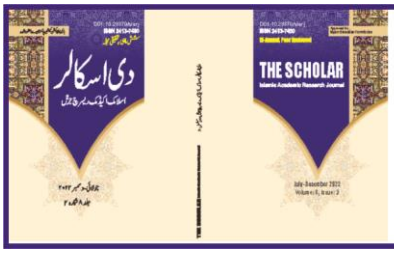

THE SCHOLAR
ISLAMIC ACADEMIC RESEARCH JOURNAL
ISSN: 2413-7480 (Print)
ISSN: 2617-4308 (Online)
DOI: 10.29370/siarj
Journal home page: https://siarj.com

TEACHERS' RELIGIOUS RESPONSIBILITY IN SMOOTH ASSESSMENT: AN ANALYSIS OF ITEM DISCRIMINATION OF EXAMS CONDUCTED AT UNIVERSITY LEVEL

1-Sheeza
Lecturer English Works Program, University of Education, Faisalabad
Email: msf2000362@ue.edu.pk
ORCID ID:
https://orcid.org/0000-0002-4420-5257

2- Asif Iqbal
Associate Professor of Education, University of Education Lahore
Email: asif.iqbal@ue.edu.pk
ORCID ID:
https://orcid.org/0000-0003-1488-8197

3-Aniqa Riaz
Lecturer Department of Humanities and Linguistics, University of Agriculture Faisalabad
Email: Aniqa.riaz@gmail.com
ORCID ID:
https://orcid.org/0009-0008-9527-5504

To cite this article:
Sheeza, Asif Iqbal, and Aniqa Riaz. "TEACHERS' RELIGIOUS RESPONSIBILITY IN SMOOTH ASSESSMENT: AN ANALYSIS OF ITEM DISCRIMINATION OF EXAMS CONDUCTED AT UNIVERSITY LEVEL." The Scholar Islamic Academic Research Journal 9, no. 2 (November 24, 2023)
To link to this article: https://doi.org/10.29370/siarj/issue17ar3

Journal	The Scholar Islamic Academic Research Journal Vol. 9, No. 2||July –December 2023 || P. 33- 58
DOI:	10.29370/siarj/issue17ar3
License:	Copyright c 2017 NC-SA 4.0
Journal homepage	www.siarj.com
Published online:	2023-11-24
Journal Indexed by:	DOAJ |AIL| Almanhal |National Library of Australia |Academia, |DRJI |WorldCat |SCILIT |Gale| The Internet Archive|10-A Digital Library||Harvard Library E-Journals|-Library | University of Ottawa| ScienceGate| NAVER Academic, Asian Digital Library| Tehqeeqat, |SEMANTIC SCHOLAR | Publon| Repository |Globethics| EuroPub database| Cornell University Library | Advanced Sciences Index

TEACHERS' RELIGIOUS RESPONSIBILITY IN SMOOTH ASSESSMENT: AN ANALYSIS OF ITEM DISCRIMINATION OF EXAMS CONDUCTED AT UNIVERSITY LEVEL

Sheeza, Asif Iqbal, Aniqa Riaz

ABSTRACT:

Assessment is an essential part of all learning systems. Consideration of discrimination is a crucial step in test formation. No assessment may be claimed as standard without this important technique. The purpose of this study was to examine the discrimination index of exams conducted at university level. The assessment of item discrimination in semester exams administered at the university level is explored in this descriptive study. Convenient sampling techniques were used to collect data for analysis from a sample consisting of 12 mid-term papers of social sciences disciplines. To evaluate the test items' ability to distinguish between students who perform well and those who do not, the research focuses on analyzing the differences in performance of individual test items. Exam items' discriminatory power was assessed using the statistical technique of the item discrimination index. The results provided important information about how well exam items measure students' comprehension and knowledge in a range of academic settings. Moreover, they lacked validity as they were unable to discriminate between high and low achievers. Recommendations derived from the analysis suggested to improve the quality and discriminatory ability of examination items for more accurate evaluations in the academic setting.

KEYWORDS: Discrimination index, Item analysis, Students' assessment

1. Introduction

Humans are Allah's creatures, with certain roles based on their nature and temperament. Humans are considered spiritual creatures because of their fitrah (Ar-Rum: 30) creation, inclination towards kindness and truth, and yearning to hold and profess the truth of religion (Islam). Kadir [1]asserts that actual daily conduct must express religiousness, not just theoretical understanding, or normative concepts. The realization of religious components, such as knowledge, belief, attitude, and action, indicates a person's commitment to their faith. Religious components are interrelated, their manifestations are non-linear and vary among individuals[2].

Islam is a faith based on moral aspects that guide people through all aspects of life[3]. One of Islam's special institutions is accountability, which makes the community's social structure more accountable for achieving the goals and objectives based on the religion's prescribed and revealed practices. In its purest form, accountability refers to holding public servants accountable to a higher authority for their actions[4]. Islamic administrative accountability refers to moral leadership in which a Muslim administrator or ruler acts in the best interests of the Islamic welfare state,

---

[1] Kadir, H. M. A. (2003). Applied Islamic Science: Initiating the Practice Paradigm in Islam. Student Library and STAIN Kudus.

[2] Fuaduddin, Bisri, & Hasan, C. (2002). ISLAMIC RELIGIOUS EDUCATION, Dynamics of Islamic Thought in Higher Education: Discourse on Islamic Religious Education/editors Fuaduddin & Cik Hasan Bisri (Vol. 2002). Jakarta: Logos.

[3] Chandio, A. R., Iqbal Ameen Brohi, S. A., Ali, M., & Gadhi, G. N. (2019). Administrative ethics with special reference to Islam. *GSJ, 7*(7).

[4] Chandio, A. R. (2021a). Public Administration and its Principles: With Special Reference to Islam. *Public Administration, 5*(5), 41- 48.

holding them accountable, deserving, and responsible for God's rules. We must first deem someone accountable for their decisions, actions, and authority. The Holy Quran has highlighted the depth of accountability and uses the word "Hesab" to describe it more than eight times in various passages[5].

Teachers are challenged to apply assessment techniques appropriately. An evaluation of students' learning improves the teaching-learning process as a part of the assessment. The connection between assessment and learning is acknowledged by university academia who participate in curriculum development. Therefore, a crucial component of successful curriculum development is creating a comprehensive assessment plan[6]. Tests with multiple choice options are the most effective type of written evaluation since they are legitimate and dependable due to their thorough covering of course material. It is the most preferred objective technique for knowledge domain evaluation for evaluating students across all disciplines. Item analysis is a technique for evaluating an exam's quality by examining its items. Item difficulty, item discrimination, and distracter analysis are useful essential components of

[5] Askary, S., & Clarke, F. (1997). Accounting in the Koranic Verses. Proceedings of International Conference, The Vehicle for Exploring and Implementing Shariah Islami'iah in Accounting, Commerce and Finance, Macarthur: University of Western Sydney.

[6] Alsubaie, M. A. (2016). Curriculum Development: Teacher Involvement in Curriculum Development. *Journal of Education and Practice, 7*(9), 106-107.

test production and course evaluation that employ statistical approaches to offer indications that measure the quality of tests[7].

A test's success is determined by two parameters. How well a test assesses the predetermined goals is the first consideration, followed by the question of whether or not each test item can distinguish between brilliant and average pupils. A test cannot be deemed effective if it cannot distinguish between pupils and measure the stated objectives in a meaningful way. Therefore, checking the test items and choosing just those that are sufficient, trustworthy, and genuine in order to achieve the objectives is a crucial duty for test developers. Item analysis refers to the entire procedure. Regarding item analysis, two factors should be prioritized while doing item analysis: first, the difficulty level of the items, and second, the discriminating power of the test items. The item-analysis approach must be used before the test's final format is created. The elements that make up a test determine its characteristics, including mean-variance, score distribution, reliability, and validity. Item analysis is the process used to evaluate an item's quality. Test analysis is a broad word that refers to a variety of notes on statistical and analytical methods that can be used to enhance tests, exams, quizzes, and other mental assessment instruments[8].

---

[7] Musa, A., Shaheen, S., Elmardi, A., & Ahmed, A. (2018). Item difficulty and item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*, *11*(2), 1477-1486.

[8] Cechova, I., Neubauer, J., & Sedlacik, M. (2014, October). Computer-adaptive testing: Item analysis and statistics for effective testing. In *European Conference on e-Learning* (p. 106). Academic Conferences International Limited.

Effective test items may be distinguished between performers and non-performers and have the right amount of difficulty. The quality of test items may be determined using an item analysis based on the difficulty index, discriminating index, and distractor efficiency. The ease or complexity of a task is affected by hints and mistakes. The items on teacher-generated exams must thus be analyzed; by eliminating the things that were incorrectly and ineffectively answered, the test may be made more efficient. It's interesting to look at distractor performance on examinations made up by teachers. High-quality items require well-written options, with effective distractors. To be efficient, these options should be based on common misconceptions. Functional or efficient distractors are chosen by more than 5% of examinees. As a result, decreasing the number of non-functional distractors and framing reasonable distractions increase test quality. Since teacher-created tests are frequently used for evaluation, it's critical to analyze the test items' quality and save high-quality answers.

MCQs are used to evaluate a wide range of material and objectives on a large number of pupils. Not all academics support the application of this format. Few admit that creating MCQs to assess higher cognitive skills takes time and is difficult. Others contend that MCQs prioritize memory over understanding, application, and analysis of course-related material. Sadly, it has been found that in medical and nursing contexts, the percentage of exam items evaluating factual and conceptual concepts is significant. Over 91 percent of the MCQs utilized for educational evaluation in a baccalaureate nursing program were of the recall/ comprehension level categories. Tests must be well-designed, accurately reflect the cognitive processes being measured, and be error-free to

considered legitimate. Items with errors can't be used to evaluate the target domains, which lowers the test's efficacy. Furthermore, when the items are faulty, the performance of high achievers is not fairly reflected in comparison to the borderline pupils. By adhering to the item-writing guidelines and analyzing the item's performance, teachers may decrease item defects and enhance item quality.

The degree to which passing an exam as a whole equates to passing a particular item is known as item discrimination. Any item with negative or zero discrimination weakens the test since all test items are meant to work together to get a final test result. Unless it is so high that the item essentially replicates the information given by previous test items, positive item discrimination is often beneficial. The test's high and low-scoring groups are divided into equal-sized groups to calculate the discrimination index. Divide the difference between the number of successes by the high group and the number of successes by the low group on the item by the size of the group. It denotes that the high-ability candidates are those who have a high ability to properly answer more questions, while the low-ability candidates are those who have a low capacity to correctly answer questions (Setiawan, 2014). Nevid and McClelland (2013) found that students in a psychology course struggled to respond to questions of assessment and explanation at higher cognitive levels of Bloom's taxonomy and that these questions stood out the most between students with high and low performance. The discrimination index ranges from 1.00 to +1.00; when item discrimination is high, the best students tend to respond properly while the weaker students do so more frequently. If the coefficient is near zero, both groups of students have an equal chance of providing the correct response, and if it is close to one, only the less capable students are able to do so while the more capable ones cannot. The

causes of this might vary, but very poor item discrimination is frequently caused by imprecise test question design, necessitating their editing or complete removal from the exam.

The discrimination analysis of the test items for the application and synthesis/evaluation levels, however, was greater for the online and on-site test items than it was for the remembering and understanding levels. This study aims to determine whether the discrimination indices of test items, used in exams for various social sciences disciplines at the University of Education Faisalabad campus, vary based on cognitive levels.

1.1. Objectives

The objectives of this research are given below:

1. To find the discrimination level of test items in exams.
2. To determine the effectiveness of papers for assessing students' performance.
3. To estimate overall comparative performance as well as low and high-achiever groups of university students appearing in a multiple choice-based examination of the Department of English

1.2. Research Questions

This study answered the following questions:

1. What is the discrimination level of test items in exams conducted by the disciplines of Social Sciences?
2. How to determine the effectiveness of papers regarding assessing students' performance?
3. How to estimate the overall comparative performance of high as well as low-achiever groups of university students appearing in a multiple-choice examination of English subjects?

1.3. Delimitations

The data were collected from one university only i.e., University of Education, Faisalabad, Pakistan. It was not practically possible to collect a wide range of data from different universities due to restricted resources. Also, there are various assessment techniques to analyze a student's performance, however, this study used only two techniques to assess a learner's performance. Moreover, the topic of difficulty and discrimination index has received less attention in history as only a few studies are available that analyzed limited populations. Out of multiple studies, only a few determine both assessment techniques and their effects on learner achievement.

2. Literature Review

Assessment has a significant impact on what students learn, how they learn, and how successfully they learn because the methods used to evaluate pupils are the most important of all effects on their learning. The quantity of work examined and the caliber of the assessment kinds are two factors that must be taken into account. This research examines several assessment strategies as well as how evaluations affect how students approach their coursework. There are no significant variations in evaluation techniques[9].

Evaluation involves several different aspects at once. Simple dualities like categorization and diagnosis are not the topic here[10].

[9] Peters A., Bijan P., & Uli, S. (2018). *Generating multiple choice questions from ontologies*: Lessons learned. In OWLED, pages 73–84. Citeseer.

[10] Siddiqi, A., & Nguyen, Q. C. (2009). A cross-national comparative perspective on racial inequities in health: the USA versus Canada. *J. Epidemiol. Community Health, 64*, 29–35.

A system for characterizing assessment settings at the level of an engineering program that can discriminate between susceptible, average, and talented students was developed by the researcher. Also included was an evaluation experience form that could gauge how well the pupils were learning. This research followed both approaches to investigate various programs. A crucial part of the educational process is assessment. In schools, the most visible assessments are summaries. Summative assessments are used to measure what students have learned by the end of a unit, to motivate students, and to ensure that they meet the criteria necessary for certification of school completion or for entry into certain jobs or to select jobs. For admission to higher education. Ministries or departments of education can use summary assessments and evaluations as a way to allocate publicly funded schools to provide quality education[11].

Evaluation and judgment are used interchangeably in this study to refer to assessments of student performance and assessments of a program's or organization's efficacy. The use of data to inform instructors about upcoming national initiatives to inform and adapt to the classroom or policies should always be thought of as secondary levels of formation assessment. This study is particularly interested in how administrators and instructors develop or reinforce assessment cultures. In the evaluation culture, educators provide fresh data on the material students use and why

[11] Logie, C., James, L., Tharao, W., & Loutfy, M., (2013). Associations between HIV-related stigma, racial discrimination, gender discrimination, and depression among HIV-positive African, Caribbean, and black women in Ontario, Canada. *AIDS Patient Care STDS* 27, 114–122.

it is effective, share their findings with colleagues, and help students become more adept at meeting a variety of learning goals. An evaluation culture refers to the development of a shared language for learning and teaching purposes and a shared understanding of the purpose of evaluation to achieve these goals[12].

Data are frequently gathered and published, but what good does that do? You need to take action with the findings. As a result, assessment frequently focuses on enhancing students' learning while simultaneously giving a department or program the chance to showcase their achievements. It may be applied to enhance educational possibilities for pupils. It can also be used to recruit new students for the program. To ensure that everyone is moving in the same direction, it is crucial for instructors to collaborate on the definition of learning outcomes. Having a chat might be helpful with evaluation projects at times. Understanding how a course fits within the larger syllabus may be determined by evaluating the teacher. What skills, abilities, knowledge and attributes do they display as a result of participating in the program? There can be a mixture of direct and indirect evidence, which is usually measured using techniques such as surveys and exit questionnaires.

[12] Seng, J. S., Lopez, W. D., Sperlich, M., Hamama, L., Meldrum, C. D. R. (2012). Marginalized identities, discrimination burden, and mental health: Empirical exploration of an interpersonal-level approach to modeling intersectionality. *Social Science Med*icine, *75*, 2437–2445.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

According to Frost[13], The variables that define evaluation settings are discovered to differ greatly, and how we evaluate children has a significant influence on their learning. It is well recognized that the level of critical thinking and the style of problem-solving evaluation positively affect the caliber of learning results. Furthermore, syllabus coverage, amount and quality of feedback, use of feedback, test preparation, appropriate assessment, objectives, explicit criteria, and an in-depth approach to learning are all related to high-level formation-only assessment and high-level feedback.

Although most teachers choose items that reflect the variety of subject matter covered in the class, they frequently neglect to consider the difficulty or grade of the questions they employ. Unfortunately, multiple-choice assessments are frequently prepared merely for identification or memory in instructor manuals and textbooks[14]. Multiple-choice questions (MCQs) are utilized more often to evaluate student performance. As a result, there is rising doubt over the validity of the assessment exams. MCQs are used to evaluate a wide range of material and objectives on a large number of pupils. The usage of this format is not endorsed by all instructors. Few acknowledge that creating MCQs to assess higher cognitive skills takes time and is difficult. Others contend that MCQs

[13] Frost, D. M. (2017). The benefits and challenges of health disparities and social stress frameworks for research on sexual and gender minority health. *Journal of Social Issues, 73*, 462–476.

[14] Kato, K., Moen, R., & Thurlow, M. (2009). Differentials of a state reading assessment: Item functioning, distractor functioning, and omission frequency for disability categories. *Educational Measurement: Issues and Practice, 28*(2), 28–40.

stress memory over understanding, application, and analysis of course-related material. Over 91 percent of the MCQs utilized for educational evaluation in a baccalaureate nursing program were of the recall/comprehension level categories. Tests must be well-designed, accurately reflect the cognitive processes being measured, and be error-free in order to be considered legitimate. Items with errors can't be used to evaluate the target domains, which lowers the test's efficacy. Additionally, when the items are faulty, the performance of high-achieving children is not accurately reflected in contrast to the borderline pupils. By sticking to the guidelines for item drafting and performance analysis, teachers may minimize item defects and maximize item quality[15].

Effective MCQs can distinguish between performers and non-performers and have the right amount of difficulty. The quality of MCQs may be assessed using an item analysis based on the difficulty index (DIF I), discriminating index (DI), and distractor efficiency (DE). Errors and hints affect how easy or difficult a MCQ is. There is a widespread consensus in the literature that adequate intervention is necessary when assessing (scoring) MC questions; otherwise, the dimensions of the individual will be overstated, and the item parameters will be skewed. According to one study, elimination scoring and correction for scoring are effective approaches to lessen the bias brought on by conjecture[16].

[15] Bock, R. D. (2002). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29-51.

[16] Wu, Q., De Laet, T., & Janssen, R. (2018). Elimination scoring versus correction for guessing: A simulation study. In M. Wiberg, S.

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

3. Research methodology:

The research is descriptive. A quantitative approach-based cross-sectional survey was conducted to collect data from Social Sciences disciplines at the University of Education, Faisalabad Campus. The population of the research study was all the teachers who conducted mid-term exams in 2023 at the University of Education, Faisalabad campus. Moreover, the data was collected manually from mid-term papers of the students of bachelor and masters appearing in the examination of different social sciences courses offered where each paper consisted of six MCQs. Then excel sheets were developed by applying the calculation formulas of discrimination and difficulty indices and the mean score was calculated. The data taken from Mid-Term exams was uploaded on Excel sheets to calculate the discriminating and difficulty efficiencies of each item in a course along with an interpretation of data.

3.. Interpretation of Discrimination index

The discriminating index of each question item was analyzed and interpreted by the researcher using the following table.

Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), Quantitative psychology (pp. 183–193). *Springer*.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

Discrimination Index	Item Evaluation	Recommendation
Negative	Worst/Defective Item	Discard
<0.20	Marginal Item, not discriminating	Revise/Discard
0.20-0.29	Moderately discriminating, fair item	Keep
0.30-0.39	Discriminating item, good item	Keep
>0.04	Very discriminating, very good item	keep

4. Data analysis and interpretation

This section deals with the detection of the Discrimination index and difficulty index of mid-term papers. The tabulation of data is evaluated through frequency analysis.

Table 4.1: ***Faculty of Social Sciences Code 1***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	1.00	Very Easy	-0.06	Deletion
Q2	0.89	Very Easy	0.18	poor item
Q3	1.00	Very Easy	-0.06	Deletion
Q4	1.00	Very Easy	-0.06	Deletion
Q5	0.97	Very Easy	0.00	Deletion
Q6	0.97	Very Easy	0.00	Deletion
Total	0.97	Very Easy	0.00	Deletion

According to table 4.1, the paper was constructed poorly because it showed 0.00 mean discrimination value which showed it didn't discriminate among higher and low achievers.

Table 4.2: ***Faculty of Social sciences Code 2***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.85	Very Easy	0.10	poor item
Q2	0.40	Difficult	0.20	fair item

Q3	0.85	Very Easy	0.10	poor item
Q4	0.50	Average	0.80	Very Good item
Q5	0.20	Very Difficult	0.20	fair item
Q6	0.25	Difficult	0.30	Good item
Total	0.51	Average	0.28	fair item

Mean: According to table 4.2, the mean value of the discrimination index depicts that grossly the paper had been fairly constructed as its discrimination index value is 0.28[17].

Table 4.3: ***Faculty of Social sciences Code 3***

Q#	**Difficulty Index (p)**		**Discrimination Index (D)**	
Q1	0.96	Very Easy	0.00	Deletion
Q2	0.78	Easy	0.38	Good item
Q3	0.56	Average	0.38	Good item
Q4	0.56	Average	-0.08	Deletion
Q5	0.96	Very Easy	0.00	Deletion
Q6	0.63	Easy	0.54	Very Good item
Total	0.74	Easy	0.21	fair item

Mean

With the reference of table 4.3, it can be seen that the paper was distinguishing between exceeding and emerging students of the class as its discrimination index is showing that paper fairly differ among high and low ability students with mean discrimination value of 0.21.

Table 4.4: ***Faculty of Social sciences Code 4***

[17] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.62	Easy	0.52	Very Good item
Q2	0.93	Very Easy	0.17	poor item
Q3	0.82	Very Easy	0.30	Good item
Q4	0.36	Difficult	0.09	poor item
Q5	0.76	Easy	0.52	Very Good item
Q6	0.69	Easy	0.48	Very Good item
Total	0.70	Easy	0.35	Good item

Mean

Conforming to table 4.4, it is analyzed that paper was easy with reference to its difficulty as it secured the difficulty index of 0.70. However, it was capable of distinguishing between high achievers and low achievers of the class by getting the mean discrimination index value of 0.35.

Table 4.5: ***Faculty of Social sciences Code 5***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.91	Very Easy	0.00	Deletion
Q2	0.95	Very Easy	0.04	Deletion
Q3	0.82	Very Easy	0.36	Good item
Q4	0.59	Average	0.09	poor item
Q5	0.73	Easy	0.18	poor item
Q6	0.91	Very Easy	0.18	poor item
Total	0.82	Very Easy	0.14	poor item

Mean

In accord with table 4.5, this paper was very easy as the average range of items would stand between value.41- .60[18]. The authenticity of any paper lies on its capability of being challenging and discriminating. According to the mean graph, the mean discrimination index had a calculated value of

[18] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

0.14. The calculations had shown that the paper developed poorly not keeping up with standards of test formation.

Table 4.6: ***Faculty of Social Sciences Code 6***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	1.00	Very Easy	-0.09	Deletion
Q2	1.00	Very Easy	-0.09	Deletion
Q3	0.95	Very Easy	0.00	Deletion
Q4	0.95	Very Easy	0.00	Deletion
Q5	0.95	Very Easy	0.00	Deletion
Q6	0.95	Very Easy	0.00	Deletion
Total	0.97	Very Easy	-0.03	Deletion

Mean

With the reference of table 3.38, it is concluded that all the items were very easy. According to the calculated mean value of difficulty index, it is extracted that the paper was an easy paper with a secured mean difficulty index of 0.97. This paper was very easy as the average range of items would stand between value.41- .60[19]. This shows that it is poorly formulated or constructed. As per the mean discrimination index is concerned, the paper should be revised or deleted as its mean discrimination index is -0.03. The paper was not able to differentiate among high achievers and low achievers as most of the low achievers marked the right options instead of high achievers.

Table 4.7: ***Faculty of Social Sciences code 7***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.65	Easy	0.59	Very Good item

[19] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

Q2	0.79	Easy	0.29	Good item
Q3	0.97	Very Easy	0.06	Deletion
Q4	0.91	Very Easy	0.06	Deletion
Q5	0.88	Very Easy	0.24	fair item
Q6	0.76	Easy	0.12	poor item
Total	0.83	Very Easy	0.23	fair item

Mean: According to the table 4.39, it has been concluded all the 6 MCQs questions of the paper this were quite easy. Majority of the students were able to mark the correct option of the each MCQ. This paper was very easy as the average range of item would stand between value.41- .60[20]. As it displayed in the mean chart, the mean calculated value of discrimination index is 0.23. Overall, the paper was poorly challenging and discriminate among high and low achievers.

Table 4.8: ***Faculty of Social sciences Code 8***

Q#	**Difficulty Index (p)**		**Discrimination Index (D)**	
Q1	0.89	Very Easy	0.15	poor item
Q2	0.89	Very Easy	0.15	poor item
Q3	0.63	Easy	0.54	Very Good item
Q4	1.00	Very Easy	-0.08	Deletion
Q5	1.00	Very Easy	-0.08	Deletion
Q6	1.00	Very Easy	-0.08	Deletion
Total	0.90	Very Easy	0.10	poor item

Mean

According to table 4.8, the mean value of discrimination index is 0.10. This concluded that items were easy and inefficient to distinguish high ability learners and low ability learners.

[20] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.86	Very Easy	0.18	poor item
Q2	0.43	Average	0.09	poor item
Q3	0.90	Very Easy	0.09	poor item
Q4	0.48	Average	0.73	Very Good item
Q5	0.29	Difficult	0.36	Good item
Q6	0.29	Difficult	0.55	Very Good item
Total	0.54	Average	0.33	Good item

Table 4.9: ***Faculty of Social Sciences Code 9***

Mean

According to table 4.9, the calculated difficulty value showed that overall paper was an average paper with the mean difficulty value of 0.54. The mean chart depicted the mean discrimination value of 0.33. This showed that the paper was well constructed according to the parameters of test development and discriminated the high and low achievers.

Table 4.10: ***Faculty of Social Sciences Code 10***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.86	Very Easy	0.18	poor item
Q2	0.95	Very Easy	0.18	poor item
Q3	0.86	Very Easy	0.18	poor item
Q4	0.76	Easy	0.55	Very Good item
Q5	0.90	Very Easy	0.27	fair item
Q6	0.76	Easy	0.55	Very Good item
Total	0.85	Very Easy	0.32	Good item

Mean

According to table 4.35, most of the questions were easy as were attempted by the majority of students. In accord with the mean chart of discipline the discipline has a good calculated mean value of discrimination index. Few of the items like question 1, 2 and 3 were

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

poorly constructed. However, on the whole paper was able to differ among a variety of learners.

Table 4.11: ***Faculty of Social Sciences Code 11***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.86	Very Easy	0.29	fair item
Q2	0.89	Very Easy	0.21	fair item
Q3	0.79	Easy	0.43	Very Good item
Q4	0.79	Easy	0.43	Very Good item
Q5	0.89	Very Easy	0.21	fair item
Q6	0.96	Very Easy	0.07	Deletion
Total	0.86	Very Easy	0.27	fair item

Mean:

With the reference of table 4.11, all the items were easy. This depicts that the paper was not challenging, which is the core characteristic of any paper. It can be seen from the mean chart that the mean value of difficulty index is 0.86. However, the ideal value of difficulty must lie among 0.40-0.60[21]. The mean value of the discrimination index is 0.27. In a nutshell, the paper was too easy as the majority of learners were able to attempt it. This declared that it was not developed in accordance with the standards of assessment and paper development criteria.

[21] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

The Scholar Islamic Academic Research Journal
Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

Table 4.12: ***Faculty of Social Sciences Code 12***

Q#	Difficulty Index (p)		Discrimination Index (D)	
Q1	0.87	Very Easy	0.36	Good item
Q2	0.91	Very Easy	0.27	fair item
Q3	0.91	Very Easy	0.27	fair item
Q4	0.57	Average	1.00	Very Good item
Q5	0.96	Very Easy	0.18	poor item
Q6	1.00	Very Easy	0.09	poor item
Total	0.87	Very Easy	0.36	Good item

Mean

With reference to Table 4.12, all the questions were easy. The mean value of the discrimination index is 0.36. The average value of the discrimination index must be 0.40 or more than it. In a whole, the paper was able to discriminate between high and low achievers, but it was too easy.

5. Findings

The major findings of the study are described in the following section.

1. According to table 4.1, it is concluded that the paper of ''Study of Al Quran' was constructed poorly as it showed a mean discrimination value of 0.00 which concluded it didn't discriminate among higher and low achievers and needed deletion.
2. According to Table 4.2, the mean value of the discrimination index depicted that grossly the paper had been fairly constructed as its discrimination index value is 0.28. The paper was constructed according to the standards of test formation and assessment was also done on standards.
3. With reference of Table 4.3, it can be seen that the paper was distinguishing between exceeding and emerging students of the class as its value of mean discrimination index is 0.21.

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

4. Conforming to table 4.4, it is analyzed that paper of 'General Science' was capable of distinguishing between high achievers and low achievers of the class having the mean discrimination index value of 0.35.
5. In accordance with table 4.5, it had been analyzed that in the course 'Contents and methods of teaching' 82 % of the learners were able to mark the correct answer which clearly showed that the question paper was too easy. According to the mean graph, the mean discrimination index had a calculated value of 0.14 so it should be discarded. The calculations had shown that the paper was developed poorly not keeping up with standards of test formation.
6. With the reference of table 4.6, it is concluded that the mean discrimination index is negative. The paper should be revised or deleted as its mean discrimination index is -0.03. The paper was not able to differentiate among high achievers and low achievers as most of the low achievers marked the right options instead of high achievers.
7. According to table 4.7, the paper of 'Educational Psychology' calculated the mean value of discrimination index was 0.23. Overall, the paper was very easy or poorly challenging and was able to discriminate among high and low achievers.
8. According to table 4.8, the mean value of discrimination index is 0.10. This concluded that items were easy and inefficient to distinguish high ability learners and low ability learners. The paper should be revised or discarded with reference to its inability to

differ among high and low achievers[22].

9. According to table 4.9, the calculated values showed that the overall paper of 'Inclusive education' was an average paper with a mean difficulty value of 0.54. The paper had a mean discrimination value of 0.33. This showed that the paper was well constructed according to the parameters of test development and discriminated against the high and low achievers.
10. According to table 4.10, this paper was very easy as the average range of items would stand between value.41- .60 (Sharma, 2021). The discipline has a good calculated mean value of discrimination index 0.32. Overall, the paper was very easy considering its difficulty and was able to discriminate between learners of different abilities.
11. With the reference of table 4.11, the mean discrimination value of the paper of 'Guidance and counseling' was 0.27. In a nutshell, the paper had performed fairly to differ among high achievers and low achievers and the paper was too easy as the majority of learners were able to attempt it. This declared that it was not developed in accordance with the standards of assessment and paper development criteria.
12. With the reference of table 4.12, the mean calculated value of the discrimination index was 0.36. In a whole, the paper was able to discriminate between high and low achievers but it was too easy so it should be revised with reference to its difficulty.

[22] Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple-choice questions of speech sounds of English. *International Research Journal of MMC (IRJMMC),* 2(1), 15-28.

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

6. Conclusion:

The calculation of the difficulty index, discrimination index, and distractor efficiency during item analysis is a typical method used to determine the reliability and validity of an item or test following the assessment. The majority of the study's question items did not meet the standards for fair discrimination which showed that the MCQs selected were of low quality. The purpose of the study was to evaluate the discrimination index of university-level midterm exams. Data for each research table was gathered from faculty members, and calculations were made on Excel sheets using descriptive analysis. The study reveals that the papers constructed by teachers did not satisfy the needs of assessment as they were not developed according to the standards of test formation. This fact proved a hurdle in meeting the standards of assessment. The study concluded that the majority of the papers taken at the university level are easy regarding the mean difficulty index. A negative correlation between the difficulty and discrimination index showed that the discrimination index decreased as the difficulty index increased. The discrimination index drops as the exam items get simpler, making it impossible to distinguish between poor and strong pupils. Though many papers fairly discriminate among high and low achievers, it is not satisfactory as the worth or functionality of any assessment tool can only be measured by its ability to differentiate

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

high and low achievers. The instructor should develop the papers by assessment needs and requirements.

7. Recommendations:

The following suggestions are made in light of the findings to address the issue and help students become more competent through evaluation procedures:

1. Teachers should be dedicated to use a variety of evaluation and paper development techniques that encourage student's creativity, productivity, and knowledge.
2. Assessment is a crucial step in the educational process; thus it must be carried out with great caution and attention. The concerns with assessment techniques in the classroom must be carefully considered by educational professionals, and they must take action to continually raise the standard of the practice and upgrade it.
3. Systems for Question Banks should be developed. In this approach, a significant number of questions from each topic or unit of the syllabus are prepared in advance, and when the moment arrives for the exam or test, a smaller number of questions from each topic or unit of the curriculum are drawn from the pool. They should be printed separately, and these questions are used to help administer the test. Making a question bank must be a routine exercise in that a set number of various types of questions are consistently created by specialists and

Vol. 9, No. 2 || July-December 2023 || P. 33-58
https://doi.org/10.29370/siarj/issue17ar2

added to the bank. All of these questions should be standardized using a rigorous process of item analysis, difficulty index, and discrimination index processes. So that before giving the test to the candidates, its overall reliability and validity may be determined.

4. For having a quality assessment process, the conduction of a faculty development program or pieces of training can be helpful to enhance the learning and performance of faculty for the development of new standardized MCQs.

5. future researchers could collect data from various universities of different cities to measure the validity of assessment procedures.