Item Analysis and Internal Reliability of Entry Tests for prospective Electrical Engineering and Computer Science students in a University Setting

By

Malik Shahzad Iqbal
NUST201463053MSEECS61414S

Supervised By

Ms. Farzana Ahmed
Department of ITE SEECS, NUST

This thesis is submitted in partial fulfillment of the requirements for the degree of Masters of Science in Innovative Technologies in Education (MS ITE)

School of Electrical Engineering and Computer Science (SEECS),
National University of Sciences and Technology (NUST), Islamabad, Pakistan

(January 2016)

Approval

It is certified that the contents and form of the thesis entitled **"Item Analysis and Internal Reliability of Entry Tests for prospective Electrical Engineering and Computer Science students in a University Setting** "submitted by **Malik Shahzad Iqbal** have been found satisfactory for the requirement of the degree.

Advisor: Ms. Farzana Ahmed

Signature: _____

Date: _____

Committee Member1: Dr. Aisha Mehmood

Signature_____

Date: _____

Committee Member2: Mr. Jaudat Mamoon

Signature: _____

Date: _____

Committee Member3: Ms. Erum Afzal

Signature: _____

Date: _____

Dedication

I dedicate this thesis work to my parents who from the start encouraged and supported me. Also to my advisor who is an endless source of motivation and guidance.

It is also dedicated to my brothers, sisters, and friends and to my teachers with whom I have an exceptional and valuable relationship.

Certificate of Originality

I hereby declare that the research entitled **"Item Analysis and Internal Reliability of Entry Tests for prospective Electrical Engineering and Computer Science students in a University Setting"** is my work and to the best of my knowledge it contains no materials previously published or written by any other person, nor material that to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my work, except for the assistance from others in the project's design and conception or in style, presentation, and linguistics that has been acknowledged.

Author Name: Malik Shahzad Iqbal

Signature: _____

Date: _____

Acknowledgment

First and foremost, praises be to Allah, the Almighty, on whom ultimately we depend for direction and guidance. I wish to express my deep sense of gratitude to my supervisor Ms. Farzana Ahmed and my Committee member and co-supervisor Dr. Aisha Mahmood for giving me a chance to work with her, in a research environment and polish my skills.

Words are inadequate in offering my thanks to the members of Advisory Committee for their valuable input whenever required. Finally, yet importantly, I would like to express my heartfelt thanks to my colleagues and fellow students who helped me directly or indirectly

List of Contents

List of Figure

List of Tables

ABSTRACT

Choosing a degree course at tertiary level of education is not an easy task. It requires careful analysis of one's desire for a particular field and the skillset that's required for that field of study. Due to growing population rate, it is becoming more competitive to get admission in reputable university. Entrance test is the primary tool that is used by most of the universities in Pakistan to filter out the students with desired skill and knowledge set and selects the best pool of students, across the country. For that reason, universities should continuously improve the process of selection and the construction of Entrance test items. The purpose of this research is to check the reliability of Entrance tests format that were consistently taken over last two years by a private sector university in Pakistan. We analyzed, difficulty index, discrimination index distractors functionality and distractor strength for all the items in last two year's data of about one thousand students. The Tests were paper based; Multiple Choice Questions (MCQs) type and the responses from answer sheets were extracted using Image processing techniques. Moreover, we analyzed the internal consistency of the test using Cronbach's alpha. The results of item analysis showed that the difficulty levels and discriminability are not homogeneous and a strong criticism emerged from the quality of distractors. The items with least reliability were removed and recommendations were given to improve the internal consistency, reliability and how to balance the distractors strength to make the test more convergent.

Chapter 1. Introduction

A nation's social prosperity and economic growth depends upon the human resource development of Engineers. Engineering colleges play an important role in this development process. Millions of students apply in the universities of Pakistan, and only a few of them get selected (UNESCO, 2012) based on a particular criteria. According to UNESCO report, total enrollment in the higher education in Pakistan is 1.319 million. Out of which 86% are in public and 14% are in private Universities. The Learning of the students and the results associated to that learning are based on how they got admitted to the University, A study by Chinese Education Society (Hsiao-Fang, 2012) shows that students who are selected through an admission/entrance test are far better than students selected through any other selection criteria. The practice of conducting the reliability entrance tests is lacking around the whole Pakistan. Most of the entry tests consist of the Multiple Choice Questions (MCQs). Despite of heavy criticism (Wood 1991) using these tests is still common. Many educational institutes in Pakistan uses the same MCQ's tests as the instrument for selection of the students. The main criticism about the MCQs is that they cannot measure the cognitive outcomes, as it can be done with subjective questions. Multiple Choice Questions (MCQs) is a technique that measures a specific concept and we should not expect it to measure overall capabilities of the students. (Wood, 1991).One of the main advantages that MCQs have on subjective tests is that they can test the certain domain extensively (Wiersma, 1990).

The candidates in the Engineering disciplines should be selected through credible, reliable, transparent and fair methods especially when we have a large number of eligible participants who can take entry test and get admission. The competition of getting admitted to good universities is increasing in the world including Pakistan. (Khan, Tabasum, & Mukhtar, 2013). Every year almost 35000 students who get more than 60% marks in Intermediate and are eligible to appear in the admission test in different universities of Punjab, Pakistan. So, in such a competitive environment and situation, the instrument (entry test) used to select the students for the university should be reasonably reliable. As mentioned above, most of the universities use the multiple choice question in the entry test, so the reliability of these tests is crucial. The reliability includes the item difficulty,

item discrimination power, distractor effectiveness, functional and non-functional distractors, items efficiency and the internal reliability of the items.

In this study, we have analyzed the entry tests of a university that requested to be kept anonymous. So henceforth, the said University have been referred to as Uni. A. The entry tests of the year 2013 and 2015 were analyzed. The study was conducted to help teachers and examiners of the Uni.A in constructing a reliable test and figuring out the reliability and efficiency of the previous items in the tests. Item analysis is used to calculate different properties related to items; the difficulty index, discrimination index, distractor analysis, functional and non-functional distractors and items efficiency. The internal consistency of the test is calculated through Cronbach's alpha. By the calculated measures, we extracted the strength and effectiveness of the items in the entry tests of 2013 and 2015.

1.1. Motivation

Pakistan has always been struggling in developing a coherent formal selection criteria in the educational system. Every university has their entry test, conditions, and procedure for selecting the students in different departments. Thousands of students apply for the admission in top universities of Pakistan, and only a few hundreds of them get selected every year. The applicants come from rural and urban areas of Punjab, Sindh, Baluchistan and Sindh. The main difference between those selected and rejected students is "The performance in the Entrance Test". Now the question arises, "Is that entrance test reliable to select or reject the students?" So the motivation behind this study was to examine the effectiveness and reliability of the entry test for Uni.A. In Uni.A, every year a large number of students apply from every corner of Pakistan; few of them got selected, so it is an interesting study to figure out the reliability of their rejections or acceptances?

1.2. Scope of the Study

The study have focused on the reliability of the entry test of the 2013 and 2015. This study have focused on the identification of problems in the test items, that effects the reliability of the items in particular and entrance tests in general. The entry test of 2013 includes data of 500 students while the entry test of 2015 has data of 800 participants who appeared in the test. The research is

conducted to get answers to the following questions; are the items very difficult or easy? Do they have the power to discriminate the high and low achievers? How many are functional or non-functional distractors? What is the efficiency of the questions?

1.3. Research Questions

1. To what extent the entry Test of Uni.A is reliable in selecting students for Electrical Engineering and Computer Science in the year 2013?
2. To what extent entry Test of Uni.A is reliable in selecting students for Electrical Engineering and Computer Science in the year 2015?

1.4. Research Hypothesis

1. Test items are in the acceptable range of "difficulty" for the year of 2013 and for the year of 2015 Entrance Tests.
2. Test items are in the acceptance range of "Discrimination Power".
3. Distractors of the items in entry test for the year of 2013 and for the year of 2015 are Effective.
4. Distractors are functional in the entry test for the year of 2013 and 2015.
5. The value of Cronbach alpha is greater than 0.6 for the entry test of 2013 and 2015

1.5. Structure of the Study

We have discussed the previous or related studies in Literature Review (Chapter 2); the methodology and research design have been discussed in Methodology (Chapter 3). The data analysis, and the different tests and techniques applied to the data have been discussed in the Results (Chapter 4). The discussion on the results have been discussed in the Discussion (Chapter 5). The next chapter have discussed Conclusion, Recommendations, References and Appendices.

Chapter 2. Literature Review

2.1. Item Analysis

The quality of the test can be decided by the quality of the items in the test.(Singamaneni, 2011) Item analysis is a statistical technique for instructors that helps them to identify the reliability and strength of the questions (items) in a test. Classical item analysis helps in improving the quality of tests by revising and improving the elements in the test. Item analysis plays a significant role in the fairness of the test when a teacher is developing quality assessment and especially effective multiple choice questions. Teachers can identify the content areas that are problematic for students, and are not able to measure the abilities of the pupils. Item analysis works more efficiently when the sample size of the examinees exceeds 50. Item analysis can be used to identify the mistakes in scoring, ambiguous questions and the distractors that are not functional. The main purpose of item analysis is to analyze the items by conducting test consisting of the MCQ's. After conducting the tests, responses for every question are analyzed. On the basis of the results question may be improved. In extreme case, where the acceptability of the questions is less than minimal, the questions may get discarded. Experts having an experience of relevant material carry out qualitative reviews during the development of items from different domain experts. Although these experts carry out a rigorous review during the development process, however there is still a possibility of non funcional items. This happens not only due to the negligence of review board experts, but mainly due to the different and multidisciplinary nature of the content on the test. Another reason can be different demographic characteristics of the those candidates who are appearing in the test. Afore mentioned reasons increases the possibility of bad test items that ultimately effects the reliability of the test. In such a case, statistical analysis comes into play, using this analysis the defective or problematic items on the standard test, which escaped the attention of an expert, can be identified in a multiple ways. Therefore after the successful conduction of the test, the quantitative analysis is applied on the responses of the test. The main objective of these assessments whether quantitative or qualitative is the same, which is assessing

the items' quality. One important point of realization is that there are multiple reasons for an item on a standardized test to fail in meeting the minimum quality standards, irrespective of what the set standards are. Most commonly these errors or flaws in test items come due to the following reasons

- The item has flaws in itself.
- The instructional content has flaws.
- The difficulty of the item is either too high or too low.
- The distractors of an item are weak.
- The item might not be differentiating the high and low achievers

2.2. Difficulty Index

Difficulty index is also known as item difficulty or p-values. The proportion of the students who answered the item correctly and the total number of students answering that question (Matlock-Hetzel, 1997). Item difficulty is defined as relative frequency with which those are taking the test, respond to the question with the right option. The difficulty index value is usually presented as the proportion in the range of 0.0 to 1.0. When the value is 0.0, it means no student has answered the question correctly, while when it is 1.0, it means that all the participants have answered the question correctly. In the same way, a value of 0.5 shows that half of the students answered the question right. The higher the difficulty index is, the easier the item is understood to be (Wood, 1960; Awodele, Faremi, & Adetunji, 2013) The items with moderate item difficulty are preferred to those having low or high difficulty index (Boopathiraj, 2013). To calculate the item difficulty, the sum of people answering the item correctly in the upper and lower group is divided by a total number of participants who attempt that question in the upper and lower group. The difficulty index is donated by **p** (Boopathiraj, 2013; Escudero & Reyna, 2000).

The following formula is used to find difficulty level. (Boopathiraj, 2013)

$$\mathbf{p} = (\mathbf{Ru} + \mathbf{Rl})/(\mathbf{Nu} + \mathbf{Nl})$$

Where,

Ru = the number of right answers in upper group

Rl = the number of right answers in the lower group

Nu= Number of students in the upper group

Nl= Number of students in the lower group

2.3. Discrimination Index

Tests are developed to assess the abilities of high and low achievers. If so then the participants who are good should perform better in the test and should score high and participants who are weak should perform low in the tests. Item discrimination can differentiate the high and low achievers, by what they know about the item. Item discrimination of a test refers to the degree to which success or failure of an item to differentiate the high and low achievers can be measured. It shows the power of an item to discriminate the examinees. (Boopathiraj, 2013).

So, the difference between the percentage of high achievers and the percentage of low achievers is called Discrimination Index. The high and low achievers are separated by calculating the scores of the participants, and the upper and lower groups are created. The upper group is created by 33% of the high achievers, and the lower group is created by, the lower 33% (Krishnan & Ph, 2013). The value ranges between 0.0 and 1.00. The higher value means the item has more discrimination power. A high discrimination value indicates that students who performed well in the test got this item correct, and the students performed badly in the test got this item incorrect. On the other hand if the discrimination value is small or even negative, it means that the examinees who performed well in the test got the item incorrectly, and the examinees who performed poorly in the test got that item right. Items having negative discrimination are rejected. If an item has discrimination index above .20 are regarded satisfactory for the use in any academic test or examination. (Aggarwal, 1986)

The formula for discrimination Index (DI) (Boopathiraj, 2013)is

$$\mathbf{DI} = \frac{\mathbf{RU} - \mathbf{RL}}{\mathbf{NU}\ (\mathbf{or})\mathbf{NL}}$$

Where,

DI = Discrimination Index

Ru = the number of right answers in upper group

Rl = the number of right answers in the lower group

Nu= Number of students in the upper group

Nl= Number of students in the lower group

2.4. Item Distractors and Distractor Efficiency

There is another important part of item analysis that is an analysis of item distractors. As distractors show an important relationship between the distractor chose and the total score of a participant, so they are very important part of item analysis. The performance of the participant is dependent on, how good the distractors are designed. (Dufresne RJ, Leonard WJ, 2000;Abdulghani, Ahmad, Aldrees, Khalil, & Ponnamperuma, 2014). To check the distractor effectiveness, that how good it has been designed to check student's concept or how bad it has been designed that it has no effect on the performance of the student, the concept we use is known as distractor efficiency. Any distractor, if chosen by less than 10% of the students is non-functioning distractor (NF-D). (Tarrant, Ware, & Mohammed, 2009) Ideally the students who are weak in a particular subject/concept should choose the distractor more frequently than the right answer from the distractors of an item. Distractor efficiency helps us to identify the errors in the distractors, and it becomes easy to revise, remove or replace the non-functional distractors (Gronlund NE, Linn RL, 1990; Hingorjo & Jaleel, 2012).

Tarrant and Ware stated an interesting stat about the flawed MCQs, according to them, flawed MCQs can effect the perferomance of high achievers more than the low achievers.Therefore the construction of good MCQ's (with maximum functioning distractor) is very important, and more importantly it helps to address students' concerns about getting acceptable grades in tests and help the faculty to differentiate between the high and low achievers. (Tarrant M, Ware J, 2008)

The Distractor Efficiency (DE) of an item can be calculated by calculating the Discrimination Index (DI) for every distractor of an item. First of all, lower (33%) and upper (33%) groups should be taken by arranging the score for an item in ascending order. The p and DI values should be calculated through the following formula.

$$\mathbf{DI_i} = \frac{\mathbf{Ru_i - Rl_i}}{\mathbf{Nu_i\ (or) Nl_i}}$$

Where,

DI = Discrimination Index for distractor i

Ru = the number of students selected i distractor in upper group

Rl = the number of students selected i distractor in lower group

Nu= Number of students in the upper group

Nl= Number of students in the lower group

After calculating the DI, the non-functioning distractors will be those selected by less than 10% of the students. Distractor Efficiency ranged from 0.0% to 100.0%. The DE can be determined by the number of the non-functional distractors in an item.(Hingorjo & Jaleel, 2012)

- If NF-D = 4 than DE= 0%
- If NF-D = 3 then DE = 25%
- If NF-D = 2 then DE = 50%
- If NF-D =1 then DE = 75%
- If NF-D = 0 then DE = 100%

2.5. Reliability

An unfair and distortion-free assessment can produce the efficient and effectives results. Reliability is a term that refers to the extent to which assessments produce the consistent results. There are four types of test/assessment reliabilities (Krishnan & Ph, 2013). 1) **Inter-Rater Reliability**: it is used to check the consistency when two different rater/observers give an estimation of the same phenomena. 2) **Test-Retest Reliability**: to assess the consistency when a test is taken at different times. 3) **Parallel-Forms Reliability**: used to access the consistency of the tests that are constructed at a different time but in the same way for same domain knowledge. 4) **Internal Consistency Reliability**: used to access the coherence of a single test that has been held one time.

2.5.1. Internal Consistency Reliability

Internal consistency is the reliability of the test components (test items/questions). It measures how consistently a set of questions measures a specific concept or behavior. Internal consistency reliability is used to measure the consistency of the instrument. The test is considered internally reliable when the items of the tests are strongly inter-correlated. The most well-known method used for testing the Internal Reliability is coefficient alpha. The usefulness of coefficient alpha was first recognized by Cronbach (1951), so now it is known as Cronbach Alpha.

2.6. Relevant Research around the World

Now we will discuss some of the relevant studies conducted previously.

2.6.1. Examination of the Quality of Multiple-choice Items on Classroom Tests (DiBattista & Kurzawa, 2011)

A Canadian study was carried out to assess the quality of the items that were used in classroom tests by using the technique of statistical item analysis. These tests were analyzed as the multiple choice questions based tests are widely used in higher education; this asks for the assessing of quality of items of questionnaires on a regular basis to ensure the quality of the tests. In this study, 240 different undergraduate courses were selected that were offered in the winter, fall and spring at University of Ontario, Canada. After that 12 random courses at each of the four levels within five faculties were selected. A letter was sent to the instructors inviting them to complete the survey. In the survey the instructors were asked different information about the test, and filled MCQs' response sheets. Thirty eight instructors returned the MCQs' response sheets, which were then sent to University Information Technology Service Department for optical scoring and scanning. After that, the difficulty index, discrimination index and distractor efficiency was calculated. The analytical analysis was done through SPSS 18.0.0 (SPSS Inc., Chicago Illinois). In the study, 1198 multiple choice questions were examined, the responses of undergraduate students were examined against these 1198 MCQs. These questions were administered to students in various disciplines in sixteen classroom tests. More than 30 percent of the items had unsatisfactory coefficient less than the standard of +0.20. Whereas the mean of the item discrimination came out to be +0.25. From the 3819 distractors, the 45 percent of these were flawed because of their positive correlation with test score rather than the negative correlation with the test score. Another reason of this 45% flaw was that, the examinees did not select more than 95 percent distractors and they selected only 5 %. The discrimination coefficient of more than 40

percent items in three tests was unsatisfactory, and half of the distractors were found un-functional. The study also revealed that a dramatic suffering in discriminatory power depending on the positive correlation of selection with the test scores, however for the distractors the selection was made by examinees (less than 10%) was minimally affected. Finally, the findings from the study suggest that there is quite a big room for improvement in the quality of the multiple-choice questions. Another suggestion the study put forward is that the teachers and instructors should improve the quality of multiple choice standard tests, and this is possible when the item analysis of these tests is carried out on regular basis. Additionally, this can be improved by the modification of distractors blamed for the impairment of the items' discriminatory power. Research Ethics Board of the University had reviewed and approved all the Results.

2.6.2. Analysis Of Test Items On Difficulty Level And Discrimination Index In The Test For Research In Education (Boopathiraj, 2013)

In this study, test items were analyzed that were designed by researchers for the student-teacher of Master of Education in the subject of Research in Education. Item difficulty and item discrimination were calculated for the tests. The data were collected through a test using multiple choice questions in different colleges of Education. A sample of 200 student-teachers taken randomly. The sample includes both gender. Calculating Item difficulty and item discrimination index of the Multiple Choice questions was the main objective of this work, as it helped to determine the quality of the MCQs.

The test composed of 60 MCQs items was used for the data collection. The test was developed under the supervision of Tamilnadu Teachers Education University by subject experts and researcher in the field of Education. The framework used for the test construction was Bloom's revised taxonomy. The English language was used to administer the test. The study was conducted for the students and teachers who were enrolled in the Master of Education in Tamilnadu University. These students and teachers were studying in different colleges that were associated with the university. The pupils and teachers were males and females. Researchers were happy and excited to conduct the test for the data collection and administration of the relative colleges of education supported them to the maximum extents.

After the successful conduction of the tests in different colleges, the tests were marked, and total scores of students were calculated. Microsoft Excel was used for data processing, the total

score of the students were entered in the sheet. The scores were then arranged in the descending order. The upper and lower groups of high and low achievers were created based on descending order. The item analysis was conducted on the created groups of the high and low achievers. The mediocre students were neglected as they behaved similarly to answer the items. The difficulty (p) and discrimination index (DI) were calculated through following formulae.

$$\mathbf{p} = \frac{\mathbf{Ru+Rl}}{\mathbf{Nu+Nl}} \qquad \mathbf{DI} = \frac{\mathbf{RU-RL}}{\mathbf{NU\ (or)NL}}$$

The test items should not be too easy so that every student can answer it, and shouldn't be too difficult so that no one can answer it. Similarly, the test items should effectively discriminate the high and low achievers and measure the level of the expertise and grip of knowledge, in particular, concepts. These both concepts are interrelated, so both of them are important for any test to be a useful measuring instrument. The items with difficulty index between 21% and 81% are good to be part of the test (Singh. Y.K. 2012). Researchers had chosen the items according to the above criteria. Seven items had discrimination power of 80%, and they were the best items in the test. Following table states the items that were selected by difficulty index and discrimination index.

Table 2.1 Results of Difficulty and Discrimination Index ((Boopathiraj, 2013)

	Not Good	Average	Good	Excellent
Difficulty Index (DI)	16%	35%	42%	7%
Discrimination Index	-	19%	20%	46%

Most of the questions were falling in the range of the difficulty and discrimination index, so they got accepted but few questions got rejected. The difficulty index and discrimination index of 13 questions out of 60 questions were not up to the mark, so they were rejected. The questions that were accepted without any revision or rewriting were 35 out of 60. There were few questions (1-2) that were accepted after the revision or rewriting.

This study is very helpful for the students and teachers and even more helpful to the people who are responsible for the test development and conduction. Teachers can conduct the similar study in their classrooms to improve the tests and can create a pool of effective and good questions for the upcoming years. All the stakeholders of test development should be very careful as there are

other important factors that can affect the efficiency of the tests for example; Test length, item difficulty and discrimination indexes ranges and the test purpose. The main function of a test in the educational field is to measure the student's abilities and concepts, so these types of studies can help teachers to make the assessment more effective and reliable.

2.6.3. The Entry Test To The Degree Course In Science Of Education (Marzano, Tammaro, & Notti, 2012)

The admission test of a university can be of different types, for example, survey tests, analytical tests or selective tests. The entry test of a university will determine the new students who will be studying the courses next year as there are hopelessly limited seats available for admission to any university. There is vital need to check the reliability and validity of the entry tests. An unreliable test will produce unreliable results, (Steven et al., 1990; 1991) which will not only affect the students but also the faculty. So, to fulfill the need and importance of the reliability of the entry test, every possible measure should be taken to build reliable and valid tests.

This piece of work was conducted in Italy, the objective of the research was to check the reliability and the validity of the entry tests that were designed for the course in Science of Education. The analytical analysis was run on the entry tests and their items. It was examined, 'Is there any problem with the reliability if the tests?' After the analysis, some suggestions were made to make entry tests more effective and consistent to check the concepts for what they were conducted. The permission was taken from the authorities for the study above, after that the question papers in the electronic form to the students and the answer sheets with the results were attained. The statistical analysis was run on the results of the entry test to check its reliability and validity and internal consistency. The test was selective, and it was not a difficult test, many unreliable items were part of the entry test. The results showed that the participants of the test were facing more difficulties in two areas "Literature and Linguistics" and "Geography". The distractors' quality was not up to the mark as many non-functional distractors were part of the test.

The study was conducted for the entry tests of the academic year 2008/2009 for the courses of Science of Education in University of Salerno (Italy). The participants in the entry test 2013 were 1133. The test was used to identify the weak and strong students in different areas, and an interesting statement written on the top of the test was "This test is for both of us (students and

faculty) to know the areas where we are lacking, so don't use the improper means to attempt the test." The following results emerged from the data.

Male Students = 6.75%

Female Students = 93.25%

The entry test had 80 multiple choice questions, and every question has three distractors and one right answer. For the right answer students were awarded 1 point and 0.25 points were detected, if student gave the wrong answer. 80 questions were divided into four areas. Geography has 12 questions, Linguistics and Literature part has 20 questions History part has 15 questions, and General Culture has 33 questions.

The data was pre-processed using Excel sheets and further it was analyzed in SPSS 19.0 using the modified data after the pre-processing. Item analysis was used to check the reliability and validity of the items. The difficulty index, discrimination index, distractor analysis and internal consistency of the test were calculated. The difficulty index can be defined as the resistance to select the right answer from the available options. The difficult index is directly proportional to the numbers of wrong answers for a specific item. The difficulty index will rise as the number of wrong answers increased. It is possible to figure out that either the question is very easy or difficult, either the distractors of the questions are effective, or non-functional, are the questions good discriminates between the high and low achievers. The value of the difficulty index can be between 0 (means all the students gave wrong answers) and 1(all the students choose right options). The value of the discrimination index can be between -1 and +1. The discrimination index allows the teachers to distinguish between the students with good skills and grip on knowledge from the students who are weak in studies. Cronbach's alpha was used to calculate the internal consistency of the test. The value of the Cronbach's alpha can range between 0 and +1. The criteria for Cronbach's alpha was

If $\alpha > 0.81$, then the internal consistency is ideal

If $\alpha > 0.71$ and $\alpha < 0.81$ then the internal consistency is very good

If $\alpha > 0.61$ and $\alpha < 0.71$ then the internal consistency is satisfactory

If $\alpha < 0.6$ then the internal consistency is problematic

The item reliability can be calculated by multiplying the discrimination index and difficulty index. The item reliability index values lie between -1 and 1. The item will be reliable if the values go in the positive range. This can only be calculated for that item; those have the acceptable difficulty index and discrimination index values. The results of the following study are mentioned below. Researchers were interested in investigating the quality of the test of the academic year 2008/2009 for the courses of Science of Education in University of Salerno (Italy).

Table 2.2 Item Analysis Results (Marzano, Tammaro, & Notti, 2012)

	Parameters	Values
Difficulty Index	Easy Items	33.75%
	Medium-Hard Items	66.25%
Discrimination Index	Bad-Average Discriminators	48.45%
	Good-Excellent Discriminators	51.25%
Reliability	Not reliable items	62.5%
	Reliable Items	37.5%

There were 33.75% items that were very easy and almost all the students gave the right answers to them. Other 66.25% items were good by difficulty index. 48.45% items were not able to discriminate the upper and lower students by their skills and knowledge. 51.25 items were reasonably good to discriminate the high and low achievers. The main concern was the reliability of the questions, 62.50% questions were under the bottom-line. Only 37.50% items were reliable. There was another concern about the quality of the distractors. The distractors used for an item should be equally weak or strong. The final result given by the researchers was that this test has the quality to be placed in the reliable rankings after some modifications.

2.6.4. Quantitative Analysis For Discrimination Index, Difficulty Index And Distractor Analysis of MCQs (Hingorjo & Jaleel, 2012)

In MCQs tests student are allowed to select one right option from the available three, four or five options. Most of the educational institutes are following these types of assessments. MCQs are very effective, in accessing large scale of the curriculum by minimizing the burden on the students in the assessment phase. Although it takes much time and effort of the examiners to design

MCQs based assessments instead of subjective tests. These tests are very effective in assessing the abilities and weaknesses of the students and helpful for the teachers to figure out the learning problems, in particular, area for a particular student. (Tan LT, McAleer JJ, 2008). If MCQ's items are properly constructed and designed they are best in assessing the high-level cognitive processes, like interpreting concepts, knowledge grip and reflection of the concepts.

Item Analysis is a simple and classical procedure that helps teachers to investigate the reliability and performance of a test item after the conduction of the test. It helps teachers and examiners to figure out, the difficulty index (means how much an item is difficult) and discrimination Index (how much an item is powerful to distinguish the high and low achievers) and distractor analysis (the analysis of the options other than the right one). The advantages of the item analysis are, examiners can modify the items, and even they can remove the items those are below power. Difficulty index (p-values) tells, how many students gave the right answers for the specific item, its value lies between 0% and 100%. The accepted range for the difficulty index 30% - 70%. If the value is under 30% or greater than 70%, then the items are either too easy or too difficult. It is suggested that easy items should be placed at the start of the test, and then items with moderate difficulty and finally the items with high difficulty index. Discrimination Index (DI) is the power of an item to distinguish high and low achievers; its values lie between -1 and +1. High achieving students select more right answers than the low achievers if the DI values in positive than it means that the students, who are high achievers selected the right answer more than the students who are low achievers. Similarly, if the DI value is in negative, then it means that the students with low scores have selected the right answers more than the high achievers. The discrimination and difficulty index are often related as inversely proportional, but it may not be always right, for example that questions having low difficulty indices can discriminate the students effectively and vice versa.

Distractor Analysis is also an essential part in Item Analysis. They are the key components of a MCQs. The designer of the MCQs item is directly related to the performance of the students. To check the distractor effectiveness, that how good it has been designed to check student's concept or how bad it has been designed that it has no effect on the performance of the student, the concept we use is known as distractor efficiency. Any distractor, if chosen by less than 10% of the students is non-functioning distractor (NF-D) (Tarrant et al., 2009). Ideally the students who

are weak in a particular subject/concept should choose the distractor more frequently than the right answer from the distractors of an item. Distractor efficiency helps us to identify the errors in the distractors, and it becomes easy to revise, remove or replace the non-functioning distractors. (Gronlund NE, Linn RL, 1990).

The study was conducted at Fatima Jinnah College for dentists, in Karachi for the year 2009. The students who participated in the studied were from the first year, and their age was around 17 to 20 years. After teaching different topics: kidney, nerves, muscle and blood, the test was conducted on the subject of physiology. The test consisted of 50 MCQs. The questions have possible five options, including one right answer. The time for the test was 60 minutes. The author calculated the p-values, DI values, distractor efficiency for the 60 test items. The answer sheets were provided to students to write their choices, one right answer was awarded 1 point, and a zero point was awarded to incorrect answers, the blank spaces were considered as wrong answers. So the maximum score for the test was 50 and minimum could be zero. After receiving the answer sheets, the scores were entered in the Excel sheets and arranged in the ascending order. The two group's high achievers and low achievers were created. The DI and p-values for every item were calculated by following formulae.

$$p_i = \frac{H + L}{N} \times 100$$

$$DI_i = 2 \times \left(\frac{H - L}{N}\right) \times 100$$

N is the sum of the students in upper and lower groups, L and H are the correct responses in lower and upper groups respectively. The ideal items were, those having p-values between 30 and 70 and Di value greater than 0.24. The distractors selected by less than 10% students were declared as non-functional. The distractor efficiency was calculated through the non-functional distractors NFDs. The values can be between 0% and 100%. If a question has 0 NFDs, then it has 100% distractor efficiency, and if it has 4 NFDs, then the distractor efficiency will be 0.

102 students appeared in the test, as mentioned earlier the test had 50 MCQS from the physiology. 27.31±5.75 was the mean score, as 50 could have been the maximum score. The mean score for the upper group was35.85 ± 2.09, for the middle it was 27.63 ± 2.05 and for the lower group it was 18.25 ± 3.26. The highest score was 40, and the lowest score was 6. The DI, DE, and

p-values were calculated for every item. There were 2 items that have negative DI's. The difficulty index of 2 items were 66 (mean they were very easy) and 21 (mean very difficult). The DE was calculated for the 50 MCQs/. 78% items were in good range as the mean of their p-values were 51.44 ± 11.11. Similarly, 62% items had very good DI as the mean for their discrimination index was (0.465 ± 0.083). There are very few items, having poor p and Di values.

The deeper study revealed that 64% (32 items) were the ideal items according to the criteria mentioned earlier. (Having p-value between 30 and 70 and DI value greater than 0.24). The numbers of the distractors for the 50 items were 200. 23.5% (47 distractors) were nonfunctional distractors. 58% (29 items) items had one or many non-functional distractors, 42% (21 items) had functional distractors. Items with 0,1,2 NFD had average p-value ranging from 44.38 to 62.66 and DI value ranging from 0.35 to 0.42 while items are having 3 NFD were showing the p-value of 77.5% and DI of 0.16.

Table 2.3 Difficulty Index of Items (Hingorjo & Jaleel, 2012)

P values	Criteria	Mean p-values	DE%	No of Items
Less than 30	Very difficulty	18.3	100	3 items
In-between 30 and 70	Average	51.4	81.4	39 items
Greater than 70	Very Easy	80.7	43.7	8 items

Table 2.4 Discrimination Index of Items (Hingorjo & Jaleel, 2012)

DI values	Criteria	Mean DI	DE%	No of Items
Less than 0.15	Poor	0.004	58.3	6 items
In-between 0.15 and 0.24	Modification needed	0.196	66.6	6 items
In-between 0.25 and 0.34	Good	0.285	71.8	7 items
Greater than 0.35	Excellent	0.469	83.0	31 items

Table 2.5 Item Analysis with respect to upper and lower groups (Hingorjo & Jaleel, 2012)

Item	Group	A	B	C	D	E	DI, p, DE
1	Lower Upper	0 0	3 0	1 1	23 27	23 27	DI = 0.14 p=89 DE=25%
2	Lower Upper	2 0	1 0	0 0	20 17	20 17	DI=0.1 p=66 DE=25%
13	Lower Upper	8 0	4 2	6 23	2 1	2 1	DI=0.57 p=54 DE=75%
43	Lower Upper	10 4	4 1	1 3	5 16	5 16	DI=-0.14 p= 21 DE= 100%

The conclusion of this study was that the item with 03 distractors are the best for discriminating the high and low achievers, and they can be more useful than 04 0r 05 distractors. Moreover, items with 2 non-functional distractors are good discriminators than an item with no non-functional distractors.

Distractor Analysis helps the teachers to check the response of the students for every single item. The table above showed the results of the distractor analysis. Total 4 question analysis is mentioned in the table. The first question in the table is actually the number one question in the entry test. The analysis of the first question shows that it is a very easy to question as 95% students had chosen the right answer. Moreover, the item couldn't discriminate the high and low achievers as a student from both groups had chosen the same answer from the options. The distractor efficiency for the item was 25%, means it has only 1 functional distractor. For the second item in the table, DI value was negative, it means that more students from lower group students selected the right answer than the students from the upper group. Distractor Efficiency for this was also

25% as distractor C and B were considered as non-functional distractors. The third MCQ in the table had a p-value of 54, Discrimination value of 0.57, and Distractor Efficiency of 75%. The values show that the item was sufficiently difficult, had the power to discriminate the students and had only one non-functioning distractor. The last item in the table had the p-value of 21; discrimination index is 0.14 and distractor efficiency of 100%. It shows that the item was very difficult, although there were no non-functional distractors, this might be due to wrong interpretation of the distractors.

Chapter 3. Methodology

In this chapter, we have described the design of the research and different techniques to extract data from the scanned entry tests sheets and the preprocess of the data before the statistical analysis.

3.1. Quantitative Research

Based on the analysis of the data, Social Science research can be categorized into three classes 1) Quantitative 2) Qualitative 3) Mixed Method. When we think of the quantitative research the statistics, numbers come in our mind. The definition of Quantitative Research by (Aliaga and Gunderson) is "Quantitative research is 'Explaining phenomena by collecting numerical data that are analyzed using mathematically based methods (in particular statistics)'." The first part of the definition in explaining the phenomena. This is the key part of any research, either it is quantitative or qualitative. For example, we want to explain why students are more interested in computer science field than any other? What are the key factors that are influencing the student achievements? , and many more questions like these. In our research, we have figured out the difficulty index, discrimination index and distractor power of the items in entry tests of a couple of years. Moreover, we have calculated the internal reliability of the items. So these are the few phenomena we want to explain or investigate.

The second part is about collecting data in numerical form. There are many situations in which, we might not get the quantitative data directly. For example, we want to gather data for students liking or disliking the classroom environment. We can design the questioner in a way that it should give us the quantitative data. For example a question in the questioner might be, classrooms are boring? Moreover, the possible answer to the question can strongly disagree, disagree, agree, and strongly agree. So we need to measure them in the form of numbers to get the numeric data, which can be like (1 for strongly disagree, 2 for disagree, 3 for agree, and 4 for strongly agree). Our data consists of the students reply to all the items of the entry tests. As an entry, test items have four or five option from which students have to select their answer. The possible options for every item were options A, Option B, Option C, and Option D. These options

were matched to the numbers. For Option A the number was 1, for Option B, it will be 2 for option C it is 3 and for option D, it is 4.

Once we have the quantitative data, we have to choose some statistical or mathematical methods that help us to analyze the data. This is the most important part of the research as this section gives us the results and help us to examine any phenomena. In our study, the item analysis have been used as a statistical method to analyze the data gathered from the answer sheets of the entry tests.

3.2. Image Processing

It was hard to type manually the responses of the students and then analyze them. The total students were around 1300, and every student answered 100 questions, so it makes 1300K responses that we have to type. We used the image processing to get the responses automatically. We were unable to use the available software to get the answers as the answer sheets resolutions were different in our case. So I have explained step by step what and how we extract the data or the responses of students from the data. The figure below shows the flow of the Image processing.

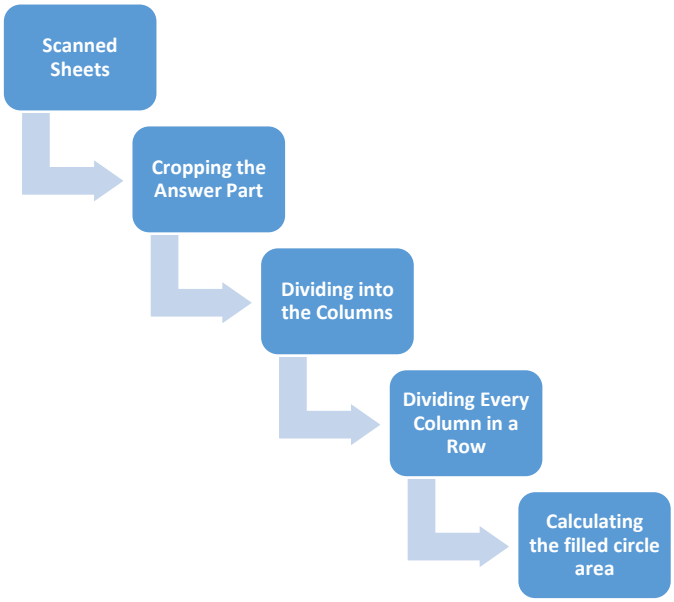


Figure 3.1 Image Processing Flow

Following was the answer sheet that was used to get the answers of the pupils.

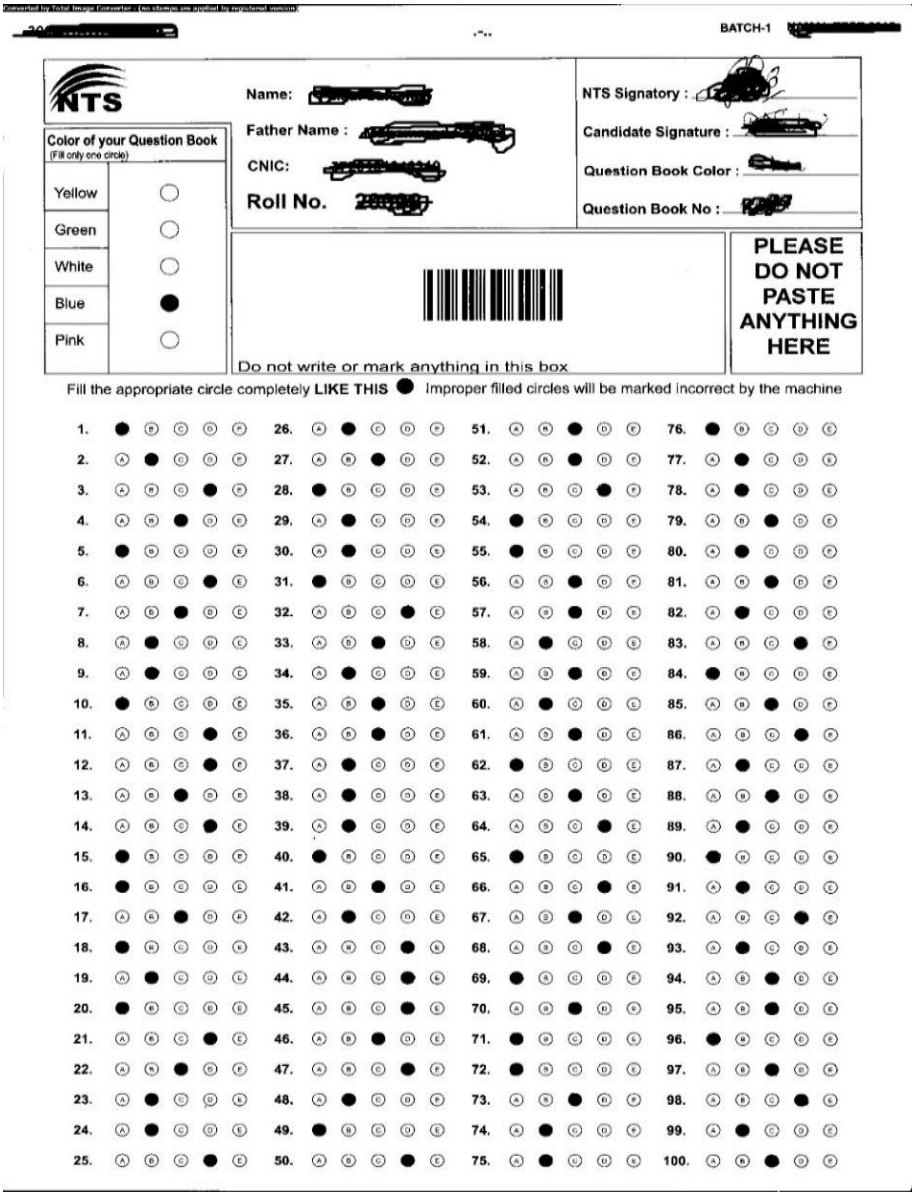


Figure 3.2 Sample Answer Sheet used in Entry Test of Uni.A

3.2.1. Cropping the Answer part

In the first step, we were interested in the answer part of the sheet, so we have cropped the answer part from this sheet so that we have correct part of the image. The data in which we were interested could be extracted from this part easily. The new image was following then.

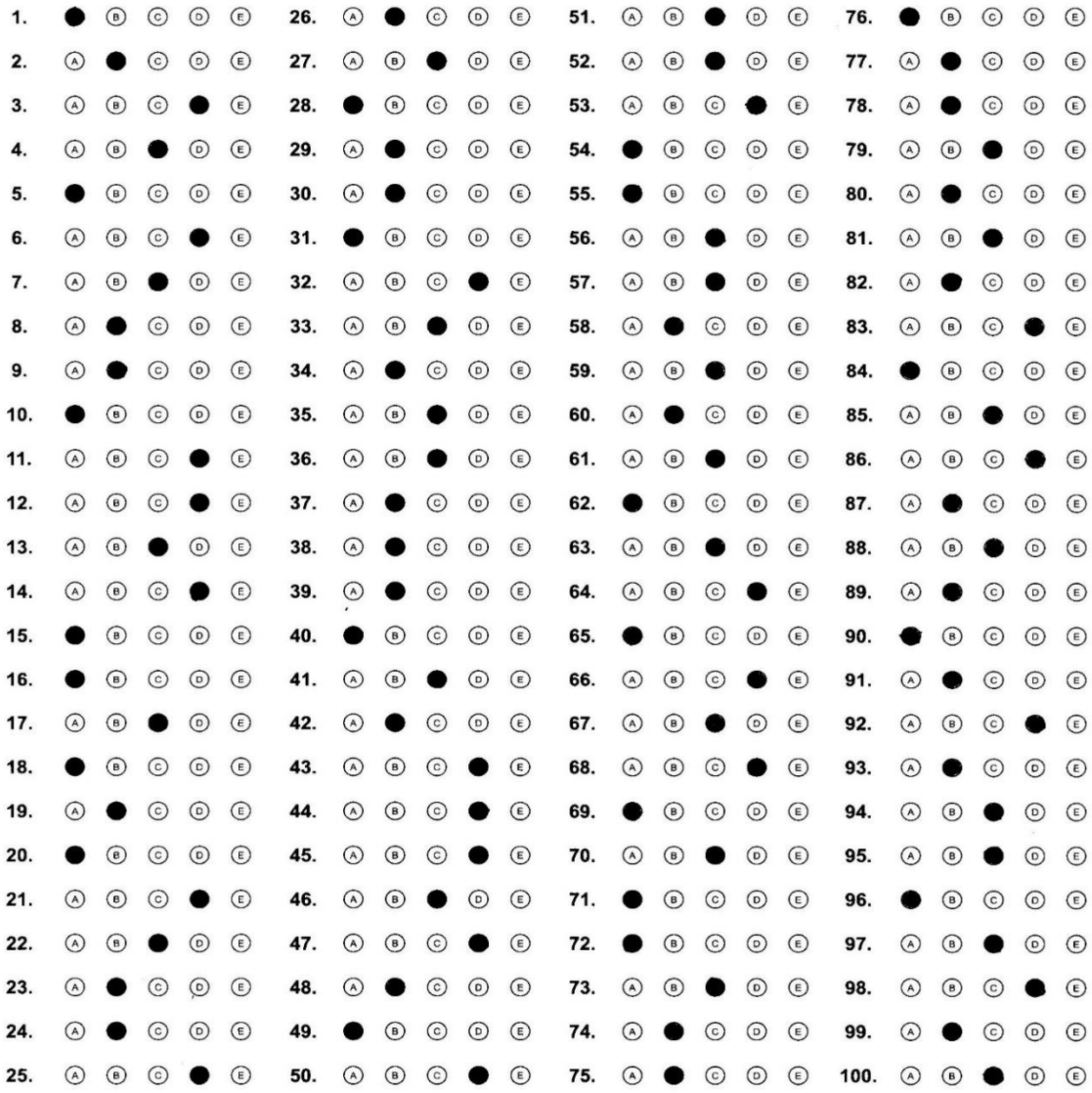


Figure 3.3 Answer Part of the Sample Sheet used in Uni.A

3.2.2. Answer Part Division

Once we have cropped the answer part from the image, we have four columns, and every column has 25 question's responses. In the next step, we have divided this picture into four images, every image is one of the columns from the four columns of the answer sheet.

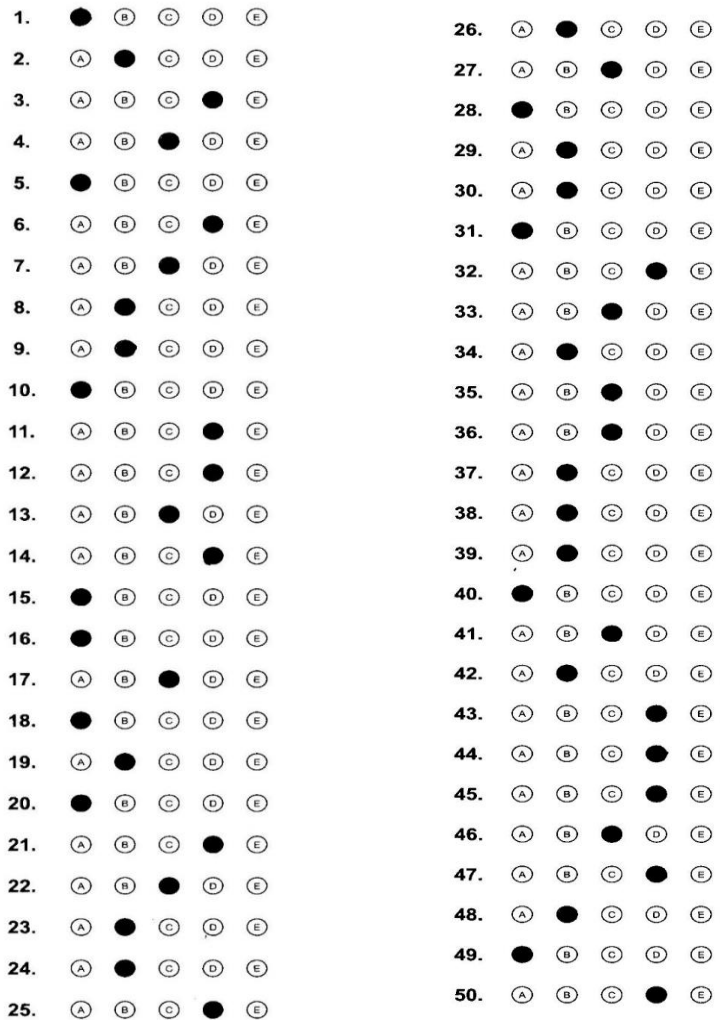


Figure 3.4 Column A and B of the Answer Sheet

3.2.3. Dividing the Column into Rows

The next thing that we have done was that we divided each column into 25 different rows, and each row contains the information about the single question. Now as, we divided the complete answer sheet and reached to single question data. Now it is easy to track the response to every question. The rows have been shown in the figure below.

Figure 3.5 Rows representing Questions in a Single Row

3.2.4. Detecting the values of filled and unfilled circles

As the students were expected to fill the circle with the answers, so if detect the filled circle we are done. We have taken smallest part of the image as shown above and figure out the filled circles. We manually set the circles and the pixels to figure out which circle is filled. For example, we have started from the left and started figuring out. If it is like 10 pixels from the beginning of the image, then it has to be option A's circle, and if it is 50 pixels away, it has to be option E's circle. We have stored the results of the filled or unfilled results in an excel sheet. Excel sheet had five columns; each column have stored value for one option. Excel sheet was like, as the table shown below.

Table 3.1 Values for the Filled and Un-Filled Circles

Q-No	A	B	C	D	E
1	150	120	**1650**	160	140
2	130	**1130**	130	140	130
3	110	140	140	**1560**	140
4	**1120**	130	130	150	150
5	120	130	**1590**	160	140

As shown in the table although these values are not coming from the an original sheet, the

table displays the examplary data and shows how Excel sheet looked like after detecting the filled circles.

3.2.5. Detecting the Right Option

As now we have the values for all the circles, we can detect the option selected by the candidate, as we can see from the table above that there was a significant difference between the values of the filled and unfilled circles. So there was a threshold of around 600, which was set to figure out the right answer. If the two options were selected by the students, it would be count as zero.

3.2.6. Image Processing Results:

The results for the exemplary sheet are shown in the table below. It has values of the filled area in the circles at the place of A, B, C, D and E. The next column shows the highest value of the possible options with the position in the next column. The last column was inserted manually to see the automated algorithm working correctly.

Table 3.2 Comparison of the Automatic and Manual Results for Answer Sheets

Questions	A	B	C	D	E	Highest Value	Automatic Generated Answers	Manual Answers Typed
1	1179	350	331	336	334	1179	1	1
2	330	1171	334	333	343	1171	2	2
3	328	343	334	1143	342	1143	4	4
4	331	352	1164	341	355	1164	3	3
5	1148	346	337	337	363	1148	1	1
6	329	358	327	1134	354	1134	4	4
7	328	364	1170	345	344	1170	3	3
8	333	1173	318	346	330	1173	2	2
9	323	1179	320	344	336	1179	2	2
10	1143	365	319	355	345	1143	1	1
11	332	360	323	1140	336	1140	4	4

Questions	A	B	C	D	E	Highest Value	Automatic Generated Answers	Manual Answers Typed
12	330	362	329	1146	342	1146	4	4
13	336	358	1166	351	348	1166	3	3
14	328	353	332	1185	345	1185	4	4
15	1152	347	317	352	338	1152	1	1
16	1120	349	309	327	336	1120	1	1
17	323	365	1173	346	353	1173	3	3
18	1149	359	316	333	340	1149	1	1
19	329	1156	317	338	335	1156	2	2
20	1137	355	319	354	339	1137	1	1
21	331	361	315	1140	342	1140	4	4
22	334	355	1150	352	337	1150	3	3
23	328	1165	325	360	335	1165	2	2
24	330	1154	322	345	351	1154	2	2
25	327	358	325	1145	350	1145	4	4
26	334	1166	335	350	343	1166	2	2
27	340	381	1173	350	350	1173	3	3
28	1170	367	346	345	348	1170	1	1
29	339	1179	349	348	359	1179	2	2
30	360	1167	345	343	370	1167	2	2
31	1164	356	331	346	344	1164	1	1
32	348	371	333	1178	352	1178	4	4
33	338	365	1157	351	345	1157	3	3
34	346	1167	325	347	336	1167	2	2
35	340	365	1171	342	336	1171	3	3
36	339	368	1154	353	341	1154	3	3
37	353	1165	341	357	342	1165	2	2

Questions	A	B	C	D	E	Highest Value	Automatic Generated Answers	Manual Answers Typed
38	344	1176	342	356	345	1176	2	2
39	336	1170	333	354	336	1170	2	2
40	1190	369	332	353	338	1190	1	1
41	323	356	1124	344	325	1124	3	3
42	340	1168	330	350	340	1168	2	2
43	345	359	334	1162	369	1162	4	4
44	347	368	338	1187	358	1187	4	4
45	342	375	334	1162	352	1162	4	4
46	347	379	1162	363	348	1162	3	3
47	352	377	331	1167	341	1167	4	4
48	345	1176	338	360	357	1176	2	2
49	1164	376	337	364	360	1164	1	1
50	344	366	339	1165	351	1165	4	4
51	337	344	1169	397	442	1169	3	3
52	330	357	1167	392	446	1167	3	3
53	338	351	334	1212	717	1212	4	4
54	1159	361	338	372	461	1159	1	1
55	1171	349	334	362	447	1171	1	1
56	340	347	1173	412	448	1173	3	3
57	331	353	1166	413	470	1166	3	3
58	332	1159	331	375	476	1159	2	2
59	334	352	1149	413	468	1149	3	3
60	332	1153	330	383	464	1153	2	2
61	340	358	1158	418	454	1158	3	3
62	1150	362	325	371	461	1150	1	1
63	343	359	1159	408	469	1159	3	3

Questions	A	B	C	D	E	Highest Value	Automatic Generated Answers	Manual Answers Typed
64	346	371	334	1167	753	1167	4	4
65	1171	371	331	365	476	1171	1	1
66	329	347	326	1126	716	1126	4	4
67	341	358	1152	409	465	1152	3	3
68	336	359	334	1178	742	1178	4	4
69	1160	362	343	377	461	1160	1	1
70	346	366	1169	412	466	1169	3	3
71	1157	367	348	383	469	1157	1	1
72	1174	372	346	378	470	1174	1	1
73	342	368	1190	420	472	1190	3	3
74	349	1150	343	375	465	1150	2	2
75	347	1166	343	383	466	1166	2	2
76	1179	364	336	345	357	1179	1	1
77	337	1163	351	364	356	1163	2	2
78	337	1172	336	360	372	1172	2	2
79	337	373	1168	352	369	1168	3	3
80	333	1163	337	358	369	1163	2	2
81	328	360	1167	358	354	1167	3	3
82	327	1200	328	356	338	1200	2	2
83	328	373	353	1152	357	1152	4	4
84	1141	362	344	365	353	1141	1	1
85	347	359	1152	365	352	1152	3	3
86	348	367	348	1157	354	1157	4	4
87	337	1156	342	354	353	1156	2	2
88	338	373	1172	364	353	1172	3	3
89	350	1154	354	363	361	1154	2	2

Questions	A	B	C	D	E	Highest Value	Automatic Generated Answers	Manual Answers Typed
90	1205	363	352	356	350	1205	1	1
91	331	1110	339	342	338	1110	2	2
92	336	350	336	1163	346	1163	4	4
93	341	1146	336	369	360	1146	2	2
94	334	361	1170	364	350	1170	3	3
95	347	369	1179	366	349	1179	3	3
96	1158	369	347	363	363	1158	1	1
97	331	356	1166	358	370	1166	3	3
98	332	361	350	1184	371	1184	4	4
99	338	1164	336	368	363	1164	2	2
100	337	382	1202	364	359	1202	3	3

3.3. Data Pre-Processing

Now as we got the data, extracted from all the scanned answer sheets, the data was in the raw form, and it needs to get pre-processed before we analyze the data and calculate different parameters from that data. Following figure shows the flow of the data pre-processing.

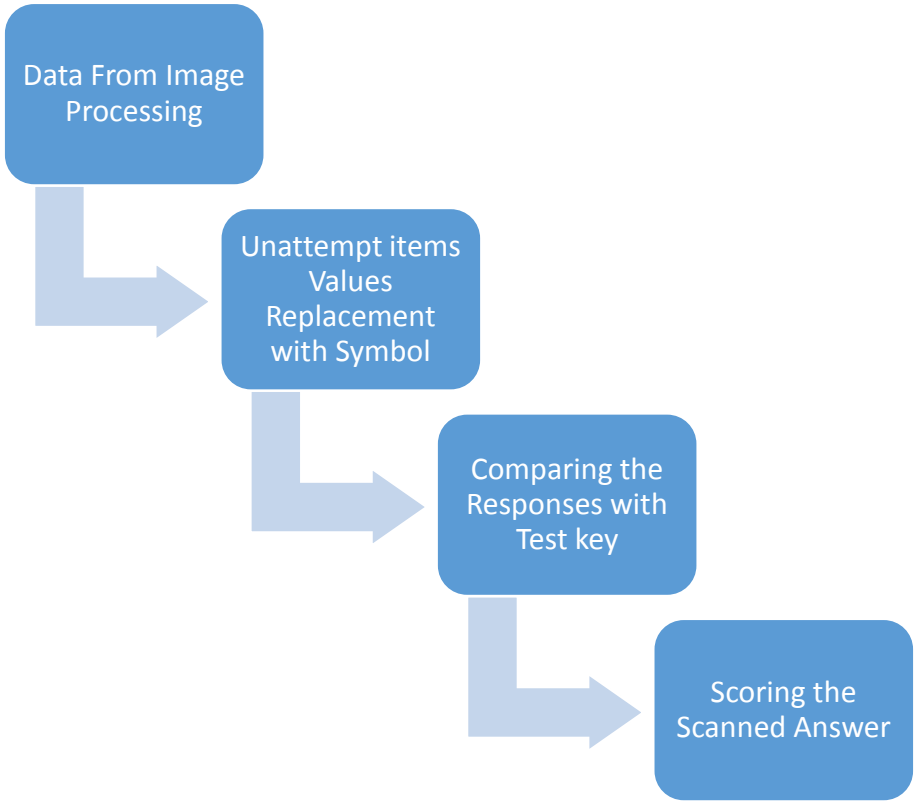


Figure 3.6 Flow Chart for Data Pre-Processing

3.3.1. Data Extracted from Image Processing

The combined data was inserted into the excel sheet, in which every column shows the answers from a single student. The data in the table below has been shown for first 30 questions and five students.

Table 3.3 Students Responses generated by Image processing

	S1	S2	S3	S4	S5		S1	S2	S3	S4	S5
Q1	1	3	2	3	1	Q16	1	3	1	2	4
Q2	2	2	1	3	2	Q17	3	3	3	3	3
Q3	4	1	1	4	1	Q18	1	4	2	1	4
Q4	3	3	2	2	1	Q19	2	3	2	2	3
Q5	1	2	4	1	3	Q20	1	1	1	1	1
Q6	4	1	4	4	1	Q21	4	4	4	4	4
Q7	3	3	4	3	3	Q22	3	3	3	3	3
Q8	2	2	2	2	2	Q23	2	2	2	1	2

Q9	2	3	3	1	4	Q24	2	3	1	1	4
Q10	1	1	3	2	3	Q25	4	2	4	3	4
Q11	4	4	2	3	1	Q26	2	2	3	3	2
Q12	4	1	1	1	3	Q27	3	1	3	3	3
Q13	3	3	4	3	4	Q28	1	3	1	3	1
Q14	4	4	1	1	1	Q29	2	4	4	2	4
Q15	1	2	3	1	4	Q30	2	3	3	1	2

3.3.2. Replacing unattempted questions with a specific Symbol

There is a possibility that students might have left the questions empty or selected multiple options. For this, our algorithms were designed to give a zero (0) value. So, there were many zero values present many questions. We replaced them with (–) symbol to avoid confusion in the data and possible options a student can select. We will explain why we have done it in the next session when we will be marking the scores.

Table 3.4 Unattempt Questions answers replaced by Symbol

	S1	**S2**	**S3**	**S4**	**S5**		**S1**	**S2**	**S3**	**S4**	**S5**
Q31	1	3	-	4	3	Q46	3	4	4	3	3
Q32	4	4	4	1	1	Q47	4	3	4	1	2
Q33	3	3	-	3	3	Q48	2	4	1	3	1
Q34	2	2	2	4	4	Q49	1	1	1	1	1
Q35	3	3	1	3	1	Q50	4	2	-	2	4
Q36	3	3	-	3	2	Q51	3	3	2	3	-
Q37	2	2	-	1	2	Q52	3	3	-	4	-
Q38	2	2	-	2	4	Q53	4	2	4	4	-
Q39	2	2	-	3	4	Q54	1	1	2	3	-
Q40	1	4	-	3	3	Q55	1	3	1	2	1
Q41	3	1	-	2	1	Q56	3	3	-	4	-
Q42	2	4	-	2	2	Q57	3	2	2	2	-
Q43	4	3	-	2	3	Q58	2	3	-	2	-
Q44	4	1	2	3	4	Q59	3	2	-	3	-
Q45	4	3	-	3	1	Q60	2	2	-	3	-

3.3.3. Comparing the Responses with Test key

We have matched the responses of students with the key provided by the Uni.A administration, and replaced it with zero and then with empty spaces. The key was placed in the last row in the excel sheet, while **if formula** was run against every cell matching the corresponding cell and replacing it with zero.

3.3.4. Marking the Output:

Once the formula for zeros got executed, then zeros were replaced with Blanks. **Countblanks** formula was used to count the right answers that were eventually the final score for the student. The table below shows all the procedure that was done, while marking the scores.

	Before Scoring		After Scoring		Key Numeric Key	
	Student 1	Student 2	Student 1	Student 2		
Q1	3	3			c	3
Q2	2	3		3	b	2
Q3	4	1	4	1	c	3
Q4	1	3	1	3	b	2
Q5	1	2		2	a	1
Q6	4	4			d	4
Q7	3	3			c	3
Q8	2	2			b	2
Q9	3	2	3	2	a	1
Q10	1	4		4	a	1
Q11	1	2	1	2	c	3
Q12	1	1			a	1
Q13	4	3	4		c	3
Q14	1	1			a	1
Q15	2	2			b	2
Q16	2	2			b	2
Q17	3	3			c	3
Q18	1	1			a	1
Q19	2	2			b	2
Q20	1	1			a	1
Q21	3	4	3		d	4
Q22	3	3			c	3
Q23	2	2			b	2

	Before Scoring		After Scoring		Key Numeric Key	
	Student 1	Student 2	Student 1	Student 2		
Q24	2	4	2	4	a	1
Q25	1	2		2	a	1
Q26	3	2		2	c	3
Q27	2	4	2	4	c	3
Q28	1	4	1	4	b	2
Q29	1	3	1		c	3
Q30	1	2		2	a	1
Q31	1	4	1	4	c	3
Q32	4	4			d	4
Q33	3	4		4	c	3
Q34	2	1	2	1	c	3
Q35	3	1	3	1	b	2
Q36	4	1	4	1	c	3
Q37	3	2	3	2	a	1
Q38	2	3		3	b	2
Q39	2	1		1	b	2
Q40	3	1		1	c	3
Q41	2	4		4	b	2
Q42	4	3	4		c	3
Q43	3	2		2	c	3
Q44	4	3		3	d	4
Q45	1	4		4	a	1
Q46	4	3	4		c	3
Q47	4	3		3	d	4
Q48	1	4		4	a	1
Q49	1	2		2	a	1
Q50	2	4		4	b	2
Q51	2	2	2	2	a	1
Q52	3	2		2	c	3
Q53	4	4	4	4	c	3
Q54	1	2		2	a	1
Q55	4	4	4	4	c	3
Q56	3	4	3	4	b	2
Q57	1	2	1	2	d	4
Q58	4	1	4	1	c	3
Q59	1	2		2	a	1

	Before Scoring		After Scoring		Key Numeric Key	
	Student 1	Student 2	Student 1	Student 2		
Q60	2	3		3	b	2
Q61	1	2	1	2	c	3
Q62	4	2		2	d	4
Q63	1	1	1	1	b	2
Q64	4	1	4	1	c	3
Q65	4	4			d	4
Q66	4	1	4	1	b	2
Q67	2	1	2	1	c	3
Q68	2	1	2		a	1
Q69	4	4	4	4	a	1
Q70	1	3	1		c	3
Q71	2	1	2	1	c	3
Q72	4	2	4		b	2
Q73	2	3	2	3	a	1
Q74	3	1	3	1	b	2
Q75	2	3	2		c	3
Q76	1	1	1	1	d	4
Q77	3	1	3	1	b	2
Q78	1	2		2	a	1
Q79	1	4	1	4	c	3
Q80	1	3		3	a	1
Q81	2	3		3	b	2
Q82	3	2		2	c	3
Q83	4	3		3	d	4
Q84	3	3			c	3
Q85	4	1	4		a	1
Q86	3	3			c	3
Q87	1	3	1		c	3
Q88	4	4			d	4
Q89	4	2	4	2	c	3
Q90	4	1	4		a	1
Q91	1	4	1		d	4
Q92	4	4			d	4
Q93	1	2	1		b	2
Q94	3	3	3	3	b	2
Q95	2	2			B	2

	Before Scoring		After Scoring		Key Numeric Key	
	Student 1	Student 2	Student 1	Student 2		
Q96	1	3	1	3	D	4
Q97	3	4	3		D	4
Q98	4	4			D	4
Q99	2	2	2	2	A	1
Q100	3	4		4	C	3
Result (Total Marks out of 100)			51	37		

3.4. Final Data to Analyze:

The data was pre-prosed for the entry tests of 2013 and 2015. The data was arranged, and scoring was done in the Excel sheet. The picture below shows the data. The columns show the responses of all students for a single question and every row is indicating the response of single student against all the questions.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
1	4		1	1	2				2	4	1
2	4		1	1	2				2	4	1
3	4		1	1	2				2	4	1
4	1		2						3	3	4
5	1		1	1	2		2		2	4	
6	4		1						2		
7	1			3	2				3	3	1
8			4	1					3		1
9		3	1	3	2				2	4	2
1		5		1		5		5	3	4	
11	1		2	1	2	2	1			3	
12	4		1	1	2			1	2	3	1
13	4		1						3	4	4
14			1						2		1
15	4		1						3	3	1
16	1		1	1	2			4	3	3	1

Figure 3.7 Data after Pre-Processing

Chapter 4. Results

In this chapter we have discussed the data analysis and the results obtained. As discussed before, we have calculated Discrimination Index, Difficulty Index, Distractor Efficiency Distractor Functionality and Reliability Index.

Data Analysis

All the data was collected from the scanned answers sheets (provided by the Uni.A administration) through image processing. Which was pre-processed in the Excel sheet to get the precise data. The data was arranged according to the requirements of the Item Analysis. The item analysis was carried out for checking the Discrimination Index, Difficulty Index, distractor efficiency and distractor functionality of the items.

4.1. Data Sets

We examined the reliability of the entry tests for the year 2013 and 2015. The scanned answer sheets were provided by the university administration. There was 100 question in the entry test 2013, that were divided into three parts that are English section, it has 40 MCQs questions, and the second part was of Analytical that consists of 20 MCQs questions the last part was for Mathematics having 40 MCQs questions. The entry test of 2105 consisted of four different tests and were differentiated by the different colored question papers. The colored question papers were blue, green, pink and yellow. All the tests consisted of 100 MCQ's questions and were divided into 25 Questions of Analytical, 45 questions for Mathematics and 30 questions were of English. All the colored booklets had same questions, but the arrangements of the questions were different.

4.2. Entry Test 2013 Analysis

The total number of students appeared in the Entry Test of 2013 were 500, we used SPSS 20.0 and MS Excel for data analysis. The tests that we have applied in SPSS are for the Reliability and Normality of the data while we have calculated different parameters, like difficulty index, discrimination index, and distractor analysis in the MS Excel. As stated earlier, the entry test of 2013 was divided into three parts. We have analyzed all the three parts individually so that the problems related to the effectiveness and reliability of every part can be identified individually.

4.2.1. English Part 2013

The English part of the entry test of 2013 consisted of 40 MCQs question, and which is almost 40% of the entire test. 500 students attempted the test. The total numbers for this part were 40, one point for a right answer and zero points for wrong answers, there was no negative marking.

4.2.1.1. Mean and Standard Deviation

The mean and standard deviation of the score achieved by the students are shown in the table below. The highest score was 33, and the lowest score was 2.

Table 4.1 Mean and Standard Deviation of English Part 2013

Mean and Standard Deviation		
Mean	N	Std. Deviation
15.23	500	4.573

4.2.1.2. Dividing into Upper and Lower Groups

The Total participant in English test were divided into three groups, the high group (170 high achievers), the lower group (170 low achievers) and the middle group (160 middle achievers). The distribution was done by the scores they got in the English part.

	39	288	154	148	339	342	352	370	377	378	383	432	427	431	441	448	478	480	3	22	105	20	186	139
Q1	5	4	1	1	4	2	1		2	1	1	2	4	2	4		1	4						
Q2	5	1	4	2	1		2		2	2		2	4			1	4	1	1	1	4	1		
Q3	5	1						1										1						
Q4	5	3		4			2	3	4	2	2	3		2	3	2		2		4			3	
Q5	5	4	1	1	4	1		4	2		2		1	4	1	4	1			2	2			
Q6	5	3	5					3	3	1	3	4	3		1	3		1		1		1		
Q18	5		5																					
Q19	5	4	5	2	4	2	1	2		2	2	1			4	4	4	1		1		4		
Q20	5	4	5	1	4	4	4	4	1	4	4	1	4		4	4		1	4	2			4	
Q26	5	3	5	3	2			2	3		3			2	4				3	3	3			3
Q27	5		5	4				2	4	4	1	4	4	1	4	4		4						4
Q28	5	2	5	1			2		3		2	3							2	1				
Q29	5	4	5	4		4			3	3		2	3			3			3					
Q30	5	2	5	2	1	4	2	2		4	2	1		5		4	1	1	4	4				
Q31	5	2	5	1		1	2			4	2	4	4	2			2	4	4	2	4	4	2	4
Q36	5	5	5	4	5	4	4	5	3		3	5	4	3	5	4		5	5	2				
Q37	5	4	5	1	1	4	5	4		4	1	4	1		2	5	4	1		5		4		
Q38	5	1	5	4	1	1	4	4					5	2		1	2	4				1	1	5
Q39	5	4	5	4	5	1	5			1	1		5	4		4	4	1	5	3	3		1	
Q40	5	5	5	2		2	2		3	3		3			4	4		2	2	4				
	2	3	4	5	13	13	13	13	13	13	13	13	16	16	16	16	16	16	17	17	29	30	30	33

Figure 4.1 Upper and Lower Groups of Students on Basis of Marks

4.2.1.3. Difficulty Index (p-value)

The difficulty index (p-value) for the English part was calculated by adding the right answers in the high achievers and the right answer in the low achievers and dividing them by the total number of the students in the high and low achievers. The results are shown in the table, and graph below.

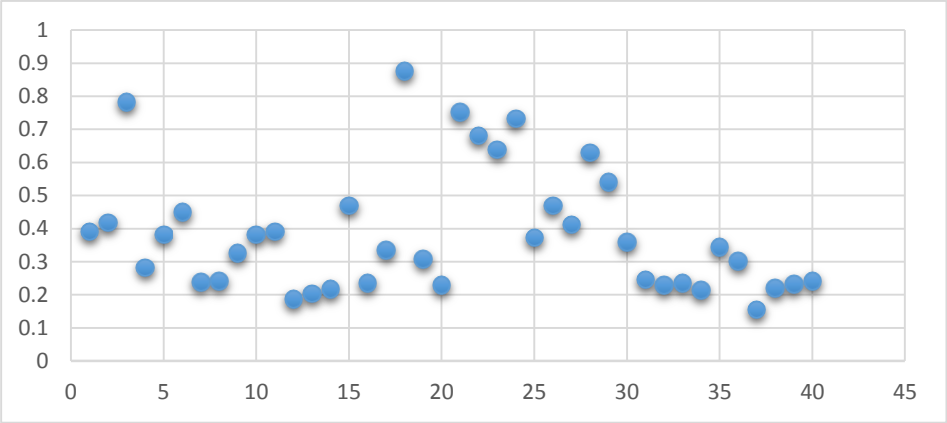


Figure 4.2 Difficulty Index Graph for English Section for Entry Test 2013

The results for the difficulty of the English section shows that 16 questions were difficult to their difficulty index (p-value) was less than 0.3, the four questions were too easy as most of the students from both groups gave the right answers to them. The acceptable p-value range is between 0.3 and 0.7, and 20 items fall in this range. To conclude, the 16 questions were difficult, four questions were very easy, and rest of the 20 questions were right to be included in the test.

Table 4.2 Results for Difficulty Index of English Section for Entry Test 2013

Item Difficulty Results	
Criteria	No of Questions
p-value < 0.3	16
0.3< p-value <0.7	20
p-value > 0.7	4

4.2.1.4. Discrimination Index

Discrimination Index (DI) was calculated by subtracting the right answers in the low group from the high question and dividing them by a total number of student in one of the group. The discrimination Index tells the power of the question to discriminate the high and low achievers. The results of the DI values are shown in the table below.

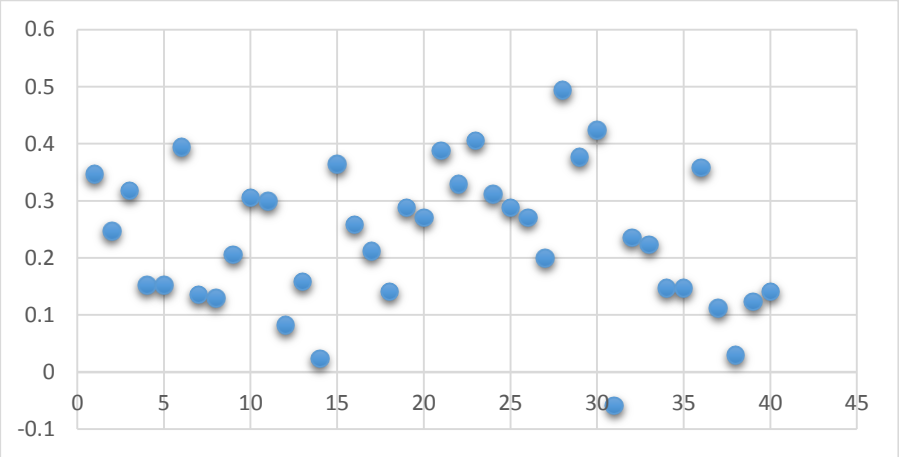


Figure 4.3 Discrimination Index of English Section for Entry Test 2013

The results showed that there were 21 questions on the test that were acceptable as far as discrimination index was concerned, as their DI value was greater than 0.24. 19 questions were poorly discriminating as their DI value was less than 0.24. There was a question (no 31) that has DI value in the negative (-0.06), which means that the number of right answers for that question in the lower group is greater than the number of correct responses in the high group.

Table 4.3 Discrimination Index of English Section for Entry Test 2013

Item Discrimination Result	
Criteria	No of Questions
DI < 0.24	19
DI >0.24	21

4.1.2.5.Distractor Efficiency

Distractors are critical parts for any MCQs; they directly affect the efficiency and effectiveness of the items. In the table below the discrimination index for every distractor was calculated. The empty cells are the answer to that question.

	A	B	C	D		A	B	C	D
Q1	-0.15	-0.10		-0.09	Q21	-0.09		-0.25	-0.04
Q2	0.04	-0.19		-0.08	Q22	-0.15		-0.16	-0.01
Q3	-0.19	-0.08	-0.04		Q23	-0.34		-0.03	-0.03
Q4		-0.25	0.18	-0.09	Q24		-0.25	-0.05	-0.01
Q5	-0.12	0.03		-0.05	Q25	-0.22	0.13	-0.18	
Q6	-0.01		-0.30	-0.08	Q26		-0.08	-0.02	-0.13
Q7	-0.04		0.01	-0.09	Q27	-0.08	-0.17		0.06
Q8	0.00	-0.03		-0.09	Q28	-0.18	-0.15	-0.15	
Q9	-0.24	0.00		0.05	Q29		-0.12	-0.11	-0.13
Q10		-0.17	-0.09	-0.04	Q30	-0.16	-0.17		-0.08
Q11		-0.05	-0.15	-0.09	Q31	-0.12	0.04		0.16
Q12	-0.12	-0.06	0.11		Q32	-0.03	-0.02	-0.15	
Q13	-0.06	-0.10		0.02	Q33	-0.10	-0.02	-0.02	
Q14		-0.04	-0.01	0.04	Q34	0.06	-0.12	-0.05	-0.04
Q15		-0.05	-0.15	-0.15	Q35	0.08	-0.09	-0.03	-0.10
Q16		-0.02	-0.16	-0.06	Q36		-0.09	-0.14	-0.10
Q17	0.03	-0.13	-0.11		Q37	-0.08	-0.04		0.08
Q18	-0.03		-0.09	-0.01	Q38	-0.03	-0.02		0.01
Q19	-0.05	-0.11		-0.12	Q39	0.01		0.02	-0.03
Q20	-0.10	-0.05		-0.11	Q40		0.01	-0.12	-0.01

Color coding was used to differentiate the effectiveness of the distractors. The green colored are the good distractors (around 83%), as they are chosen more by the low achievers than the high achievers. The red colored boxes show the wrong distractor (around 17%), as they are chosen more by high achievers than the low achievers as an answer. The empty spaces were the actual response to the question. In short there were total 120 distractors for 40 questions, 100 distractors were good as their DI value was negative, and 20 distractors were bad as their DI value was in positive.

4.1.2.6.Functional/non-functional Distractors

Another perimeter that can measure against the distractors is functional/non-functional distractors. The non-functional distractor is the one that is chosen by less than 10% of the students from the total students. The value can be calculated by adding the number of pupils who choose this distractor in both the upper and lower groups, then divide it by a total number of students in the both group and finally multiply the answer with 100 to get the value in percent.

Table 4.4 Functional and Non-Functional distractors for Entry Test 2013

	A	B	C	D		A	B	C	D
Q1	28	13	0	20	Q21	8	0	15	2
Q2	28	14	0	15	Q22	16	0	9	6
Q3	15	4	3	0	Q23	31	0	3	1
Q4	0	32	29	9	Q24	0	19	4	2
Q5	18	31	0	13	Q25	16	32	13	0
Q6	17	0	28	10	Q26	0	11	32	8
Q7	16	0	17	41	Q27	8	20	0	29
Q8	53	11	0	11	Q28	13	12	11	0
Q9	31	19	0	16	Q29	0	11	18	16
Q10	0	12	17	31	Q30	24	16	0	24
Q11	0	21	21	17	Q31	21	22	0	31
Q12	16	50	14	0	Q32	24	20	27	0
Q13	19	24	0	34	Q33	20	22	16	0
Q14	0	21	24	31	Q34	14	30	22	13
Q15	0	12	20	21	Q35	20	10	21	15
Q16	0	27	17	32	Q36	0	10	17	9
Q17	18	24	23	0	Q37	20	12	0	40
Q18	3	0	6	3	Q38	23	16	0	22
Q19	24	27	0	17	Q39	28	0	19	15
Q20	21	16	0	40	Q40	0	32	12	24

The yellow color shows the right answer for the questions; the green colored boxes shows the (102) distractors (around 85%) which are functional, mean they are good to distract the students if their concepts are weak. The red colored boxes show 18 distractors (around 15%) that are non-functional and need to be changed or reviewed as less than 10% students have chosen it as an answer.

4.1.2.7.Internal Reliability Analysis

The internal reliability of the English test was measured by Cronbach alpha. The Cronbach's alpha tells us that how closely the items of the test are related to a group.

Table 4.5 Internal Reliability for English 2013

Internal Reliability for English 2013	
Cronbach's Alpha	N of Items
.623	40

As shown in the table above the Cronbach's Alpha value is 0.623 that means that the English part is internally consistent.

4.1.3. Analytical Part 2013

The analytical part was analyzed in the same way as it was done for English part. We have stated the outputs of the analysis. There were 20 total questions. The tables and detail results can be seen in the Appendices.

Table 4.6 Data Analysis for Analytical Part 2013

	Criteria	Total Questions (20)
Difficulty Index (P-value)	p-value<0.3	3
	0.3<p-value<0.7	**14**
	p-value>0.7	3
Discrimination Index (DI)	Di<0.24	4
	DI>0.24	**16**
Distractor Efficiency	DI>0	7
	DI<0	**53**
Distractor Functionality	Selected by less than 10%	19
	Selected by more than 10%	**41**
Internal Reliability	Cronbach Alpha	0.96

There were total 20 questions. Out of these, three questions had p-values less than 0.3 and three questions had p-value more than 0.7. It shows that 14 questions were in the acceptable range

of Difficulty, as they fell in the range of 0.3 – 0.7. As far as Discrimination Index was concerned, four questions had the discrimination power less than 0.24 and 16 questions had the discrimination power more than 0.24. Five distractors had DI value equals to zero, two questions had a positive value and 53 distractors had negative DI value. 19 distractors were chosen by less than 10% of the students, other 41 distractors were chosen by more than 10% of the students. The Cronbach alpha for this part was .916.

4.1.4. Mathematics Part 2013

There were total 40 questions in the mathematical part, the data analysis statistics have be reported below, more detail in the form of graphs and tables is attached in the appendices.

Table 4.7 Data Analysis for Mathematical Part 2013

	Criteria	Total Questions (40)
Difficulty Index (P-value)	p-value<0.3	11
	0.3<p-value<0.7	**29**
	p-value>0.7	0
Discrimination Index (DI)	Di<0.24	7
	DI>0.24	**33**
Distractor Efficiency	DI>0	12
	DI<0	**108**
Distractor Functionality	Selected by less than 10%	21
	Selected by more than 10%	**99**
Internal Reliability	Cronbach Alpha	0.95

Difficulty Index; Out of 40 questions 11 questions have p-values less than 0.3 and 29 questions have p-value higher than 0.3 and less than 0.7. There was no question that has the p-value greater than 0.7. Discrimination Index; 3 questions that have the discrimination power in negatives and 7 questions that have DI value less than 0.24 and 33 questions have the discrimination power more than 0.24. Distractor Efficiency; 12 distractors that have a positive DI value, and 108 distractors have negative DI value. Functional/Non-Functional Distractors; 21 distractors were chosen by less than 10% of the students, other 99 distractors are chosen by more than 10% of the students. Internal Reliability Analysis; the Cronbach alpha for this test was .95.

4.3. Combined Results for Entry Test 2103

The results for the whole test was combined and shown in the table below, the number of reliable questions on the basis of different parameters are highlighted and shown in the table below.

Table 4.8 Results for Item Analytical for Entry Test 2013

			English	Analytical	Math's	Total
	Parameters	**Criteria**	**40**	**20**	**40**	**100**
1	Item Difficulty (p-values)	p-value<0.3	16	3	11	30
		0.3<p-value<0.7	**20**	**14**	**29**	**63**
		p-value>0.7	4	3	0	7
2	Discrimination Index (DI)	Di<0.24	19	4	7	30
		DI>0.24	**21**	**16**	**33**	**70**
3	Distractor Efficiency	DI>0	20	7	12	39
		DI<0	**100**	**53**	**108**	**261**
4	Functional/Non Functional Distractors	Selected by less than 10%	18	19	21	58
		Selected by more than 10%	**102**	**41**	**99**	**242**
5	Internal Reliability		0.623	0.91	0.95	

4.4. Entry Test 2015 Analysis

The total number of student appeared in the Entry Test of 2015 were 800, we have used SPSS 20.0 and MS Excel as tools for the data analysis. The analysis tests that we applied in the SPSS are for the Reliability and Normality of the data while we have calculated different parameters, like difficulty index, discrimination index, and distractor analysis in the MS Excel. There were total four-color of questions booklet used in Entry Test 2015, in every booklet the arrangements of the questions were different, and overall the questions were same in all the booklets. Every booklet was divided into three parts English, Analytical, and Mathematics. We have analyzed all the booklets and three different parts in it individually so that we can identify the problems related to the effectiveness and reliability of tests effectively.

4.4.1. Blue Booklet

The blue booklet consist of 100 questions, English part contains 30 questions while Mathematics and Analytical contain 45 and 25 respectively.

4.4.1.1. English Part 2015

The English part of the entry test of 2015 consisted of 30 MCQs, and that is almost 30% of the entire test. The total numbers for this part were 30, it means that one point for one right answer and zero points for wrong answer, there was no negative marking. Discrimination Index, Difficulty Index, distractor analysis, and Internal Reliability was calculated for the English part.

4.4.1.1.1. Difficulty Index

The difficulty index (p-value) for the English part was calculated by adding the right answers in the high achievers and the right answer in the low achievers and dividing them by the total number of the students in the high and low achievers. The results were not good and showed that most of the questions lies either above or below the expectable range. 13 questions lied in the expectable range while 10 questions were very easy and 7 were tough as their p-value were less than 0.3 and high than 0.7 respectively.

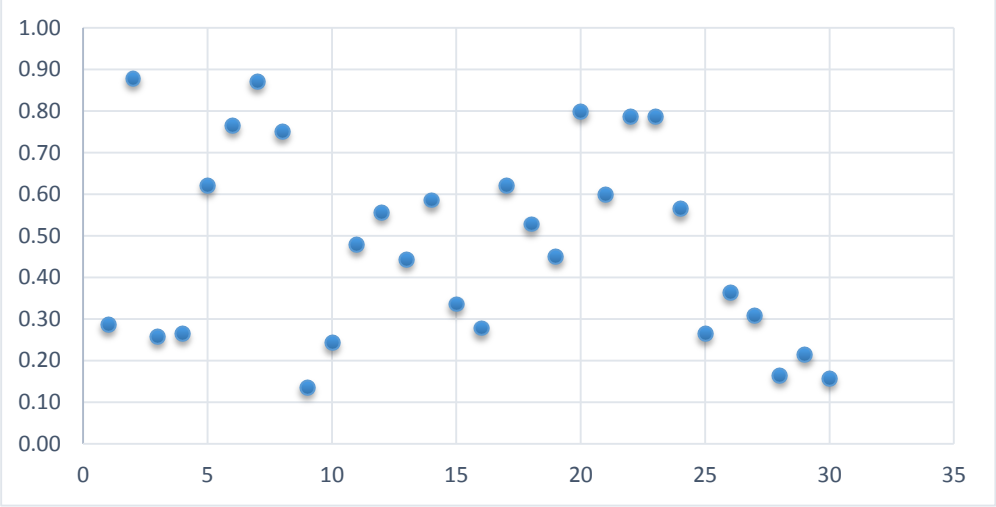


Figure 4.4 Difficulty Index_EntryTest2015_BlueBooklet_EnglishPart

4.4.1.1.2. Discrimination Index

Discrimination Index (DI) was calculated by subtracting the right answers in the low group from the high question and dividing them by a total number of students in one of the group. The discrimination Index tells the power of the question to discriminate the high and low achievers. There were total 30 questions in this part and out of 30, 16 question lies in the acceptable range of difficulty, that is greater than 0.24 while 14 questions were not right to discriminate the high and low achievers.

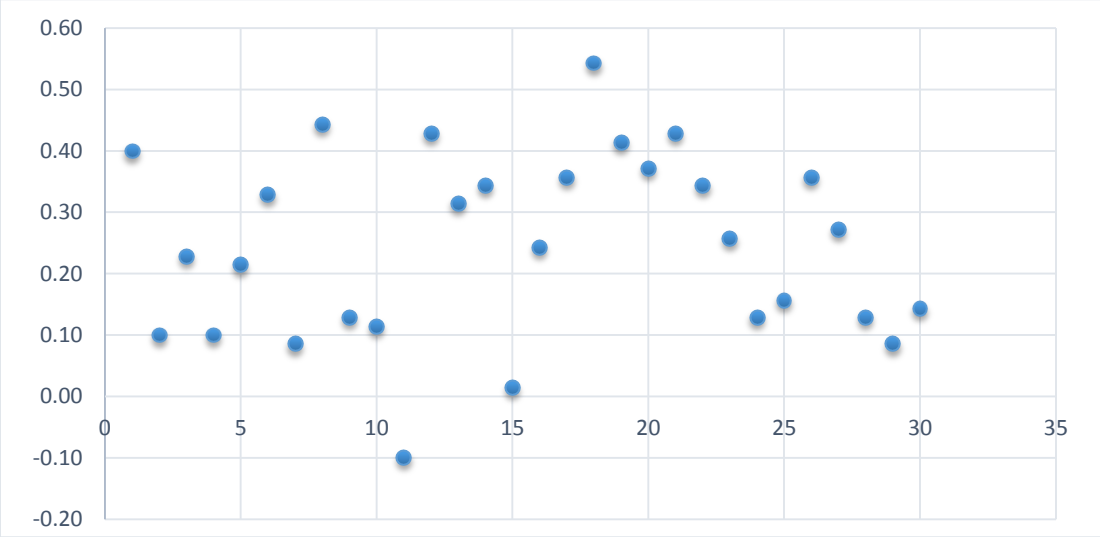


Figure 4.5 DiscriminationIndex_EntryTest2015_BlueBooklet_ English Part

4.4.1.1.3. Distractor Efficiency

Distractors are a critical part of the MCQs; they directly affect the efficiency and effectiveness of the items. Students with the weak concept always go for the distractors as an answer. So for the useful item, the distractors should be plausible. In the table below the discrimination index for every distractor was calculated. The empty cells are the answer to that question.

Table 4.9 Distractor Analysis_ EntryTest2015_BlueBooklet_ English Part

	A	B	C	D		A	B	C	D
Q1	-0.14	-0.14		-0.13	Q16	-0.09		-0.14	-0.01
Q2	-0.01		-0.04	-0.03	Q17	-0.16	-0.17		-0.03
Q3	0.03	-0.06		-0.20	Q18		-0.14	-0.33	-0.09
Q4	-0.26		0.23	-0.06	Q19	-0.19		-0.10	-0.13
Q5		-0.16	-0.01	-0.04	Q20		-0.20	-0.09	-0.09
Q6	-0.20	-0.09	-0.03		Q21	-0.21	-0.11	-0.10	
Q7	-0.10	0.03		-0.01	Q22	-0.13	-0.11		-0.10
Q8	-0.33		-0.04	-0.06	Q23	-0.09		-0.07	-0.10
Q9		0.06	-0.07	-0.13	Q24		-0.01	0.00	-0.09
Q10		-0.01	-0.17	0.04	Q25		-0.04	-0.10	0.00
Q11	0.17	-0.07		0.01	Q26	-0.06	-0.13		-0.14
Q12		-0.03	-0.36	-0.03	Q27	-0.20	-0.14		0.09
Q13	-0.09	-0.14		-0.10	Q28	-0.10		-0.01	0.01
Q14		-0.10	-0.09	-0.16	Q29	-0.04	0.17		-0.19
Q15	0.14		-0.09	-0.09	Q30		-0.03	-0.01	-0.07

Color coding was used to differentiate the effectiveness of the distractors. The green color is the good distractors as they are chosen more by the low achievers than the high achievers. The red color shows the wrong distractor, as they are chosen more by high achievers than the low achievers as an answer. The empty spaces were the actual response to the question. In short there were total 90 distractors for 30 questions, 79 distractors were good as their DI value was negative, and 11 distractors were bad as their DI value was in positive.

4.4.1.1.4. Functional/Non-functional Distractors

Distractors can either functional (selected by more than 10% of the students), or non-functional (chosen by less than 10% of the pupils). The value can be calculated by adding the number of students who choose this distractor in both the upper and lower groups, then divide it by a total number of students in the both group and finally multiply the answer with 100 to get the value in percentage.

Table 4.10 Functional/ Non Functional Distractors_ EntryTest2015_BlueBooklet_ English Part

	A	B	C	D		A	B	C	D
Q1	25.71	17.14	0.00	27.86	Q16	15.71	0.00	18.57	37.86
Q2	2.14	0.00	6.43	2.86	Q17	23.57	10.00	0.00	4.29
Q3	51.43	4.29	0.00	18.57	Q18	0.00	10.00	22.14	14.29
Q4	42.86	0.00	24.29	5.71	Q19	15.00	0.00	19.29	20.71
Q5	0.00	32.14	3.57	2.14	Q20	0.00	10.00	5.71	4.29
Q6	12.86	7.14	1.43	0.00	Q21	16.43	8.57	15.00	0.00
Q7	7.86	4.29	0.00	0.71	Q22	9.29	7.14	0.00	5.00
Q8	19.29	0.00	2.14	2.86	Q23	12.86	0.00	3.57	5.00
Q9	0.00	32.86	40.71	12.14	Q24	0.00	26.43	1.43	14.29
Q10	0.00	15.00	34.29	25.00	Q25	0.00	32.14	19.29	20.00
Q11	32.86	6.43	0.00	12.14	Q26	14.29	35.00	0.00	12.86
Q12	0.00	4.29	32.14	7.14	Q27	18.57	25.71	0.00	22.86
Q13	10.00	10.00	0.00	33.57	Q28	30.71	0.00	20.71	30.71
Q14	0.00	12.14	5.71	23.57	Q29	12.14	42.86	0.00	22.14
Q15	35.71	0.00	18.57	11.43	Q30	0.00	34.29	39.29	9.29

The green color shows the functional distractors while the pink color shows the non-functional distractors and the yellow color shows the answers to the questions. For the English part, 76 distractors were functional while 14 distractors are non-functional.

4.4.1.1.5. Internal Reliability Analysis

The internal reliability of the English test was measured by Cronbach alpha. The Cronbach's alpha tells us that how closely the items of the test are related to a group.

Table 4.11 Internal Reliability for EntryTest2015_BlueBooklet_EnglsihPart

Internal Reliability for EntryTest2015_BlueBooklet_EnglsihPart	
Cronbach's Alpha	N of Items
0.66	30

As shown in the table above the Cronbach’s Alpha value is 0.66 that means that the English part is internally consistent.

4.4.1.2. Analytical Part 2015

The analytical part was analyzed in the same way as it was done for English part. We have writen the outputs of the analysis. There were 25 total questions. The details of the results can be seen in the Appendices.

Table 4.12 Results_EntryTest2015_BlueBooklet_Analytical Part

	Criteria	Total Questions (25)
Difficulty Index (P-value)	p-value<0.3	5
	0.3<p-value<0.7	**19**
	p-value>0.7	1
Discrimination Index (DI)	Di<0.24	14
	DI>0.24	**11**
Distractor Efficiency	DI>0	13
	DI<0	**62**
Distractor Functionality	Selected by less than 10%	21
	Selected by more than 10%	**54**
Internal Reliability	Cronbach Alpha	0.71

Difficulty Index; There were total 25 questions, five questions had p-values less than 0.3, and there was one questions having p-value more than 0.7. 19 questions were in the range of 0.3 – 0.7. Discrimination Index; 14 questions had the discrimination power less than 0.24 and 11 questions had the discrimination power more than 0.24. Distractor Efficiency; 13 distractors had DI value greater than zero that means that they were not effective distractors while 62 distractors had DI value less than zero that means they were effective distractors. Functional/Non-Functional Distractors; 21 distractors were chosen by less than 10% of the students, other 54 distractors were chosen by more than 10% of the students. Internal Reliability Analysis; the Cronbach alpha for this part was .71

4.4.1.3. Mathematics Part 2015

There were total 45 questions in the mathematical part, the data analysis statistics have reported below, more detail in the form of graphs and tables is attached in the appendices.

Table 4.13 Results_EntryTest2015_BlueBooklet_Mathematics Part

	Criteria	Total Questions (45)
Difficulty Index (P-value)	p-value<0.3	18
	0.3<p-value<0.7	**26**
	p-value>0.7	1
Discrimination Index (DI)	Di<0.24	21
	DI>0.24	**24**
Distractor Efficiency	DI>0	15
	DI<0	**120**
Distractor Functionality	Selected by less than 10%	22
	Selected by more than 10%	**113**
Internal Reliability	Cronbach Alpha	0.73

Difficulty Index; There were total 45 questions, 18 questions had p-values less than 0.3, and there was one questions having p-value more than 0.7. 26 questions were in the range of 0.3 – 0.7. Discrimination Index; 21 questions had the discrimination power less than 0.24 and 24 questions had the discrimination power more than 0.24. Distractor Efficiency; 15 distractors had DI value greater than zero that means that they were not effective distractors while 120 distractors had DI value less than zero that means they were effective distractors. Functional/Non-Functional Distractors; 22 distractors were chosen by less than 10% of the students, other 113 distractors were chosen by more than 10% of the students. Internal Reliability Analysis; the Cronbach alpha for this part was .73

4.4.2. Green Booklet

The blue booklet consist of 100 questions, English part contains 30 questions while Mathematics and Analytical contain 45 and 25 respectively.

Table 4.14 Complete Results_EntryTest2015_Green Booklet

			English	Mathematics	Analytical	Total
	Parameters	**Criteria**	**30**	**45**	**25**	**100**
1	Item Difficulty (p-values)	p-value<0.3	7	14	1	22
		0.3<p-value<0.7	**18**	**30**	**19**	**67**
		p-value>0.7	5	1	5	11
2	Discrimination Index (DI)	Di<0.24	11	23	10	44
		DI>0.24	**19**	**22**	**15**	**56**
3	Distractor Efficiency	DI>0	10	28	11	49
		DI<0	**80**	**107**	**64**	**251**
4	Functional/Non Functional Distractors	Selected by less than 10%	24	27	11	62
		Selected by more than 10%	**66**	**108**	**64**	**238**
5	Internal Reliability		0.65	0.72	0.76	

Difficulty Index; There were total 100 questions, 22 questions had p-values less than 0.3, and there were 11 questions having p-value more than 0.7. 67 questions were in the range of 0.3 – 0.7. Discrimination Index; 44 questions had the discrimination power less than 0.24 and 56 questions had the discrimination power more than 0.24. Distractor Efficiency; 49 distractors had DI value greater than zero that means that they were not effective distractors while 251 distractors had DI value less than zero that means they were effective distractors. Functional/Non-Functional Distractors; 62 distractors were chosen by less than 10% of the students, other 238 distractors were chosen by more than 10% of the students.

4.4.3. Yellow Booklet

The yellow booklet consist of 100 questions, English part contains 30 questions while Mathematics and Analytical contain 45 and 25 respectively.

			English	Mathematics	Analytical	Total
	Parameters	**Criteria**	**30**	**45**	**25**	**100**
1	Item Difficulty (p-values)	p-value<0.3	7	10	6	23
		0.3<p-value<0.7	**18**	**34**	**19**	**71**
		p-value>0.7	5	1	0	6
2	Discrimination Index (DI)	Di<0.24	15	20	10	45
		DI>0.24	**15**	**25**	**15**	**55**
3	Distractor Efficiency	DI>0	14	22	13	49
		DI<0	**76**	**113**	**62**	**251**
4	Functional/Non Functional Distractors	Selected by less than 10%	23	18	8	49
		Selected by more than 10%	**67**	**117**	**67**	**251**
5	Internal Reliability		0.61	0.73	0.78	

Difficulty Index; There were total 100 questions, 23 questions had p-values less than 0.3, and there were 6 questions having p-value more than 0.7. 71 questions were in the range of 0.3 – 0.7. Discrimination Index; 45 questions had the discrimination power less than 0.24 and 55 questions had the discrimination power more than 0.24. Distractor Efficiency; 49 distractors had DI value greater than zero that means that they were not effective distractors while 251 distractors had DI value less than zero that means they were effective distractors. Functional/Non-Functional Distractors; 49 distractors were chosen by less than 10% of the students, other 251 distractors were chosen by more than 10% of the students.

4.4.4. Pink Booklet

The Pink booklet consist of 100 questions, English part contains 30 questions while Mathematics and Analytical contain 45 and 25 respectively.

			English	Mathematics	Analytical	Total
	Parameters	**Criteria**	**30**	**45**	**25**	**100**
1	Item Difficulty (p-values)	p-value<0.3	8	10	8	26
		0.3<p-value<0.7	**15**	**35**	**16**	**66**
		p-value>0.7	7	0	1	8
2	Discrimination Index (DI)	Di<0.24	13	24	14	51
		DI>0.24	**17**	**21**	**11**	**49**
3	Distractor Efficiency	DI>0	11	27	11	49
		DI<0	**79**	**108**	**64**	**251**
4	Functional/Non Functional Distractors	Selected by less than 10%	30	22	8	60
		Selected by more than 10%	**60**	**113**	**67**	**240**
5	Internal Reliability		0.66	0.71	0.79	

Difficulty Index; There were total 100 questions, 26 questions had p-values less than 0.3, and there were 8 questions having p-value more than 0.7. 66 questions were in the range of 0.3 – 0.7. Discrimination Index; 51 questions had the discrimination power less than 0.24 and 49 questions had the discrimination power more than 0.24. Distractor Efficiency; 49 distractors had DI value greater than zero that means that they were not effective distractors while 251 distractors had DI value less than zero that means they were effective distractors. Functional/Non-Functional Distractors; 60 distractors were chosen by less than 10% of the students, other 240 distractors were chosen by more than 10% of the students.

Chapter 5. Discussion

In this chapter, we have discussed the results and have answered the research questions and effectiveness of the research.

5.1. Results Entry Test 2013

In this section we have discussed the results for the reliability of the Entrance Test for 2013, there were total 100 questions, and the test was divided into three parts that are English, Mathematics, and Analytical. The results were satisfactorily enough.

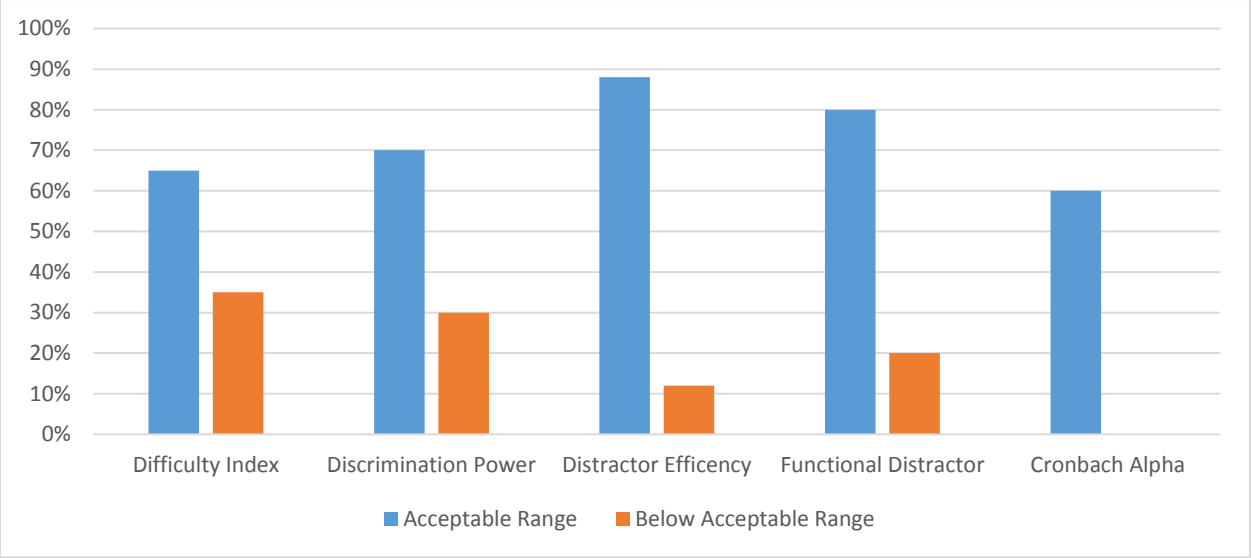


Figure 5.1 Results Interpretation for Entry Test 2013

There were total 65 questions out of 100 that were in the acceptable range for the difficulty index, as far as the discrimination power is a concern, 70 questions out of 100 were discriminating the high and low achievers. 88% distractors were efficient; it means that out of 300 distractors for 100 questions 261 distractors were effective, out of 300 distractors approximately 80% (245) distractors were functional. The Cronbach Alpha for Entry Test 2013 was 0.6, and it was in the acceptable range.

Now we will discuss the research questions, and the research hypothesis answers those question one by one.

5.1.1. Research Question 1:

To what extent the entry Test of Uni.A. is reliable in selecting students for Electrical Engineering and Computer Science in 2013?

5.1.1.1. Hypothesis 1:

The first hypothesis for Entry Test 2013 was "Test items are in the acceptable range of "difficulty" for 2013". The results show that 65% of the question are in the acceptable range that means that the alternative hypothesis is failed to reject.

5.1.1.2. Hypothesis 2:

"Items in the entry test of 2013 are well written to discriminate the high and low achievers" was the second hypothesis for the Research question, to what extent the entry Test of Uni.A. is reliable in selecting students for Electrical Engineering and Computer Science in 2013? The results for discrimination power of the questions were satisfactory, and almost 70% of the question were laying in the acceptable range of DI (Discrimination Index), and that was enough that alternative hypothesis is failed to reject.

5.1.1.3. Hypothesis 3:

"The distractors of the items in entry test of 2013 are effective" was the third hypothesis and results showed that out of 300 distractors 261 distractors were efficient, that means that almost 88% of the distractors were effective. So the hypothesis was true, and the alternative hypothesis failed to reject.

5.1.1.4. Hypothesis 4:

"The distractors of the items in entry test of 2013 are functional" was the fourth hypothesis and results showed that out of 300 distractors 245 distractors were functional, that means that almost 81% of the distractors were effective. So the hypothesis was true, and the alternative hypothesis failed to reject.

5.1.1.5. Hypothesis 5:

"The entry test of 2013 is internally consistent? " was the fifth hypothesis for the first research question and the value of Cronbach's alpha for the entry test 2013 was calculated, and it is greater than 0.6, that means that entry test for 2013 was internally consistent.

5.2. Results Entry Test 2015

In this section we have discussed the results for the reliability of the Entrance Test for 2015, there were total 100 questions, and the test was divided into three parts that are English, Mathematics, and Analytical. The results were satisfactorily enough.

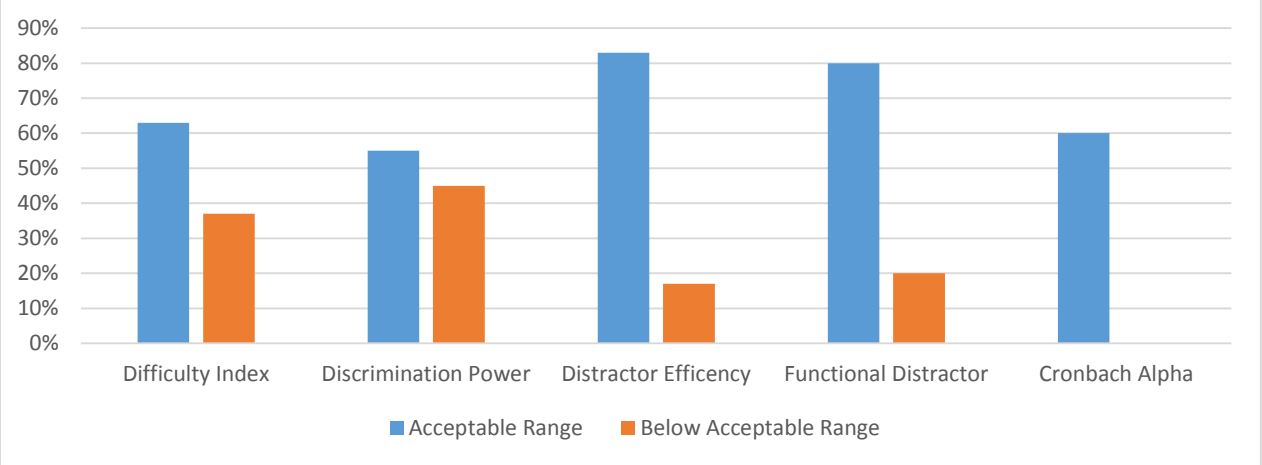


Figure 5.2 Results Interpretation for Entry Test 2015

There were total 63 questions out of 100 that were in the acceptable range for the difficulty index, as far as the discrimination power is a concern, 55 questions out of 100 were discriminating the high and low achievers. 83% distractors were efficient; it means that out of 300 distractors for 100 questions 250 distractors were effective, out of 300 distractors approximately 80% (242) distractors were functional. The Cronbach Alpha for Entry Test 2013 was greater than 0.6, and it was in the acceptable range.

Now we will discuss the research questions and the research hypothesis and answer those questions for Entry Test 2015.

5.2.1. Research Question 2:

To what extent the entry Test of Uni.A. is reliable in selecting students for Electrical Engineering and Computer Science in 2015?

5.2.1.1. Hypothesis 1:

The first hypothesis for Entry Test 2015 was “Test items are in the acceptable range of “difficulty” for 2015”. The results show that 65% of the question are in the acceptable range that means that the alternative hypothesis is failed to reject.

5.2.1.2. Hypothesis 2:

“Items in the entry test of 2015 are well written to discriminate the high and low achievers” was the second hypothesis for the Research question, to what extent the entry Test of Uni.A. is reliable in selecting students for Electrical Engineering and Computer Science in 2015? The results for discrimination power of the questions were satisfactory, and almost 55% of the question were laying in the acceptable range of DI (Discrimination Index), and that was enough that alternative hypothesis is failed to reject.

5.2.1.3. Hypothesis 3:

“The distractors of the items in entry test of 2015 are effective” was the third hypothesis and results showed that out of 300 distractors 250 distractors were efficient, that means that almost 83% of the distractors were effective. So the hypothesis was true, and the alternative hypothesis failed to reject.

5.2.1.4. Hypothesis 4

“The distractors of the items in entry test of 2015 are functional” was the fourth hypothesis and results showed that out of 300 distractors 242 distractors were functional, that means that almost 80% of the distractors were effective. So the hypothesis was true, and the alternative hypothesis failed to reject.

5.2.1.5. Hypothesis 5:

“The entry test of 2015 is internally consistent? “ was the fifth hypothesis for the first research question and the value of Cronbach’s alpha for the entry test 2015 was calculated, and it is greater than 0.6, that means that entry test for 2015 was internally consistent

Chapter 6. Conclusions and Recommendations

In this chapter we have discuss and conclude the aims and the findings of the study conducted. Section 6.1 revisits the aim of the study, Section 6.2 highlights the methodologies, and discuss how the research was carried out. Section 6.3 discuss the findings of the study. Section 6.4 discuss the recommendations and Section 6.5 will cover future work.

6.1. Revisiting the Aims of the Study

The study focused on the reliability of the entry tests used in Uni.A. The study focused on the identification of problems in the test items, that effects the reliability of the items in particular and entrance tests in general. There were two entrance tests that were analyzed, test held for 2013 admissions had data for 500 students while the entry test of 2015 had data for 800 participants. The research was conducted to get answers to the following questions; are the items very difficult or easy? Do they have the power to discriminate the high and low achievers? How many distractors are functional or non-functional? To what extent the questions were efficient?

6.2. Revisiting the Flow of the Research

It was Quantitative research, in which the data was extracted from the scanned answer sheets provided by the Uni.A administration. The data was extracted using Image Processing techniques. Once the data was extracted from the sheets, it was preprocessed in Microsoft Excel. The questions were marked with the help of the key provided by the concerned University. The students were separated into two groups on the basis of the marks, the low achievers and the high achievers. The data was analyzed in the SPSS 20.0 to find out different parameters related to the reliability of the items. The difficulty Index (p-values) was calculated for the items in both tests. The items in the range of 0.3 to 0.7 were considered good items. The next parameter was discrimination Index (DI value) of the items, the items with DI value greater than 0.24 were considered as good discriminators, means that they effectively discriminate the high achievers from the low achievers. The distractors efficiency was also calculated, the distractors with negative efficiency value were considered good distractors as they were chosen more by the low achievers than the high achievers. The distractor functionality was also calculated, the distractor chosen by

more than 10% students were considered as functional otherwise non-functional. The Cronbach alpha was also calculated, the test with value greater than 0.6 was considered reliable.

6.3. Findings of the Research

In the entrance test of 2013, there were total 65 questions out of 100 were in the acceptable range for the difficulty index, as far as the discrimination power is a concern, 70 questions out of 100 were discriminating the high and low achievers. 88% distractors were efficient; it means that out of 300 distractors for 100 questions 261 distractors were effective, out of 300 distractors approximately 80% (245) distractors were functional. The Cronbach Alpha for Entry Test 2013 was 0.6, and it was in the acceptable range.

For entrance test of 2015, there were total 63 questions out of 100 were in the acceptable range for the difficulty index, as far as the discrimination power is a concern, 55 questions out of 100 were discriminating the high and low achievers. 83% distractors were efficient; it means that out of 300 distractors for 100 questions 250 distractors were effective, out of 300 distractors approximately 80% (242) distractors were functional. The Cronbach Alpha for Entry Test 2013 was greater than 0.6, and it was in the acceptable range.

6.4. Recommendations

According to the analysis carried out on the given datasets, the entry tests were fairly reliable, improvement on the quality of distractors is recommended. For the test designers it is recommended that distractors should be plausiable, as many of them were not plausiable and should be designed in a way that they can identify misconceptions of the students. The students with misconceptions should feel it difficult to select the right option..By improving the distractors, quality of questions will also impove. The proper training should be conducted for the designers before they prepare the tests.

6.5. Future Research

This study is just the initiative to design the best and effective possible entry tests, the entry tests in the coming years should also go through this process. Research should be continued for the standardization of Entry Test Process. This case study calls for a Longitudinal study and it can carried out for next few years that will help to standerized the process of the Entrance Test.

References

Abdulghani, H., Ahmad, F., Aldrees, A., Khalil, M., & Ponnamperuma, G. (2014). The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. *Journal of Health Specialties*, *2*(4), 148. http://doi.org/10.4103/1658-600X.142784

Awodele, B. A., Faremi, Y. A., & Adetunji, A. A. (2013). Difficulty and Discrimination Indices of Logical-Choice Weight and Confidence Scoring Methods on Chemistry Multiple Choice Test in Nigeria Secondary Schools . *Journal of Education Adn Practices ISSN 2222-1735*, *4*(6), 114–120.

Boopathiraj, C. (2013). ANALYSIS OF TEST ITEMS ON DIFFICULTY LEVEL AND DISCRIMINATION INDEX IN THE TEST FOR RESEARCH IN EDUCATION. *International Journal of Social Science & Interdisciplinary Research*, *2*(2), 189–193.

Bruno JE, Dirkzwager A: Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. Educ Psychol Meas 1995, 55(6):959-966.

Cizek GJ, O'Day DM: Further investigations of nonfunctioning options in multiple-choice test items. Educ Psychol Meas 1994, 54(4):861-872.

DiBattista, D., & Kurzawa, L. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests. *Canadian Journal for the Scholarship of Teaching and Learning*, *2*(2). http://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Downing SM: Assessment of knowledge with written test forms. In International handbook of research in medical education Vol- ume II. Edited by: Norman GR, Van der Vleuten C, Newble DI. Dor- crecht: Kluwer Academic Publishers; 2002:647-672.

Educ Psychol Meas 1993, 53(4):999-1010.

Escudero, E. B., & Reyna, N. L. (2000). The Level of Difficulty and Discrimination Power of the Basic Knowledge and Skills Examination (EXHCOBA), *2*, 1–16.

Farley JK: The multiple-choice test: writing the questions. Nurse Educ 1989, 14(6):10-12.

Haladyna TM, Downing SM: Validity of a taxonomy of multiple- choice item-writing rules. Appl Meas Educ 1989, 2(1):51-78.

Haladyna TM, Downing SM: How many options is enough for a multiple-choice test item?

References

Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *Journal of the Pakistan Medical Association*, *62*(2), 142–147.

Hsiao-Fang, L. (2012). The Impact of Taiwan's University Multiple-Channel Entrance Policy on Student Learning Outcomes. *Chinese Education & Society*, *45*(5), 59–69. http://doi.org/10.2753/CED1061-1932450505

Khan, J. S., Tabasum, S., & Mukhtar, O. (2013). Comparison of pre-medical academic achievement, entrance test and aptitude test scores in admission selection process. *Journal of the Pakistan Medical Association*, *63*(5), 552–557.

Krishnan, V., & Ph, D. (2013). The Early Child Development Instrument (EDI): An item analysis using Classical Test Theory (CTT) on Alberta ' s data, (January).

Lord FM: Optimal number of choices per item – A compari- son of four approaches. J Educ Meas 1977, 14:33-38.

McCoubrie P: Improving the fairness of multiple-choice ques- tions: a literature review. Med Teach 2004, 26(8):709-712.

Marzano, A., Tammaro, R., & Notti, A. M. (2012). THE ENTRY TEST TO THE DEGREE COURSE IN SCIENCE OF EDUCATION, (264), 286–295.

Matlock-Hetzel, S. (1997). Basic Concepts in Item and Test Analysis.

Schuwirth LWT, Vleuten CPM van der: Different written assess- ment methods: what can be said about their strengths and weaknesses? Med Educ 2004, 38(9):974-979.

Singamaneni, S. (2011). Exploring content validity , item level analysis and predictive validity for two algebra progress monitoring measures.

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*, *9*, 40. http://doi.org/10.1186/1472-6920-9-40

Tversky A: On the optimal number of alternatives at a choice point.

Appendices

Code for Checking Answer Sheets

```
clear all
close all
clc
excelsheet='none';
srcFiles = dir('C:\Users\Malik Shahzad Iqbal\Desktop\test\*.jpg');  % the folder in which ur images exists
for i = 1 : length(srcFiles)
    excelsheet=srcFiles(i).name;
    filename = strcat('C:\Users\Malik Shahzad Iqbal\Desktop\test\',srcFiles(i).name);
    image = imread(filename);
%image=imread('C:/Users/Ambreen/Downloads/developer/Sample NTS.jpg');
    bwImage=im2bw(image);
    [rows,columns,numberOfColorBands] = size(bwImage);

% personal imformation region
personal_info=bwImage(1:rows/3,1:end);%figure;imshow(personal_info);

% MCQ answers region
answer_part=bwImage((rows/3)+45:end,1:end);
%
 white_band_columns_start=180;
 white_band_columns_end=160;
 end_white_band=90;
 %figure;imshow(answer_part);
```

```
%
% % extract 4 columns from the answer parts
 [row_1,col_1,numberOfColorBand]=size(answer_part);
 Col1=answer_part(1:end-end_white_band+20,white_band_columns_start-
50:fix(col_1/4)+50);%figure,imshow(Col1);
 Col2=answer_part(1:((end-end_white_band+20)),fix(col_1/4)+60:fix(col_1/2)-
30);%figure,imshow(Col2);
 Col3=answer_part(1:end-end_white_band+20,fix(col_1/2)-30:fix(col_1/4)*3-
109);%figure,imshow(Col3);
 Col4=answer_part(1:end-end_white_band+20,fix(col_1/4)*3-50:col_1-
white_band_columns_end);%figure,imshow(Col4);

% struct for everything about answers
sum=0; % sum for all the pixels in square
choice=0; % from 1-4
image = [];
field='answer';
value={sum,choice,image};
answers=struct(field,value);
% to find columns for each of the stripped image
total_col1=0;
total_col2=0;
total_col3=0;
total_col4=0;
for j=1:4
   if j==1
      [rows,cols]=size(Col1);
      total_col1=cols;
      var=1;
      for k= 1:25
```

```
        %var=i
        answers(k).image=Col1(var:rows/25+var-1,1:end);
        %figure;imshow(answers(k).image);
        var=rows/25+var;
    end

elseif j==2
    [rows,cols]=size(Col2);
    total_col2=cols;
    var=1;
    for k= 26:50
        %var=i
        answers(k).image=Col2(var:rows/25+var-1,1:end);
        %figure;imshow(answers(k).image);
        var=rows/25+var;
    end

elseif j==3
    [rows,cols]=size(Col3);
    total_col3=cols;
    var=1;
    for k= 51:75
        %var=i
        answers(k).image=Col3(var:rows/25+var-1,1:end);
        var=rows/25+var;
    end

else
```

```
        [rows,cols]=size(Col4);
        total_col4=cols;
        var=1;
        for k= 76:100
            %var=i
            answers(k).image=Col4(var:rows/25+var-1,1:end);
            %figure;imshow(answers(k).image);
            var=rows/25+var;
        end

    end
end

fileExcel= strcat(excelsheet,'.xls');
% create an excel sheet to write results
filename = 'result_sheet.mat';
%fileExcel= excelsheet;
sheet=1;
xlRange='A1';
title={'Questions ','A','B','C','D','E';};
ExcelArray(1,:)=title;
Array=zeros(101,5);
final=zeros(100:2);
%Array(1,:)=[];

for i =1:100
   % q=Col1/25;
   % now scan and identify the choice
```

```
  % scan complete row and find sum
  % break into choices and identify choice
%    for j=1:5
   if (i<=25)
      char_part=170;
      choice_col=fix((total_col1-char_part)/5);
      threshold=5;
   elseif (i<=50 && i>25)
      char_part=130;
      choice_col=fix((total_col2-char_part)/5);
      threshold=0;
   elseif (i<=75 && i>50)
      char_part =130;
      choice_col=fix((total_col3-char_part)/5);
      threshold=10;
   elseif(i<=100 && i>75)
      char_part=100;
      choice_col=fix((total_col4-char_part)/5);
      threshold=10;
   end


    char=answers(i).image(1:end,1:char_part);%figure;subplot(2,3,1);imshow(char);
   % 1-25



choice1=answers(i).image(1:end,1+char_part:(char_part+choice_col));%subplot(2,3,2);imshow(
choice1);


choice2=answers(i).image(1:end,(char_part+choice_col)+1:(char_part+choice_col+choice_col));
%subplot(2,3,3);imshow(choice2);
```

```
choice3=answers(i).image(1:end,(char_part+choice_col+choice_col):(char_part+choice_col+cho
ice_col+choice_col));%subplot(2,3,4);imshow(choice3);

    choice4=answers(i).image(1:end,(char_part+choice_col+choice_col+choice_col)+1-
threshold:(char_part+choice_col+choice_col+choice_col+choice_col-
threshold));%subplot(2,3,5);imshow(choice4);

    choice5=answers(i).image(1:end,(char_part+choice_col+choice_col+choice_col+choice_col)-
threshold*2:end);%subplot(2,3,6);imshow(choice5);

    % decide question solved ot unsolved if solved which choice is filled,

    % also store results for multiple choices marked and consider them wrong

    %display(choice1);

    clear sum;


   S1= sum(choice1(:)==0);

  %xslwrite(filename,S1)

 % display(S1);

   S2= (sum(choice2(:)==0));

  %display(S2);

   S3= (sum(choice3(:)==0));

  %display(S3);

   S4= (sum(choice4(:)==0));

  %display(S4);

   S5= sum(choice5(:)==0);

   %xslwrite(filename,S5)

  %display(S5);

   S=[S1,S2,S3,S4,S5];

  % find minimum of sum

  [MAX,I]=max(S);

  marked_difference=600;
```

```
  if (MAX>marked_difference)
     if (I==1)
        answers(i).choice='A';
        answers(i).sum=S1;
     elseif (I==2)
         answers(i).choice='B';
         answers(i).sum=S2;
     elseif (I==3)
        answers(i).choice='C';
        answers(i).sum=S3;
     elseif (I==4)
        answers(i).choice='D';
        answers(i).sum=S4;
     elseif (I==5)
        answers(i).choice='E';
        answers(i).sum=S5;
     end
  else
     answers(i).choice=' ';
     answers(i).sum=0;
     I=0;
  end
   final(i,:)=[MAX,I,answers(i).choice];
   %save('finalanswer.mat','final');
 % figure; %imshow(answers(i).image);
 ExcelArray{i+1,1}=i;
 ExcelArray{i+1,2}=S1;
 ExcelArray{i+1,3}=S2;
```

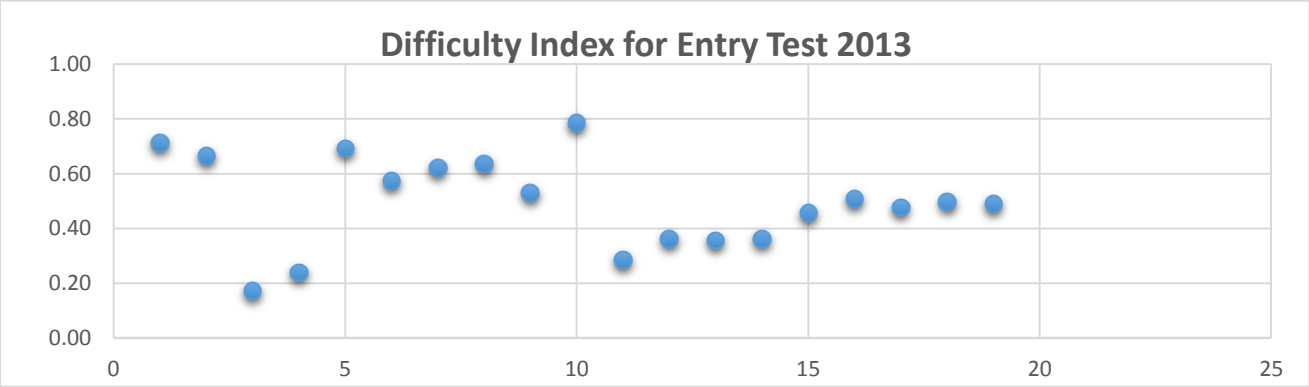
```
 ExcelArray{i+1,4}=S3;
 ExcelArray{i+1,5}=S4;
 ExcelArray{i+1,6}=S5;

end
 %ExcelArray={title;S;};
 %save(filename,'Array');
 xlswrite(fileExcel,ExcelArray,sheet);
 sheet=2;
 xlswrite(fileExcel,final,sheet);
end
```

Entry Test 2013- Analytical Part

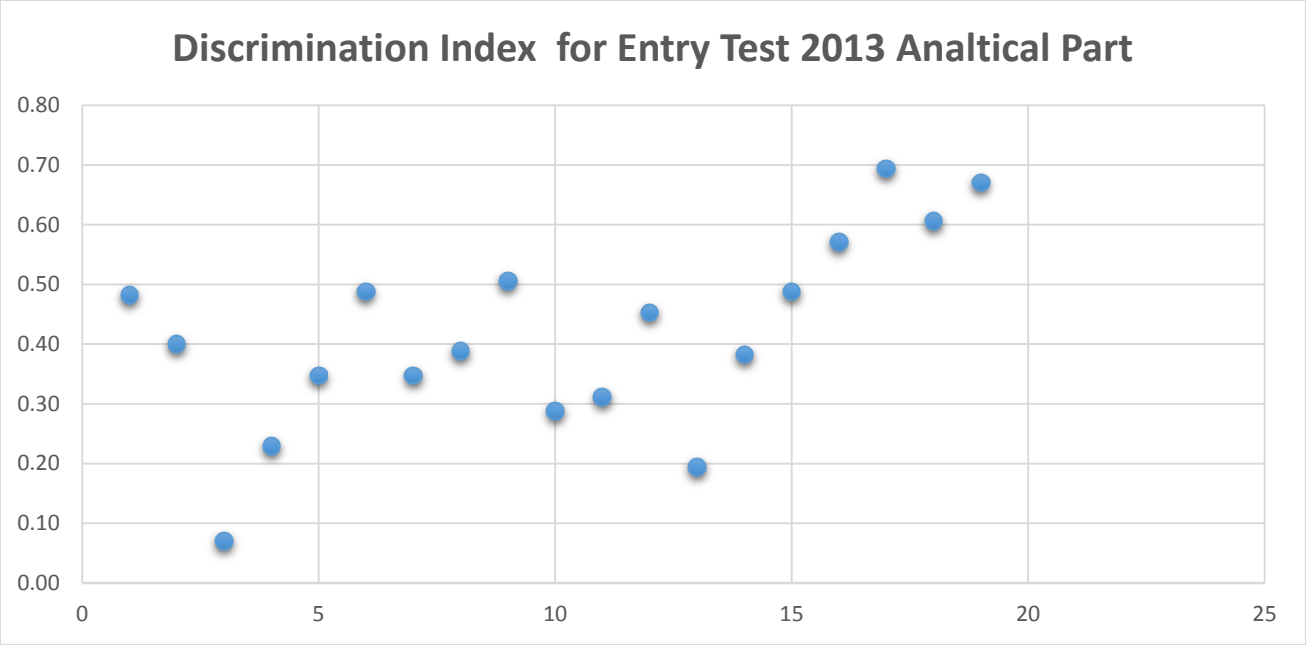
Distractors Analysis for Analytical Part Entry Test 2013				
	A	B	C	D
Q41	-0.11	-0.19		0.28
Q42	-0.12	-0.06	-0.03	
Q43	-0.08	0.03		-0.21
Q44		-0.02	-0.11	-0.04
Q45	-0.18	-0.18	-0.06	
Q46	-0.10		-0.08	-0.23
Q47	-0.12		-0.19	-0.08
Q48	0.00	-0.38	-0.17	
Q49	0.00		-0.29	-0.21
Q50	0.00	-0.16		-0.29
Q51		-0.15	-0.15	-0.25
Q52	-0.18	-0.15	-0.08	

Q53	-0.26	-0.19		-0.14
Q54	-0.08	-0.14	-0.06	
Q55	-0.10		-0.15	-0.15
Q56	0.00	-0.19		-0.06
Q57		-0.16	-0.07	-0.08
Q58		-0.12	-0.15	-0.19
Q59	0.00		-0.15	-0.05
Q60	-0.15	-0.04		-0.04



Functional- Non Functional for Entry Test 2013 Analytical Part				
Q41	3.53	2.94	0.00	2.06
Q42	0.00	7.94	10.00	10.29
Q43	7.94	14.41	0.00	8.82
Q44	10.88	21.18	0.00	50.00
Q45	28.24	25.59	19.12	0.00
Q46	0.00	10.29	7.65	6.76
Q47	13.24	15.88	0.00	10.59
Q48	8.53	9.41	18.82	0.00
Q49	8.82	0.00	10.88	15.59
Q50	29.12	9.12	8.24	0.00
Q51	7.35	0.00	10.29	2.35
Q52	7.94	30.88	0.00	30.88

Q53	20.88	22.06	19.12	0.00
Q54	0.00	21.47	24.71	16.76
Q55	30.88	0.00	7.06	24.41
Q56	16.76	0.00	17.35	18.82
Q57	0.00	18.53	12.35	16.76
Q58	6.76	30.29	13.24	0.00
Q59	8.24	14.12	0.00	26.47
Q60	10.59	0.00	22.94	15.59



Entry Test 2013- Mathematical Part

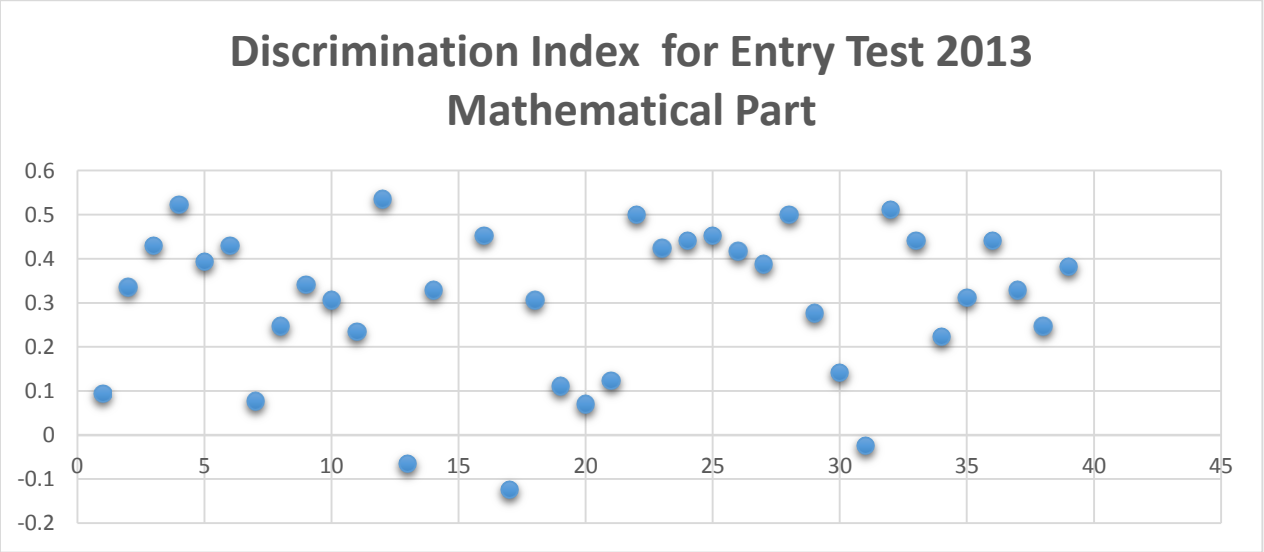
Distractors Strength Entry Test 2013- Mathematical Part				
Q61	-0.21	-0.11		-0.08
Q62		0.06	0.02	-0.12
Q63	-0.15	-0.18		0.01
Q64	-0.18	-0.11		-0.12
Q65	-0.18		-0.22	-0.10
Q66	-0.12		-0.21	-0.02

Q67	-0.18		-0.16	-0.06
Q68		-0.06	-0.04	0.02
Q69	-0.01	-0.15	-0.06	
Q70	-0.05		-0.21	-0.06
Q71		-0.12	-0.11	-0.06
Q72	-0.16	-0.09		0.06
Q73	-0.18	-0.25		-0.09
Q74	-0.07		-0.15	0.29
Q75	-0.06	-0.19		-0.02
Q76	0.01	0.05	0.10	0.11
Q77		-0.08	-0.21	-0.15
Q78	-0.02	-0.10		0.28
Q79	-0.05		-0.18	-0.05
Q80	0.02	-0.08		-0.04
Q81	-0.15	-0.09		0.22
Q82	0.07	-0.11		-0.03
Q83		-0.12	-0.16	-0.18
Q84	-0.08	-0.23	-0.11	
Q85		-0.07	-0.18	-0.16
Q86		-0.19	-0.16	-0.08
Q87	-0.10	-0.21		-0.09
Q88	-0.11	-0.22		-0.02
Q89	-0.28	-0.11	-0.09	
Q90	-0.10		-0.11	-0.05
Q91		-0.05	-0.07	0.00
Q92	-0.07		0.16	-0.05
Q93	-0.18	-0.21		-0.09
Q94	-0.05		-0.22	-0.11

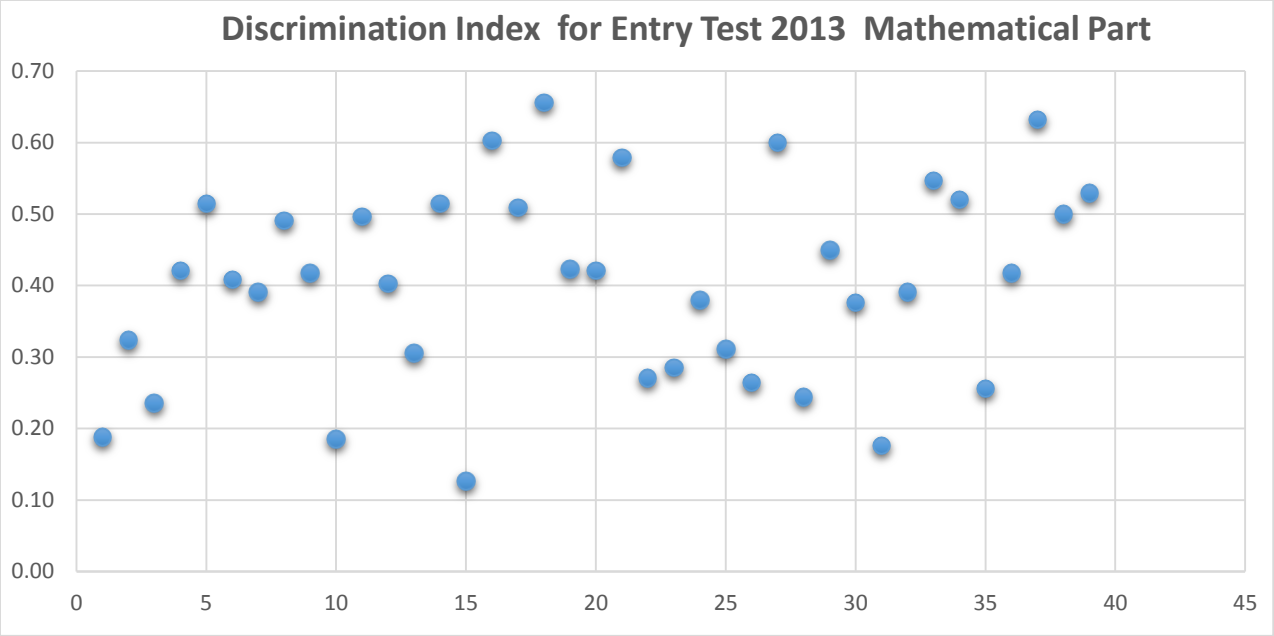
Q95	-0.06	0.06	-0.19	
Q96	-0.02		-0.22	-0.06
Q97	-0.09	-0.21		-0.12
Q98		-0.21	-0.05	-0.04
Q99		-0.11	-0.06	-0.05
Q100		-0.21	-0.06	-0.08

Functional- Non Functional for Entry Test 2013 Mathematical Part				
Q61	5.88	0.00	70.00	4.41
Q62	0.00	7.65	21.76	36.47
Q63	0.00	16.47	33.82	22.65
Q64	26.47	17.94	0.00	12.35
Q65	28.82	7.65	11.18	0.00
Q66	16.76	0.00	28.24	12.94
Q67	7.35	41.76	0.00	11.18
Q68	0.00	20.29	11.18	17.65
Q69	11.47	30.29	15.29	0.00
Q70	12.94	0.00	16.18	51.18
Q71	0.00	19.12	10.00	19.41
Q72	15.29	14.71	0.00	27.35
Q73	0.00	31.47	20.00	15.88
Q74	0.00	14.71	19.71	12.94
Q75	0.00	25.29	27.94	30.59
Q76	9.71	13.82	0.00	14.71
Q77	15.59	23.53	0.00	7.94
Q78	17.94	0.00	10.29	5.00
Q79	33.82	0.00	14.41	7.65
Q80	24.41	21.76	0.00	10.59
Q81	14.71	0.00	18.53	7.35
Q82	16.18	15.88	0.00	38.24
Q83	45.29	12.35	0.00	12.65
Q84	25.59	0.00	18.82	16.18
Q85	45.29	16.47	5.00	0.00
Q86	10.59	42.65	0.00	17.94
Q87	0.00	23.53	9.71	5.59
Q88	3.53	24.41	0.00	45.59
Q89	9.12	0.00	29.71	14.12

Q90	9.12	28.24	0.00	21.76
Q91	22.35	42.06	15.29	0.00
Q92	0.00	14.41	33.53	10.59
Q93	13.24	7.06	0.00	22.06
Q94	18.82	0.00	14.12	11.76
Q95	25.88	24.71	0.00	19.71
Q96	8.24	35.59	0.00	11.76
Q97	0.00	5.59	18.53	11.76
Q98	0.00	25.00	6.18	15.88
Q99	4.71	0.00	25.59	15.29
Q100	14.12	13.82	0.00	16.18



Appendices


Discrimination Index for Entry Test 2013 Mathematical Part
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
5
10
15
20
25
30
35
40
45