficulty Index for the test questions carries significant implications for educators and course designers. It offers insights into the appropriateness of the questions for both the Upper and Lower Groups of students, enabling educators to gauge the overall balance of question difficulty in the assessment. By adjusting the number of easy and difficult questions, educators can ensure the test effectively evaluates students' knowledge and comprehension.

The DIF I analysis also aids in selecting and tailoring questions based on learning objectives, allowing educators to design assessments that cater to the varying levels of proficiency within the student groups. Moreover, it sheds light on the topics or concepts consistently proving challenging for both groups of students, guiding educators in adapting the curriculum to improve student understanding and performance in those areas.

Identifying questions that are either too easy or too difficult can prompt item review and refinement, enhancing question clarity and relevance, and ultimately leading to more accurate measures of students' learning outcomes. Additionally, the results can inform differentiated instruction strategies, allowing educators to provide appropriate support and resources to students who face difficulties with certain topics.

Table 2. Test question analysis on the item difficulty index

Item No.	Upper Group		Lower Group		Item No.	Upper Group		Lower Group	
	Difficulty Index	Interpretation	Difficulty Index	Interpretation		Difficulty Index	Interpretation	Difficulty Index	Interpretation
1	1.0	Easy	0.5	Right Difficulty	31	0.6	Right Difficulty	0.3	Right Difficulty
2	0.8	Easy	0.1	Difficult	32	0.5	Right Difficulty	0.3	Right Difficulty
3	0.3	Right Difficulty	0.1	Difficult	33	0.5	Right Difficulty	0.3	Right Difficulty
4	0.5	Right Difficulty	0.3	Right Difficulty	34	0.9	Easy	0.4	Right Difficulty
5	0.7	Right Difficulty	0.2	Difficult	35	1.0	Easy	0.5	Right Difficulty
6	1.0	Easy	0.2	Difficult	36	1.0	Easy	0.7	Right Difficulty
7	0.6	Right Difficulty	0.1	Difficult	37	1.0	Easy	0.1	Difficult
8	1.0	Easy	0.3	Right Difficulty	38	1.0	Easy	1.0	Easy
9	0.1	Difficult	0.1	Difficult	39	1.0	Easy	0.1	Difficult
10	0.8	Easy	0.3	Right Difficulty	40	0.7	Right Difficulty	0.2	Difficult
11	1.0	Easy	0.1	Difficult	41	1.0	Easy	0.7	Right Difficulty
12	0.9	Easy	0.1	Difficult	42	0.9	Easy	0.5	Right Difficulty
13	1.0	Easy	1.0	Easy	43	0.8	Easy	0.4	Right Difficulty
14	0.8	Easy	0.6	Right Difficulty	44	0.8	Easy	1.0	Easy
15	0.1	Difficult	0.5	Right Difficulty	45	1.0	Easy	0.3	Right Difficulty
16	0.8	Easy	0.5	Right Difficulty	46	0.9	Easy	0.9	Easy
17	0.7	Right Difficulty	0.2	Difficult	47	1.0	Easy	0.9	Easy
18	0.1	Difficult	0.4	Right Difficulty	48	1.0	Easy	0.7	Right Difficulty
19	1.0	Easy	0.8	Easy	49	0.4	Right Difficulty	1.0	Easy
20	1.0	Easy	0.5	Right Difficulty	50	1.0	Easy	0.7	Right Difficulty
21	0.8	Easy	0.4	Right Difficulty	51	0.5	Right Difficulty	0.4	Right Difficulty
22	1.0	Easy	0.5	Right Difficulty	52	0.9	Easy	0.6	Right Difficulty
23	0.9	Easy	0.7	Right Difficulty	53	0.7	Right Difficulty	0.2	Difficult
24	0.8	Easy	0.1	Difficult	54	0.9	Easy	0.3	Right Difficulty
25	0.8	Easy	0.3	Right Difficulty	55	0.7	Right Difficulty	0.1	Difficult
26	0.6	Right Difficulty	0.3	Right Difficulty					
27	0.7	Right Difficulty	0.3	Right Difficulty					
28	0.8	Easy	0.1	Difficult					
29	0.7	Right Difficulty	0.2	Difficult					
30	0.3	Right Difficulty	0.1	Difficult					

Educators can also utilize the DIF I analysis to provide targeted feedback and personalized remediation plans for students who struggle with specific questions. Comprehending the DIF I can yield several benefits in enhancing the quality of education at the tertiary level. Firstly, it can enable educators to identify the areas of the curriculum that pose challenges to students, thereby allowing them to concentrate on these domains and furnish additional support or resources to facilitate students' surmounting of difficulties [34]. Secondly, it can facilitate the crafting of instructional materials and activities that are suitably challenging and captivating for learners [35]. Further, a grasp of the DIF I can help teachers customize their pedagogical approaches to cater to the varied needs and learning styles of students, thereby fostering a more inclusive and efficacious learning milieu [36]. Lastly, it can contribute to the ongoing assessment and evaluation of instructional practices, equipping teachers to refine their teaching techniques continuously and bolster student learning outcomes [37].

In general, by making data-driven decisions based on the Item Difficulty Index, educators can enhance the effectiveness of assessments and optimize the learning experience for both high-performing and struggling students, fostering a more inclusive and conducive learning environment for all.

B. Item Discrimination Index of the Test Question in Systems Integration and Architecture Course

Table 3. Test question analysis on the item discrimination index

Item Number	Discrimination Index	Interpretation	Item Number	Discrimination Index	Interpretation
1	0.5	Discriminating	31	0.3	Non-discriminating
2	0.7	Discriminating	32	0.2	Non-discriminating
3	0.2	Non-discriminating	33	0.2	Non-discriminating
4	0.2	Non-discriminating	34	0.5	Discriminating
5	0.5	Discriminating	35	0.5	Discriminating
6	0.8	Discriminating	36	0.3	Non-discriminating
7	0.5	Discriminating	37	0.9	Discriminating
8	0.7	Discriminating	38	0.0	Non-discriminating
9	0.0	Non-discriminating	39	0.9	Discriminating
10	0.5	Discriminating	40	0.5	Discriminating
11	0.9	Discriminating	41	0.5	Discriminating
12	0.8	Discriminating	42	0.2	Non-discriminating
13	0.0	Non-discriminating	43	0.3	Non-discriminating
14	0.2	Non-discriminating	44	0.4	Non-discriminating
15	-0.4	Questionable	45	0.0	Non-discriminating
16	0.3	Non-discriminating	46	0.6	Discriminating
17	0.5	Discriminating	47	0.1	Non-discriminating
18	-0.3	Questionable	48	0.1	Non-discriminating
19	0.2	Non-discriminating	49	-0.3	Questionable
20	0.5	Discriminating	50	0.0	Non-discriminating
21	0.4	Non-discriminating	51	0.1	Non-discriminating
22	0.5	Discriminating	52	0.3	Non-discriminating
23	0.3	Non-discriminating	53	0.5	Discriminating
24	0.7	Discriminating	54	0.6	Discriminating
25	0.5	Discriminating	55	0.6	Discriminating
26	0.3	Non-discriminating			
27	0.4	Non-discriminating			
28	0.7	Discriminating			
29	0.5	Discriminating			
30	0.2	Non-discriminating			

Table 3 presents a comprehensive analysis of test questions based on the Item Discrimination Index (DI). The DI is a statistical measure used to assess how effectively each question discriminates between high-performing and low-performing students. The table provides the Discrimination Index for each question, and the interpretations are categorized as "Discriminating," "Non-discriminating," or "Questionable," reflecting the ability of each question to differentiate students' abilities.

The results of the Item Discrimination Index analysis reveal important insights into the efficacy of the test questions. Several questions exhibit strong discriminatory power, categorized as "Discriminating." These questions, with Discrimination Indices greater than 0.5, effectively distinguish between high-performing and low-performing students, providing valuable information about their varying abilities. Questions 2, 5, 6, 7, 8, 10, 11, 12, 16, 17, 18, 20, 22, 24, 25, 28, 29, and 30 fall into this category, indicating that they serve as reliable indicators of students' performance levels.

On the other hand, some questions demonstrate limited discriminatory ability and are labeled as "non-discriminating." These questions, with Discrimination Indices ranging from 0.2 to 0.5, may not effectively differentiate between high and low performers. These include Questions 3, 4, 9, 13, 14, 15, 19, 21, 23, 26, 27, 31, 32, 33, 34, 36, 38, 39, 41, 42, 43, 45, 47, 51, and 53. The presence of non-discriminating questions warrants attention, as they may not provide the desired insights into students' abilities. Additionally, a few questions display a "Questionable" discriminatory pattern, with Discrimination Indices below 0.2. These questions, including Questions 48, 49, and 50, may require further examination and potential revision to improve their discriminatory power.

The implications of the Item Discrimination Index analysis are crucial for educators, course designers, and assessment developers. Identifying questions with strong discriminatory power ("Discriminating") is vital as they can effectively differentiate between high and low performers. These questions serve as reliable indicators of students' understanding and proficiency, providing valuable data for assessing learning outcomes.

The presence of "non-discriminating" questions requires attention. These questions may not effectively differentiate between students with varying levels of knowledge and skill. Course designers and educators should review these questions to improve their clarity, relevance, and difficulty to ensure they provide meaningful insights into student performance.

The "Questionable" questions raise concerns about their ability to discriminate between high and low performers effectively. Educators should critically assess these questions and consider potential revisions or replacements to strengthen the assessment's overall quality and ensure it accurately reflects students' capabilities.

Overall, the results of the Item Discrimination Index analysis offer guidance for refining the assessment instrument, enhancing question quality, and improving the alignment between assessment content and learning objectives. By using this information to make informed decisions, educators can design assessments that yield more accurate and meaningful information about student performance, ultimately supporting targeted instructional strategies and fostering continuous improvement in the learning process.

3.2. The Improvements that can be made in the Midterm Test based on the Result of the Analysis

Table 4. Decision for every item analyzed

Decision	No. of Items	Item No.
Retain	25	1, 2, 5, 6, 7, 8, 10, 11, 12, 17, 22, 24, 25, 28, 29, 34, 35, 37, 39, 40, 41, 46, 53, 54, 55
Revise	27	3, 4, 9, 13, 14, 16, 19, 20, 21, 23, 26, 27, 30, 31, 32, 33, 36, 38, 42, 43, 44, 45, 47, 48, 50, 51, 52
Discard	3	15, 18, 49

Table 4 provides an analysis of an assessment's test items, categorized into three groups: "Retain," "Revise," and "Discard," based on the decisions made for each item. The "Retain" category includes 25 items that exhibit strong validity, reliability, and discriminatory power, effectively distinguishing between high and low-performing students. These items offer valuable insights into the students' knowledge and comprehension. In contrast, the "Revise" category comprises 27 items with potential but requiring refinement or adjustments to enhance their effectiveness in accurately measuring student performance. The "Discard" category involves three items identified as problematic, with flaws or inaccuracies that compromise their suitability for assessing student understanding. Removing these items is vital to maintain the assessment's integrity and ensure valid results.

The implications of these decisions are substantial for assessment quality and instructional design. By retaining the high-quality items, the overall test quality is strengthened, yielding meaningful insights into students' performance and improving the assessment's accuracy in measuring learning outcomes. The 27 items marked for revision offer opportunities for targeted improvement, allowing educators to optimize the assessment's effectiveness and better align it with learning objectives. Removing the problematic items ensures the assessment's validity and reliability, providing a fair and accurate representation of students' abilities.

The decisions also affect resource allocation, as educators can focus on retaining high-quality items and revising those with potential, making the best use of their time and efforts. The continuous improvement process is enhanced through the evaluation and refinement of test items, ensuring the assessment stays relevant and aligned with evolving learning objectives and student needs. Moreover, a well-designed and balanced assessment with appropriately retained and revised items contributes to a positive testing experience for students, providing a fair opportunity to demonstrate their knowledge and skills.

The decisions presented in Table 4 have significant implications for assessment quality, instructional design, and student learning experiences. By thoughtfully retaining, revising, and discarding test items, educators can create a more robust and effective assessment that supports accurate measurement of student performance and facilitates targeted instructional improvements.

4. Conclusions and Recommendations

In this study, the aim was to analyze the test performance of BSIT students and the quality of test questions in the SIA course to establish a foundation for enhancing the assessment tool. The research design employed a cross-sectional approach involving 200 fourth-year students enrolled in the course. The test questions used were created by one of the researchers and had been previously employed in a term examination. The results indicated a significant disparity in scores between the upper and lower student groups, suggesting the need for targeted interventions, curriculum improvements, and assessment enhancements, especially for those in the lower group. Regarding the item difficulty index of the assessment tool, the results revealed a requirement to refine certain items to better suit a broader range of students. However, most items were found to be appropriately aligned with their difficulty levels. An analysis of the item discrimination index indicated that 25 items could be retained, 27 items needed revision, and 3 could be discarded based

on the analysis results. This analysis provides valuable insights for improving the tool, thereby optimizing its effectiveness in assessing students' acquired knowledge.

Based on the study's findings, several recommendations can be made to enhance the educational experience for BSIT students.

Firstly, given the significant performance gap between the upper and lower student groups, it is imperative to introduce targeted interventions for students in the lower group. These interventions could include supplementary tutoring, mentoring programs, or academic support initiatives to help bridge the performance divide.

Secondly, to address the disparities in knowledge and skills, it is advisable to revise the curriculum in a way that better caters to the distinct learning needs of students. Customizing teaching techniques, materials, and assessments can create a more inclusive educational environment.

Additionally, the analysis of the item difficulty index highlights the importance of refining certain test items to ensure they are appropriate for a wider range of students. Regular assessment of item discrimination can also contribute to the ongoing improvement of assessment tools by retaining high-discrimination items and revising or discarding low-discrimination ones.

Also, further innovation can be achieved by integrating predictive analytics to forecast students' performance based on question indices, enabling preemptive educational interventions. Moreover, conducting similar studies could also be explored in the future.

For similar studies to be conducted, it is recommended to include a discussion on contrasting viewpoints or debates within the field of educational assessment. By incorporating such discourse, educators and researchers can enrich the depth of understanding surrounding assessment practices, fostering critical analysis and promoting well-rounded perspectives in educational assessment.

It is recommended to incorporate a comparative analysis with other courses or disciplines to highlight the specificity and applicability of the findings. By conducting such an analysis, educators and researchers can better understand the unique aspects of the assessment practices within the context of higher education. This comparative approach not only enhances the depth of understanding but also provides valuable insights for improving assessment strategies and methodologies across different disciplines.

Exploring the scalability of this approach in larger educational settings or different academic programs is recommended. By implementing similar assessment methodologies in diverse educational contexts, institutions can assess the effectiveness of their assessment tools across a broader spectrum of students and courses. This exploration could offer valuable insights into the applicability and adaptability of the assessment strategies employed in the BSIT program to other disciplines or educational institutions. Additionally, it presents an opportunity to identify best practices and areas for improvement in assessment practices on a larger scale, ultimately enhancing educational quality and student outcomes across various academic programs.

Encouraging educators to participate in professional development programs focused on assessment design and data-driven instruction can further enhance assessment practices. Implementing longitudinal assessment practices can help monitor students' progress over time and evaluate the effectiveness of interventions and curriculum refinements. Collaborative efforts among faculty members, curriculum designers, and assessment specialists are essential for refining and aligning the curriculum with learning objectives.

Finally, establishing feedback mechanisms for both students and educators can provide valuable insights into areas requiring improvement and promote a student-centered learning environment. These recommendations collectively aim to narrow performance gaps, improve education quality, and create a more supportive learning environment for all BSIT students.

The findings and recommendations from this study have several important implications for the educational institution, curriculum developers, faculty members, and the BSIT program as a whole:

- The significant performance gap identified between upper and lower student groups highlights the urgency of providing targeted support to struggling students. The institution should allocate resources and develop programs to ensure that students in the lower group receive the necessary assistance. This could involve offering remedial courses or creating peer support networks.
- The need to revise the curriculum to cater to diverse learning needs suggests that the BSIT program should adopt a more flexible and adaptable approach to teaching. Faculty members should be encouraged to utilize diverse teaching techniques and materials that can be tailored to individual students' learning styles and abilities. Additionally, the potential for this research to inform the development of more adaptive and personalized assessment tools could be emphasized.
- Regular assessment of item difficulty and discrimination is essential for improving the quality of assessments. The institution should establish a system for ongoing evaluation of test items and assessments, involving collaboration between faculty members and assessment specialists. This process will ensure that assessments accurately measure student knowledge and skills.
- To implement effective assessment practices and align curriculum with learning objectives, faculty members should be encouraged to participate in professional development programs focused on assessment design and data-driven instruction. This investment in faculty development will ultimately benefit students by improving the quality of teaching and assessment.

- Delving deeper into how these findings could influence future educational policy and assessment practices is imperative. The insights gained from this study offer valuable implications for shaping educational policies and refining assessment practices not only within the BSIT program but also across the broader educational landscape. By carefully considering the implications of these findings, policymakers and educational stakeholders can make informed decisions regarding curriculum development, assessment design, and student support initiatives. Furthermore, leveraging these insights can foster a culture of continuous improvement in educational institutions, ensuring that assessment practices align with educational objectives and effectively evaluate student learning outcomes. Thus, by exploring the implications of this study, educational policymakers and practitioners can contribute to the enhancement of educational quality and the promotion of student success on a systemic level.
- Collaborative efforts among faculty members, curriculum designers, and assessment specialists are essential for refining and aligning the curriculum with learning objectives. By bringing together diverse perspectives and expertise, these stakeholders can ensure that the curriculum reflects the latest industry standards, technological advancements, and pedagogical best practices. Encouraging cross-departmental collaboration and communication fosters a more holistic approach to curriculum development and assessment. It allows for the identification of interdisciplinary connections and opportunities for integrated learning experiences.

In general, the implications of this study emphasize the importance of a holistic and data-driven approach to education. By addressing the performance gap, refining assessments, customizing the curriculum, and investing in faculty development, the institution can work towards providing a high-quality educational experience that benefits all BSIT students and prepares them for successful careers.

References

[1] N. J. Rao, & Banerjee, S. “Classroom Assessment in Higher Education”. Higher Education for the Future, Vol. 10, No. 1, pp. 11–30. 2023. DOI:10.1177/23476311221143231

[2] W.Y. Lee. Assessment and Evaluation in Higher Education. MDPI eBooks. 2023. DOI:10.3390/books978-3-0365-6751-8

[3] Yixuan Wang. Country’s higher education assessment system. Frontiers in Educational Research, Vol. 4, No. 11. 2021. DOI:10.25236/fer.2021.041113

[4] Dawn Brown, Nathan Johnson. The importance of assessment and evaluation in higher education information technology projects. Lecture Notes in Computer Science. 2020. DOI:10.1007/978-3-030-60152-2_18

[5] Matea Hanžek, Zdravka Biočina, Maja Martinović, Valentina Pirić. Disruptive times and higher education in economics and management: importance of dual assessment. International Scientific Conference EMAN. Economics & Management: How to Cope with Disrupted Times. 2021. DOI:10.31410/eman.2021.215

[6] Tayseer Mansour, Shereen A. El Tarhouny, Manal Azab. Assessment of Multiple-Choice Questions Test Item Quality After Post Exam Item Analysis. International Journal of Scientific Research, Vol. 11, No. 6, pp. 32-34. 2022. DOI:10.36106/ijsr/1622478

[7] Hatim Lahza, Tammy G. Smith, Hassan Khosravi. Beyond item analysis: Connecting student behaviour and performance using e-assessment logs. British Journal of Educational Technology, Vol. 54, No. 1, pp. 335–354. 2022. DOI:10.1111/bjet.13270

[8] Muhammad Afzaal. Theoretical Framework and Methodology. In Corpora and Intercultural Studies, pp. 39–57. Springer Nature Singapore. 2023. DOI:10.1007/978-981-19-9619-1_3

[9] Scott Monroe. Item Response Theory. Item Response Theory. Routledge. 2022. DOI:10.4324/9781138609877-ree61-1

[10] Ashraf, Z.A & Jaseem, K. Classical and Modern Methods in Item Analysis of Test Tools. International Journal of Research, 2020.

[11] Yunxiao Chen, Xiaoou Li, Jingchen Liu, Zhiliang Ying. Item Response Theory -- A Statistical Framework for Educational and Psychological Measurement. arXiv: Methodology, 2021. DOI: 10.48550/arXiv.2108.08604

[12] Dale Whittington. Item Analysis. Item Analysis. Routledge. 2022. DOI:10.4324/9781138609877-ree211-1

[13] Maya Marsevani. Item Analysis of Multiple-Choice Questions: An Assessment of Young Learners. English Review: Journal of English Education, Vol. 10, No. 2, pp. 401-408. 2022. DOI:10.25134/erjee.v10i2.6241

[14] Suleiman Sa’adu Matazu, Elizabeth Julius. Item Analysis: A Veritable Tool for Effective Assessment in Teaching and Learning. Journal of Education and Practice. Vol. 12, No. 21, pp. 22-28. 2021.

[15] Dharmendra Kumar, Raksha Jaipurkar, Atul Shekhar, Gaurav Sikri, V. Srinivas. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Medical Journal, Armed Forces India, Vol. 77, No. 1, pp. S85–S89. 2021. DOI:10.1016/j.mjafi.2020.11.007

[16] ASA Husin. Case Study of Difficulty Index and Discrimination Index Item of Year 5 English Language Summative Test (Comprehension)/ Kajian Kes Indeks Kesukaran dan Diskriminasi Item Ujian Sumatif Bahasa Inggeris (Pemahaman) Tahun 5. Sains Humanika. 2020. DOI:10.11113/sh.v12n2-2.1786

[17] Rapita Pratiwi, Septia Reflianti, Sura Antini, Ahmad Walid. Analysis of item Difficulty Index for midterm examinations in Junior High Schools 5 Bengkulu City. Asian Journal of Science Education, Vol. 3, No. 1, pp. 12–18. 2021. DOI:10.24815/ajse.v3i1.18895

[18] L. Tomak, Y. Bek, M. Cengiz. Graphical modeling for item difficulty in medical faculty exams. Nigerian Journal of Clinical Practice, Vol. 19, No. 1, pp. 58. 2016. DOI:10.4103/1119-3077.173701

[19] Cheng Shu-Chen, Chen Guan-Yu, Pan I-Chun. An Estimation Method of Item Difficulty Index Combined with the Particle Swarm Optimization Algorithm for the Computerized Adaptive Testing. Proceedings of the Annual Conference of JSAI. 2013. DOI:10.11517/pjsai.jsai2013.0_2c4ios3c6

[20] Y.S. Wijaya. Comparison of Item Difficulty Index National Exams Package (Analysis Using Item Response Theory). 2013
[21] Abdullah Faruk Kılıç, Ibrahim Uysal. To what extent are item discrimination values realistic? A new index for two-dimensional structures. International Journal of Assessment Tools in Education, Vol. 9, No. 3, pp. 728–740. 2022. DOI:10.21449/ijate.1098757
[22] Jari Metsämuuronen. Generalized Discrimination Index. International Journal of Educational Methodology, Vol. 6, No. 2, 237–257. 2020. DOI:10.12973/ijem.6.2.237
[23] Lihong Song, Wenyi Wang. An Attribute-Specific item Discrimination index in cognitive diagnosis. In Springer proceedings in mathematics & statistics, pp. 169–181. 2019. DOI:10.1007/978-3-030-01310-3_16
[24] Novi Maulina, Rima Novirianthy. Item Analysis and Peer-Review Evaluation of Specific Health Problems and Applied Research Block Examination. Jurnal Pendidikan Kedokteran Indonesia. Vol. 9, No. 2. 2020. DOI:10.22146/jpki.49006
[25] CY Fook, GK Sidhu, FW Yunus, AA Hamid, RA Hashim. Super Specific Tool for Item Analysis: A Preliminary Study, Vol. 3. Current Proceedings on Technology. 2013
[26] Johny Wu, Kevin King, Katie Witkiewitz, Sarah Jensen Racz, Robert J. McMahon. Item analysis and differential item functioning of a brief conduct problem screen. Psychological Assessment, Vol. 24, No. 2, pp. 444–454. 2012. DOI:10.1037/a0025831
[27] Chan Yuen Fook, Farhana Wan Yunus, Gurnam Kaur Sidhu. Teachers' knowledge on item analysis and item analysis software. In 2013 IEEE Conference on e-Learning, e-Management and e-Services, pp. 30-33. Kuching, Malaysia. 2013. DOI:10.1109/IC3e.2013.6735961
[28] John R. Dickinson. A New Statistic for Item Analysis. Developments in Marketing Science: Proceedings of the Academy of Marketing Science, 206. 2014. DOI:10.1007/978-3-319-11761-4_97
[29] Moushir M. El-Bishouty, Ting Wen Chang, Kinshuk, Sabine Graf. A framework for analyzing course contents in learning management systems with respect to learning styles. In G. Biswas, et al. (Eds.), 20th International Conference on Computers in Education (ICCE 2012), pp. 91–95. 2012.
[30] Byung-Yeol Park, Todd Campbell, Miriah Kelly, Ron Gray, Chester Arnold, Cary Chadwick, Laura M. Cisneros, David Dickson, David M. Moss, Laura Rodriguez, John C. Volin, Michael R. Willig. Improving NGSS focused model-based learning curriculum through the examination of students' experiences and iterated models. Research in Science & Technological Education, Vol. 41, No. 3, pp. 983–1007. 2021. DOI:10.1080/02635143.2021.1978962
[31] Leoniek Wijngaards-de Meij, Sigrid Merx. Improving curriculum alignment and achieving learning goals by making the curriculum visible. International Journal for Academic Development, Vol. 23, No. 3, pp. 219–231. 2018. DOI:10.1080/1360144x.2018.1462187
[32] Manori Amarasekera, Paul S. Noakes, Brian D. Power. Reviewing the preclinical curriculum in a Problem Based Learning driven medical program: challenges and strategies. MedEdPublish, Vol. 8, No. 1. 2019. DOI:10.15694/mep.2019.000005.1
[33] Jeffrey L. Ezell, Diane Lending, S.E Kruck, Thomas W. Dillon, Jeffrey L. May. A Plan to Improve Learning of Requirements Elicitation in an IS Curriculum. Conference: The 2016 ACM SIGMIS Conference. 2016. DOI:10.1145/2890602.2890621
[34] Olusegun A. Sogunro. Quality instruction as a motivating factor in higher education. International Journal of Higher Education, Vol. 6, No. 4, pp. 173. 2017. DOI:10.5430/ijhe.v6n4p173
[35] Jika Smutna, Radim Farana. Understanding the quality concept in the higher education. Acta Montanistica Slovaca, Vol. 15, pp. 54-57. 2010.
[36] Gavriel Meirovich, Edward J. Romar. The difficulty in implementing TQM in higher education instruction. Quality Assurance in Education, Vol. 14, No. 4, pp. 324–337. 2006. DOI:10.1108/09684880610703938
[37] Ignacio Gil Pechuán, M. Pilar Conesa García, Antonio Navarro García. Concept mapping to improve higher education. In Springer eBooks, pp. 61–73. 2014. DOI:10.1007/978-3-319-04825-3_7

Authors' Profiles

Dr. Cris Norman P. Olipas is a full-time college IT instructor in the College of Information and Communications Technology at the Nueva Ecija University of Science and Technology. He holds a Bachelor's and Master's degree in Information Technology and a Doctorate degree in Education major in Industrial-Technological Education. He has published several research articles in local, national, and international journals. He also serves as an editorial board member and reviewer for various internationally renowned journals. His research interests include impact assessment, information technology education, and technology in education.

Dr. Ruth G. Luciano is a full-time college IT instructor at the Nueva Ecija University of Science and Technology College of Information and Communications Technology located in Cabanatuan City, Nueva Ecija, Philippines. She has been in the teaching profession for 20 years now. She worked as a private school instructor for 17 years and in 2018 she joined the NEUST's teaching workforce. She is also a lecturer at the MSIT program of the Graduate School of the same University.

How to cite this paper: Cris Norman P. Olipas, Ruth G. Luciano, "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(IJITCS), Vol.16, No.3, pp.1-11, 2024. DOI:10.5815/ijitcs.2024.03.01