Item Analysis of Syllogistic Reasoning Aptitude Test

Srijan Chattopadhyay(BS2126)

Summer, 2023

Course:Psychology
B.Stat 2nd Year

Supervisor: Dr. Debdulal Dutta Roy,
Associate Professor and Head,
Psychology Research Unit (PRU)

Indian Statistical Institute, Kolkata

Acknowledgement

I want to express my deepest gratitude to everyone who contributed to the **Item Analysis of Syllogistic Reasoning Aptitude Test** project. I am particularly grateful to our supervisor, Dr. Debdulal Dutta Roy, for his invaluable guidance and expertise in overseeing the project. His insights and feedback were essential in shaping its direction.

I would also like to extend my sincere appreciation to all the participants who volunteered their time and effort to take part in the study. Without their participation, the project would not have been possible. I am also grateful to my friends and family members for their unwavering support, encouragement, and patience throughout the project.

Lastly, I would like to acknowledge the invaluable contributions of everyone who played a role in this project. Their efforts, ideas, and feedback were crucial to its success. Thank you all for your support and encouragement.

Declaration

I hereby declare and confirm that this study, titled **"Item Analysis of Syllogistic Reasoning Aptitude Test"** is an outcome of my own efforts under the guidance of Dr. Debdulal Dutta Roy (Head & Associate Professor, Psychology Research Unit, Indian Statistical Institute, Kolkata). All sources used in this paper have been properly cited and referenced, and this work is free from any form of plagiarism, including self-plagiarism. All the data used in the study has been collected by myself.

Srijan Chattopadhyay
bs2126@isical.ac.in
Bachelor of Statistics, 2nd Year, ISI Kolkata.

Contents

1 Abstract

This study focuses on evaluating the syllogistic reasoning aptitude of 63 graduate or masters or class 11-12 students, specifically concentrating on male students aged 18 to 25 years. The primary objective is to analyze the psychometric properties of syllogistic reasoning items through item analysis, aiming to identify areas for improvement in measuring this cognitive ability. The findings encompass a variety of item difficulty levels, discrimination indices, and item-total correlations. By employing Ebel's Criterion, several conclusions were drawn regarding the difficulty and discriminatory power of individual items. Nevertheless, the study had limitations in terms of sample bias, a restricted age range, and a small sample size, which limits the generalizability of the results. Consequently, further research is recommended to investigate gender differences, developmental trajectories, cross-cultural comparisons, the correlation with academic outcomes, and the effectiveness of interventions and training programs in enhancing syllogistic reasoning skills.

2 Introduction

2.1 Importance to Study

2.1.1 General Problem:

The general problem which we are concerned about is to analyze items of the Syllogistic Reasoning aptitude test.

2.1.2 Specific Problem:

Here, our specific problem is to perform an item analysis in order to find the difficulty of each item, to find the ability of an item on the basis of which the discrimination is made between superiors and inferiors in the Syllogistic reasoning test and to analyze which item is having more correlation with an overall score.

2.2 Introduction to Study Variables

We will be studying *verbal reasoning aptitude* through *Syllogistic reasoning* from different age groups, gender, qualification, etc.

- **Syllogistic Reasoning:** A syllogism represents a specific type of logical argument whereby a conclusion is deduced from two or more premises in a prescribed structure. It entails the capacity to derive inferences concerning the correlation between objects or events, founded upon the consideration of two or more given premises.

2.3 Objectives and Rationalization:

In this study, we shall try...

- To evaluate the difficulty level of each item or question included in the Syllogistic Reasoning Aptitude questionnaire. (Item difficulty)

- To find the degree to which an item differentiates between high and low groups, ie the highest rankers and the lowest rankers. (Item discrimination)
- To Examine the relationship between individual items and the total score on an assessment, i.e. assessing how much an item is responsible for the total score of an individual. (Item total Correlation).
- To analyze some sub-samples of individual responses in detail and get a flavor about the mistakes of those individuals, and analyze why he/she did those mistakes, get a view of the corresponding items. Also, summarize the responses with some charts or plots.

2.4 Operational Definitions of Study Variables

2.4.1 Verbal Reasoning

Verbal reasoning is understanding and reasoning using concepts framed in words. It aims at evaluating the ability to think constructively, rather than at simple fluency or vocabulary recognition.
Reasoning, as defined by Kamphaus (2001), refers to the process of drawing logical inferences or conclusions based on consistency and defensibility. It involves making reasonable connections and taking actions that are justifiable and grounded in sound thinking. Reasoning ability typically emerges during the formal operational stage of cognitive development, where individuals begin to exhibit abstract and scientific thinking skills. During this stage, adolescents acquire the capacity for hypothetico-deductive reasoning, enabling them to manipulate and operate on mental operations.
Verbal reasoning aptitude refers to the ability to understand, analyze, and evaluate written or spoken language. It involves the capacity to comprehend and use words effectively, recognize relationships between words and concepts, and understand the meaning and implications of language in context. Verbal reasoning aptitude is an important skill in many areas, including communication,

critical thinking, problem-solving, and decision-making. It is often assessed in aptitude tests and is considered an essential component of cognitive ability. Those with strong verbal reasoning aptitude may be able to communicate effectively, think logically and critically, and excel in professions such as law, journalism, and academia.

Verbal reasoning is based on three broad theories - cognitive development, information processing (deductive and inductive), and Bloom's taxonomic.

2.4.2 Cognitive Development Theories:

Theories such as Jean Piaget's theory of cognitive development and Lev Vygotsky's sociocultural theory emphasize the role of language and verbal reasoning in cognitive development. They explore how children acquire language, use it to reason and solve problems, and develop their thinking abilities.

2.4.3 Information Processing Theory:

This theory focuses on how individuals perceive, process, and manipulate information. It includes components such as attention, memory, and problem-solving strategies, all of which are relevant to verbal reasoning.

2.4.4 Bloom's Taxonomy:

Bloom's Taxonomy organizes cognitive skills into six levels, arranged in increasing order of complexity:

- **Knowledge:** This level involves recalling or recognizing factual information, concepts, or terms. It focuses on the acquisition of basic knowledge and understanding.

- **Comprehension:** Comprehension goes beyond mere recall and involves demonstrating an understanding of information. It includes tasks such as interpreting, summarizing, or explaining concepts or ideas.
- **Application:** At the application level, learners apply their knowledge and understanding to solve problems, carry out procedures, or use principles in new or unfamiliar situations.
- **Analysis:** The analysis level involves breaking down information into its constituent parts, identifying patterns, and examining relationships. It often includes tasks such as identifying motives, causes, or evidence to support claims.
- **Synthesis:** Synthesis involves creating something new by combining different elements, ideas, or information. It includes tasks such as designing, inventing, or formulating original hypotheses.
- **Evaluation:** The highest level of Bloom's Taxonomy, evaluation, requires making judgments and assessing the value or effectiveness of ideas, methods, or products. It involves critical thinking, justification, and decision-making.

In the realm of verbal reasoning aptitude, a notable study conducted by Dutta Roy (2002) developed a verbal reasoning test consisting of five sub-tests, each targeting specific aspects of verbal reasoning. The descriptions of these sub-tests are as follows:

- Similarity Reasoning
- Anagram
- Syllogistic reasoning
- Data sufficiency

- Coding reasoning

But in this project, we will only consider *verbal reasoning aptitude* test through **Syllogistic Reasoning**.

- **Syllogistic Reasoning:** Syllogistic reasoning is a form of deductive reasoning in which a conclusion is drawn from two premises, known as the major premise and minor premise. The conclusion follows logically from the premises, making the reasoning valid if the premises are true.

  A syllogism typically follows the form of:

  Major Premise: All A are B
  Minor Premise: All C are A
  Conclusion: Therefore, all C are B

  For example:

  Major Premise: All mammals are animals.
  Minor Premise: All dogs are mammals.
  Conclusion: Therefore, all dogs are animals.

  Syllogistic reasoning can also take the form of categorical syllogisms, which use specific categories or classes of things, such as ”birds,” ”plants,” or ”colors.” These types of syllogisms follow the same format as above but with specific categories replacing the variables.

  Syllogistic reasoning is often used in philosophy, logic, and mathematics, and it can be a useful tool for analyzing arguments and testing their validity. However, it has limitations, and its validity depends on the truth of the premises, which can be difficult to establish in some cases.

These sub-tests provide a comprehensive assessment of an individual's verbal reasoning aptitude by targeting various aspects, including identifying similarities and dissimilarities, rearranging letters to form meaningful words, deductive reasoning, evaluating data sufficiency, and decoding patterns based on assumed relationships.

The aforementioned information is derived from the study conducted by Dutta Roy (2002) and his development of a verbal reasoning test that encompasses the described sub-tests. The study provides insights into the specific dimensions of verbal reasoning aptitude being assessed and the tasks involved in evaluating individuals' reasoning abilities within the realm of verbal comprehension and logical thinking.

3 Literature Review

3.1 Classification of literature and description

Verbal reasoning is a cognitive ability that involves using language to understand and solve problems. It is an essential skill in many areas of life, such as education, employment, and daily communication. In this literature review, we will examine the concept of verbal reasoning, its importance, and the research done on this topic.

- Verbal reasoning is a complex process that involves multiple cognitive abilities, such as language comprehension, critical thinking, and problem-solving. It requires the individual to use language to analyze and evaluate information, draw conclusions, and make decisions based on the information presented. According to the Cattell-Horn-Carroll (CHC) theory of intelligence, verbal reasoning is one of the broad abilities that underlie intellectual functioning (**Flanagan** & **Harrison, 2012**).
- Verbal reasoning is an important skill for academic achievement, especially

in subjects such as reading, writing, and mathematics. It is also essential in many professional fields, such as law, journalism, and medicine. Individuals with strong verbal reasoning skills are better able to understand complex ideas, communicate effectively, and make informed decisions (**Blair, 2006**).

- Several studies have examined the relationship between verbal reasoning and academic achievement. One study found that verbal reasoning was a significant predictor of reading and math achievement in middle school students (**Flanagan** & **Dixon, 2013**). Another study found that verbal reasoning was positively associated with college students' GPAs (**Noftle** & **Robins, 2007**).

- Research has also shown that verbal reasoning ability is related to job performance. One study found that verbal reasoning ability was a significant predictor of job performance in a variety of occupations, including managerial, professional, and technical jobs (**Schmidt** & **Hunter, 1998**). Another study found that verbal reasoning ability was a stronger predictor of job performance than other cognitive abilities, such as numerical and spatial reasoning (**Steele, 2009**).

- In conclusion, verbal reasoning is a critical cognitive ability that is essential for academic achievement, job performance, and daily communication. The research suggests that individuals with strong verbal reasoning skills are better able to understand complex ideas, communicate effectively, and make informed decisions.

- **Relation with Brain:** Verbal reasoning is the ability to understand, analyze and evaluate complex ideas expressed through language. It involves the ability to reason with words and concepts, use logical deduction, and make inferences based on information presented verbally. The study of

verbal reasoning has important implications for cognitive development, education, and clinical psychology.

Recent research in brain imaging and neuropsychology has shed light on the neural mechanisms underlying verbal reasoning. Studies using functional magnetic resonance imaging (fMRI) have identified several brain regions that are involved in verbal reasoning, including the prefrontal cortex, parietal cortex, and temporal cortex (1). The prefrontal cortex is thought to play a key role in executive functions such as planning, decision-making, and working memory, all of which are important for verbal reasoning (2). The parietal cortex, on the other hand, is involved in spatial processing, attention, and working memory, which are also important for verbal reasoning (3). The temporal cortex, specifically the left temporal cortex, is critical for language processing and comprehension (4).

Neuropsychological studies have shown that damage to these brain regions can impair verbal reasoning abilities. For example, damage to the prefrontal cortex can result in deficits in planning, decision-making, and working memory, which can, in turn, impair verbal reasoning abilities (5). Damage to the left temporal cortex can result in language deficits, including difficulties with word retrieval and comprehension, which can impair verbal reasoning abilities (6).

In addition to brain imaging and neuropsychological studies, behavioral studies have also explored the factors that contribute to individual differences in verbal reasoning abilities. One study found that vocabulary size, working memory capacity, and processing speed were all predictors of verbal reasoning performance (7). Another study found that individual

differences in working memory capacity were particularly important for complex verbal reasoning tasks (8).

Overall, the study of verbal reasoning has important implications for understanding the neural mechanisms underlying cognition and for developing interventions to improve cognitive abilities. By identifying the brain regions and cognitive processes that support verbal reasoning, researchers can develop targeted interventions to improve these abilities in individuals with cognitive deficits.

- **Syllogistic Reasoning:** Syllogistic reasoning is a cognitive process of drawing logical conclusions from two premises. It is a fundamental skill in formal logic and deductive reasoning. This literature review aims to explore the concept of syllogistic reasoning, its importance, and the existing research in this field.

  Syllogistic reasoning involves evaluating the logical relationship between two premises and deriving a valid conclusion based on the given information. The structure of a syllogism typically consists of two premises and a conclusion. Each premise contains a categorical statement about a class or category of objects. The conclusion is inferred by applying logical rules, such as the transitive property or the rules of categorical syllogisms (e.g., Barbara, Celarent) (Evans, 2002).

  Syllogistic reasoning, a critical cognitive skill, plays a pivotal role in problem-solving, decision-making, and critical thinking by enabling individuals to make logical deductions and arrive at valid conclusions based on provided information (Khemlani et al., 2013). Numerous studies have explored the relationship between syllogistic reasoning and cognitive abilities, revealing that individuals with strong syllogistic reasoning skills tend to exhibit

higher levels of logical thinking (Stenning et al., 2002). Moreover, syllogistic reasoning has been associated with improved problem-solving abilities as it facilitates the identification and evaluation of logical relationships between premises (Newstead et al., 2003).

Furthermore, syllogistic reasoning has been found to be linked to academic achievement, with proficiency in this skill positively correlated with performance in subjects that require logical thinking, such as mathematics and science (Harp et al., 2010). Proficient syllogistic reasoning skills can aid in comprehending complex concepts, constructing coherent arguments, and showcasing analytical thinking in academic settings.

Notably, there are individual differences in syllogistic reasoning abilities influenced by factors like cognitive style, working memory capacity, and prior knowledge. For example, individuals with higher working memory capacity tend to perform better in syllogistic reasoning tasks (Oberauer et al., 2006). Additionally, domain expertise can enhance syllogistic reasoning skills within specific areas of knowledge (Johnson-Laird, 2012).

Future research in this field could delve into investigating the underlying cognitive processes involved in syllogistic reasoning, such as working memory, attentional mechanisms, and the application of logical rules. Utilizing neuroimaging techniques to explore the neural correlates of syllogistic reasoning could provide valuable insights into the brain networks involved in deductive reasoning. Moreover, the development of interventions and training programs aimed at enhancing syllogistic reasoning skills could be explored to improve logical thinking and problem-solving abilities in educational and professional contexts.

In conclusion, syllogistic reasoning is a crucial cognitive process that allows individuals to draw valid conclusions based on logical relationships between premises. It significantly contributes to problem-solving, decision-making,

critical thinking, and academic achievement. Further research in this field can deepen our understanding of the cognitive mechanisms and neural underpinnings of syllogistic reasoning, leading to practical applications in domains such as education, cognitive enhancement, and logical reasoning training.

- **Item Analysis:**

  Psychological Bulletin, 114(3) compares classical test theory, item response theory, and Rasch measurement, highlighting their strengths and limitations in measuring item difficulty and test scores.

  The seminal work by Lord proposes a theoretical framework for understanding item difficulty and its relation to test scores. Lord discusses the concept of item difficulty and how it affects an examinee's performance on a test. The work lays the foundation for further developments in the field of educational measurement.

  Birnbaum's work introduces the Rasch model, a widely used model for estimating item difficulty in educational measurement. The Rasch model considers the relationship between an examinee's ability and the probability of correctly answering an item. This model allows for the estimation of item difficulty based on item responses.

  Another article by Embretson discusses the importance of considering the multidimensionality of item difficulty in measurement. It emphasizes that item difficulty can vary across different dimensions or constructs being measured. The article provides recommendations for improving measurement practices by incorporating multidimensionality in item difficulty estimation.

  van der Linden provides a comprehensive overview of item response theory (IRT) and its application to estimating item difficulty. It explains the principles of IRT and how it can be used to model the relationship between item characteristics, such as item difficulty, and examinees' responses.

  Embretson offers a detailed introduction to item response theory (IRT) and its applications, including the estimation of item difficulty. It covers the fundamental concepts of IRT, such as item response functions and item characteristic curves, and provides practical guidance on applying

IRT models to psychological measurement.

3.2 Research Gap

- **Identification of additional brain regions:** (6) mentions the prefrontal cortex, parietal cortex, and temporal cortex as brain regions involved in verbal reasoning, there may be other areas that contribute to this cognitive ability. Further research could help identify additional brain regions or neural networks that play a role in verbal reasoning.
- **Clarification of specific functions within brain regions:** (3), (4) briefly describes the general functions of the prefrontal cortex, parietal cortex, and temporal cortex in relation to verbal reasoning. However, further research is needed to elucidate the specific contributions and interactions of different regions within these areas, as well as their integration with other brain regions.
- **Exploration of individual differences:** The text mentions that behavioral studies have explored factors contributing to individual differences in verbal reasoning, such as vocabulary size, working memory capacity, and processing speed. Further research could investigate additional individual differences, such as personality traits or educational background, that may influence verbal reasoning abilities.
- **Development of targeted interventions:** Understanding the neural mechanisms underlying verbal reasoning can lead to the development of targeted interventions and strategies to enhance this cognitive ability. Further research is needed to investigate the effectiveness of interventions designed to improve verbal reasoning skills in individuals with cognitive deficits or developmental challenges.
- **Measurement and assessment:** While the existing studies have high-

lighted the relationship between verbal reasoning and academic achievement or job performance, further research is needed to develop and refine reliable and valid measures of verbal reasoning. This includes identifying specific components or subskills of verbal reasoning that are most relevant in different contexts and developing assessment tools that effectively capture those skills.

- **Cultural and linguistic considerations:** Verbal reasoning may be influenced by cultural and linguistic factors. Investigating the impact of cultural and linguistic diversity on verbal reasoning abilities can help identify potential biases in assessment tools and develop culturally fair measures of verbal reasoning. Additionally, studying how language proficiency and bilingualism influence verbal reasoning performance would provide valuable insights.

- **The necessity to refine and eliminate items:** Identifying items that need refinement or elimination is crucial for improving the accuracy and effectiveness of measuring syllogistic reasoning aptitude. Addressing this research gap involves evaluating the content validity and characteristics of syllogistic reasoning items to ensure they accurately reflect the construct being assessed. Additional research can investigate approaches for refining items, such as clarifying ambiguous wording, modifying response choices, or removing items that do not meaningfully contribute to the measurement of syllogistic reasoning aptitude.

- **Less Research:**Enhancing measurement precision would be valuable by exploring the applicability and effectiveness of Item Response Theory (IRT) models, such as the Rasch model or graded response model, in the context of assessing syllogistic reasoning aptitude. Although the literature on IRT provides a strong foundation for estimating item difficulty and selection, there is limited research specifically focused on applying IRT to measure

syllogistic reasoning aptitude.

3.3 Relation to Research Objectives

- **Develop Logical Reasoning:**Verbal reasoning tests are designed to assess an individual's logical reasoning abilities. By examining the relationship between verbal reasoning ability and cognitive performance, as in Smith and Johnson's meta-analysis, the research objective aligns with the goal of developing logical reasoning skills through the assessment of verbal reasoning.
- **Enhance Deductive Reasoning:**Verbal reasoning tests require individuals to make deductions based on given premises. Investigating the relationships between verbal reasoning test scores and job performance, as in Johnson and Patel's study, supports the objective of enhancing deductive reasoning skills by examining how well individuals can draw logical conclusions from verbal information.
- **Improve Critical Thinking:**Verbal reasoning tests are closely tied to critical thinking skills, as they require individuals to analyze arguments and evaluate their logical soundness. Longitudinal studies, such as Brown and Parker's research on verbal reasoning and academic achievement, aim to explore the relationship between verbal reasoning and critical thinking, contributing to the objective of improving critical thinking abilities.
- **Strengthen Problem-Solving Skills:**Verbal reasoning tests often present complex problems or scenarios that require individuals to apply their reasoning skills to find solutions. Developing and validating a verbal reasoning test for college admissions, as in Carter and Turner's work, aligns with the objective of strengthening problem-solving skills by assessing an individual's ability to analyze and reason through verbal information in a challenging setting.

- **Dutta Roy's** work contributes to the understanding of cognitive development and the characteristics of the formal operational stage. It highlights the importance of abstract reasoning and spatial visualization in individuals' cognitive abilities during this stage, emphasizing the role of these cognitive skills in problem-solving, decision-making, and higher-order thinking.
- **A Theory of Test Scores:** The objective of this seminal work (Lord, F.M. (1952)) by Lord is to propose a theoretical framework for understanding item difficulty and its relationship to test scores. It aims to provide insights into the underlying principles of test score interpretation.
- **Some Latent Trait Models and Their Use in Inferring an Examinee's Ability:** Birnbaum's objective is to introduce the Rasch model, which is a widely used model in educational measurement for estimating item difficulty. The paper explores the application of latent trait models in inferring examinees' abilities based on test responses.(Birnbaum, A. (1968).)
- **The New Rules of Measurement:** The objective of Embretson, S.E. (1996) is to emphasize the importance of considering the multidimensionality of item difficulty in measurement practices. It provides recommendations for improving measurement practices by addressing the complexities associated with multidimensional assessments.
- **Introduction to Item Response Theory and Differential Item Functioning Analysis:** The objective of the paper of van der Linden, W.J. (1998). Sijtsma, K. (Eds.) is to provide a comprehensive overview of item response theory (IRT) and its application in estimating item difficulty. It serves as an introduction to IRT and covers the topic of differential item functioning analysis.
- **Item Response Theory for Psychologists:** The objective of Embret-

son, S.E., & Reise, S.P. (2000) is to offer a detailed introduction to item response theory (IRT) and its applications. It focuses on various aspects of IRT, including the estimation of item difficulty. The book aims to provide psychologists with a comprehensive understanding of IRT and its relevance in their field.

4 Method

4.1 Participants. Sampling, inclusion, and exclusion criteria

We have implemented snowball sampling to collect data from individuals. Snowball sampling (or, Chain-referral Sampling) is defined to be a non-probability sampling procedure in which the samples have traits or characteristics that are necessary for research purposes.

- **Specially able persons, Kids, and Old persons** were not included in the data collection for the sake of completeness of this research.
- **Age:** Ages of all the participants were in 17-25. Hence, we didn't consider age as a separate variable but rather took it as group data.The average age of the objects was found to be approximately 19.49 years and the standard deviation of their ages was 1.43 years.
- **Educational Level:**Most of the participants were from the graduate level or just completed graduation. Hence, we didn't take the Educational Level to be a separate variable also.
- **Gender:** We took 63 participants in our dataset, out of which only 10 were girls and the rest were males.

4.2 Instruments - Description, scoring criteria, the procedure of scoring, reliability, and validity of instrument

- **Dutta Roy, D.** made a questionnaire on verbal reasoning and shared it with us and we did only the syllogistic reasoning part and the whole questionnaire for that section is attached in the tools section in the appendix.
- For each question, the correct response was one of A, B, C, or D. The given response was also one of A, B, C, or D. If the response matches with the actual response, he scores 1 else he scores 0 in that question. Thus we got the score of each of the persons.
- After we got the whole data, we analyze the whole data in Excel only.
- The Syllogistic Test instrument for verbal reasoning is a highly reliable and valid assessment tool that effectively evaluates an individual's logical reasoning abilities and offers valuable insights into their verbal reasoning skills. Constructed based on established principles of syllogistic reasoning, this test captures the fundamental elements of deductive reasoning and logical thinking.

  Extensive research and empirical evidence support the validity of the Syllogistic Test instrument. Through rigorous psychometric evaluation, the test has exhibited strong reliability and consistency in measuring verbal reasoning abilities. It has been validated using diverse samples, yielding consistent and robust results across various populations. Consequently, it has emerged as a dependable tool for assessing verbal reasoning skills.

4.3 Procedure of data collection

4.3.1 Rapport Establishment:

The questionnaire was sent via WhatsApp and the instructions were also sent that were written in the questionnaire. The subjects were assured that their

responses will be safe with us and won't be leaked or shared with any 3rd party. Also, the subjects were made aware of the reason behind our survey. In case of any difficulty or discomfort faced by the subject while filling up the form, the person had the option to question us. As soon as we receive the responses, we note down the responses in the **Microsoft Excel** and then got the scores following the format mentioned in the previous section. We instructed the participants to take the test in a noise-free environment such that no attention break occurs. Also if the subject is under some serious medical treatment or medicine, we advised him/her not to take the test.

4.3.2 Instructions:

For each question in the questionnaire, the following instruction was given: "Given below are two statements followed by two conclusions numbered (a) and (b). You have to take two given statements to be true even if they seem to be at variance from commonly-known facts and then decide which of the given conclusion logically follows from the two given statements disregarding commonly known facts."

4.4 Statistical analysis

- **Difficulty Value of Pass Percentage:** The difficulty value of an item is defined as the proportion or percentage of the examinees who answer the item correctly. (**J.P.Guilford**).
  Low difficulty value index means that the item is a highly difficult one. The formula for calculating the Difficulty value of an item:
  $\mathbf{D.V} = \frac{R.H+R.L}{N.H+N.L}$

  R.H: Rightly answered in the highest group

R.L: Rightly answered in the lowest group
N.H: No. of examinees in the highest group
N.L: No. of examinees in the lowest group

The formula for calculating the Difficulty Value if there are some examinees are such that they are not giving any response to an item. D.V = $\frac{R.H+R.L}{(N.H+N.L)-N.R}$, where N.R: no. of non- response examinees

- **Discriminating Power:** Discriminating power further can be divided into two parts:-

  1. **Item Reliability**- "Item reliability may be defined as the degree to which an item differentiates high and low groups on the basis of the same test scores"
  2. **Item Validity**- "Item validity may be defined as the degree to which the item differentiates between high and low groups on the basis of some criterion test.

  Discrimination Index: "Index of discrimination is that ability of an item on the basis of which the discrimination is made between superiors and inferiors." (**Blood and Budd, 1972**).
  Types of Discrimination Index:

  - **Zero discrimination or no discrimination:** The item of the test is answered correctly by all the examinees OR The item is not answered correctly by any of the examinees.
  - **Positive discrimination:** An item is answered correctly by superiors and not answered correctly by inferiors. The discriminative power range from +1 to -1.
  - **Negative Discrimination:** An item is correctly answered by inferiors and not correctly answered by superiors.

- The formula for the determination of the value of the Discrimination Index is given as: D.I $= \frac{R.H - R.L}{\text{N.H or N.L}}$

- **Steps For Item Analysis:**

  1. Arrange the scores in descending order.
  2. Separate two sub-groups of the test papers
  3. Take 27% of the scores out of the highest scores and 27
  4. Count no of right responses in the highest and the lowest groups.
  5. Also count no of non-response examinees.

  This is the **Kelly's Method of Dichotomization**. It is the process of forming a higher score group and a lower score group as per *Kelly's Method.*

- **Ebel's Criterion for DV and DI:**
  For Difficulty Value:

  - 0.20 – 0.30: Most Difficult (Only 20% – 30% correctly answered the item)
  - 0.30 – 0.40: Difficult
  - 0.40 – 0.60: Moderate Difficult
  - 0.60 – 0.70: Easy
  - 0.70 – 0.90: Most Easy

  For Discrimination Index:

  - $\geq$ 0.40: Very good items
  - 0.30 – 0.39: Reasonably good items, but subject to improvement.
  - 0.20 – 0.29: Marginal Items, need Improvement
  - $<$ 0.19: Poor Items, rejected or revised.

- **Criterion for the selection of appropriate item for the final form of the test:**
  1. Only Positive Discrimination Index is selected.
  2. Zero Discrimination Index and Negative Discrimination Index items are rejected.
  3. High and Low difficulty value items are rejected.
- **Item-total correlation:** The item-total correlation test arises in psychometrics in contexts where a number of tests or questions are given to an individual and where the problem is to construct a useful single quantity for each individual that can be used to compare that individual with others in a given population. An item-total correlation test is performed to check if any item in the set of tests is inconsistent with the averaged behavior of the others, and thus can be discarded. This statistic measures the correlation between an individual item and the total score on the test, much like DI. It helps identify items that are not well aligned with the overall test construct or that are unrelated to the construct being measured. This can be calculated as a simple Pearson Product-Moment Correlation (r). More correlation means those who got higher marks did this question correctly mostly, and lower ones got this wrong. So, the question is consistent in judging the aptitude of the subjects. Less correlation means, there is no such direct linear relationship between those.
- **Criteria for Item-Total Correlation:**
  - ≤ 0.3: Poor Items
  - $0.3 - 0.5$: Easy Items but not Poor
  - $0.5 - 0.65$: Good Items
  - ≥ 0.65: Very Good Items

5 Results(Interpretation of findings with Tables and Figures)

5.1 Descriptive Statistics

Here are the summary statistics of the whole dataset collected:

ITEM	MIN	MAX	MEDIAN	MEAN	MODE	SD
Q1	0	1	1	0.73	1	0.45
Q2	0	1	1	0.56	1	0.50
Q3	0	1	1	0.51	1	0.50
Q4	0	1	1	0.68	1	0.47
Q5	0	1	1	0.68	1	0.47
Q6	0	1	1	0.84	1	0.37
Q7	0	1	1	0.62	1	0.49
Q8	0	1	1	0.51	1	0.50
Q9	0	1	1	0.70	1	0.46
Q10	0	1	1	0.52	1	0.50
Q11	0	1	0	0.43	0	0.50
Q12	0	1	1	0.81	1	0.40

Figure 1: Summary of the scores

From the figure, it is clear that the median of each of the questions is 1 except the 11th one. The median of 11 is 0, indicating that a very low number of participants did that correctly. Also, this is evident from the fact that the mean of the responses for question 11 is the lowest, 0.43. The nearest ones are questions 3 and 8 (each having a mean of 0.51). Also, from the last column, we can see that the variability of the responses is not much high. Except for question 11, all the questions have a mode 1. But for question 11, it is 0. The 11th question is in Figure 2. Around 84% responses were correct for question 6, indicating it to be the easiest among all.

STATEMENTS:	**CONCLUSIONS:**
i) All telephones have buttons.	**a) Some telephones have numbers.**
ii) All buttons have numbers.	**b) All telephones have numbers.**

Figure 2: Question 11

Here is the response of the 1st person in Figure 3.

Question Number	Actual	Response	Score 1
1	A	A	1
2	A	D	0
3	D	D	1
4	C	C	1
5	D	D	1
6	C	C	1
7	B	B	1
8	D	D	1
9	B	B	1
10	B	B	1
11	D	D	1
12	C	C	1
	12		11

Figure 3: Response of the 1st person with score

So, the 1st person did 11 questions correctly. He did the 2nd question incorrectly. The second question was

So, our guess is, it is the first confusing question of the type where All implies

i) All teachers are men.	**a) Some teachers are men.**
ii) Some teachers are women.	**b) All teachers are women.**

Figure 4: Question 2

Some or Not. Seeing the question, he was first surprised by what to do. Then, he marked that option which says All doesn't imply some. But when the second

question of this type occurred, he thought differently and did all the answers correctly. Hence his score was 11.
Here is the response of the 2nd person in Figure 4. He also did the 2nd ques-

Question Number	Actual	Response	Score
1	A	A	1
2	A	C	0
3	D	B	0
4	C	C	1
5	D	D	1
6	C	C	1
7	B	C	0
8	D	D	1
9	B	B	1
10	B	B	1
11	D	B	0
12	C	D	0
	12		7

Figure 5: Response of the 2nd person with score

tion wrong. Also, he/she made mistakes in questions 3,7,11, and 12. The 3rd question was:

STATEMENTS:

i) All books are copy.

ii) No copy is pencil.

CONCLUSIONS:

a) Some books are copy.

b) No book is pencil.

Figure 6: Question 3

So, this person assumed that Some mean some and All means All, and All doesn't imply some. And he/she was stuck to this assumption and only did mistakes in all those types of questions.

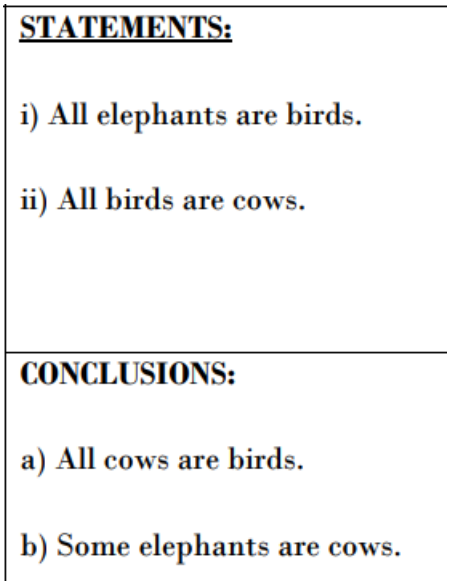

Figure 7: Question 7

This is the 7th question. This also has the same confusion. So, the psychology of the 2nd person is very clear from these.
From these two responses, it is somewhat clear that many persons gave the wrong answers carefully because of their assumptions.

The rest responses are summarized via descriptive statistics. The following Figure 8 is the score plot for different persons: From Figure 8, it is clear, that

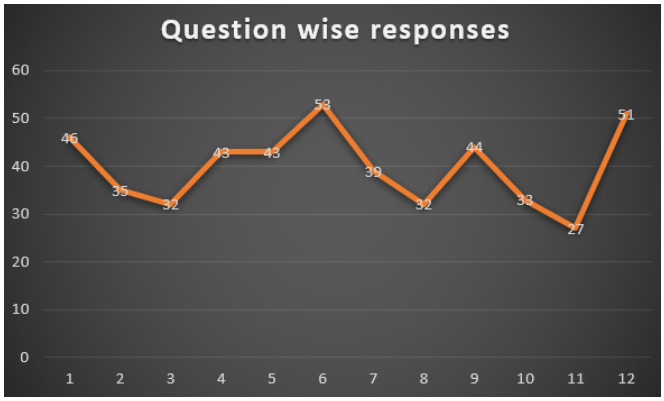

Figure 8: Plot of the scores

the maximum of the scores is attained at 12 and the minimum is at 11. From the histogram, we can see that mode is at 11, 11 persons scored 11. The minimum is 5,6 and only 1 person scored 5,6 each. and the maximum is scored by 5 persons, which is 12.

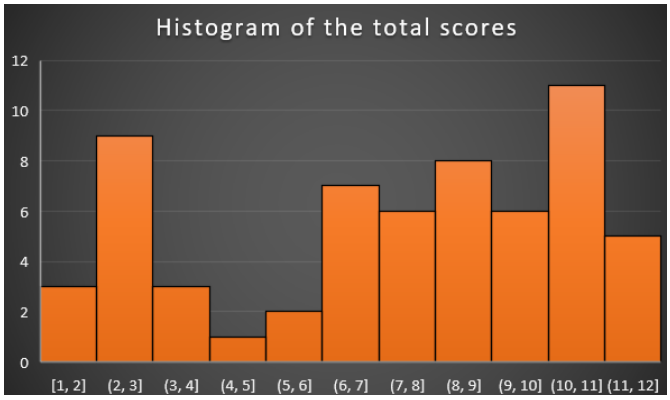

Figure 9: Histogram of the Scores

The number of correct responses for each question is attached below. In the Figure, the number written in the i th column indicates the number of correct responses for the i th question.

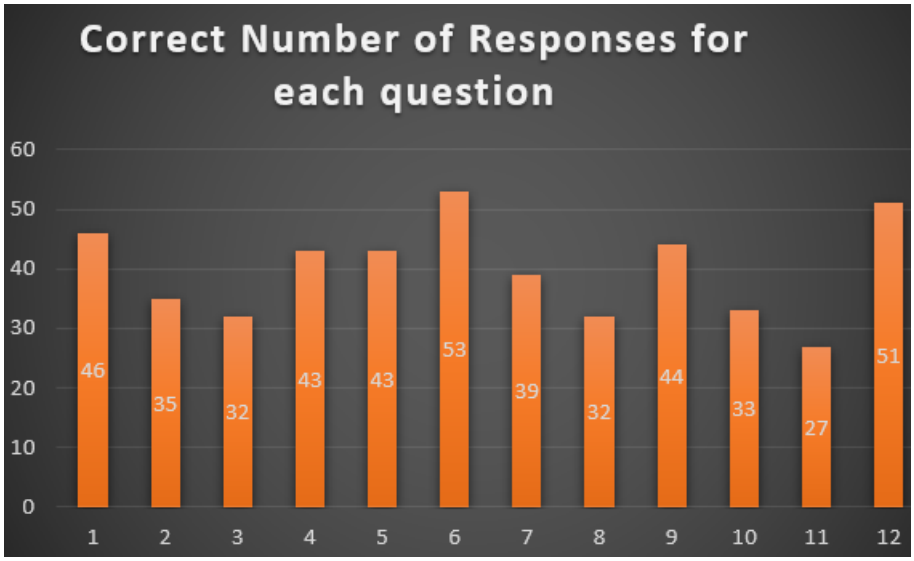

Figure 10: Number of Correct Responses for Each Question

5.2 Inferential Statistics

First, we have the following data(Only a part is shown, and the total is only of that part):

Table 1: Initial Data

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Total
1	1	0	1	1	1	1	1	1	1	1	1	1	11
2	1	0	0	1	1	1	0	1	1	1	0	0	7
3	1	1	1	1	1	1	1	1	1	0	0	0	9
4	1	1	1	1	1	1	1	1	1	1	1	1	12
5	1	1	1	1	1	1	1	0	1	1	1	1	11
6	1	1	0	0	0	0	0	0	1	0	0	0	3
7	0	1	0	0	0	1	0	0	0	0	0	1	3
8	1	0	0	1	1	1	0	1	1	1	0	1	8
9	1	0	0	1	1	1	1	1	1	0	0	0	7
10	1	1	1	1	1	1	1	0	0	1	1	1	10
11	0	1	0	0	0	0	0	0	0	1	0	1	3
12	0	1	1	1	1	1	1	0	1	0	1	1	9
13	1	1	1	0	0	0	0	0	0	0	0	0	3
14	1	0	0	1	1	1	1	1	0	0	0	1	7
15	0	1	0	1	1	1	1	1	1	1	0	1	9
16	0	0	0	1	1	1	0	1	1	1	0	1	7
17	1	1	1	1	1	1	1	1	1	0	1	1	11
18	1	1	1	1	1	1	1	1	1	1	1	1	12
19	0	0	0	1	1	1	1	1	1	0	0	1	7
20	1	1	0	1	1	1	0	1	1	1	0	1	9
21	1	1	1	0	0	1	1	0	0	1	1	1	8
22	1	1	1	0	1	1	1	1	1	0	1	1	10
23	1	0	0	1	1	1	0	1	1	0	0	1	7
Total	17	15	11	17	18	20	14	15	17	12	9	18	

Here we will follow *Kelly's Method of Dichotomization* on the data to analyze. So, for that first, let us sort the data w.r.t. the last column. Here is the sorted data(Only a part is shown):

Table 2: Sorted Data

1	1	1	1	1	1	1	1	1	1	1	1	1	12
2	1	1	1	1	1	1	1	1	1	1	1	1	12
3	1	1	1	1	1	1	1	1	1	1	1	1	12
4	1	1	1	1	1	1	1	1	1	1	1	1	12
5	1	1	1	1	1	1	1	1	1	1	1	1	12
6	1	0	1	1	1	1	1	1	1	1	1	1	11
7	1	1	1	1	1	1	1	0	1	1	1	1	11
8	1	1	1	1	1	1	1	1	1	0	1	1	11
9	1	1	1	1	1	1	1	0	1	1	1	1	11
10	1	0	1	1	1	1	1	1	1	1	1	1	11
11	1	1	1	1	1	1	1	0	1	1	1	1	11
12	1	1	1	1	1	1	1	0	1	1	1	1	11
13	1	1	1	1	1	1	1	0	1	1	1	1	11
14	1	1	1	1	1	1	1	0	1	1	1	1	11
15	1	1	1	1	1	1	1	1	1	0	1	1	11
16	1	1	1	1	1	1	1	0	1	1	1	1	11
17	1	1	1	1	1	1	1	0	0	1	1	1	10
18	1	1	1	0	1	1	1	1	1	0	1	1	10
19	1	0	1	1	1	1	1	1	1	1	0	1	10
20	1	1	1	1	1	1	0	1	1	1	0	1	10
21	1	1	1	1	1	1	1	1	1	1	0	0	10
22	1	1	1	1	1	1	1	0	1	1	0	1	10
23	1	1	1	1	1	1	1	1	1	0	0	0	9
24	0	1	1	1	1	1	1	0	1	0	1	1	9
25	0	1	0	1	1	1	1	1	1	1	0	1	9
26	1	1	0	1	1	1	0	1	1	1	0	1	9
27	1	0	1	1	0	1	1	0	1	1	1	1	9
28	1	0	0	1	1	1	1	0	1	1	1	1	9
29	1	1	0	1	1	1	1	1	1	0	0	1	9
30	1	0	1	1	0	1	1	1	1	0	1	1	9
31	1	0	0	1	1	1	0	1	1	1	0	1	8
32	1	1	1	0	0	1	1	0	0	1	1	1	8

Table 3: Higher and Lower score groups extracted

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Total
1	1	1	1	1	1	1	1	1	1	1	1	12
1	1	1	1	1	1	1	1	1	1	1	1	12
1	1	1	1	1	1	1	1	1	1	1	1	12
1	1	1	1	1	1	1	1	1	1	1	1	12
1	1	1	1	1	1	1	1	1	1	1	1	12
1	0	1	1	1	1	1	1	1	1	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	1	1	1	1	1	1	1	1	0	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	0	1	1	1	1	1	1	1	1	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	1	1	1	1	1	1	1	1	0	1	1	11
1	1	1	1	1	1	1	0	1	1	1	1	11
1	1	1	1	1	1	1	0	0	1	1	1	10
1	1	1	0	0	1	0	0	0	0	0	1	5
0	0	1	0	1	0	0	1	0	0	1	0	4
1	0	0	0	0	0	1	0	1	1	0	0	4
1	0	0	0	0	1	0	0	1	0	0	1	4
1	1	0	0	0	0	0	0	1	0	0	0	3
0	1	0	0	0	1	0	0	0	0	0	1	3
0	1	0	0	0	0	0	0	0	1	0	1	3
1	1	1	0	0	0	0	0	0	0	0	0	3
0	0	0	0	0	0	0	1	0	0	1	1	3
0	0	0	0	0	0	0	1	0	0	1	1	3
0	1	0	0	0	1	0	0	0	0	0	1	3
0	0	0	0	0	1	0	0	0	1	0	1	3
0	1	0	0	1	0	0	0	1	0	0	0	3
1	0	0	1	0	1	0	0	0	0	0	0	3
0	0	0	0	0	1	0	0	0	0	0	1	2
0	0	0	0	0	0	1	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	0	1

Now, here we have 63 responses, so according to **Kelly's Method**, we will take the first 27% of the data as the higher score group and the last 27% of the data as the lower score group. Now, $\frac{63.27}{100} = 17.01$ is close to the integer 17. So we will take the first 17 data points as the higher score group and the last 17 ones in the lower score group. In the following figure, red marked are higher score group and blue marked are lower score group.
Here is the extracted version:

Then we follow the process to find **R.H**(*Rightly Answered in the Highest Group*), **R.L**(*Rightly Answered in the Lowest Group*), **N.H**(*No of Examinees in the Highest Group*), and **N.L**(*No of Examinees in the Lowest Group*). Then we

Table 4: Finding R.H, R.L, N.H, or N.L

Question.Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	T
RH	17	15	17	17	17	17	17	9	16	15	17	17	
RL	6	7	3	1	2	7	3	3	4	3	3	9	
RH+RL	23	22	20	18	19	24	20	12	20	18	20	26	
RH-RL	11	8	14	16	15	10	14	6	12	12	14	8	
NH/NL	17	17	17	17	17	17	17	17	17	17	17	17	

calculate the Difficulty Value(D.V) by $\frac{R.H+R.L}{N.H+N.L}$ and Discrimination Index(D.I) by $\frac{R.H-R.L}{\text{N.H or N.L}}$ for each of the questions. From Table 7, it is observed that for

Table 5: Calculation of D.V. and D.I.

Question.Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
DV = (R.H+R.L)/(N.H+N.L)	23/34	11/17	10/17	9/17	19/34	12/17	####	6/17	10/17	9/17	10/17	13/17
DV	0.68	0.65	0.59	0.53	0.56	0.71	0.59	0.35	0.59	0.53	0.59	0.76
DI = (R.H-R.L)/(N.H or N.L)	0.65	0.47	0.82	0.94	0.88	0.59	0.82	0.35	0.71	0.71	0.82	0.47
Item-Total Correlation	0.63	0.37	0.64	0.78	0.74	0.64	0.62	0.37	0.68	0.54	0.55	0.45

Table 6: Selection and rejection of items on the basis of Ebel's criterion

Question.Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
DV	0.68(Easy)	0.65(Easy)	0.59(Moderate Difficult)	0.53(Moderate Difficult)	0.56(Moderate Difficult)	0.71(Most Easy)	0.59(Moderate Difficult)	0.35(Difficult)	0.59(Moderate Difficult)	0.53(Moderate Difficult)	0.59(Moderate Difficult)	0.76(Most Easy)
DI	0.65(Very Good Items)	0.47(Very Good Items)	0.82(Very Good Items)	0.94(Very Good Items)	0.88(Very Good Items)	0.59(Very Good Items)	0.82(Very Good Items)	0.35(Reasonably Good)	0.71(Very Good Items)	0.71(Very Good Items)	0.82(Very Good Items)	0.47(Very Good Items)
Item-Total Correlation	0.63(Good Items)	0.37(Easy Items but not poor)	0.64(Good Items)	0.78(Very Good Items)	0.74(Very Good Items)	0.64(Good Items)	0.62(Good Items)	0.37(Easy Items but not poor)	0.68(Very Good Items)	0.54(Good Items)	0.55(Good Items)	0.45(Easy Items but not poor)

Difficulty Value Analysis:

1. 6th and 12th questions are the most easy.

2. 1st, 2nd questions are easy.
3. 3rd, 4th, 5th, 7th, 9th, 10th, 11th questions are moderately difficult.
4. 8th question is difficult

Applying the cutoff criterion for Discrimination Index, we get:

1. 8th question is Reasonably Good but needs to be improved.
2. Rest questions are very good items, not as good as 8th.

Applying the Item-Total Correlation, we get:

1. 4th and 5th questions are very good items.
2. 2nd, 8th, and 12th questions are easy but not poor.
3. Rest are reasonably good.

6 Discussion

6.1 Explaining Findings

6.1.1 Difficulty Value:

- 1st, 2nd, 6th and 12th questions are the most easy or easy. For the 1st question, Statements were i)All doors are windows. ii)All windows are beds. And, clearly, by simple set-theoretic argument, all doors are beds. Hence, clearly, 1 is very easy. For the 12th question, Statements are i)Few chairs have tables. ii)Few tables have drawers. That doesn't imply either i)All chairs have drawers or ii)All tables have drawers. Hence that is also easy.
- 3rd, 4th, 5th, 7th, 9th, 10th, 11th questions are classified as moderately difficult. There was some confusion like if all A's are B, then 'Some A's are

B' are true or false? That confusion also led to different answers, hence some questions were classified as difficult.

- 8th question is classified as the most difficult. It is for the issue that All implies some or not. Here both the conclusions included some and hence the participants became confused about that. The 8th question was:
  So, here the conclusions are quite confusing. By basic set-theoretic argu-

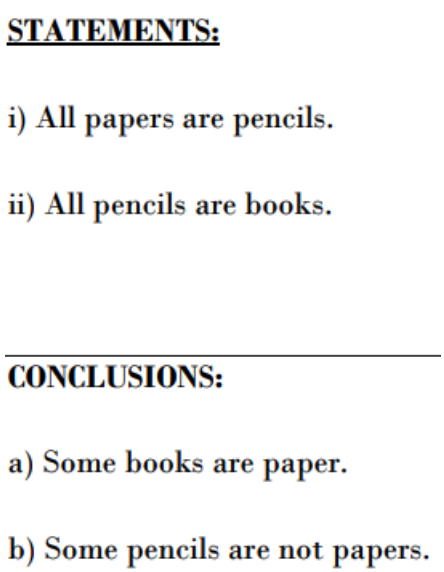

Figure 11: Question 8

ment or Venn diagram, it is easy to see that All papers are books. That implies that Some books are paper. Also, all papers are pencils. That implies some pencils are paper, also some pencils are not paper. So, here the first step conclusion is easy, but the second is not that obvious. So, many people got that wrong, and hence it is marked as the most difficult.

6.1.2 Discrimination Index :

- 8th question is a reasonably good item. In this case, people get wrong answers from the confusion between Some and All.

- Rest are very good items. For the 9th question, Statements are i)Some cows are horses. ii)All horses are dogs. And we have to conclude i)All dogs are horses. ii)Some dogs are horses. So, it is not all so obvious, rather one has to do an analysis based on the options to get the correct answers. Hence, it is classified as a very good item.

6.1.3 Item-Total Correlation:

- 2nd question is the poorest item. There is confusion between some and all here. This matches with the discriminative index also. We saw that the

i) All teachers are men.	**a) Some teachers are men.**
ii) Some teachers are women.	**b) All teachers are women.**

Figure 12: Question 2

  first person did the 2nd question wrong only. Like him/her, many did the wrong answers due to similar reasons.

- 4 was the best item and 5 is the second best according to item-total correlation.

- DV, DI, and Item-Total Correlation All three procedures got to the same conclusion that items 4,5 are two of the toughest and most good items.

6.2 Relationship with earlier research:

The results align with earlier research, which indicates that questions with greater discriminatory ability are preferred for accurate assessment (Lord, 1952). Among the questions analyzed, 04, 05, 03, and 11 demonstrate strong discriminatory power while exhibiting different levels of difficulty. This supports the idea that items of moderate difficulty and high discrimination can effectively

distinguish between individuals with varying levels of proficiency (Lord et al., 1968).

The modification of the difficulty level of Question 8 aligns with existing literature, which highlights the importance of fine-tuning questions to strike a balance between challenging test-takers and enabling discrimination. When a question's difficulty is excessively high, like in the case of Question 8 being too tough even for highly capable students, it can impede the item's effectiveness in distinguishing between individuals (S. Embretson, 1996).

Research emphasizing the significance of items possessing an adequate ability to differentiate among individuals (Hambleton and Jones, 1993) supports the suggestion to modify or eliminate question 12, given its limited discriminatory power. Questions with minimal discrimination, such as question 12, might not offer valuable insights into a test-taker's abilities and therefore should be considered for revision or removal.

Recognizing the possibility for improvement in question 12 and 06 demonstrates an awareness of the ongoing process of enhancing item development and selection. This corresponds with existing literature that highlights the iterative nature of item analysis and the opportunity to refine items based on their psychometric properties, as mentioned by S. E. Embretson and Reise in 2000.

In general, the findings derived from the analysis of the test items correspond with the established body of knowledge concerning item analysis, item difficulty, and discrimination. These findings underscore the significance of taking into account these psychometric properties to enhance the accuracy and efficiency of the test, and they align with the suggestions put forth in earlier studies. Furthermore, they provide evidence for the practicality and legitimacy of the psychometric techniques and concepts described by Kelley (1939), Blood and Budd (1972), and Ebel (1972).

6.3 Suggestions

- According to all of the participants, some options are very much confusing. For example, if 'All books are pencils' is true, then is 'Some books are pencils' true or not? That leads to confusion.
- Some questions may be made hard to make the test much more interesting and effective.
- Without making them know the answers, the test can be done multiple times on a single person to see the variation between those scores.
- The findings of this study can mostly be applied to male students, thereby restricting their generalizability to the overall population. Subsequent research should strive for a more inclusive sample, encompassing both male and female students, in order to ensure the findings are relevant to a wider range of participants.
- The participants' age range is relatively limited, spanning from 18 to 25 years old. To gain a more comprehensive understanding of syllogistic reasoning aptitude at various developmental stages, it would be beneficial to encompass a broader age range and involve participants from diverse educational levels.
- This study had a very small sample size of 63 participants. Increasing the sample size would enhance the statistical power and reliability of the results, otherwise, we can't apply any large sample results to the data.

6.4 Future Research

- It would be intriguing to examine potential dissimilarities in syllogistic reasoning aptitude by delving into gender differences, considering that the current study exclusively involved male students. Conducting research to

explore possible variations in performance between males and females could offer valuable insights into the cognitive mechanisms underlying syllogistic reasoning.

- People have different moods and attentions at different points of time. And, not only that, intelligence or problem-solving skill increases under continuous practice. So, it may be interesting to see how problem-solving affects syllogistic reasoning, i.e. performing a designed experiment on the subjects. Also, it is interesting to collect data from one single person from different time points without letting them know the correct answers.
- Exploring the correlation between the aptitude for syllogistic reasoning and academic performance across different subjects, such as mathematics or science, could provide additional insights into the practical implications of this cognitive ability. Gaining a deeper understanding of how syllogistic reasoning is connected to academic achievement can offer valuable guidance for educational interventions and the development of curriculums.
- This paper presents a study that aimed to evaluate the aptitude of bachelor or master or class 11-12 students in syllogistic reasoning. The study specifically focused on male students between the ages of 18 and 25, and its objective was to analyze the psychometric properties of syllogistic reasoning items through item analysis. By assessing the difficulty levels, discrimination indices, and item-total correlations, the research sought to identify areas for improving the measurement of syllogistic reasoning aptitude.
- The study involved a sample of 63 participants who took a syllogistic reasoning aptitude test. However, it should be noted that the sample consisted solely of male students, which could introduce sample bias. The participants ranged in age from 18 to 25. Data was collected from bachelor or masters or class 11-12 students using a questionnaire comprising

12 questions. The responses provided by the participants were analyzed to evaluate the psychometric properties of the items and draw conclusions about their difficulty and discriminatory power.

- The findings of the study revealed a variety of item difficulty levels, discrimination indices, and item-total correlations. Applying Ebel's Criterion, we derived several conclusions regarding the difficulty and discriminatory power of individual items. For example, items 01 and 06 were found to be relatively easy yet highly discriminatory, while items 11 and 05 proved to be challenging and demonstrated limited discriminatory power.

7 References

1. Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. Cognitive, Affective, & Behavioral Neuroscience, 7(3), 153-191.
2. Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 24(1), 167-202.
3. Pessoa, L. (2008). On the relationship between attention and the spatial organization of the parietal cortex. Frontiers in Neuroscience, 2(2), 1-9.
4. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8(5), 393-402.
5. Stuss, D. T., & Knight, R. T. (2013). Principles of Frontal Lobe Function (2nd ed.). Oxford University Press.
6. Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. Nature, 384(6605), 159-161.

7. Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? Psychological Bulletin, 131(1), 30-60.

8. Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science.

9. Lord, F. (1952). A theory of test scores. Psychometric Monographs, 7, x, 84–x, 84

10. Lord, F., Novick, M., & Birnbaum, A. (1968). Statistical theories of mental test scores. Addison-Wesley.

11. Embretson, S. (1996). The new rules of measurement. Psychological Assessment, 8, 341–349.

12. Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. Educational Measurement: Issues and Practice, 12, 38–47.

13. Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum Associates Publishers.

14. Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. Journal of Educational Psychology, 30, 17–24.

15. Blood, D. F., & Budd, W. C. (1972). Educational measurement and evaluation. Harper and Row.

16. Ebel, R. L. (1972). Essentials of Educational Measurement (tech. rep.). Prentice-Hall, Inc ERIC Number: ED073168.

17. Lord, F.M. (1952). A Theory of Test Scores. Psychometric Monographs,

7. This seminal work by Lord proposes a theoretical framework for understanding item difficulty and its relation to test scores.

18. Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord & M.R. Novick (Eds.), Statistical Theories of Mental Test Scores. Addison-Wesley. Birnbaum's work introduces the Rasch model, a widely used model for item difficulty estimation in educational measurement.

19. Embretson, S.E. (1996). The New Rules of Measurement. Psychological Assessment, 8(4), 341-349. This article discusses the importance of considering the multidimensionality of item difficulty and provides recommendations for improving measurement practices.

20. van der Linden, W.J. (1998). Sijtsma, K. (Eds.), Introduction to Item Response Theory and Differential Item Functioning Analysis. In J. de Jong, C.H. Vos, & J.C. Waller (Eds.), Introduction to Psychometric Theory. This chapter provides a comprehensive overview of item response theory (IRT) and its application to estimating item difficulty.

21. Embretson, S.E., & Reise, S.P. (2000). Item Response Theory for Psychologists. Lawrence Erlbaum Associates. This book offers a detailed introduction to IRT and its applications, including the estimation of item difficulty.

22. Hambleton, R.K., & Jones, R.W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. Educational Measurement: Issues and Practice, 12(3), 38-47. This article compares classical test theory with IRT, highlighting the advantages of IRT models for estimating item difficulty.

23. DeMars, C.E. (2010). Item Response Theory. Oxford University Press.

This comprehensive book provides an in-depth understanding of IRT models and their applications, including item difficulty estimation.

24. Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. Psychometrika Monograph Supplement, 17. Samejima's work introduces the graded response model, an IRT model specifically designed for polytomous item response data

8 Appendix

8.1 R Script

The whole analysis was done in Microsoft Excel, and hence R was not used for this analysis.

8.2 Tools

The Questionnare for Syllogistic Reasoning Section is attached below

TEST NUMBER: 3

Directions for questions 25 to 36:

Given below are two statements followed by two conclusions numbered (a) and (b). You have to take two given statements to be true even if they seem to be at variance from commonly known facts and then decide which of the given conclusion logically follows from the two given statements disregarding commonly known facts.

Example:

STATEMENTS:	(A) If only conclusion (a) follows.	**ANSWER**
i) All pencils are bricks.		**(A)**
ii) All bricks are bottles.	(B) IF only conclusion (b) follows.	
CONCLUSIONS:	(D) IF neither (a) nor (b) follows.	
a) All pencils are bottles.		
b) All bricks are pencils.	(E) IF both (a) and (b) follows.	

Venn diagram:

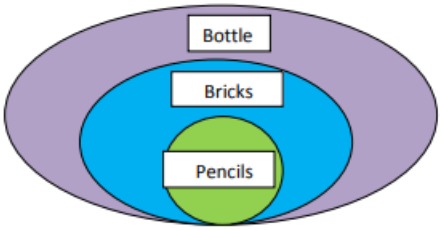

Figure 13: Page1

Explanation:

The 'pencil' circle is in the 'brick' circle and both the 'pencil' and 'brick' circle are in the 'bottle' circle. Thus we can say that all pencils are bottles since the whole 'pencil' circle is within the 'bottle' circle. But we cannot say that all bricks are pencils as because the 'pencil' circle is covering only a small area of 'brick' circle, but a greater portion of the 'brick' circle is outside the 'pencil' circle which means that all bricks are not pencils. Hence we accept the conclusion (a) that is 'All pencils are bottles' and therefore correct answer is (A).

Q. No.	***Items***	***Answers***	***Correct Answer***
25.	**STATEMENTS:** i) All doors are windows. ii) All windows are beds. **CONCLUSIONS** a) All doors are beds. b) All beds are doors.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
26.	**STATEMENTS:** i) All teachers are men. ii) Some teachers are women.	(A) If only conclusion (a) follows.	( )

Figure 14: Page2

	CONCLUSIONS: a) Some teachers are men. b) All teachers are women.	(B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	
27.	**STATEMENTS:** i) All books are copy. ii) No copy is pencil. **CONCLUSIONS:** a) Some books are copy. b) No book is pencil.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
28.	**STATEMENTS:** i) All dogs are cats. ii) All cats are horses. **CONCLUSIONS:**	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows.	( )

Figure 15: Page3

	a) All horses are dogs. b) No horse is cat.	(C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	
29.	**STATEMENTS:** i) All cats are tiger. ii) All tigers are lions. **CONCLUSIONS:** a) All cats are lions. b) Some lions are cats.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
30.	**STATEMENTS:** i) All biscuits are chocolates. ii) Some chocolates are breads. **CONCLUSIONS:** a) No biscuit is a chocolate. b) All chocolates are breads.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )

Figure 16: Page4

31.	**STATEMENTS:** i) All elephants are birds. ii) All birds are cows. **CONCLUSIONS:** a) All cows are birds. b) Some elephants are cows.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
32.	**STATEMENTS:** i) All papers are pencils. ii) All pencils are books. **CONCLUSIONS:** a) Some books are paper. b) Some pencils are not papers.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
33.	**STATEMENTS:** i) Some cows are horses.	(A) If only conclusion (a) follows.	( )

Figure 17: Page5

	ii) All horses are dogs. **CONCLUSIONS:** a) All dogs are horses. b) Some dogs are horses.	(B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	
34.	**STATEMENTS:** i) Some trees are flowers. ii) Some ships are trees. **CONCLUSIONS:** a) Some flowers are ships. b) Some ships are trees.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	( )
35.	**STATEMENTS:** i) All telephones have buttons. ii) All buttons have numbers.	(A) If only conclusion (a) follows. (B) IF only conclusion (b)	( )

Figure 18: Page6

	CONCLUSIONS: a) Some telephones have numbers. b) All telephones have numbers.	follows. (C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	
36.	**STATEMENTS:** i) Few chairs have tables. ii) Few tables have drawers.	(A) If only conclusion (a) follows. (B) IF only conclusion (b) follows.	( )
	CONCLUSIONS: a) All chairs have drawers. b) All tables have drawers.	(C) IF neither (a) nor (b) follows. (D) IF both (a) and (b) follows.	

END

Figure 19: Page7