1: OVERVIEW

1.1. Item analysis

In the discussion below, a stimulus is a multiple-choice test question—also called item— and the response is the choice selected by a respondent (i.e., a person taking the test). Item analysis is the study of response-based validation of the stimuli with respect to the objectives of the test. The process involves statistical analysis of the responses of a sufficiently large group.

Large scale recruitment, diverse job descriptions and a variety of job specifications have given rise to the need for a stringent selection process. Multiple-choice tests have been found to be time-efficient scrutiny techniques for these purposes. Item analysis has a critical role in maintaining and enhancing the quality of such tests.

Item analysis helps in identifying behavioral traits of the recipients of the stimuli and enables better selection that best meets the intended objective of the test. Given a set of responses, item analysis works as a statistical data mining tool to generate feedback on the quality of the stimuli, which in turn can help to enhance the test. The underlying assumption is that there is a specific relationship between the stimulus and the response. The relationship plays the role of criterion for selection. If the relationship is observed then the respondent is supposed to be a suitable candidate to be selected. When analyzed, responses and the stimuli together can generate useful feedback to the respondents.

If everyone's response to a stimulus is the same then the item is not a good discriminator, and if most of the responses are not in line with the expected one then it could be a case of an ambiguous item. These are the examples of heuristics employed in generating item level feedback. There are well-defined metrics to compute the difficulty or discrimination ability of an item, overall reliability of the test, etc.

The potential of item analysis has been explored in diverse areas. Remarkable success of this technique has been reported in projects like cancer patients'

responses to the Multidimensional Affect and Pain Survey (MAPS) [Scrolla] and visualizing country profiles of scientific competence in [Svein08].

The research reported here is essentially an attempt to define an integrated solution, based on item analysis, to the issue of quality of tests generally employed in any selection process. The system concept, user-friendly interface, flexible control for customization, adaptive learning, and extended scope of test definition (from multiple-choice to descriptive answers) are novel features of this research. A significant amount of work has been reported in the contemporary literature on item analysis based solutions. To the best of our understanding there is no mention in commercial or academic research of a system with these features [Tools]. A comprehensive framework for a solution based on item analysis is a central contribution of this work.

1.2 Examination audit system (EAS)

The development of a decision support system (DSS) called Examination Audit System (EAS) is reported in this thesis. The primary purpose of EAS is to evaluate the worth of a test or examination conducted in the context of some selection process. The question paper, the key to the paper and the students' responses along with their profiles are the inputs. The system serves various actors involved in the process of the examination, (namely, the controller of examinations, paper setters, teachers, data entry operators and the students) by reporting infirmities like (i) miskey—the standard key used in the evaluation is incorrect, (ii) multikey—more than one alternatives are found correct and key states only one, (iii) ambiguous key—the given key is not the unique answer to the question and / or the given key lacks strong support in the literature, (iv) misprint—typographical errors, incomplete question or answers. The system analyses data to uncover the worth of an item for its discrimination and proper selection ability. The system also reports negative discrimination and wrong selection cases if any. The suitability of the time constraint is validated by the system. An analysis of the responses of a specific group, e.g., gender-specific responses or responses of ethnic groups, is possible.

1.3 Significance of the work

The proposed system is a part of post-examination processing and could be categorically labeled as an examination audit system. While processing the result, the system may not necessarily follow the scoring scheme suggested by the examiner. Rather, a 'suitable' scheme would be devised by the system. As all the students' responses are considered in the input and nothing other than this is a significant input to the system, the system ensures objective evaluation. Speed and transparency are the obvious consequences of computerization. However user-friendliness is a value-adding unique feature of the system.

The work could be treated as a contribution to the world of e-education. The system can be tailored to either a traditional or a modern on-line examination system with the primary goal of fair evaluation of the conduct of an examination. It raises alerts for the examination authorities in case of potential discrepancies. Being an integrated solution it maintains the students' record along with its responses that can be used by both teachers and students as feedback and also for trend analysis. The infirmities reported by the system can provide the clues to the paper setters for further enhancements. A by-product of this system is an 'item-bank' that can be used to educate the users for the proper effective reuse of the examination data. The framework suggested here has the flexibility to extend the scope by employing adaptive learning and hence to converge to a system that generates tests for selection, given the job specification and job descriptions data.

1.4 Thesis organization

The thesis is organized as follows:

Chapter 2 provides the foundation knowledge for the novice reader of item analysis. The traditional and contemporary trends in item analysis are discussed along with the basic nomenclature in the domain. The material presented includes the three main streams of item analysis (namely, Item Response Theory, the Rasch method and the classical method of item analysis); the parameters of item analysis (item difficulty, discrepancy, distracter, reliability and standard error of measurement) and advanced trends (multilevel item analysis, test bias and item bias, visual item analysis and tools available in the market).

Chapter 3 describes the Examination Audit System (EAS) model that uses item analysis as its basis. The first part of the Chapter is intended for developing an understanding about the standard DSS, decision making model and DSS architecture. The design and features of EAS are described in the later part of this chapter.

Chapter 4 is on the methodology followed in this work. The hypothesis is corroborated with a series of experiments. Interesting outcomes of our experiment include: (i). Miskeys were identified in 7 out of 10 test papers in an examination. In another case a miskeys were found for 20 questions in sequence. The error was due to the verified fact that the data entry operator missed the correct sequence of the key somewhere in mid-entry. (ii). Attractive distracters and cases of false or low discriminators were identified and reported almost in all the test papers. (iii) Annual test papers in a certain examinations were processed, complexity characteristic of a test paper over the years was studied and fluctuations were brought to the notice of the concerned people so that corrective measures were formulated for the next year's examination. (iv) The weak and strong areas of the clusters of the candidates who participated in selection process were identified. The report was found to be useful for designing further training programmes for the candidates. (v) A report was generated on the impact of negative marking schemes and similar constraints on students' behavior, and their contribution towards proper selection.

The interpretation and discussion of the important results of our research are compiled in **Chapter 5** and some directions for future work are provided. Our hypothesis, that EAS could play a promising role in the enhancement of the conduct of examinations, is validated. We hope that the self-learning model of EAS discussed in this work will mature into a system that generates a good quality test paper dynamically.