# Study of Public Examination System and Proposed E-examination to Control Malpractices and Evaluation Anomalies

A Thesis submitted to

**Goa University**

for the award of the Degree of

**Doctor of Philosophy**

in

**Computer Science**

By

**Gauns Dessai Kissan Ganesh**

Under the guidance of

**Prof. Venkatesh V. Kamat**
Department of Computer Science & Technology
Goa University, Goa, India

**2018**

# Certificate

This is to certify that the thesis entitled **"Study of Public Examination System and Proposed E-examination to Control Malpractices and Evaluation Anomalies"** submitted by **Gauns Dessai Kissan Ganesh** to **Goa University, Taleigao Pleateu, Goa** for the award of the degree of **Doctor of Philosophy in Computer Science** is a bonafide record of research work carried out by him under my supervision and guidance.

The contents embodied in the thesis have not been submitted in part or full to any other Institution or University for the award of any Degree/Diploma or any other similar title.

**Prof. Venkatesh V. Kamat**

Research Guide

Department of Computer Science & Technology

Goa University, Goa

*Date :*

*Place: Goa University*

# Declaration

I, **Gauns Dessai Kissan Ganesh** hereby declare that the thesis entitled **"Study of Public Examination System and Proposed E-examination to Control Malpractices and Evaluation Anomalies"**, submitted to Goa University for the award of the degree of **Doctor of Philosophy in Computer Science**, is a record of original research work done by me under the supervision and guidance of **Prof. Venkatesh V. Kamat**, Department of Computer Science & Technology, Goa University and that it has not formed the basis for the award of any Degree/Diploma or any other similar title.

The literature related to the problem investigated has been cited.

**Gauns Dessai Kissan Ganesh**

Research Scholar

*Date :*

*Place: Goa University*

# Acknowledgements

First and foremost, I thank the Almighty God for blessing me with the ability and strength to complete this research.

I owe my deep gratitude to my esteemed Guide, **Prof. Venkatesh V. Kamat**, Professor, Department of Computer Science and Technology, Goa University. His unflinching motivation, incisive comments, visionary ideas and invaluable guidance have helped and inspired me profoundly in this entire endeavour. I offer my sincere thanks to my Guide for constantly triggering and nourishing my intellectual ability and engaging me in intelligent and novel ideas.

I express my profound thanks to the members of **Faculty Research Committee (FRC)** for their valuable feedback toward improving my work.

I humbly grab this opportunity to acknowledge reverentially, the Director of Higher Education, **Shri Prasad Lolayenkar** for his varied contributions in assorted ways. He was instrumental in giving a tremendous boost to me and acted as a catalyst for timely completion of this research work.

I acknowledge wholeheartedly the invaluable support rendered by **Dr. Joydeep Bhattacharjee**, Principal, Government College of Arts, Science and Commerce, Quepem-Goa in facilitating my research work. His continual assistance and encouragement helped me immensely during this memorable journey of research.

I am most grateful to **Prof. Ramrao S. Wagh**, Associate Professor, Department of Computer Science and Technology, Goa University for providing all the assistance, motivation and unstinted support to me.

I am thankful to all **Professors from the Department of Computer Science & Technology, Goa University**, who were the driving force right from my days as a MCA student to-date.

I would like to thank my ex Principals, **Dr. Gervasio S.F.L. Mendes** and **Dr. Lucy James** at Government College of Arts, Science and Commerce, Sanquelim, Goa for their constant support, words of wisdom and encouragement.

I am grateful to the **Management, Principal and all my ex colleagues from Shree Damodar College of Commerce and Economics, Margao-Goa** for their impetus and support throughout my tenure in Shree Damodar College of Commerce and Economics, Margao-Goa.

Special thanks goes to all my **colleagues from Department of Computer Science at Government College, Sanquelim and Government College, Quepem** for all the encouragement and providing a friendly and conducive research atmosphere at work. I would also like to thank all my **colleagues from Government Colleges of Quepem and Sanquelim** for their constant support and motivation.

I appreciate all the **administrative staff of Goa University** together with the **support staff in the Department of Computer Science and Technology, Goa University**. I am also thankful for the assistance offered by all the **support staff of Government College Quepem and Sanquelim**.

My heartfelt thanks to **Prof. Florence Rebello**, my ex-colleague from Shree Damodar College of Commerce and Economics, Margao-Goa in completing the proofreading meticulously and tirelessly. I gratefully acknowledge the efforts taken by her for improving overall presentation of this thesis.

I am ever indebted to my **parents** who deserve special mention for their indissoluble support and prayers and for everything they sacrificed in their life for me. I would also like to express my deepest gratitude to my **wife, Seema** and **daughter, Prajakta** for their unfailing emotional support.

I offer my sincere thanks to all of those people who supported me and offered their gracious help directly or indirectly during the completion of this research work.

Thank you all.

## Abstract

Summative examinations are a major and universal means to assess achievement and qualification of the examinees. Due to the pressures associated with such assessment and achievement, summative examinations are often marred by the plethora of academic misconducts, malpractices and evaluation anomalies.

Literature as well as the media are rife with evidence that the intensity and pervasiveness of malpractices in summative examinations have reached alarming levels and have taken sophisticated and techno-centric dimensions. Therefore, it is necessary to take appropriate measures to deter and detect such malpractices and to uphold the academic honesty and the integrity of the examination system. The menace of malpractices such as question paper leakage and collusion/plagiarism can be curbed to a great extent by generating a unique question paper Just-In-Time (JIT) for each examinee/group of examinees.

Apart from malpractices, evaluation anomalies such as errors / lapses / intra/inter examiner heterogeneity also greatly mar the examination system. Evaluation anomalies mainly crop up due to the tendency of the examiners for carelessness or inconsistency or severity or leniency during the evaluation. Institutions conducting examinations seek succour from a number of remedial approaches to control evaluation anomalies. However, in manual

and multi-examiner evaluation of a large number of answer-scripts, achieving consistency and uniformity with precision and perfection is a tall order.

In this research, we examine and explore some of the inadequacies found in the 'Public examination system' under the ambit of summative examinations. We investigate the process of question paper/answer-scripts delivery between the examination authority and the examinees/examiners. We also propose protocols for exchange of question paper and answer-scripts between the examination authority and the examinees in an E-examination. The goal of the proposed protocols is to provide anonymity and an unambiguous binding of exchanged question papers and the answer-scripts between the communicating entities. We use the ProVerif tool/mathematical proofs to gauge the correctness of proposed security properties.

We also investigate the evaluation anomalies such as errors/lapses and intra/inter examiner variation in evaluation and present two approaches for controlling these evaluation anomalies. In the first approach, we present a method of computer-assisted evaluation using rubrics for subjective answer-scripts. This method helps in reducing/eliminating the evaluation errors, wastage of time and examiner variability. In the second approach, we make use of machine learning techniques to build an E-moderation model. This enables classification of each answer evaluation as negligent or normal. Further, the E-moderation model builds up a tuned marks predictor to

minimize the examiner subjectivity and the ‘Hawk-Dove effect’ found in the intra/inter examiner evaluation.

The proposed approaches, can be a great boon, to the education system in general and the examinee community in particular. As they result in a fair and reliable evaluation measured on a uniform and unambiguous scale.

**Keywords:** Summative Examination, Examination, Examination Malpractices, Evaluation Anomalies, Disguised Public Key, Associativity, Anonymity, Applied $\pi$ Calculus, ProVerif, Plagiarism, Collusion, Examiner Subjectivity, E-moderation.

## Publications

1. Dessai Kissan, Kamat Venkatesh (2018). "Computer Assisted Evaluation using Rubrics for Reduction of Errors and Inter and Intra Examiner Heterogeneity". *IGI Global. ISSN No: 2155-5605.* International Journal of Information and Communication Technology Education (IGI Global). Volume 14(4), pp. 49-65, October-December 2018.

2. Dessai Kissan, Kamat Venkatesh (2018). "E-moderation of Answer-scripts Evaluation for Controlling Intra/Inter Examiner Heterogeneity". *9th International Conf. On Tech. for Education, T4E 2018. IEEE.*. Chennai, India, Dec. 11-13, 2018.

3. Dessai Kissan, Kamat Venkatesh (2018). "Security Analysis of Conventional/ Electronic Summative Assessments". *2nd International Conference in Computing and Data Sciences, ICACDS 2018. Springer, CCIS. 905. ISBN No: 978-981-13-1810-8.* Uttarakhand, India, April 20-21, 2018.

4. Dessai Kissan, Kamat Venkatesh (2016). "A Framework for Analysing Associativity and Anonymity in Conventional and Electronic Summative Examinations". *12th International Conference, ICISS 2016. Springer, LNCS.10063. pp.303-323. ISSN No: 0302-9743*. Jaipur, India,

December 16-20, 2016.

5. Dessai Kissan, Kamat Venkatesh (2014). “Effective Use of Rubrics in Computer Assisted Subjective Answer-script Evaluation”. *6th IEEE International Conference on Technology for Education (T4E) 2014. ISSN No: 978-1-4799-6489-5/14*. Kerala, January 12-15, 2014.

6. Dessai Kissan, Kamat Venkatesh (2014). “Fair and Non-repudiable Protocol for Exchange of E-Examination Question Paper and Answer-scripts”. *6th IEEE International Conference on Technology for Education (T4E) 2014. ISSN No: 978-1-4799-6489-5/14*. Kerala, January 12-15, 2014.

# Table of Contents

# List of Figures

# List of Tables

# CHAPTER 1
# Introduction

Summative examinations are conducted worldwide to record and report an estimate of the examinees' achievements [MO99]. Summative examinations are normally high-stake assessments since the outcome of these assessments have an enormous impact on academic as well as career prospects of examinees [Rov00]. Thus, reliability is a central concern to summative assessments.

Public examinations are held by institutions for promotion, placement, certification, and accountability come under the ambit of summative examinations. They are prone to a plethora of misconducts and malpractices, due to the fact that the public at large consider them as keys to success/qualifications [TERS05].

The CBSE Board paper leak in Class X and Class XII examinations held in March, 2018, is a direct testimony to serious issues related to public examinations. Any such eventuality badly affects lakhs of examinees resulting in a huge expense to the exchequer. The intensity and pervasiveness of the problem of malpractices are constantly on the rise due to a strong nexus between examinees and other stakeholders along with the external agents. The endemic and ingenious incidents of malpractices in summative examinations encompass collusion, impersonation, leakage of question papers, plagiarism, altering answer-scripts, misconduct in examination centres, influencing supervisors / examiners, making false entries in the award lists/ assessment registers and issuing fake certificates/degrees, etc. [Eck03, Mah11].

Some of the approaches in controlling malpractices in the conventional setup of examinations are listed in Table 1.1.

Table 1.1: Approaches to control Examination Malpractices

| Sr. | Approach | Security purpose | Shortcomings |
|---|---|---|---|
| 1 | Appointment of multiple paper setters per subject/course paper. | Preservation of secrecy of the question paper and establishing the anonymity of paper setters. | Question paper is susceptible to leakage due to the manual process of final question paper selection, production and transportation. |
| 2 | Submission of a sealed hard copy of the manuscript of question paper. | Protect integrity of question paper. | Absence of the original manuscript for verification during the conduct of the examination prevents verification and fraud detection. |
| 3 | Monitoring and supervising the examination conduct from start to finish. | Control examinee acts of academic dishonesty | Large examination blocks makes human monitoring ineffective and prone to lapses, cheating/copying. |
| 4 | Maintenance of examinee attendance record. | Prevent denial of committed action of stakeholders. | One way attendance record helps in safeguarding only one communicating entity, i.e. examination authority. |
| 5 | Hiding examinee identity from answer-books and assigning a code mapped to examinee identity. | Keep identity of examinee anonymous. | The manual coding process can reveal examinee identity before marks announcement. |

The currently employed steps in controlling malpractices in summative examinations have many shortcomings [DKW14], some are more significant than the others (refer Table 1.1). Malpractices during examinations appear to be on the rise and have taken incredible, sophisticated and techno-centric dimensions. This is due to ineffective security regulations and the non-availability of the means of implementing these regulations universally [Eck03]. The electronic assessment has the potential to curb most of the shortcomings associated with

the conventional assessment. Some of the strategies for controlling malpractices in the electronic counterpart are listed below:

1. Generation of the question paper just before the commencement of the examination, i.e., Just-In-Time (JIT) to prevent question paper leakage [Var14].
2. Use of question bank to generate a unique question paper for each examinee.
3. Encryption of the question paper using symmetric/asymmetric encryption techniques for the preservation of secrecy of the question papers [CRHJDJ06].
4. Message digest/hashing technique to preserve the integrity of the question papers/answer-scripts [SB00].
5. Digital signatures for non-repudiation, i.e. to prevent the denial of action committed by entities engaged in the communication [CRHJDJ06, Wei05].
6. Mixnet servers to establish anonymity of the examinees/examiners [GLR14].

Coupled with malpractices, the evaluation of subjective answer-scripts suffers from a variety of evaluation anomalies such as errors/lapses/heterogeneity. Evaluation of subjective answer-scripts is a highly human intensive task. It needs a focused and an unbiased intervention of human resources such as examiners, moderators and verifiers. However, in manual evaluation of a large number of answer-scripts, a certain degree of lapse, error as well as variance in perception is bound to occur giving rise to evaluation anomalies. Institutions rely upon a number of corrective approaches at different stages of examination to control the evaluation anomalies. Some of the approaches for countering evaluation anomalies are described in Table 1.2.

Table 1.2: Approaches to Control Evaluation Anomalies

| Sr. | Approach | Purpose | Shortcomings |
|---|---|---|---|
| 1 | Moderation of evaluated answer-scripts | Ensure equitable treatment to all the examinees and to judge them on merit by reducing the 'examiner subjectivity' to the extent possible [Blo09]. | Several examiners do not follow the agreed norms of evaluation as their personal perception of strictness and leniency impinges on their evaluation. This can lead to an inclination to be erratic as well as at times careless [Blo09]. |
| 2 | In-house verification | Detect and correct evaluation errors. | Not suitable to address the lapses or intra and inter examiner variation in the evaluation [DKW14]. |
| 3 | Personal verification | Provide an opportunity for the examinees concerned to independently verify the evaluated answer-scripts to detect any errors in evaluation. | Does not address examiner subjectivity and variation [DKW14]. |
| 4 | Re-evaluation | Independent re-evaluation of answer-scripts to detect any errors in evaluation | Not suitable to control examiner subjectivity and lapses [DKW14]. |
| 5 | Scaling | Control general variations in evaluation. | Scaling is not suitable to control examiner variability arising from the 'Hawk-Dove effect' (strict/liberal evaluation)[GC88] and evaluation errors/lapses. |

## 1.1 Motivation

The management of a plethora of activities of summative examinations is a circuitous process and is prone to anomalies and security breaches [BS11].

A variety of cases of malpractices are frequently reported and recorded in the public domain. Together with this fact, the analysis of the current conventional/ electronic public examination system has revealed that the assessment process needs further enhancements and security interventions. The currently adopted security practices in conventional/electronic assessments are insufficient. They do not provide relief from all the security concerns of each and every stakeholder. This is apparent from the ever-increasing cases of breaches in the assessment security and of malpractices.

With the commencement of each examination season, cases of examination-related malpractices are reported repeatedly in Educational Institutes, Boards and Universities from across the country. The repeated incidents of question paper leakage in CBSE examination held in March, 2018 for Class X and XII affecting lakhs of examinees, is a glaring example of the malaise. As per the information released by CBSE and the information gleaned from local sources, in every examination, there is a steep rise in the number of cases of malpractices detected each year.

The malpractices such as question paper leakage and rampant collusion/plagiarism practices can be controlled substantially by generating a unique question paper, JIT to each examinee/group of examinees. However, if a unique question paper is provided to each examinee/group of examinees, we require a mechanism to establish an irrefutable link between the examinee identity and the question paper. It is also necessary to associate unambiguously the unique question paper received by the examinee to the corresponding answer-script produced by the examinee. The established association needs to be strong enough to prevent both the sender and the receiver from repudiating their action in the future. The binding of the unique question paper and the answer-script, needs to be done in such a way that, it satisfies the following security requirements:

1. Answer-scripts produced by the examinees are kept hidden from the examination

authority.

2. Answer-scripts produced by the examinees are available to the examiner, but the identity of the examinees is hidden from the examiner.

In order to achieve the above security requirements, we need a security scheme for establishing an unbreakable association between the question paper and the answer-scripts while revealing only the necessary information/identity to the communicating entities.

Secondly, in an examination system with a large number of answer-scripts pertaining to each course paper/subject, it is not possible to get all the answer-scripts evaluated by one examiner. Multi-examiner evaluation of the answer-scripts suffers from the subjectivity of each examiner. Examinees are also aggrieved by 'Hawk-Dove effect' [MMT06]. Herein, some examiners are strict and are prone to assign less marks even to an excellent answer, whereas, some other examiners are liberal and tend to allot marks leniently even to an average answer. The ultimate victims of the anomalous and heterogeneous evaluation are the examinees. The serious flaws in the current evaluation are apparent from the verification/re-evaluation results where the majority of the grievances of the examinees are converging into significant changes in the overall marks.

Thus, it is evident that examination malpractices, fraudulent acts and evaluation anomalies badly affect the reliability, uniformity and consistency of assessment and in general the entire examination system. Ultimately, it is the examinees who suffer due to the anomalies in evaluation. Therefore, the methods that are employed in evaluation need to be consistent, fair and error free for all the examinees.

## 1.2 Research Aims and Objectives

This research aims to devise mechanisms for controlling some of the malpractices/ anomalies that impair and mar the public examination system severely. The findings of this research will contribute to the development of a framework to support the delivery of examination content (question papers and answer-scripts) securely. This research proposes to delve into various dimensions of anomalies which crop up during evaluation and offers an E-moderation model for effective detection and correction of evaluation anomalies. This conceptual framework underpins the core research objectives, which are as follows:

1. To devise a mechanism for:
   (a) Achieving anonymity of examinees and examiners in answer-scripts exchange.
   (b) Creating an inseparable association/bonding between the unique question paper and the answer-scripts exchanged between the examination authority and the examinees. This would result in irrefutable identification of the question paper and answer-scripts pair.
2. To design a series of protocols that meet the defined security requirements for delivering examination content, namely question papers/answer-scripts amongst entities concerned.
3. To measure the evaluation anomalies in a specific conventional examination system and develop a unified approach to ensure an error-free and a uniform evaluation.
4. To devise an E-moderation model for:
   (a) Classification of each evaluation as anomalous or normal.
   (b) Prediction of tuned marks in an attempt to control intra/inter examiner variation in evaluation.

## 1.3 Thesis Focus and Key Contributions

This research advances the state-of-the-art in the design and analysis of secure and uniform public examination system with original contributions as listed below:

- The first contribution is the use of a cryptographic scheme to disguise the public key of the recipient from the sender, based on the concept of blind signature [Cha83] to achieve anonymity.
- The second contribution is a security property for linking the question paper and answer-scripts associated with the examinee and revealing only the selective and essential part of the aggregated information to the recipient based on the concept of dual signature [OPT97] .
- The third contribution is a formal framework for the security analysis of examination protocols. In this framework, a series of examination content delivery protocols are built. An inseparable association/bonding between the question paper and answer-script exchanged between the examination authority and the examinees are provided. This is done while maintaining the anonymity of the examinees and the examiners from each other.

We, in this research also identify different types of evaluation anomalies and quantify examiner subjectivity and the errors/lapses in the evaluation. In the context of evaluation anomalies in summative examinations, we offer the following original contributions:

- The design of computer-assisted evaluation using rubrics (CAER) to control errors in evaluation, reduce examiner subjectivity(intra/ inter examiner variation) and eliminate wastage of time.
- The design of E-moderation model comprising of:

1. Evaluation classifier to classify the given evaluation as negligent or normal using Support Vector Machine (SVM) classifier.
2. Tuned marks predictor to predict the normalized marks in the multi-examiner evaluation to control the intra/ inter examiner variation in evaluation using Artificial Neural Network (ANN).

The two methods proposed in this research, namely CAER (refer Chapter 5) and E-moderation (refer Chapter 6) offer two independent solutions for controlling examiner subjectivity (refer section 5.2.3).

## 1.4 Thesis Organization

**Chapter 1 : Introduction**

The first Chapter provides an introduction to the problem along with the motivation and the objective of the work.

**Chapter 2 : Summative Examinations**

This Chapter presents a detailed overview of summative examinations and its related components. It also discusses the methods of conducting summative examinations and the associated lacunae. Chapter 2 provides an overall understanding of the current research in handling security requirements and evaluation anomalies in summative examinations. This Chapter concludes with the catalogue of the threats faced by summative examinations and current approaches for countering those threats.

**Chapter 3 : Security Technologies for E-examination**

This Chapter provides the theoretical background of existing information security technologies and security protocol specification tool, namely Applied $\pi$ calculus. The user security model adopted in summative E-assessments is introduced and the limitations of the

model are highlighted. The Chapter ends with a brief review of security challenges in E-assessments.

**Chapter 4 : Security Protocols in E-examination against Malpractices of Collusion/Plagiarism**

In this Chapter, we introduce a dual purpose cryptographic scheme, namely 'disguised public key' to achieve anonymity and confidentiality in E-examinations. This Chapter also discusses the protocols for secure exchange of question papers/answer-scripts between examination authority and examinees for controlling malpractices of coercion/collusion/plagiarism. In addition, the security analysis of the proposed question papers/answer-scripts protocol is provided.

**Chapter 5 : Computer-Assisted Evaluation using Rubrics**

This Chapter discusses a solution in the form of Computer-Assisted Evaluation using Rubrics (CAER) for controlling examiner subjectivity and errors in evaluation. We validate the effectiveness of the proposed CAER approach over existing manual evaluation by comparing and contrasting the two approaches.

**Chapter 6 : E-Moderation for Detection and Correction of Evaluation Anomalies**

This Chapter reports the experiments carried out to control evaluation anomalies and inter/ intra examiner heterogeneity in summative examination. We discuss the machine learning techniques adopted to classify the evaluation as negligent or normal and further predict the tuned marks in an attempt to control the heterogeneity observed in the intra and inter examiner evaluation.

**Chapter 7 : Conclusion and Future Work**

This Chapter concludes the thesis with a summary of its contributions and an overview of directions for future work.

# CHAPTER 2
# Summative Examinations

Summative examination is the most significant and universally used instrument for measuring the level of knowledge and learning outcome of examinees [TVK05, Sad05, BF06]. A non-comprehensive list of assessment methods used in summative examinations to test progress include an objective or subjective tests/examinations, assignments, presentations, viva-voce examinations, coursework, group assessment and peer assessment. According to [Elt04], it is essential to conduct regular assessments to:

1. Provide support and feedback to examinees and improve their ongoing learning.
2. Report on what examinees have already achieved in the form of a grade or marks.

## 2.1 Types of Assessments

The purpose of assessment can be broadly categorized into assessment for learning and evaluation for decision making. Assessment for learning is achieved using formative assessments and evaluation for decision making and grading is done using summative assessments [MO99].

### 2.1.1 Formative Assessment

Formative assessment refers to the range of assessment procedures used by teachers during the teaching/learning process to provide guided feedback to the learners. It is a tool to monitor

examinees' learning to provide an ongoing feedback that can be used by instructors or teachers to enhance their teaching and by examinees to improve their learning [Ram83, Sad89, BW98]. Formative assessment includes tools for helping examinees to shape their learning, and bolster their abilities to take ownership of their learning and make them understand that the goal is to improve learning, not simply score marks [TL13]. More specifically, formative assessments take place while a class is ongoing and it continuously monitors examinee progress. Formative assessments are generally low stake assessments, as they are used to identify strengths and weaknesses of examinees and recommend further efforts for improvement.

**Examples of formative assessment:** Summary of the topic covered in the class by the examinee, self-assessment quiz, homework assignment, clicker questions with the examinee response system, etc.

### 2.1.2 Summative Assessment

Summative assessment refers to the range of assessment procedures used by assessors at the conclusion of a defined instructional unit (typically at the end of a project unit, course, semester, program or a school year) to measure progress and achievements of the examinee [MO99]. It is used to evaluate examinee learning at the end of an instructional unit by comparing it against some standard or benchmark [Sad89, Lin08]. The main purpose of summative assessment is grading, certification and placement [New78, Har05]. Summative assessment results are often recorded as scores or grades that are then factored into an examinee's permanent academic record. Summative assessments are high-stake assessments as they are used for promotion, placement, certification, and accountability [Rov00]. Thus, summative assessments need to be conducted in a manner that increases its robustness and reliability [Kni02]; since the results may have an enormous impact on an examinees' academic future/career prospects.

**Examples of summative assessments**: Tests, Semester End Examination, project assessment, viva voce examination, etc.

## 2.2 Framework of Summative Examinations

In this research, we focus on the standardized Public Summative Examinations. Such examinations are normally conducted at the end of the semester/term and are subjective in nature. In this section, we discuss the key components and structure of such summative examinations conducted within the framework of well-defined rules and regulations.

### 2.2.1 Examination Stakeholders

In general, following entities form part of the summative examination system:

1. **Question Paper Setter (P)** is an entity who sets the questions based on pre-defined syllabus. A subset of such questions are used in the examination based on the requirement of the question paper (QP).

2. **Examinee (E)** is an entity who appears for the examination and answers the given QP pertaining to each enrolled course as per the predefined schedule.

3. **Supervisor (S)** is an entity who is responsible for controlling and monitoring the examination during the conduct phase of the examination.

4. **Examiner (X)** is an entity who evaluates the answer-scripts of examinees at the end of the examination and allots the marks/grades based on a marking scheme.

5. **Examination authority/Controller of Examination (C)** is an entity, responsible for conducting an examination in a fair manner. Examination authority is responsible for

appointment of paper setters, supervisors, examiners and producing QP, delivering QP and collecting answer-scripts (AS), getting answer-scripts evaluated and at the end marks collection, entry and processing of marks and the declaration of results.

### 2.2.2 Examination Assets

1. **Question Paper (QP)** is a document, having a set of questions organized as per the predefined format based on the approved course curriculum. QP comprises of header and content. The header of a QP contains a unique code of the QP, the name of the examination, duration of time, maximum marks, subject/paper name, programme name and instructions for examinees. A question paper content part contains the collection of questions. A typical question paper consists of:

   (i) 'M' number of the main questions. All 'M' questions could be compulsory or an examinee needs to attempt any 'N' questions out of 'M'. Some other variations are also possible.

   (ii) Usually, each main question comprises of 'n' number of sub-questions. In this all 'n' questions could be compulsory or examinees need to attempt any 'p' questions from 'n' questions.

   (iii) Sometimes a choice between the two questions, i.e. 'A' or 'B' is permitted.

2. **Answer-Script (AS)** is a document produced/written by an examinee, carrying answers to the set of questions contained in the QP.

3. **Marks (MK)** represent the numerical value assigned by the examiner to each answer after assessing the content of the AS.

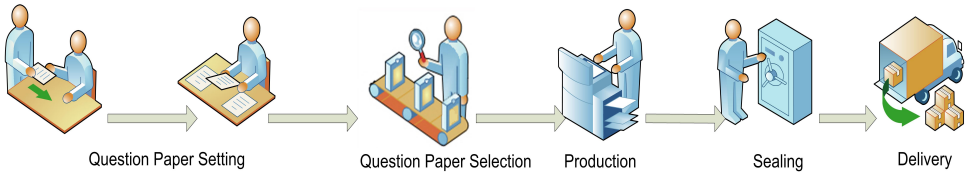
### 2.2.3 Examination Stages

The multitude of tasks involved in summative examinations can be broadly classified into three stages. They are the pre-conduct stage, the conduct stage and the post-conduct stage.

#### Pre-conduct Stage

The pre-conduct stage of the examination identifies and establishes the basic requirements necessary for conducting an examination efficiently. The two main activities in this stage are registration of eligible examinees (refer Fig. 2.1a) and question paper production involving the appointment of question paper setters, question paper setting, question paper printing, sealing and delivery of question papers to respective examination centres (refer Fig. 2.1b).


(a) Examinee Registration




(b) Question Paper Production and Delivery

Fig. 2.1: Pre-Conduct Stage of the Examination

1. **Examinee Registration**

   Examinee registration process identifies the eligible examinees interested in appearing for the examination. Such eligible examinees are issued a unique examination seat number.

2. **Question Paper Finalization**

   Question paper finalization involves the following activities:

- Appointment of paper setters.
- Setting of sets of question papers for each offered course through an appointed panel of paper setters.
- Verification of manuscript of question paper for any errors/out of syllabus question.
- Sealing of question papers and delivering them to the examination authority for final printing.

3. **Question Paper Production**

   The Controller of Examination randomly selects one of the sealed manuscript of the question paper and sends it for printing. As per the number of examinees registered, the question papers are produced, course-wise and sealed in separate packets.

4. **Question Paper Delivery**

   The Controller of Examination delivers the requisite number of sealed copies of question papers to the examination centres, well in advance, before the conduct of the examination.

5. **Appointment of Supervisors**

   Examination is conducted in a supervised environment to monitor the actions of examinees and to control malpractices during the examination. Supervisors are entrusted with the task of providing free and fair environment to all the examinees during the answering of examination.

6. **Appointment of Moderators and Examiners**

   The moderators and panel of examiners are selected and appointed. The number of examiners to be appointed depend on the number of answer-scripts to be evaluated. The paper setters of a particular question paper are by default appointed in the panel of examiners.

Besides these activities, in the pre-conduct stage of examination many other miscellaneous activities are also carried out. These include acquiring stationery, blank answer-books, preparation of seating arrangement and preparation of schedule of examination.

**Conduct Stage**

In the conduct stage, a conducive environment for examinees while answering the examination is provided. Together with this, it is also essential for a strict monitoring of the examinee activities/discipline in the examination hall. A series of steps have to be taken during the conduct stage of an examination (refer Fig. 2.2).

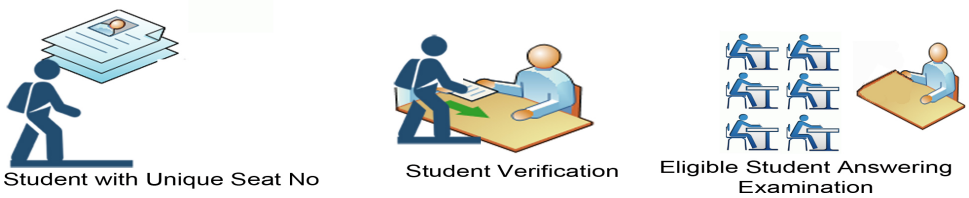



Fig. 2.2: Conduct Stage of the Examination

1. The block supervisor at the examination centre provides the examinee with the blank answer-book. He ensures that the blank answer-book provided to the examinee bears his signature. This is done to ensure that the answer-book issued to the examinee is fresh and no answer-book is smuggled inside the examination hall. The examination authority supervises the entire process of examination conduct to deal with any untoward incident.

2. The examinee occupies the seat marked for him in the examination hall.

3. The seating arrangement for examinee is made in such a manner so as to ensure that the content of the examinees answer-book is concealed from other examinees answering the question paper.

4. The question papers are distributed to each examinee as per the schedule of the examination.

5. Verification of identity of examinees is undertaken.

6. An attendance record of the examinees answering the examination is maintained. An examinees' signature is obtained on an attendance sheet. Supervisors maintain a report indicating the details of examinees present and examinees absent for the examination. However, in this entire process, no receipt of whatsoever is provided to examinees confirming that the examinee has answered the particular examination question paper.

7. No examinee is permitted to enter the examination hall after 15 minutes from the commencement of the examination.

8. Similarly, no examinee is permitted to leave the examination hall during the last 10 minutes of the examination.This is essential to prevent the last minute chaos and malpractices and permit supervisor to focus on the task of supervision more diligently.

9. Finally, the examinees answer-scripts are collected.

**Post-conduct Stage**

The post-conduct stage of the examination involves the following activities (refer Fig. 2.3) :


Fig. 2.3: Post-Conduct Stage of the Examination

1. **Tallying**

   In the tallying process, the collected answer-books are tallied with the supervisors' report to cross check any lapse in the collection of answer-scripts. After verification, answer-books are sealed along with the Supervisors report. Sealed packets of answer-scripts are then forwarded to the respective examiners for evaluation.

2. **Evaluation**

   Each appointed examiner gets the examinees' answer-scripts for evaluation. The identity of the examinee is kept hidden from the examiners to prevent any coercion between examinees and examiners.

   If there are more than one examiners, then the moderator discusses with them collectively the modalities for evaluation. They all agree on one pattern of assigning marks as per the content. The number of answer-scripts to be evaluated per day is strictly enforced and monitored.

   There is also a provision of moderation of evaluated answer-scripts. The moderator picks up randomly some percentage of answer-scripts evaluated by examiner 'X' and evaluates them independently. Examiner 'X' has/must evaluate all the answer-scripts again, if major variations are observed.

3. **Result preparation**

   The result preparation task involves entry of marks obtained by each examinee in each course paper in the computer system and then tabulation of those marks for preparation of statement of marks indicating the examinee performance.

4. **Personal Verification/Re-evaluation**

   Examinees with grievances pertaining to the evaluation, have the option of applying for personal verification/re-evaluation. In the personal verification process, the entire evaluated answer-script is verified again by the examinee concerned, to check whether

any answer is unassessed or there are any errors in the total or while transferring the marks on the main statement of marks. Some examinees opt for re-evaluation where the entire answer-script of the examinee concerned is re-evaluated by a different examiner. If there is a major difference of marks between the evaluation of the first and the second examiner, then, the examinee gets the benefit of the higher marks.

This stage also deals with the scrutiny of the unfair means, verification, re-evaluation, tabulation of the marks and generation of the statement of marks indicating the performance of the examinees in the examination.

## 2.3 Methods of Conducting Summative Examinations

The methods that are employed in conducting the summative examinations needs to be fair and accurate irrespective of the type of assessment. The two widely used methods of conducting summative examinations are, Conventional Paper/Pen examination and Electronic (Information Communication Technology enabled) examination.

### 2.3.1 Conventional Paper/Pen based Examination

Conventional examination is "an assessment delivered to the examinee in the form of the physical question paper and where the examinee answer the examination on physical paper answer-book", [BRW06]. Conventional examination is often the first choice of institution(s) for conducting summative subjective assessments as it does not require any sophisticated tools and technology.

The main advantages of Paper/Pen based examinations are:

- It offers familiar and well understood examination environment.

- It can be used in any setting without the need of any sophisticated tools/technology or expertise.
- It allows a teacher to assess the strengths and the weaknesses of examinees coupled with reflections, and the thinking processes of examinees.

Some of the disadvantages of Paper/Pen based examinations are:

- Inaccessibility

  Administering Paper/Pen based examinations in a class-room environment sometimes excludes differently-abled examinees from participating.
- Limited Flexibility

  It offers only a fixed set of questions to all examinees and customization/randomization of questions is a rare activity.
- Lengthy Process

  It takes a huge amount of time and human resources to complete the entire process.
- High Cost

  The manual process of conducting examinations and transporting the answer-scripts to the respective marking centres/locations, is not only time consuming, but also requires significant human resources, at various levels. This results in increased effort and costs while increasing the chances of errors / loss / damage.
- Non Eco-friendly

  A lot of paper is needlessly wasted in the Conventional examination. Wasting such a valuable environmental resource often goes against the institution's 'green' initiatives.
- Evaluation errors

  Manual evaluation can lead to calculation errors or unchecked questions due to

evaluators' oversight that can be detrimental to an examinee's career as well as impact the institution's reputation.

### 2.3.2 Electronic Examination

Electronic examination (E-examination) refers to the "use of Information Communication Technology (ICT) to deliver assessments to candidates and manage assessment related tasks", [BRW06]. E-examinations are popularly used for conducting objective tests suitable for formative/summative assessments. E-examination offers many more advantages than the Conventional Paper/Pen examinations as follows:

- It significantly reduces the logistics cost associated with conventional examinations like printing of papers, manual evaluation and need for additional resources for re-evaluation, verification etc.
- E-examinations can be configured for 24/7 availability.
- E-examinations can be easily scaled to large examinee population over a wide spread of locations.
- E-examinations can be randomized to present questions in a predetermined or random order or a different question paper to each examinee [HM99].
- Digitization and automation can best provide the needed security for transporting answer-scripts back and forth from the examiners to the Controller of Examination with proper control over access.
- Real time monitoring of answer-scripts ensures the complete elimination of administrative errors which include the incorrect calculation of marks, un-evaluated questions and as a result, incorrect awarding of marks.

- E-examination provides the option of digital storage and handling of the answer-scripts which reduces the risk of misplacement, mutilation and misuse.

Some of the prominent disadvantages associated with E-examinations are:

- E-examinations are not fully suitable for subjective assessments.
- The security and confidentiality of summative E-examinations are often at stake compromising the standard of the examination and may result in a cancellation or in a retake of the examination.
- Technology is not always reliable. Information can be lost, if a system breaks down.

## 2.4 E-evaluation in Summative Examinations

The evaluation of answer-scripts is the most crucial and integral part of summative examinations. A typical summative examination system comprises of a large number of answer-scripts pertaining to each subject/course paper. Evaluation of these subjective answer-scripts is a highly human intensive task and needs focused and unbiased intervention of human resources such as examiners, moderators and verifiers. However, in the manual evaluation of a large number of answer-scripts, achieving consistency and uniformity with a precision and perfection is a tall order.

E-evaluation is perceived to mitigate majority of the shortcomings associated with manual evaluation. E-evaluation enables examiners to evaluate the answer-scripts on a computer screen or automatic marking of answer-scripts with the aid of artificial intelligence and computer recognition systems.

### 2.4.1 Types of E-evaluation approaches

Some of the E-evaluation approaches for the evaluation of subjective answer-scripts are as follows:

**Marginally Automated E-evaluation (MAE)**

In MAE approach, first the hand-written answer-scripts are scanned and transformed into digital documents. Examiners are provided with the scanned copies of answer-scripts and electronic marks recording system for evaluation.

As answer-scripts are stored digitally, the need for physical storage is totally eliminated. Thereby, ensuring an error-free tabulation of marks, rapid evaluation or re-evaluations, safety of answer-scripts, retrieval of answer-scripts with great ease based on date/paper, quick compiling and collating evaluations of any examination.

However, MAE approach demands huge initial investment in acquiring machinery. Also, in MAE approach, there is a recurring and tedious process of manual scanning, arranging and storing of a large number of answer-scripts inviting scope for errors and threats.

Example: Online Marking System (OSM)
An online marker sees a scanned version of the examinee's handwritten answer-script on their computer screen. Examiners read and judge the answers against the marking scheme as usual, but marks are awarded with a mouse click rather than the examiners' traditional red pen.

**Semi Automated E-evaluation (SAE)**

SAE is an extension of the first type of evaluation. An electronic marks recording system, equipped with an answer-key, rubrics, checklist, or other form of scoring guide accompanies this type of assessment.

SAE is a much viable option to realize the significant process improvements while retaining the adaptability/human intelligence involved in manual examiners input.

Some of the tools/research exploring the variety of approaches for effective semi-automated evaluation include:

1. 'ALOHA' - It offers a semi-automatic grading mechanism with the aid of computer-assisted rubrics for programming courses combined with providing the feedback to students [AR06].

2. 'DES' - It is a system that addresses the error-prone tasks of allocating the marks and calculation of marks with the help of web-based testing system [RSS09].

3. Learning Management Systems (LMS) such as Moodle (https://moodle.org/) and Blackboard (http://www.blackboard.com) allows the use of rubrics to assess different aspects of the assessment.

4. 'Evalcomix' - It uses rubrics mechanism for definition of assessment instruments and also enabling peer and self assessment approaches (http://evalcomix.uca.es/).

The assessment process in these systems can benefit from some automation, but some aspects of the subjective assessment such as ambiguity and volume are too difficult and expensive to fully automate reliably.

**Fully Automated E-evaluation (FAE)**

In the FAE, the evaluation of answer-scripts are carried electronically without any human intervention. The automatic E-evaluation is mainly based on keyword match, sequence match and quantitative analysis, but semantic analysis of descriptive answer is still in its primitive stage. In the semantic analysis of descriptive answer, Natural Language Processing (NLP)

tasks and applications are used. It involves parsing of answer text to find the semantic meaning of examinee answer and finally compare it with answer provided by the expert and assign the final scores. At this juncture, it is imperative to state that, although there are several fully automated evaluation systems in existence, yet they are not fully operational [KS04, HTLC10, KKP03].

### 2.4.2 Essential Qualities of Evaluation

#### Accuracy

The accuracy in evaluation refers to the degree to which the result of a measurement/calculation of examinee performance conforms to the correct or the actual performance of the examinee. The evaluation of the answer-scripts need to be error-free and without any anomalies.

#### Fair

The evaluation process under no circumstances should provide any opportunities to examinees or other stakeholders to engage in malpractices. The fairness in evaluation advocates assessing all examinees in a standardized manner using identical assessment methods, administration, scoring, and interpretation procedures. In an evaluation, every examinee deserves an equal opportunity to demonstrate what he or she understands, knows, and can perform. The evaluation bias occurs when a group of examinees has an unfair advantage over an item or group of items. e.g. An examinee who has a good handwriting stands to benefit than an examinee with an illegible handwriting.

### Reliability

Reliability in evaluation refers to the consistency of marking/grading within and across examiners and repeatability of the outcome when presented with the same assessing factors [Kni02]. An assessment is considered reliable when an examinee's grade/marks does not vary, regardless of who did the evaluation and when the evaluation was carried. Reliability in an assessment is important because assessments provide information about an examinee's achievement and progress. Assessors need to come up with the same/similar results when more than one assessor is involved in the evaluation or when the same assessor assesses the same work on different occasions. The former is referred to as inter-examiner reliability and latter is referred to as intra examiner reliability.

### Validity

An assessment is considered valid when it is fit for the purpose required or when it measures what it is planned to measure [New78]. An assessment may become invalid when factors that are irrelevant to the learning outcomes are accommodated. Thus, it is required that the course content of an assessment should closely match the content of the specification, it is designed to assess [Dow03].

### Standardization

Standardization is another quality of a good assessment. For an assessment to be standardized, it must consist of similar content and format, the assessment must be administered the same way and scored in the same manner for every examinee the assessment is given to.

### 2.4.3 Related Research in Summative E-assessment

We need a mechanism in evaluating the subjective answer-scripts which take care of variation and errors without much additional cost and resources. The current research in subjective answer-script evaluation approaches this issue under two main schools of thought, i.e. fully automated [KS04, KKP03, HTLC10] and semi-automatic [RSS09, AK09]. Both approaches are focused towards establishing uniformity in evaluation and reduction of errors. In this thesis we restrict our discussion to semi-automatic evaluation of subjective answer-scripts. In the web based descriptive examination system (DES) proposed by [RSS09] the error-prone tasks of allocating the marks and calculation of marks is addressed effectively. However, DES does not address the issue of examiner variation in evaluation and providing automatic feedback to the examinees, as it does not include any well defined evaluation framework or mechanism to record the compliance of examinee answer to the desired answer.

Rubric is one such mechanism which offers uniform and consistent evaluation platform [AC06]. Rubric is a tool used for assessment that identifies specific criteria to be assessed and gives numeric scores as per the quality of performance. Starting with the highest level and descending to the lowest, these levels of performance are used to assess the degree of proficiency attained by an examinee [RA10].

The computer-assisted grading using rubrics has been shown to help in solving the examiner variation and achieve objectivity in assessment [AK09] and improve the speed of assessment [AASK08]. A semi-automatic grading tool based on computer assisted rubrics is presented in [AR06] for programming courses. This tool is intended to provide a consistent and objective grading between different examiners combined with providing of the feedback to the examinees. There are tools offering semi-automated approaches to grade essays and provide feedback based on analytical assessment rubrics [And00].

Numerous electronic examination solutions have been designed with the intent of improving

the efficiency and eliminating the loopholes associated with the current examination system. Learning Management Systems (LMS) such as Blackboard (http://www.blackboard.com), Moodle (https://moodle.org/) allow the use of rubrics in assessment. However, these tools are suitable for specific type of questions/linear type of question paper such as essays [And00], assignments (http://www.blackboard.com), (https://moodle.org/), programming questions [AR06].

There is difficulty in adopting these solutions per se, in conducting our examinations as they do not correctly model our examination requirements. The question paper format of most of the examinations is generally non-linear in nature, having optional components between series of questions (refer Section 2.2.2). Examinees can attempt an exact number of questions or additional questions from optional component. When examinees attempt extra questions, examiners have to select the higher marks out of all the optional questions attempted. Also, the examination system that is under our consideration have a provision of entitlement marks to examinee due to participation in sports, cultural activities, National Cadet Corps (NCC) and National Social Service (NSS) activities. These kind of specific and customized requirements are difficult to adjust/accommodate in the existing electronic solution, as a result such solutions, have limited scope and cannot be used in the same way in present day full-fledged public examinations.

### 2.4.4 E-Evaluation Systems in Public Examinations

Public examinations consist of a large number of answer-scripts for evaluation. However, in order to speed up the process of evaluation and result declaration and to reduce evaluation errors, there are a variety of E-evaluation systems. These have already been used in public examination systems.

**On Screen Marking (OSM)**

OSM (https://www.orioninc.com) is the technological aid for evaluation of subjective answer-scripts. The main thrust of OSM technology is the improvement of efficiency, transparency, flexibility and overall quality of evaluation. It also simplifies to a great extent the scoring process for examiners. In OSM, each page of every answer-script is scanned with the help of heavy duty sophisticated scanner. The scanned soft copy of answer-scripts are offered to examiners for evaluation. Examiners evaluate the scanned answer-scripts digitally with the help of the OSM digital marking system. OSM also has an add-on feature for monitoring and validating the evaluation. Thus, it increases the reliability of the evaluation. OSM solution also lends itself to accuracy and transparency. The latter is assured by making the evaluated answer-scripts available to examinees for personal verification. The OSM method was implemented in 2014 by Central Board of Secondary Education (CBSE) to evaluate major subjects of Class X examination.

**Pros:**

1. OSM improves the quality of evaluation. It also helps to monitor the evaluation.
2. OSM is compatible for grading both objective and subjective type of answers.
3. It eliminates the risk involved while dispatching answer-scripts from examination centres and evaluation centres.
4. It ensures that the examiners evaluate only the answer-scripts assigned to them, thus preventing any coercion and intrusion.

**Cons:**

1. Scanning of each and every page of answer-script is a highly demanding task. It requires a lot of time and focus in segregating individual pages of each answer-book

for simultaneous scanning of both sides of each page. Additionally, each answer-book needs to be stitched back to restore it to its original state at the completion of the scanning process.

2. Heavy duty sophisticated scanners are expensive. The cost further inflates due to labour and the time factors.

3. It is mandatory to retain the scanned copies of the answer-scripts for a stipulated period. This results in additional cost on server space for storing huge amount of data.

4. In this system, examiners need to download scanned images of answer-scripts from a central server for evaluation. In Public examinations, thousands of examiners are simultaneously engaged in evaluation. Therefore the internet requirement will be huge, leading to further escalation of costs.

5. It is prone to the problem of mismatched answer-scripts due to human errors/lapses during the scanning.

6. Computers and other necessary infrastructure have to be installed to facilitate access to the answer-scripts for evaluation in the examination centres.

7. Security and accountability is at stake when the entire activity of scanning of answer-scripts is outsourced.

**Orion Live Ink Character Recognition (OLICR)**

OLICR (https://www.orioninc.com) is a technology for marks digitization. Herein, a re-purposed digital pen is used to capture handwritten marks in real time on a digital device. Examiners use the conventional Paper and Pen method of evaluation, and then use a digital pen to record marks on the OLICR page. The pen digitizes the marks and transfers the data in real time to the accompanying digital device such as a tablet. The digitized marks are

encrypted and then transferred to a secure cloud based server. The Council for the Indian School of Certificate Examinations (CISCE) used the OLICR solution to evaluate major subjects of Class X and XII examination in 2016.

**Pros:**

1. It is appropriate for evaluation of both subjective and objective type of answers.
2. Security is ensured as no traces of scanned data remains in the digital pen/device.
3. Tasks such as assignment of rubrics to each question, validation of maximum marks and totalling of marks are carried out by the system, it eliminates likelihood of any human error.
4. Marks allotted by the examiners to each answer are instantaneously transferred to the central servers. This facilitates the compilation and declaration of results within a short span of time.
5. The efficiency of an examiner can be easily traced, as the identity of the examiner, the date and the time of evaluation is captured by the digital pen.

**Cons:**

1. LICR documents are very expensive and in most cases beyond the budget of the examination section due to the high recurring cost.
2. LICR solution needs constant maintenance to sustain its integrity and reliability.
3. Creation of LICR documents is a time consuming process as LICR documents need to be printed according to precise standards.

**Paperless Digital Examinations (PEXA)**

PEXA (http://www.littlemoreinnovation.com) is a secure cloud-based end-to-end solution that allows examinees to answer descriptive examinations digitally, using electronic pad (Digitaal) and a stylus instead of using paper and pen. PEXA system works on a software as a service model. It handles the entire process of conducting an examination digitally, right from question paper setting, delivering the digital question papers to examinees, pushing digital answer-scripts to examiners for evaluation to declaring results. Thereby, it eliminates the necessity of paper in the entire process. Many institutions such as Tamil Nadu Agricultural University, Manipal University, IIIT, Bangalore, REVA University, VIT, Vellore, NIMHANS and others are using the PEXA technology.

**Pros:**

1. There is no need of continuous internet connectivity, AC Power, LAN Cables and power backup while conducting examinations.
2. The data is transmitted to and from the devices using cloud technology, thereby, not requiring the devices to be connected to the internet all the time.
3. The evaluation become flawless due to the features in the tablet which prevents double marking, no marking, etc.
4. Examiners can pull the answer-scripts from cloud for evaluation on their own device at the comfort of working from anywhere and send it back to the cloud.
5. Storing in the soft form in the dedicated cloud platform saves space and the retrieval of data is highly efficient and easy.
6. Examiners can get to view analytics on various aspects such as how much time was spent on a particular question, the total time spent to attempt the full question paper,

comparison of marks and much more.

**Cons:**

1. Storing important and confidential examination data on external service providers always opens up risks and security challenges on a routine basis.
2. Huge investment is necessary to acquire software as a service, digital pads and other related infrastructure.
3. Security and accountability is at a stake when the entire examination activity is outsourced.
4. Question paper needs to be set as per the precise PEXA format.
5. PEXA solution needs constant maintenance to sustain its integrity and reliability and recurring cost on data storage.
6. As cloud service providers take care of a large number of clients each day, they can get temporarily suspended. Thus, preventing access to your applications, server or data from the cloud.

## 2.5 Malpractices in Public Examinations

Conducting a public examination is an expensive affair. The expenses and revenues are so tight that all efforts are made to underplay any malpractice so that a re-examination is avoided. We, in this section discuss, some of the large scale malpractices in public examinations in the recent past indicating the methods and the technological penetration.

### 2.5.1 Atlanta Public Schools (APS) Scandal (2009)

The Atlanta Public Schools (APS) scandal was exposed in the year 2009. In APS scandal, school authorities colluded in changing examinee marks to improve their institution's rankings and get more public funds in their state-administered standardized tests.

APS scandal got unearthed almost after a decade of institutionalized corruption of standardized tests. In order to satisfy annual targets of the school, teachers and administrators adopted unfair means. They gave children answers to the questions, erased incorrect answers, hid and altered documents. Together with this, they offered monetary incentives to encourage cheating and punished employees who refused to comply.

### 2.5.2 Vyapam scam (2013)

Madhya Pradesh Professional Examination Board (MPPEB), popularly known by its Hindi acronym 'Vyapam' (Vyavsayik Pariksha Mandal), is a self-financed and autonomous body incorporated by the State Government responsible for conducting several entrance tests in the state. The Vyapam scam involved leakage of question paper, use of proxies and a collusion of undeserving candidates, who had bribed the politicians as well as the MPPEB officials through middlemen, to get high ranks in the entrance tests.

One of the methods used in cheating was impersonation. After the examinee signed, the identity card for the examination was doctored to match the bribe-giver's name with the photograph of a 'scorer' - a hired proxy. After the results were declared, the proxy's photo was replaced with that of the examinee. Supervisors for the examination were paid to ignore the mismatch between names and photographs of identity cards.

### 2.5.3 UK examinee visa tests fraud (2014)

The UK examinee visa tests fraud was perpetrated by official invigilators in collusion with all the candidates. This fraud was primarily executed using two methods. In the first method, the real candidates had just to wait to have their photographs taken - as proof that they were present for the test. They had a 'fake candidate' who appeared for both the spoken and written tests for them. In the second method, the invigilators dictated the correct answers to the registered candidates.

### 2.5.4 AIPMT scandal (2015)

The All India Pre-Medical Test (AIPMT), is one among the most important and prestigious entrance examinations for admissions to MBBS and BDS colleges. The AIPMT scandal occurred in the year 2015, where a gang of external agents used mobile phones. They used pre-paid SIM cards for passing on answer keys to examinees using vests with SIM card units and Bluetooth-enabled earpieces. A criminal network had a leaked copy of the question paper. Medical examinees and doctors were hired to solve the multiple-choice questions. Simultaneously, the answers to the candidates were sent via messaging app WhatsApp during the examination. The network had provided several candidates with tiny Bluetooth devices, vests tagged with microSIM cards, and wristwatches fitted with cameras. These answers were relayed in real-time to the candidates via wireless devices.

### 2.5.5 NEET cheating scam (2017)

The National Eligibility and Entrance Test (NEET) is a medical entrance examination for the undergraduate as well as the postgraduate examinees who are willing to pursue their career in medical courses (MBBS/BDS).

Prometric, a US-based company which conducted the NEET in December, 2016 admitted that their software could be breached. As per records and evidences, the NEET scam was carried out by a network of agents and sub agents who were in touch with aspiring examinees. After hacking the examination software, the agents then advised the candidates on which examination centre they should choose while filling the NEET forms. The agents also colluded with site supervisors. The supervisors permitted the privileged examinees to use internet to remotely connect to the computer outside the examination hall to get assistance.

### 2.5.6 CBSE Examination Paper Leak (2018)

The CBSE conducts examination for Class X and Class XII. The CBSE Board spends close to Rs. 90 crores on the Board examination process. More than 20 lakh examinees appeared for Class X and Class XII CBSE Board Examination in the year 2018. In the CBSE 2018 examination, Class X Mathematics and Class XII Economics paper were leaked on WhatsApp. Since, WhatsApp messages are end-to-end encrypted, tracing the source, was a difficult task. CBSE examination question papers are stored in bank vaults, in branches near the examination centres. The sealed copies of the question papers are handed over to the staff of individual examination centres, a couple of hours prior to the examination. A group of people, in connivance with bank and examination centre staff, are suspected to be responsible for the leakage of the CBSE question paper. In order to avoid detection, the accused allegedly made handwritten copies of the question papers.

## 2.6 Approaches for Countering Evaluation Malpractices

1. Appointment of multiple paper setters per subject/course for preservation of secrecy of question papers and establishing the anonymity of paper setters.

2. Submission of a sealed hard copy of the manuscript of question paper to protect the integrity of the question paper.

3. Monitoring and supervising the way the examination is conducted from start to the finish, for controlling acts of academic dishonesty such as collusion, plagiarism, cheating, etc.

4. Use of a unique labelled question paper and answer-script book or common question paper cum answer-book to link the question paper and answer-script together.

5. Hiding examinee identity on the answers-book and assigning a code mapped to examinee identity to keep the identity of examinees anonymous.

## 2.7 Summary

This Chapter draws on the lessons learned over the years in the conduct of summative examinations. It presents the general information applicable to all instances of summative examinations including public examinations. The discussion focuses on specific constraints in the conduct of summative examinations involving a large number of examinees.

The discussion in the Chapter begins with a brief summary of the framework within which the examinations are conducted and the different types of assessments. It then, presents the different approaches in E-evaluation and some practical E-evaluation systems. The Chapter goes on to present various intricacies of summative examinations coupled with a series of approaches for countering the evaluation complexities. The Chapter concludes with a

discussion on some of the malpractices in the recent past in the domain of public examinations.

# CHAPTER 3

# Security Technologies for E-examination

The security requirements of E-examinations to a great extent concur with the variety of E-business applications. A reliable and fair E-examination system also needs strong security mechanisms coupled with well-defined layered protocols.

Cryptography plays a key role in achieving the security in most of the electronic transactions. It provides protection to the data from being viewed and ways to detect whether data has been modified. It helps in ensuring a secure means of communication over otherwise non-secure channels. A secure computer system is built on 4 main pillars of security, namely, Confidentiality, Integrity, Availability and Authentication [Gol99, PP02]. In particular,

1. Confidentiality protects the data item from interception and from being read by intruders.
2. Integrity helps in safeguarding data from unauthorized modification.
3. Availability enables data to be free from interruption.
4. Authentication ensures that the data originates from a particular party.

This Chapter deals with details of various cryptographic schemes required to achieve information security. We also discuss, the protocol specification and analysis tools used for assessing the security of proposed protocols. In the section 3.1, a brief introduction to the cryptographic principles is provided. Section 3.2, we discuss some of the security threats

associated with current examination system, coupled with approaches and shortcomings. Section 3.3, we describe the applied $\pi$ calculus and ProVerif tool used for formal modelling and verification of security protocols. The last section 3.4, cites the intervention of various security technologies in E-examination.

## 3.1 Security Primitives

Encryption/Decryption, hashing and digital signatures are generic ingredients for achieving security in electronic communications. There are other specific cryptographic mechanisms such as blind signature and dual signature schemes which are used for maintaining anonymity and to secure electronic transactions. All these cryptographic schemes have a wide variety of applications in building a secure E-examination system.

In order to achieve the security goals in electronic communication, a combination of algorithms and practices known as cryptographic primitives, are used. A list of some of the cryptographic primitives and their uses are reflected in the Table 3.1.

### 3.1.1 Secret Key Encryption

Secret key encryption algorithms use a single secret key to encrypt and decrypt data. Secret key encryption algorithms are very fast (compared with public key algorithms) and are well suited for performing cryptographic transformations on large streams of data.

Block cipher is a type of secret key algorithm that is used to encrypt one block of data at a time. Block ciphers such as Data Encryption Standard (DES) [Cop94], TripleDES [Bar17], and Advanced Encryption Standard (AES) [DR13] cryptographically transform an input block of 'n' bytes into an output block of encrypted bytes. The most widely used symmetric key cipher is the AES.

Table 3.1: Cryptographic Primitives

| Cryptographic Primitive | Use |
|---|---|
| Secret key encryption (Symmetric cryptography) | Symmetric encryption transforms the data to keep it secret from the third parties. It uses a single shared, secret key to encrypt and decrypt data. |
| Public key encryption (Asymmetric cryptography) | Asymmetric encryption performs a transformation on data to keep it secret from third parties. It uses a public/private key pair to encrypt and decrypt data. |
| Cryptographic hashing | It maps data from any length to a fixed-length byte sequence. Hashes are statistically unique. A different two-byte sequence will not hash to the same hash value. |
| Cryptographic signing | It helps verify that data originates from a specific party by creating a digital signature that is unique to that party. This process also uses hash functions. |

The disadvantage of secret key encryption is, that, it presumes, two parties have agreed on a key and communicated their values. However, the key must be kept secret from unauthorized users. Due to these problems, secret key encryption is often used together with public key encryption to privately communicate the values of the key. In a real world scenario, either the sender or the receiver generates a secret key and uses public key (asymmetric) encryption to transfer the secret (symmetric) key to the other party.

### 3.1.2 Public Key Encryption

Public key encryption uses a private key that must be kept secret from unauthorized users and a public key that can be made public to anyone. The public key and the private key are mathematically linked. Data that is encrypted with the public key can be decrypted only with the private key. Moreover, data that is signed with the private key can be verified only with the corresponding public key. The public key is used for encrypting data to be sent to the keeper

of the private key. A basic cryptographic rule prohibits key reuse. Both the keys should be unique for each communication session.

Two parties (Alice and Bob) might use the public key encryption as follows:

1. First, Alice generates a public/private key pair. If Bob wants to send Alice an encrypted message, he asks her for her public key. Alice sends Bob her public key over a non-secure network.
2. Bob uses this key to encrypt a message.
3. Bob sends the encrypted message to Alice.
4. Alice decrypts the received message using her private key.

If Bob received Alice's key over a non-secure channel, such as a public network, Bob is open to a man-in-the-middle attack. Therefore, Bob must verify with Alice that he has the correct copy of her public key.

Asymmetric encryption algorithms such as RSA [MS13] allows both encryption and signing, but DSA [MS13] can be used only for signing, and Diffie-Hellman [DH76] can be used only for key generation. The most popular implementation of public key encryption is Pretty Good Privacy (PGP) [Zim95]. In general, public key algorithms are more limited in how much data they can encrypt as they are bound to increase the size of the data that it enciphers than private key algorithms.

### 3.1.3 Hybrid Cryptosystem

Hybrid encryption incorporates a combination of asymmetric and symmetric encryption to benefit from the strengths of each form of encryption. These strengths are speed and security respectively.

Hybrid encryption is achieved through data transfer using unique session keys along with symmetrical encryption. Public key encryption is implemented for random symmetric key encryption. The recipient then uses the public key encryption method to decrypt the symmetric key. Once the symmetric key is recovered, it is then used to decrypt the message.

**Steps of Hybrid Encryption:**

In a hybrid cryptosystem, a sender performs the following steps to encrypt a message addressed to a receiver:

1. Generates a fresh symmetric key for the data encryption.
2. Encrypts the message, using the symmetric key just generated.
3. Obtains the receiver's public key.
4. Encrypts the symmetric key with the receiver's public key.
5. Sends both of these encryptions to receiver.

To decrypt this hybrid ciphertext, the receiver operates in the following manner:

1. Uses his private key to decrypt the symmetric key.
2. Uses the symmetric key obtained in previous step to decrypt the message.

The hybrid encryption method provides an added security alongwith overall improved system performance.

### 3.1.4 Hashing

Hashing takes any arbitrary length of data (binary or text) and creates a constant-length hash representing a checksum/message digest/hash values for the data. It is designed to be a

one-way function, that is, a function which is unfeasible to invert. In other words, it is difficult to recreate the input data from an ideal cryptographic hash function's output. An important application of hash values is verification of message integrity. In order to determine, if any changes have been made to a message, it is necessary to compare the message digests calculated before, and after, transmission. If the hash is cryptographically strong, its value will change significantly for minor difference in the two messages.

Some of the popular hash algorithms are MD-4, MD-5, SHA-1 and SHA-512.

### 3.1.5 Digital Signatures

Public key algorithms are used to form digital signatures. Digital signatures authenticate the identity of a sender and help to protect the integrity of data. A digital signature is generated by combining a user's private key with the data he wishes to sign in a mathematical algorithm. Once the data is signed, the corresponding public key can be used to verify that the signature is valid.

In order to use public key cryptography to digitally sign a message, a sender first applies a hash algorithm to the message to create a message digest. The sender then encrypts the message digest with its private key to create personal signature.

Upon receiving the message and signature, a receiver decrypts the signature to recover the message digest and hashes the received message using the same hash algorithm that the sender used.

If the computed as well as the received message digest match, then, the receiver is assured of the integrity of the message.

### 3.1.6 Blind Signature

In cryptography, a blind signature as introduced by [Cha83] is a form of digital signature in which the sender disguises (blinds) the message before it is sent to the signatory for obtaining his/her signature. The blind signature scheme is normally used in applications where a sender is interested in authenticating the message from a signatory without revealing the message to the signatory. In many applications, involving anonymity, it is desirable to allow a participant to sign a message without knowing its content, e.g. E-voting, E-cash.

The steps involved in working of RSA [RSA78] based blind signature scheme are enumerated below:

1. The author of the message (m) computes the product of the message and blinding factor (r) to blind the message, i.e., $m' \equiv mr^{K_X} \pmod{n}$ . Here, the pair $(K_X, n)$ is the public key of a signer.

2. The blinded message($m'$) is passed to a signer, who then signs it. The signing authority calculates the blinded signature $s'$ as: $s' \equiv (m')^{K_X^{-1}} \pmod{n}$. Here, the pair $(K_X^{-1}, n)$ is the private signing key of the signer.

3. $s'$ is sent back to the author of the message, who can then remove the blinding factor using inverse $r^{-1}$ of $r$ to reveal $s$. The valid RSA signature of m is then represented as: $s \equiv s' \cdot r^{-1} \pmod{n}$.

4. The author of the message, can verify the correctness of the message against the signer's public key as: $m \equiv s^{K_X} \pmod{n}$.

### 3.1.7 Dual Signature

Dual signature is a cryptographic technique used to secure electronic transactions. It is used to link two messages that are intended for two different recipients.

In E-payment systems, the customer, wants to send the Order Information (OI) to the merchant and the Payment Information (PI) to the bank. The merchant need not know the customer's credit-card number, and the bank too need not know the details of the customer's order.

The customer is provided extra protection in terms of privacy by keeping these two items separate. However, the two items must be linked in a way that can be used to resolve disputes, if necessary. The link is needed so that the customer can prove that the given payment was intended for a particular order and not for some other goods or service [Sta00].

The Message Digest (MD) of the OI and the PI are independently calculated by the customer. The dual signature is the encrypted MD (with the customer's secret key) of the concatenated MD's of PI and OI. The dual signature is sent to both the merchant and the bank. protocol arranges for the merchant to see the MD of the PI without seeing the PI itself, and the bank sees the MD of the OI but not the OI itself.

### 3.1.8 Random Number Generation

Random number generation is integral to many cryptographic operations. Cryptographic random number generators must generate output that is computationally unfeasible to predict with a probability that is better than one half. Therefore, any method of predicting the next output bit must not perform better than random guessing. Cryptographic keys need to be as random as possible so that it is unfeasible to reproduce them.

Nonce is a randomly generated string which is only valid for a limited period of time. This is used in encryption protocols to prevent replay attack so as the server can check if that nonce

is valid, or expired.

### 3.1.9 Blockchain Technology

Blockchain is a data structure to create and share distributed ledger of transactions among a network of computers [CPVK16]. It allows users to make and verify transactions immediately without a central authority. It uses a peer-to-peer network of computers to validate transactions. Encrypted and distributed database doesn't allow changes to the data (ledger) once it is written, unless a consensus is achieved against it. Thus, it reduces the possibility of security breaches by even it's administrators. This makes blockchain invaluable for organizations trying to accomplish a secure system.

Firstly, blockchain can help eliminate paper. It can securely and permanently store all records, issue reliable certificates and awards, transfer credits and keep track of learning achievements across a whole lifetime. Secondly, in he blockchain all participants have ownership and control over their own data, schools and colleges. These two factors, would enable educational institutes to save money as the cost of data management as well as legal costs arising from liability issues would be 'significantly reduce'.

Documents like degree and course certificates can be secured and verified. This can be done regardless of whether a user has access to an institution's record-keeping system. Using blockchain, examinees and candidates can identify themselves online while maintaining control over the storage and management of their personal data.

## 3.2 Security issues in Summative Examination

Educational institutions invest huge amount of time and resources for the smooth conduct of the examinations. Inspite of a number of security measures, many loopholes and malpractices

breach the security of the examination system. Literature as well as the media is rife with evidence that examination malpractices have reached alarming levels [AWA10b, BGD+08]. The summative assessments, especially involving public examinations encompass collusion, impersonation, leakage of question papers, plagiarism, altering answer-books, misconduct in examination centre, approaching supervisors/ examiners, making false entries in the award list/ examination registers and issuing fake certificate/degrees, etc.

Any examination system needs to protect its two crucial assets, i.e., the question paper and answer-scripts. The secrecy of the question paper needs to be protected before the conduct of the examination. Similarly, the secrecy of the answer-scripts needs to be protected from all entities, except the examiner concerned. In addition to the secrecy, anonymity is also extremely important in an examination environment. Anonymity, refers to the state of being not identifiable to the communicating entities [PK01]. Anonymity needs to be satisfied between the following entities in a examination setup:

1. Examinee and paper setter (Examinee is not required to know who is the paper setter.).
2. Examiner and examinee (Examiner is not required to know whose answer-script he is evaluating.).
3. Examinee and examiner (Examinee should not know which examiner is evaluating which answer-script.).

In this section, we present a detailed overview of threats, countermeasures and vulnerabilities associated with the conventional/electronic assessments.

### 3.2.1 Question Paper/Answer-scripts Leakage

There are two rampant malpractices plaguing the summative examinations i.e., leakage of question papers and plagiarism of answer-scripts. A question paper is susceptible to leakage

due to over dependence on the manual system to complete the entire examination process. This involves, the question paper being exposed to several people. Examinees indulge in plagiarism, even in a supervised examination environment mainly due to the usage of a common question paper for a particular course paper.

**Countermeasures**

In the conventional examination, the predominant method used to control question paper leakage is to use three paper setters for setting 3 different question paper sets. Creating three sets of manuscripts of question paper ensures secrecy of the question paper from the paper setters themselves. The examination authority, then, randomly selects one set of question paper from the given three sets for a particular examination.

In E-assessment, question papers are generated JIT from an available question bank. The Public Key Infrastructure (PKI) is used for encryption of the question paper. Each examinee/group of examinees get(s) a unique question paper. At the end of the examination, examinees submit the encrypted answer-scripts corresponding to the question papers, to the examination authority. The examination authority in turn, sends the answer-scripts, in the encrypted form, to the examiners for evaluation.

**Vulnerabilities**

In a conventional examination, setting three unique and independent sets of question papers ensures question paper secrecy from the generator of the question paper, i.e., paper setters. However, there are many vulnerable points in the method, which can breach the confidentiality/secrecy of the question paper as given below:

1. All the 3 sets of question papers are verified by one subject expert.

2. The question paper selected is known during printing.

3. The selected question paper is seen by many people during the printing and the production phase.

4. The question paper is also exposed during the manual process of sealing of the question paper.

5. Advance transportation and delivery of question papers to the respective examination centres gives an opportunity for malpractice.

The answer-scripts written by the examinees go through a lengthy supply chain before reaching the examiner. The transportation of answer-scripts from one entity to another entity provides ample opportunities for coercion and cheating. There is a possibility that persons, involved in the question paper selection, printing, production, sealing and transportation, can leak the question paper.

In E-assessment, a question paper is generated JIT. Therefore, the problem of question paper leakage does not arise. However, there is a loophole. The answer-scripts are encrypted with the public key of the examination authority and are sent to the examination authority, who, can easily manipulate/ replace the answer-scripts.

### 3.2.2 Unauthorized Alteration

An unauthorized alteration in a question paper or in answer-scripts is possible during transportation/storage. It is not possible to detect any such alteration in the early stage as it is difficult to maintain and verify the trail log in real time.

**Countermeasures**

The integrity of the question paper in a conventional environment is ensured by submitting the sealed hard copy of the manuscript of the question paper. It is also mandatory to have the signature of the paper setter on each page of the manuscript along with initials on every modification carried out.

The integrity of the answer-scripts is achieved by default as examinees produce answer-scripts in their own handwriting. The handwriting acts as a deterrent for unauthorized modification of the answer-scripts. As a safety measure, examiners draw lines on blank portions of answer-scripts submitted by the examinees to prevent any additional matter being written later on.

The integrity of the question paper in E-examination is ensured by using digital signatures. The examination authority sends the encrypted question paper with the signed hash of the question paper. An examinee verifies the hash before answering the question paper. Similarly, the examinee sends the signed answer-scripts to the examination authority. The examination authority verifies the hash to ascertain the correctness of the answer-scripts. If any unauthorized modifications are carried to the question paper/answer-scripts, it can get detected immediately. Remedial action can then be taken.

**Vulnerabilities**

If any unauthorized modifications are carried in the question paper, it can come to the light only if the original paper setter sees the final question paper during the conduct of the examination. Even if it is detected, during the conduct phase of the examination the side effects are too many and too costly to revoke the damage caused.

However, it appears that answer-scripts tampering is comparatively easier to achieve for the supervisor/ examination authority /examiner. Supervisors can easily manipulate the answer-scripts submitted by the examinees. It is a matter of just drawing a line and cancelling a

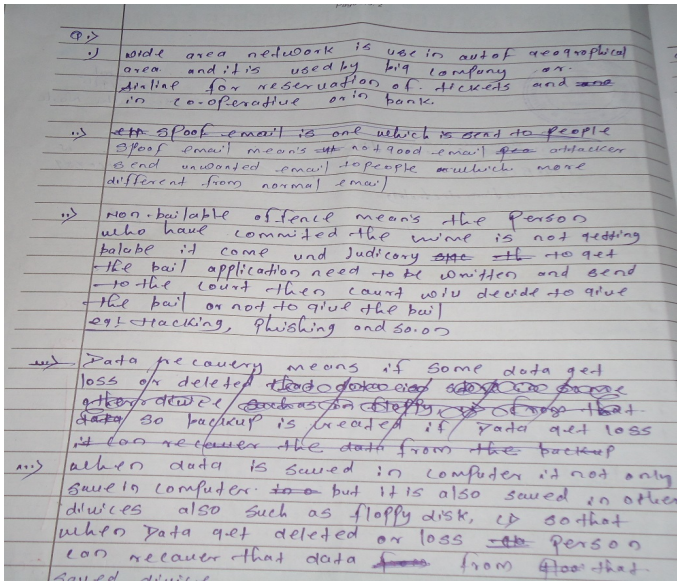


Fig. 3.1: Answer Script with Integrity Violation

particular answer or part of an answer in the answer-script. Then, the examiner will not assess those portions of the answer-scripts which are cancelled (refer Fig. 3.1). However, no evidence, whatsoever, exists to prove whether such an alteration/cancellation was done by the examinee or by somebody else.

In E-examination, tampering of question paper/answer-scripts is not possible for an unauthorized party assuming that encryption and digital signatures are unbreakable. However, as answer-scripts along with the examinee identity are available to the examination authority in an unencrypted form, there is a scope for alteration of the answer-scripts, especially, if examinee and examination authority coerce with each other.

### 3.2.3 Denial of Action

Any entity can deny its action, for taking perceived undue advantage of a vulnerability in the system. Therefore, it is extremely important to maintain non-repudiable evidence of actions performed by each entity during the conduct of an examination.

**Countermeasures**

The examination authority maintains examinee attendance records, for each course paper, during the examination. In some examinations, the supervisor also signs on the hall ticket/admission card carried by the examinee on each day of the examination. Thus, confirming the presence of the examinee for the particular course paper. Since, the answer-scripts submitted by the examinees are written in their own handwriting, there is no scope for examinees to disown their own answer-scripts.

In a fair and non-repudiable assessment system, both communicating entities maintain evidence to prevent the denial of action of the other party. The examination authority maintains the acknowledgment of the question paper sent by an examinee. The examinee maintains the acknowledgment of the answer-script sent by the examination authority. The hash value ascertains the receipt of exact content. In other words, both parties maintain Non-Repudiation of Origin (NRO) and Non-Repudiation of Receipt (NRR) to prevent each other from denying their action.

**Vulnerabilities**

If the practice of supervisor signing the hall ticket for confirming the presence of an examinee for a particular course paper is not followed, the examinee has no way of proving his presence for the said course paper. Also, the examinee is not provided with any documentary evidence to prove the exact copy of the submitted answer-script. In such a situation, if an answer-script is modified, there exists no evidence/protection for examinees to defend their case. Lack of signature plus attendance sheet can lead to a tricky situation, wherein an examinee who has not appeared for a paper can assert that he was present and had answered the paper.

In electronic exchange of information, if one party aborts the protocol before committing the receipt of the content, then, there is a risk of denial of receipt by the other party. In

E-examination also, if the examination authority aborts the protocol after receiving the answer-scripts and before sending an acknowledgment, then, the examinee is at a disadvantage. In reality, most assessment protocols favour the examination authority and are unfair to the examinees.

### 3.2.4 Favouritism, Coercion and Biased Evaluation

In summative examinations, an examinees' entire career or performance is at stake. In such an extremely challenging environment, examinees and other entities get involved in unfair means such as coercion, bribes, threats etc., to get an upper hand over others in an examination.

**Countermeasures**

If the identity of the communicating entities is hidden from each other, certain acts of favouritism, coercion, bias, threats, etc., can be controlled. In order to maintain the secrecy of the question paper and the anonymity of the paper setter, the practice of three different paper setters, setting three different question papers is followed. The identity of an examinee needs to be hidden from all till the completion of the assessment. This goal is usually achieved by hiding the Roll Number/ Seat Number recorded on the answer-book through a process called 'coding'. In the coding process, an examinees' identity is taken over by a pseudonym. The uncoding, i.e., revealing a examinees' identity is done during the declaration of the results. Examiners conduct evaluation in an impartial environment. Multiple examiners assess the answer-scripts pertaining to each course paper. Care is taken not to reveal the identity of the examiners.

The E-assessment with JIT generation of question paper from a large question bank addresses the issue of establishing paper setter anonymity. Anonymous mixnet servers are used to create pseudonyms to maintain anonymity of examinees and the examiners from each other. The

identity of the examinee is revealed only during the declaration of results. The identity of the examiner is always kept a secret.

**Vulnerabilities**

The manual process of appointment of a paper setter makes the identity of the paper setter known in advance. The coding process used for hiding the examinee identity from the examiner is naive and susceptible to disclosure of the examinees identity. If a single examiner assesses the answer-scripts, then the identity of the examiner is also known without any guesswork. Even in multi-examiner assessment, the manual appointment process reveals the identity of the examiners.

Mixnet servers successfully establish anonymity of the examinee and the examiner from each other. However, the process of generating pseudonyms through mixnet server is costly and unfeasible, especially due to the large number of examinees.

### 3.2.5 Plagiarism and Collusion

Public examinations, normally, comprise of a large number of examinees answering the examinations. In an examination system which has a single question paper, examinees easily tend to engage in the malpractice of collusion. This leads to plagiarism/copying/cheating.

**Countermeasures**

In the conventional assessment, where there is a common question paper, examinees get opportunities to engage in plagiarism. In conventional/E-examination, supervisors are appointed to control and monitor examinee behaviour and acts of collusion and plagiarism. However, with a common question paper for all examinees, supervised environment can do

very little to control acts of plagiarism. Answer-scripts plagiarism and collusion can be controlled to a great extent, if a unique question paper is provided to each examinee. In a system with unique question papers for each examinee/group of examinees, we require an unambiguous binding between a unique question paper and a examinees' answer-script.

The link between a unique question paper and an answer-script can be established by using a common question paper cum answer-book. A separate but identically labelled question paper and answer-book can also be used.

Another aspect which needs to be taken into consideration is to have a connect between the marks obtained and corresponding answer/answer-scripts. This would facilitate the verification and rechecking of both, whenever required. This connectivity is achieved in a conventional examination system by recording marks directly on the answer-scripts corresponding to the given answer. If somebody, later alters the assigned marks, it can be easily detected, when the answer-script is verified.

Leakage of question paper, plagiarism of answer-scripts and collusion can be controlled to a great extent, if a unique question paper is provided JIT to each examinee/group of examinees. In an E-assessment system with a unique question paper/papers for each examinee/group of examinees, the association between the examinee, a unique question paper and the answer-script/answer-scripts is ensured. This is done by pairing the question paper/question papers and answer-script/answer-scripts before sending it to the examination authority.

**Vulnerabilities**

One of the vulnerability of an examination is a common question paper being answered by several examinees at the same time. Dishonest examinees exploit this vulnerability and collude or plagiarize the answer-scripts of neighbouring examinees. Examinees get an opportunity for plagiarism/copying, even in a supervised environment. This occurs because

of a common question paper and large examination blocks. Therefore, supervisors find it difficult to supervise effectively. Neither the examination authority nor the examinee maintain any undeniable evidence which can prove the given answer-script is plagiarized or not. The same problem arises, when a separate answer-script is used with a unique question paper. Hence, it is not possible to fully endorse the claim of any of the communicating entities, in the event of any dispute.

An examinee can commit an intentional/unintentional error in recording Roll number/Seat number which results in two answer-scripts having identical Roll numbers. Similarly, two examinees who are hand-in-glove with each other, can write each others Seat numbers on the answer-books. This is done deliberately to benefit one of them.

The question paper and answer-script pair is available in an unencrypted form to the examination authority. They can dismantle the association. A more secure approach is to provide only the necessary part of the information to a party concerned, with the required level of associativity.

## 3.3 Tools for Formal Modelling and Verification

Security protocols need to be meticulously designed. Formal modelling and testing with analytical tools to verify the protocol security goals becomes mandatory. There are several formal methods and tools with a variety of features and sophistications for formal modelling and verification of security protocols. There are several options available under formal modelling, such as Applied $\pi$ calculus [AF01], Temporal logic [MP12], Petri nets [Mur89], Z notation [SA92], Communicating Sequential Processes (CSP) [RSGL01] and many more. Similalry, series of automated protocol verification tools are doucumented in literature, namely, AVISPA [ABB$^+$05], Athena [Son99], Scyther [Cre14], ProVerif [BAF08], Tamarin Prover [MSCB13] and Spin [Hol04] amongst others.

Applied $\pi$ calculus is one of the widely used language for describing and analysing security protocols with a great degree of sophistication. The calculus allows one to express several types of security goals and to analyse whether the protocol meets its goal or not. The analysis can sometimes be performed automatically, using the ProVerif software tool. Within this thesis, we used Applied $\pi$ calculus and ProVerif which is described in detail in the following subsections.

### 3.3.1 The Applied $\pi$ Calculus

The applied $\pi$ calculus [AF01] is a language for describing concurrent processes and their interactions.

The language is based on the $\pi$ calculus. It provides the rich syntax for detailing the actions of the participants and cryptographic operations used by security protocols. The syntax is coupled with formal semantics to allow reasoning about protocols. A wide variety of cryptographic primitives can be abstractly modelled by means of an equational theory. The applied $\pi$ calculus has been used to model security protocols in a variety of applications as follows:

1. Certified email [AB03]
2. Privacy properties [DKR09] and election-verifiability properties in electronic voting [KRS10]
3. Authentication protocols and key agreement [ABF07]

Properties of processes described in the applied $\pi$ calculus can be proved by employing manual techniques [Bla01] or by automated tools such as ProVerif [Bla04]

### 3.3.2 Protocol verification using ProVerif

ProVerif is a automatic cryptographic protocol verifier. In the formal model (so called Dolev-Yao model), a representation of the protocol properties are verified by Horn clauses [BAF08]. The salient features of ProVerif are:

1. It deals with several cryptographic primitives, including shared and public key cryptography (encryption and signatures), hash functions, and Diffie-Hellman key agreements.

2. It has a vast message space. Therefore, it can conduct an unlimited number of sessions of the protocol (even in parallel). This implies that the false attacks can be fed into the system by the verifier. When a property of the protocol cannot be proved by the tool, it executes a trace of the protocol, which falsifies the desired property as it is a recreating an attack.

3. ProVerif is sound, but not complete [BS18]. ProVerif's performance in proving correspondences, observational equivalence and reachability is sound. However, ProVerif may not be able to prove each and every property that holds.

ProVerif can prove the following properties:

1. Secrecy (the adversary cannot obtain the secret).

2. Authentication and more generally correspondence properties.

3. Strong secrecy (the adversary does not see the difference when the value of the secret changes).

4. Equivalences between processes that differ only by terms.

## 3.4 Intervention of Security Technologies in E-examination

Numerous electronic examination solutions have been designed with the intent of improving the efficiency of the current examination system and eliminating the loopholes associated with it. The Learning Management Systems (LMS) such as Moodle (https://moodle.org/), Blackboard (http://www.blackboard.com) include modules for conducting examinations. Firstly, there is a difficulty in adapting existing solutions per se, in conducting examinations as they do not correctly model the real standardized examination requirements (refer section 2.4.3). Secondly, the available solutions do not provide comprehensive security features covering security requirements of all the stakeholders. E-examinations to a great extent simplifies the entire examination process and offers many advantages over the conventional examination system.

In Chapter 1 (refer Table 1.1 ), we have identified a series of areas in the manual examination system which need security interventions. E-examination, to a great extent, simplifies the entire examination process and offers many advantages over the manual examination system. However, E-examination is susceptible to a variety of security issues. Nevertheless, the existing methods to combat the security threats, are not comprehensive.

### 3.4.1 Research Proposals Handling Security Issues

There are research proposals towards the deployment of the security goals as a solution for most data security issues in E-examinations. The confidentiality of data exchanged in E-examination is achieved using data encryption standards as introduced in [LCY+97, Wei05, CRHJDJ06]. There exists a proposal by [SB00] which uses hash functions to achieve integrity and authentication in E-examinations. One of the main security problem in online assessment is making examinees' submissions non-repudiable. The non-repudiation is achieved by [GTDPÁG06] through the use of digital signatures.

There is also an internet-based examination protocol that ensures authentication and conditional anonymity requirements with minimal trust assumption [GLB13].

Further, a formal framework in the applied $\pi$-calculus is devised to define and analyse authentication and privacy requirements for examinations through formalization of several individual and universal verifiability properties [DGK$^+$15].

There also exists a user security model for incorporating presence (and continuous presence),identity and authentication security goals against impersonation threats from examinees answering the E-examination [AWA10a, Apa10, AWA11].

An examination protocol called 'Remark!' is created that guarantees several security properties including anonymity for anonymising the examinee's test [HP10]

Most of the existing research work in this field focuses on the use of a common question paper for all the examinees answering a particular course paper in an examination. In order to address the issue of the leakage of question paper and plagiarism/copying acts by the examinees effectively, the use of multiple question papers appears to be a good solution.

The existing security approaches are insufficient when we use multiple question papers for each course paper. We need a mechanism to link the question paper answered by the examinee to the corresponding answer-script produced by the examinee unambiguously. It is also desired to keep the identity of examinee and corresponding question paper secret from the examiners assessing the answer-scripts. We also need to keep answer-scripts written by the examinees concealed from the examination authority for the purpose of better security.

Blind signature scheme have proven to be a very useful technique in applications requiring both anonymity and unforgeability, such as in e voting, e-cash and anonymous credentials. Blind signature scheme is used in e-voting to authenticate the voter without disclosing whom a voter votes for [Kuc10, IKSA03].

E-cash system based on partial blind signature, which allows the signer (the bank) to include

certain information in the blind signature of the coin, for example, the expiration date or the value of the coin [AF96].

A scheme based on cryptographic techniques such as ElGamal and blind signatures is proposed for maintenance of anonymity and double spender detection [ET11]. It is meant for untraceable off-line blind-signature-based electronic cash and possesses strong fraud control capabilities.

In the bitcoin transactions, blind signatures and Bitcoin transaction contracts (smart contracts) are used to ensure the anonymity and fairness [HBG16].

In the E-bidding scheme proposed by [FWSC13], each bidder's anonymity and all bids' unlinkability is guaranteed in order to prevent an auctioneer or any other party with malicious intention from bidding up prices of auctioned products.

In another E-auction protocol, the bidders' anonymity is maintained. There is also a new rewarding mechanism which enables winners to claim their reward without being linked to the data they contributed [DK15].

### 3.4.2 Remark! Protocol in E-examination

Remark is an electronic examination protocol for conducting summative examinations securely [GLR14]. The protocol participants are the Candidates (C), Examiner (E), Invigilator (G) and Manager (M). The role of the manager is to register eligible candidates and examiners for an examination. He has to conduct the examination to assign the test questions to the candidates and, once they have submitted their answers, to distribute the answered test to examiners and collect the marks. Finally, the manager notifies the marks to the candidates.

The examination process is broadly classified into four stages, i.e., registration, testing, grading and notification.

i) **Registration**

An eligible set of examinees and examiners are registered for the examination by issuing pseudonyms. Pseudonyms are generated by the exponentiation mixnets. The speciality of exponentiation mixnets is that each mixnet server blinds its entries by a common exponent value to provide anonymity for the candidates/examiners. A bulletin board is used to publish the pseudonyms, the questions, the tests,and the marks. The candidates/examiners can use zero-knowledge proofs to verify whether mixnet server behaves correctly and generates the correct pseudonyms.

ii) **Testing**

The manager generates the test questions and signs them with his private key. He then encrypts each test question with the help of a candidate pseudonym. In this phase, the manager authenticates each candidate. When all candidates have been authenticated, the manager publishes the encrypted test questions on the bulletin board. Once all the candidates have received their test questions, candidates are permitted to commence answering the test. At the end, each candidate submits his answer, which is signed with the candidates' private key and encrypted with the public key of the examination authority. The examination authority then, collects the test answer, checks the signature using the candidates' pseudonym, re-signs it. Then, he publishes its encryption with the corresponding candidates' pseudonym (as receipt).

iii) **Grading**

The examination authority encrypts the signed test answer with an eligible examiner pseudonym and publishes the encryption on the bulletin board. A designated examiner marks the test. The marks are appended to the signed test. The test answer and corresponding marks are signed by examiner with his private key. He then encrypts both

with the public key of the examination authority, and submits it to the examination authority.

iv) **Notification**

The manager receives the encrypted evaluation from the examiner. After, decryption and re-encryption the mixnet servers deanonymize the candidate's pseudonyms by revealing their secret exponents. The candidates anonymity is revoked, and the marks are registered. The examiners' secret exponent is not revealed to ensure his anonymity even after the examination concludes.

## 3.5 Summary

Cryptography plays a key role in securing E-examination systems. An investigation into the current research and practices in the areas of E-examination and other E-business applications was imperative. This enabled identification and summarization of the current methodologies used to overcome some of the security challenges in E-examination. Although, E-examinations solve many of the intricacies of conventional Paper/Pen examinations, yet they bring to the forefront several vulnerabilities totally unknown previously. In an endeavour to integrate further security mechanisms into the E-examination system, security protocol specification and verification tools, namely, applied $\pi$ calculus and ProVerif were investigated.

E-examination systems need to be leveraged with the state-of-the-art security technologies, inorder to minimize the possibilities of malpractices and to enhance the effectiveness of the whole process.

# CHAPTER 4

# Security Protocols in E-examination against Malpractices of Collusion/Plagiarism

The intensity and pervasiveness of malpractices in the public examinations are on the rise and is taking a shape of thriving business with the emergence of new and techno ingenious ways of cheating. Malpractices are perpetrated by examinees, examination staff and other external agents before, during and even after the examinations. Appropriate measures to deter and detect such malpractices are essential to uphold the fairness and integrity of the examination system. The menace of malpractices such as question paper leakage and collusion/plagiarism can be curbed to a great extent by generating a unique question paper JIT for each examinee/group of examinees. However, if a unique question paper is provided to each examinee /group of examinees, we require an adequate support and security service for linking the question paper and answer-script unambiguously, along with the regular security features.

In this Chapter, first we define a cryptographic scheme for achieving anonymity between the examinees and the examiner based on the concept of blind signature. Then, we define question paper and answer-scripts delivery protocols for establishing an inseparable link between the exchanged question paper and the answer-scripts between the examination authority and the examinees. We use a ProVerif tool/manual proofs to gauge the correctness of the proposed cryptographic scheme and answer-scripts delivery protocol.

This Chapter is structured as follows: In section 4.2, we discuss the threat model defining the goals of stakeholders, capabilities of adversaries and measures to tackle the adversaries. Section 4.3 describes the mechanism adopted in implementing the desired security requirements along with the overview of the proposed solution. Section 4.4 describes the cryptographic scheme, based on the concept of blind signature. Section 4.5 models the proposed protocols using applied $\pi$ calculus. Finally, section 4.6 provides the formal specification and analysis to validate the question paper/answer-scripts exchange using ProVerif and manual proof.

## 4.1 Introduction

Public examinations are often a target of a plethora of academic misconducts and malpractices as it is looked upon as stepping stone to success as well as to higher qualifications [TERS05]. The intensity and pervasiveness of the problem of malpractices can be gauged from the fact that apart from the examinees, it is also the strong nexus between other involved stakeholders along with the external agents. The malpractices in public examinations encompass collusion, impersonation, leakage of question papers, plagiarism, altering answer-books, misconduct in examination centre, influencing supervisors / examiners, making false entries in the award list/ assessment registers and issuing fake certificate/degrees, etc. [Eck03, Mah11].

Institutions seek succour from a number of remedial approaches at different stages of examination to control human errors and malpractices. Some of the well-established policies in controlling evaluation anomalies are discussed in section 5.3.1 of Chapter 5. The electronic assessment has the potential to curb most of the shortcomings associated with conventional assessment. Some of the techniques for controlling the malpractices related to conventional/electronic assessment are discussed in section 3.2 of Chapter 3. The security practices currently adopted in conventional / electronic assessments are insufficient to handle

all the security concerns. This is apparent from ever increasing cases of malpractices and successful breaches of the assessment security. The occurrence of frequent cases of malpractices as reported and recorded in the public domain and the security analysis of conventional/electronic assessment system reveals that the current assessment system needs to be streamlined with effective security interventions.

Malpractices such as question paper leakage and rampant collusion/plagiarism can be controlled to a great extent by generating a unique question paper, JIT. If the unique question paper is provided to each examinee/group of examinees, we require a mechanism to establish an unambiguous link between the examinee identity and the question paper. It is also necessary to associate the unique question paper received by the examinee to the corresponding answer-script produced by the examinee unambiguously. The established association needs to be strong enough to prevent both the sender and the receiver from denying their action in the future. The bonding of the unique question paper and the answer-script, needs to be done in such a way that, it satisfies the following security requirements:

1. The answer-script submitted by an examinee is kept hidden from the examination authority. The secrecy requirement of answer-script is necessary to prevent any fraudulent act leading to violation of answer-script integrity without getting detected.

2. The answer-script submitted by the examinee is available to the examiner, but the identity of the examinee is hidden from the examiner. This is required to prevent any dishonest act such as unfair evaluation, illicit demands, threats to obtain higher grades, etc.

The first requirement listed above, refers to the anonymity property of hiding the identity of sender/receiver from each other [PK01]. Along with the anonymity, we also need to keep the transmitted information secret from the intermediate receiver (second requirement). In

a nutshell, we need a dual purpose approach, satisfying both anonymity and confidentiality. The current approaches address anonymity requirement comprehensively, but lack the ability to maintain secrecy of answer-scripts from the examination authority.

In this Chapter, we present a cryptographic scheme to achieve the following goals:

1. De-link the receiver of the message from the sender of the message (to achieve anonymity).

2. Keep the message secret from the intermediary receiver (to achieve confidentiality).

We present a mathematical proof of the proposed cryptographic scheme to validate and support our claim. Secondly, we present a novel approach for linking the question paper and answer-script associated with the examinee and revealing only a selective and essential part of the aggregated information to the recipient. We present a protocol based on the concept of digital signature, blind signature and dual signature for establishing an unbreakable association between the question paper and the answer-script. We prove that the adversary is not able to obtain any significant information about the aggregated data (i.e., question paper-cum-answer-script) and break the association between question paper and answer-script without getting detected.

## 4.2 Threat Model

The two important assets of any examination system are the question papers and the answer-scripts. These assets need to be protected from all the entities who entertain fraudulent intent. A large number of examinees answering a particular question paper in a single examination block provide an ample opportunities for examinees to get involved in unfair means and collusion/plagiarism. In this section, first we discuss the current

answer-script delivery process in conventional/electronic examinations in order to understand the threats and vulnerabilities faced by the current examination system. In order to counter the acts of collusion/plagiarism, we propose the use of unique question paper for each examinee/group of examinees answering the examination.

### 4.2.1 Current Answer-scripts Delivery Process

In the current conventional/electronic examination environment, the answer-scripts delivery process is a communication between three entities (see Fig. 4.1).

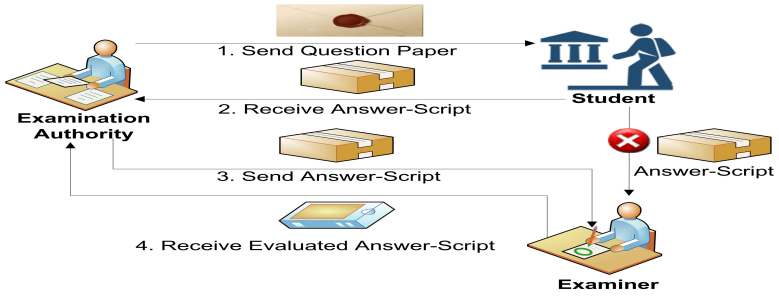



Fig. 4.1: Process of Question paper/Answer-script Exchange

The steps involved in exchange of question paper/answer-script between examination authority, examiners and examinees are listed as follows:

1. An examinee answers the examination question paper provided by the examination authority.

2. The examinee on completion of answering the examination question paper, submits the answer-script back to the examination authority (This step is required for anonymity of examinees and examiners).

3. Examination authority verifies the received answer-scripts and then forwards the answer-scripts to examiners for evaluation (This step leads to exposure of answer-scripts to many entities before evaluation).

4. The examiner returns the evaluated answer-script back to the examination authority for compilation of the final result.

### 4.2.2 Security Goals of Stakeholders

In an examination environment with unique question paper per examinee/group of examinees, we need to address the following security concerns of the entities concerned, namely, examinees and examination authority.

1. The unique question paper provided to the examinee and the answer-script submitted by the examinee need to be linked together securely.
2. The answer-script submitted by a examinee shall not be available to any person, other than the examiner concerned.
3. The identity of the examinee and answer-script produced by the examinee shall not be available together to any person (other than the examinee).
4. Provide a Non-repudiation of Receipt Service (NRR) to protect examinees from the examination authority's denial of answer-script receipt.
5. Build a Non-repudiation of Origin Service (NRO) to protect the examination authority from an examinee's denial of the answer-script origin.

### 4.2.3 Adversary Capabilities

Dolev-Yao threat model is the most widely accepted model to analyse the security protocols [DY83]. According to Dolev-Yao threat model, the capabilities of adversary include, message interception, message insertion and message alteration. However, these adversary capabilities are pertaining to the messages in transit. We also need to protect the data at the endpoints

from the internal entities, behaving as adversaries. Such adversaries possess the capability to carry deliberate insertion/manipulation of data before/after transit. Some specific internal adversary capabilities, we intend to tackle are listed below:

- **Threat Scenario 1 (TS 1) - Unauthorised alteration of answer-scripts**

  Examination authority alters the answer-script of examinee before delivering it to the examiner for evaluation. In such a scenario, there is no mechanism to detect and correct the fraudulent action of the examination authority.

- **Threat Scenario 2 (TS 2) - Coercion with examiner**

  Examinee E1, knows the examiners identity. He coerces with the dishonest examiner after the examination conduct phase. The examiner favors the said examinee E1 by evaluating the answer-scripts in an unfair manner.

- **Threat Scenario 3 (TS 3) - Plagiarism**

  Examinee E1 has the question paper QP1. He blindly copies the answers from the neighbouring examinee E2 answering question paper QP2. On detection of this malpractice, examinee E1, can claim that he had received question paper QP2 and not QP1. On the other hand, the examination authority can claim that the examinee E1 answered question paper QP1 when in reality that examinee answered question paper QP2.

- **Threat Scenario 4 (TS 4) - Coercion with examination authority**

  Examinee E1, coerces with dishonest examination authority after the examination conduct phase. The examination authority allows examinee E1 to alter/replace the submitted answer-script.

The answer-scripts delivery process described in section 4.2.1 suffers from several vulnerabilities that a dishonest entity can target to jeopardise the fairness and reliability of

examination. A system is insecure, if an attacker is able to exploit a vulnerability and compromise the integrity of the asset. Therefore, there is a need for an effective security mechanism to handle the threat scenarios.

Dolev-Yao threat model is the most widely accepted model to analyse the security protocols [DY83]. According to Dolev-Yao threat model, the capabilities of adversary include:

1. **Message Interception**

   An adversary can obtain any message passing through the network.

2. **Message Insertion**

   An adversary can send messages to any principal by impersonating another principal.

3. **Message Alteration**

   An adversary can alter the messages.

In addition, two or more stakeholders can collude with each other to disrupt the system. A dishonest entity may attempt to explore the vulnerabilities, through the defined capabilities of the adversaries as described above. The attacks would be considered successful, if the protocol fails to detect the attacks and grants undue benefit to the dishonest entity.

### 4.2.4 Adversary Counter-attack Requirements

The examination system must be strong enough to sustain the attacks by an adversary described in the examination threat model. The adversary can be handled effectively, if the following requirements are ensured:

- **Security Requirement 1 (SR 1) - Confidentiality of Question paper/ answer-scripts**

  The question papers and answer-scripts form the most crucial and confidential

documents of examination. It is extremely important to protect the confidentiality of question papers and answer-scripts.

- **Security Requirement 2 (SR 2) - Question Paper and answer-script binding**

  Herein a unique question paper needs to be provided to each examinee/group of examinees', which, then needs to be bonded securely to the answer-script submitted by each examinee. This is required to build a reliable evidence to settle any dispute arising out of plagiarism/collusion.

- **Security Requirement 3 (SR 3) - Answer-script hiding**

  The answer-scripts submitted by the examinees should be encrypted to prevent the examination authority to access the answer-scripts. Also, ensure that the examinee's identity and answer-script is not available together in any form to the examiner. This is required to prevent any coercion between examinees and examination authority/examiner with dishonest intent.

- **Security Requirement 4 (SR 4) - Question/Answer pair re-sequencing**

  Provide the question/answer pair in an altered sequence to the examiners for evaluation so that, the examinee identity is not revealed from the sequence in which questions/answers received by the examiner. This is to prevent collusion between the examiner and his favoured examinees.

- **Security Requirement 5 (SR 5) - Untraceability**

  The identity of the examiner should remain secret at all times. In other words, the identity of the examiner evaluating specific answer-scripts should not be traceable. Similarly, examiners should not be aware of whose answer-scripts they are evaluating.

- **Security Requirement 6 (SR 6) - Evidence for non-repudiation**

  Provide each communicating entity with a non-repudiable evidence, for dealing with any dispute.

A dishonest entity may attempt to explore the vulnerabilities, through the defined capabilities of the adversary. The attacks would be considered successful, if the protocol fails to detect the attacks and grants undue benefit to the dishonest entity. It is essential to enforce the security requirements defined herein to control most of the malpractices. In the next section, we provide the overview of the security solutions and how the solutions address the threats discussed in section 4.2.

## 4.3 Implementation of Security Requirements

The examination system needs to be secured from all the entities with illicit intention. Large number of examinees answering a particular question paper in a single block provide ample opportunities for examinees to get involved in unfair means, specifically plagiarism. In order to counter the answer-script plagiarism acts of examinees, we propose the use of a unique question paper for each examinee answering the examination. However, the use of multiple question papers for each course paper, functionally affects the entire security process. In this section, we give an overview of the proposed security solution for achieving the required security goals.

### 4.3.1 Assumptions

The proposed solution is on the pretext of the following assumptions:

**Assumption 1**

This is an abstract model and as such we have not considered requirements about hardware, software and networking components. However, the examination model being built is in tune and feasible with the existing available technology.

**Assumption 2**

The examinees answer the subjective examination electronically.

**Assumption 3**

The question paper is generated JIT from an existing question bank.

**Assumption 4**

The examination authority conducting the examination has no right to access the answer-scripts produced by the examinees.

**Assumption 5**

Each examinee/group of examinees is/are provided with an identical set of questions, but arranged in a unique order.

**Assumption 6**

The examiners evaluate subjective answer-scripts through a computer-aided system.

**Assumption 7**

Evaluation involving large number of answer-scripts needing multiple examiners for each subject/course paper.

**Assumption 8**

Around 10% of the answer-scripts corresponding to each subject/course paper are independently evaluated by all the subject examiners.

### 4.3.2 Basic Notations

The elementary notations used to describe the proposed cryptographic scheme and answer-scripts delivery protocol is listed in Table 4.1.

Table 4.1: Glossary of notations

| Notation | Description |
|---|---|
| $K_{A_i}, K_{A_i}^{-1}$ | Public key and private key of an entity $A_i$ |
| $SK_{A_i}$ | Secret symmetric key of entity $A_i$. |
| $(m)K_{A_i}$ | Message m is encrypted using public key of entity $A_i$ |
| $(c)K_{A_i}^{-1}$ | Cipher text c is decrypted using the private key of entity $A_i$ |
| $\mathscr{H}(m)$ | One way hash of m |
| $N_B$ | Unique random number called Nounce, generated by an entity B. |
| $r,\ r^{-1}$ | Random blind factor and its corresponding inverse. |

### 4.3.3 Basic syntax and semantics of applied $\pi$ calculus in defining security properties

Applied $\pi$ calculus offers reach repository of grammar to represent the terms and terminologies required in defining security properties. This section defines the essential syntax and sementics of applied $\pi$ calculus along with the examples to illustrate its usage.

**Definition 1.** ***(Term).*** *Given an infinite set $\mathscr{X}$ of variables and an infinite set $\mathscr{N}$ of names (used to represent atomic data, such as channel names, keys, nonces, or identities), the set of terms of the signature $\Sigma$, consisting of a finite set of function symbols, the set of finite* terms *is denoted by $\mathscr{T}(\Sigma, \mathscr{X}, \mathscr{N})$ and is inductively defined as names, variables, and function symbols applied to other terms.*

$$\begin{aligned} &\mathscr{L}, \mathscr{M}, \mathscr{N}, \mathscr{T}, \mathscr{U}, \mathscr{V} ::= && terms \\ &\quad a, b, c, ...., k, ..., m, n && names(\mathscr{N}) \\ &\quad x, y, z && variables(\mathscr{X}) \\ &\quad f(M_1, ..., M_l) && function \end{aligned}$$

*where $f$ ranges over the function signature $\Sigma$ and $l$ matches the arity of $f$*

**Example 4.3.1.** *In the context of security protocols, a standard signature is $\Sigma = \{senc, aenc, adec, pair, K_A\}$ where, senc, aenc, adec and pair are three symbols of arity*

*2, representing respectively symmetric encryption, asymmetric encryption/decryption, and concatenation, while $K_A$ is a symbol of arity 1, representing the public key associated with some private key.*

**Example 4.3.2.** *The term $t_0 = aenc(pair(a,n_a),K_A(k_a))$, where $a,n_a,k_a \in \mathcal{N}$, represents the encryption under the public key $K_A(k_a)$ of the concatenation of the identity a together with the nonce $n_a$.*

Further, applied $\pi$ calculus uses processes to define the communication between the communicating agents.

**Definition 2.** ***(Process).*** *Processes are defined as follows:*

| | |
|---|---|
| $P,\ Q,\ R ::= 0$ | *null process* |
| $P \mid Q$ | *parallel composition* |
| $!P$ | *replication* |
| $vn.P$ | *name restriction*(”*new*”) |
| $if\ M = N\ then\ P\ else\ Q$ | *conditional* |
| $u(x).P$ | *message input* |
| $\overline{u} < N > .P$ | *message output* |
| $\{M/x\}$ | *active substitution* |

*We write $\{M/x\}$ for the substitution that replaces the variable x with the term M.*

**Definition 3.** ***(Inference rule).*** *An* inference rule *is a rule of the form:*

$$\frac{u_1,\ldots,u_n}{u}$$

*with $u_1,\ldots,u_n,u$ are terms (with variables). An* inference system *is a set of* inference rules.

**Example 4.3.3.** *From the key k and the message senc(m, k), which represents the (symmetric) encryption of m over k, one can compute m. This can be represented by the rule:*

$$\frac{senc(m,k) \;\; k}{m}$$

*This inference rule states that given a message m encrypted with the secret key k and secret key k, it is possible to infer message m.*

**Definition 4.** ***(Frame).*** *A* frame *is an expression* $\varphi = \upsilon\tilde{n} \cdot \sigma = \upsilon\tilde{n} \cdot \{M_1/x_1, ..., M_n/x_n\}$ *where* $\sigma$ *is a substitution and* $\tilde{n}$ *is a set of names that are restricted in* $\varphi$*. The terms* $M_1, ..., M_n$ *represent the attacker knowledge while the names in* $\tilde{n}$ *are initially unknown to the attacker.*

Frames are static knowledge exported by a process to the execution environment. Attacker or environment learns these values.

**Example 4.3.4.** $\varphi = \upsilon\tilde{k} \cdot\ senc(m,k)/x$

*In this example, frame* $\varphi$ *represent encrypted value of m exposed to environment and hence to the attacker.*

**Definition 5.** ***(Equational Theory).*** *Let* $\Sigma$ *be a function signature. An* equational theory $\approx$ *is set of equations* $u = v$*, where u and v are terms in* $\mathscr{T}(\Sigma, \mathscr{X}, \mathscr{N})$

**Example 4.3.5.** *Let* $\Sigma = \{fst,\ snd,\ pair,\ aenc,\ adec,\ sign,\ checksign,\ blind,$ $unblind,\ hash\}$ *corresponding to first, second projection, concatenation, asymmetric encryption and decryption, sign, checksign blind and unblind signature and hash calculation. The properties of concatenation and standard encryption and blind signatures are modelled by the following set of equations:*

$fst(pair(x,y)) = x$

$snd(pair(x,y)) = y$

$adec(aenc(m,K_A),K_A^{-1}) = m$

$sdec(senc(x,k),k) = x$

$checksign(sign(m,K_A^{-1}),K_A) = m$

$unblind(blind(m,rbf),rbf^{-1}) = m$

$unblind(sign(blind(m,rbf),K_A^{-1}),rbf^{-1}) = sign(m,K_A^{-1})$

$unblind(aenc(m,blind(K_E,rbf)),rbf^{-1}) = aenc(m,K_E)$

**Example 4.3.6.** *Let $\Sigma = \{enc, dec\}$, where enc and dec are each of arity 2. Suppose a, b, c are names (perhaps representing some bitstring constants or keys), and x, y,z are variables. Then $enc(a,b)$ represents the encryption of a using the key b. The term $dec(enc(a,b),y)$ is also a term, representing the decryption by y of the result of encrypting a with b. The symbols enc and dec may be nested arbitrarily.*

Equational theories are the means by which cryptographic operations are represented. We do not model the mechanisms (whether bitstring manipulation or numerical calculation) that constitute the cryptographic operations. Rather, we model the behaviour they are designed to exhibit. Thus, stipulating the equation $dec(enc(x,y),y) = x$ models symmetric encryption.

**Definition 6.** ***(Deduction).*** *A term t is deducible from a frame $\varphi = \upsilon\tilde{n}\cdot\sigma$, if it can be deduced using $\varphi$ and any name that does not occur in $\tilde{n}$ . More formally, given an equational theory $\approx$ and a frame $\varphi = \upsilon\tilde{n}\cdot\sigma$, we write $\varphi \vdash_{\approx} t$*

### Static Equivalence

Static equivalence is introduced to define the ability of the observer to compare messages. The observer gets the data once and then conducts experiments to verify the distinguishability between the observed messages. Static equivalence is used to model the notion whether an attacker/observer can distinguish between two sequences of messages.

**Definition 7.** ***(Statistical Equivalence).*** *Two frames* $\varphi 1$ *and* $\varphi 2$ *are statically equivalent w.r.t. an equational theory* $\approx$*, denoted* $\varphi 1 \approx_s \varphi 2$*, if Dom(*$\varphi 1$*) = Dom(*$\varphi 2$*) and for any two terms* $M, N$ *we have that* $(M \approx N)\varphi 1$ *if and only if* $(M \approx N)\varphi 2$*.*

**Example 4.3.7.** *Let* $\approx$ *be the equational theory of encryption as defined in Defn.5.5. Let* $\varphi 1 = \{0/x, 1/y\}$ *and* $\varphi 2 = \{1/x, 0/y\}$*. Then* $\varphi 1 \not\approx_s \varphi 2$*. As* $(x = 0)\varphi 1$ *while* $(x \neq 0)\varphi 2$ *.*

**Example 4.3.8.** *Let* $\approx$ *be the equational theory of encryption as defined in Defn.5.5.*
*Let* $\varphi 1 = vm.\{m/x\}$ *,* $\varphi 2 = vn.\{hash(n)/x\}$ *,* $\varphi 3 = \{m/x\}$ *and* $\varphi 4 = \{hash(m)/x\}$*.*

$\varphi 1 \approx_s \varphi 2$*, since* $(x = m)\varphi 1$ *and* $(x = m')\varphi 2$ *here* $m' = hash(n)$*, where* $m$ *and* $m'$ *are indistinguishable to the observer.*
*whereas*
$\varphi 3 \not\approx_s \varphi 4$*, since* $(x = m)\varphi 3$ *and* $(x \neq m)\varphi 4$ *. Here* $\varphi 3$ *and* $\varphi 4$ *are distinguishable to the observer.*

**Observational Equivalence**

Two processes are observationally equivalent if they cannot be distinguished by an attacker. We write A $\Downarrow$ a when A can send a message on a, that is, when A $\rightarrow * C[\overline{a} < M > \cdot P]$ for some evaluation context $C[\_]$ that does not bind a.

**Definition 8.** ***(Observational Equivalence).*** *Observational equivalence (*$\approx o$*) is the largest symmetric relation* $\mathscr{R}$ *between closed extended processes with the same domain such that* $A \mathscr{R} B$ *implies:*

1. *if* $A \Downarrow a$*, then* $B \Downarrow a$*;*
2. *if A* $A \rightarrow * A'$ *, then* $B \rightarrow * B'$ *and* $A' \mathscr{R} B'$ *for some* $B'$*;*
3. $C[A] \mathscr{R} C[B]$ *for all closing evaluation contexts* $C[\_]$*.*

**Example 4.3.9.** *Let $\approx$ be the equational theory as defined in Defn. 5*

*Consider the process $\mathscr{A} = \overline{c} < enc(s_0,k) >$ and $\mathscr{B} = \overline{c} < enc(s_1,k) >$*

*We have that $vk.\mathscr{A} \approx_o vk.\mathscr{B}$. The attacker cannot distinguish between the encryption of two known values $s_0$ and $s_1$ where the encryption is by a secret key k. Technically, there is no context C which can make a distinction between them, eg. by taking some observable action based on the information made available by these processes.*

*On the other hand, if the key k is available to the attacker, we have $\mathscr{A} \not\approx_o \mathscr{B}$,*

*since the context $C[\_] = c(x).\ if\ sdec(x,k) = s_0$ , distinguishes $\mathscr{A}$ and $\mathscr{B}$*

### 4.3.4 Overview of Proposed Solution for Electronic Answer-scripts Delivery

The answer-scripts delivery is the communication between the following 3 entities.

1. The examinee - the producer of the answer-script.
2. The examination authority - the intermediary accepting the answer-scripts submitted by the examinee.
3. The examiner - the consumer of the answer-scripts sent by the examination authority.

We need a security solution that comply to the security requirements of all the stakeholders concerned (refer section 4.2.4). The proposed steps for achieving the security requirements in a transmission of single answer-script on the basis of hybrid cryptosytem are as follows:

1. The examination authority $C$ starts by selecting public key $K_X$ of the examiner $X$. Examination authority takes $K_X$ and chooses a random number $r$ producing the disguised public key $s'$ of examiner. Disguised public key is used to hide the identity

of examiner from examinee. Examination authority then encrypts $s'$ using the public key, $K_{E_i}$ of the examinee to produce $m'$.

2. The examination authority, sends the encrypted disguised public key, $m'$ to the examinee for encrypting the answer-script produced by it.

3. The examinee decrypts $m'$ using its private key, $K_{E_i}^{-1}$ to produce $s'$.

4. The examinee upon completion of answer-script $(AS)$ corresponding to the question paper $(QP)$, computes message digest of question paper and answer-script. The examinee combines message digest of question paper and answer-script and signs it with its private key to produce dual signature.

5. The examinee pairs the question paper, message digest of answer-script and dual signature and encrypts it using the public key of examination authority. Similarly, examinee select one secret key $SK_{E_i}$ and encrypt the answer-script and also encrypt the $SK_{E_i}$ and message digest of $SK_{E_i}$ pair using the disguised public key $s'$ to produce $c'$. All these details are sent to the examination authority.

6. The examination authority on receipt of the encrypted secret key $c'$, applies $r^{-1}$, the inverse of $r$ to generate $c$, the secret key encrypted using the public key of the examiner.

7. The examination authority, subsequently sends to the examiner, the answer-scripts encrypted with the secret key along with $c$.

8. The examiner first decrypt $c$ by using his private key, $K_X^{-1}$ to get secret key $SK_{E_i}$. The examiner can then decrypt answer-scripts using the secret key.

In this way, although the examinee gets the public key of the examiner, he/she is not in a position to get the identity of the examiner. Similarly, although the examination authority gets the answer-script from the examinee, it cannot view the actual answer-script as answer-scripts are encrypted using the secret key which is unknown to the examination authority. We have illustrated the working of disguised public key in Appendix A.

### 4.3.5 Safeguards against Adversary Capabilities

During the exchange of question paper and answer-scripts between examination authority, examinee and examiner, the honesty of no communicating entity can be guaranteed. In addition to alteration, insertion and interception capabilities of these entities, they also possess colluding and copying capabilities. These attacks are deemed to be successful only if protocol fails to detect any such attack and thus providing an illicit benefit to the dishonest entity.

#### Protection of Confidentiality and Integrity of Answer-scripts

An examinee submits the answer-script to examination authority by encrypting it with the secret key. Secret key is encrypted using the disguised public key of the examiner. Examination authority in such a scenario is not in a position to decrypt the answer-script as it needs private key of examiner concerned to decrypt the secret key. This ensures that the examination authority neither can see the answer-scripts not it can violate the integrity of the answer-scripts submitted to it.

#### Anonymity of Examiners and Examinees

Examination authority provides disguised public key of examiner to the examinees. The definite inference of examiner identity is not possible for examinees with this arrangement. As

examiner identity remains hidden/ anonymous from examinees, it is not possible for examiner and examinee to engage in coercion.

**Associativity of Question paper and Answer-scripts**

The linking of unique question paper received by the examinee and the corresponding answer-script created by the examinee is done with the help of dual signature. This bonding between question paper and answer-script is created to establish an undeniable evidence of question paper and answer-script related to each examinee. As each examinee is provided with an unique question paper, the attempt of copying/collusion with neighbouring examinee will result into altogether poor performance of examinee concerned. If such an examinee after committing a malpractice complains of unfair evaluation can be proved at fault in the presence of undeniable associativity of question paper and answer-script.

### 4.3.6 Shortcomings of proposed security solution and possible enhancements

The proposed solution involving electronic answer-scripts delivery appears to provide comprehensive security solution to the security requirements identified in section 4.2.4. However, the proposed solution suffers from few infirmities as listed in table 4.2.

Table 4.2: Shortcomings of proposed security solution and possible enhancements

| Shortcomings | Possible Solution |
|---|---|
| Solution is not suitable for hiding the identity of an examiner in the evaluation of answer-scripts involving a single examiner. | In the evaluation of answer-scripts involving a single examiner, hiding the identity of the examiner is a challenging task, especially, if the human element is compromised. |
| Intentional/un-intentional corruption of the answer-scripts at the recipient end. | Underlying network usually take care of transmission errors. The integrity violation of answer-scripts occurring at the recipient end can be sorted out by ascertaining retransmission of the answer-scripts. |
| Need to decide the examiner before the conduct of examination. | Usually, paper setters are the examiners, so the examiners of the answer-scripts are also known in advance. However, when there are large number of answer-scripts to be evaluated, then it is mandatory to appoint additional examiners. Therefore, adequate provision has to be made for accommodating them in the system at the pre-conduct stage of examination. |
| If an examiner dies, rejects or is removed from the evaluation panel then, assigning the answer-scripts for evaluation to a different examiner is technically difficult as the key required to decrypt answer-scripts are encrypted using the public key of examiner concerned. | Usually, in such type of computer-aided evaluation, institutions provide the electronic device to the examiners. If due to some reason the examiner concerned is unable to evaluate the answer-scripts, then, the examiner need to return back the device along with the keys and authentication details. This arrangement will ensure that evaluation will not suffer due to unavailability of any examiner for evaluation. |
| Additional computational overhead is involved in disguising the key and for the construction of dual signature. | As keys are normally of limited size, disguising it with the random factor will only add fractional overhead. Similarly, dual signature mechanism is a simple operation of addition of two hash values and then signing it. Such mechanisms are frequently used in financial transactions for added security. In other words, the proposed solution is computationally not heavy and is likely to increase the overhead marginally. |

## 4.4 Blind Signature based Cryptographic Scheme for Anonymity

In this section, we describe the proposed cryptographic scheme implementing the disguised public key in detail. We first define cryptographic primitives as an equational theory for modelling answer-script delivery protocol. We introduce a novel signature, namely, 'hide/unhide' based on the blind signature scheme to disguise the public key of the examiner from the examinees. We, then model the dual purpose cryptographic scheme based on the concept of blind signature for achieving anonymity and confidentiality.

### 4.4.1 Equational Theory

We use the following predicates for achieving the security requirements as described in subsection 4.2.4. The proposed cryptographic scheme is based on RSA public key cryptosystem [RSA78] and blind signature [Cha83]. We adopt the following signature functions to capture the cryptographic primitives used by the proposed protocol.
$\Sigma = \{aenc, adec, sign, checksign, blind, unblind, hide, unhide\}$

1. An encryption function *aenc* and the corresponding inverse *adec*, such that

$$adec(aenc(m, K_X), K_X^{-1}) = m \tag{4.1}$$

   The function aenc/adec represent asymmetric encryption / decryption as defined in the Public Key Infrastructure (PKI).

2. Message signing function *sign* and the corresponding inverse *checksign*, such that

$$checksign(sign(m, K_X^{-1}), K_X) = m \tag{4.2}$$

The function sign is used to sign the message with the private key of the sender and the function checksign is used to verify the integrity of the received message.

3. The message blinding function, *blind* and the corresponding inverse *unblind*, such that

$$unblind(blind(m,r),r^{-1}) = m \tag{4.3}$$

The function blind/unblind represents a blind signature scheme[Cha83]. In this, r is a random blind factor used to blind the message and corresponding inverse $r^{-1}$ is used to unblind the message.

4. The message blinding and unblinding function as defined in item 3 above and the message signing function *sign* as defined in item 2 above, such that

$$unblind(sign(blind(m,r),K_X^{-1}),r^{-1}) = sign(m,K_X^{-1}) \tag{4.4}$$

This predicate is used to obtain the signature of the authority on the message without revealing the message to the receiver authority. E.g. In e-voting the signature of authority is required on the vote without revealing to whom the vote is cast.

Along with the above predicates, we propose the following additional predicates in the equational theory:

5. A disguising function *hide* and the corresponding inverse *unhide*, such that:

$$unhide(aenc(m,hide(K_X,r)),r^{-1}) = aenc(m,K_X) \tag{4.5}$$

This predicate is used to hide the public key of the receiver from the sender.

6. The public key disguising function *hide* and corresponding inverse function *unhide* as

defined in item 5 above, such that

$$unhide(hide(K_X, r), r^{-1}) = K_X \tag{4.6}$$

This predicate is used to recover the disguised public key.

Examinees use *aenc* to encrypt the answer-scripts before sending it to the examination authority. Examiners use *adec* to decrypt the answer-scripts for evaluation. The function pair $hide/unhide$ (refer eq. 4.5 and eq. 4.6 ) is used by the examination authority to hide the public key of the examiner from the examinees and to remove the blind factor attached to the public key. The following section discusses in detail the disguised public key which forms the basic building block of our proposed answer-script delivery protocol.

### 4.4.2 Disguised Public Key

We now define our proposed cryptographic scheme, namely 'disguised public key' using RSA public key cryptosystem[RSA78]. Definition 9 is also applicable to other public cryptosystem such as ElGamal cryptosystem [ElG84] and Elliptic Curve Cryptography (ECC) [Kob91] as well as hybrid cryptosystem.

**Definition 9.** ***(Disguised Public Key).*** *Let E (examinee) be the producer of a message m, X (examiner) be the final consumer of a message m and C (examination authority) be the intermediary, whose task is to collect the message from E and deliver it to X. Given a public key, $(K_E, n_E)$ of the producer of the message m, a public key $(K_X, n_X)$ and a private key $(K_X^{-1}, n)$ of the consumer of the message (m). Let $(r, n_r)$ be the random blind factor and $(r^{-1}, n_r)$ be the inverse of r, selected by an intermediary, then the disguised public key produced by the intermediary is defined as $K'_X$, such that*

$$K'_X = (K_X * r)^{K_E} (mod\ n_E) \tag{4.7}$$

*having a corresponding recovery function:*

$$c = c'^{r^{-1}} (mod\ n_r) \tag{4.8}$$

*Here $c'$ represents the message encrypted by the producer of the message with the disguised public key($K'_X$) of the consumer.*

In order to effectively use disguised public key as defined above, we define following key cryptographic functions:

**Definition 10.** ***(Hide).*** *A disguising function hide refers to the process of hiding the public key, $K_X$ of the entity X using random factor r, such that the identity of the original public key holder is concealed from the user of the disguised key.*

$$hide(K_X, r) = K'_X \tag{4.9}$$

*This predicate is used to obtain the disguised public key $K'_X$.*

**Definition 11.** ***(Encrypt).*** *The encrypt function refers to encrypting the message to be transmitted to the intermediate receiver(C) using the disguised public key as defined in eq. 4.9*

$$aenc(m, K'_X) = m' \tag{4.10}$$

*This predicate is used to encrypt the message using the disguised public key of the receiver, without getting to know the receiver of the message.*

**Definition 12.** ***(Unhide).***

*The inverse function unhide corresponding to the public key disguising function hide is defined as process of restoring the message encrypted with public key by removing the*

*disguised factor.*

$$unhide(hide(K_X, r), r^{-1}) = K_X \tag{4.11}$$

*This predicate is used to recover the message encrypted with the original public key of the receiver.*

**Definition 13.** ***(Decrypt).*** *The decrypt function refers to decrypting the message(m') by the receiver (X) using the private key($K_X^{-1}$) corresponding to its public key($K_X$)*

$$adec(m', K_X^{-1}) = m \tag{4.12}$$

*This predicate is used to decrypt the message using the private key corresponding to the public key.*

**Theorem 4.4.1.** *In Public Key Cryptosystem:*

1. *Message encrypted with the disguised public key of the recipient achieves recipient anonymity.*
2. *Message encrypted with the disguised public key, achieves enforced confidentiality of a message.*
3. $unblind(aenc(m, blind(K_X, r)), r^{-1}) = aenc(m, K_X)$.
4. *Message encrypted with the disguised public key conserves the original message.*

*Proof.* Consider 3 communicating entities, viz., producer (E), intermediary (C) and consumer (X).
Let the public/private key pair of each entity derived from public key cryptosystem be represented as shown in Table 4.3:

Table 4.3: Public/Private Key pair

| Entity | Public Key | Private Key |
|---|---|---|
| E | $K_E$ | $K_E^{-1}$ |
| C | $K_C$ | $K_C^{-1}$ |
| X | $K_X$ | $K_X^{-1}$ |

Where each public key is known to the public in general and corresponding private key is known only to the owner of the private key.

Each public/private key pair satisfies the following equation as defined in equational theory (refer Section 4.4.1)

$$adec(aenc(m, K_X), K_X^{-1}) = m \tag{4.13}$$

where $K_X$ is a public key and $K_X^{-1}$ is the private key of entity $X$

As per RSA, public key cryptosystem, any message(m) is encrypted using equation:

$$c = m^{K_X} \ (\ mod\ n) \tag{4.14}$$

and the encrypted message(c) is decrypted using equation:

$$m = c^{K_X^{-1}} \ (\ mod\ n) \tag{4.15}$$

let $r$ represent a random number, having corresponding inverse $r^{-1}$ known to entity C only.

According to RSA blind signature scheme, we have message(m) blinded with the random factor (r) to obtain signature of signer as follows:

$$m' = mr^{K_X} \ (\ mod\ n) \tag{4.16}$$

In the RSA blind signature scheme, the message is blinded using equation (4.16).

We adopt a similar approach to disguise the public key of X to hide it from the entity E. By equation (4.14), C encrypts disguised public key of X, using the public key of E as follows:

$$m' = (K_X * r)^{K_E} \ (\ mod\ n) \tag{4.17}$$

Encrypted disguised public key($m'$) is sent to E. On receipt of $m'$, E decrypts $m'$ using equation (4.15) as follows:

$$s' = (m')^{K_E^{-1}} \ (\ mod\ n) \tag{4.18}$$

Using equation (4.17) it is evident that

$$s' = (K_X * r)^{K_E K_E^{-1}} \ (\ mod\ n)$$

i.e., entity E get

$$s' = K_X * r \ (\ mod\ n) \tag{4.19}$$

Now in order to prove our theorem statement,

**Message encrypted with the disguised public key of the recipient, achieves recipient anonymity**,

We need to prove that: "Given $s'$ and list of $t$ unique public keys $K_1, K_2, ...., K_t$, where $s' = K_X * r$ and one of the $K_i \equiv K_X$, $K_X$ cannot be predicted with certainty."

Based on the knowledge of E, it can try to infer the value of $K_X$ as follows,

$r_1 = \frac{s'}{K_1}$, $r_2 = \frac{s'}{K_2}$, ..., $r_t = \frac{s'}{K_t}$

It is evident from the above equations that, if we divide the given disguised public key by each of the known public key $K_i$, we get quotient $r_i$.

Let us assume that each $r_i = r$.

However, it is not possible to get identical quotient when each division is carried between common numerator and unique denominator (public keys are unique).

Such division will produce different quotient each time. In other words, our assumption that $r_i = r$ is false.

Since E is in possession of $t$ public keys and unaware of random factor $r$ used to disguise the public key $K_X$, we can say that, E can only find the public key, $K_X$ hidden in $s'$ with probability $\frac{1}{t}$.

Hence, we can state that the :

**Message encrypted with the disguised public key of recipient achieves recipient anonymity**

$E$ uses $s'$ as a key to encrypt the message (m) held by it using equation (4.14) as follows:

$$c' = (m)^{s'} \ (\ mod\ n) \tag{4.20}$$

From equation (4.19), we can simplify equation (4.20) as

$$c' = (m)^{K_X * r} \ (\ mod\ n) \tag{4.21}$$

As per RSA blind signature scheme (refer Chapter 3, Section 3.1.6), message (m) can be

blinded with random factor (r) and unblinded with corresponding inverse ($r^{-1}$) as follows:

$$unblind(sign(blind(m,r),K_X^{-1}),r^{-1}) = sign(m,K_X^{-1}) \tag{4.22}$$

E sends $c'$ to C.

C applies $r^{-1}$ the inverse of $r$ to $c'$ using the same principle as defined in equation (4.22)

Therefore,

$$c = (c')^{r^{-1}} \ (\ mod\ n) \tag{4.23}$$

From equation (4.20) and equation (4.23) we get

$c = (m)^{K_X * r * r^{-1}} \ (\ mod\ n)$

i.e., the undisguised encrypted message($c$) produced by C is

$$c = (m)^{K_X} \ (\ mod\ n) \tag{4.24}$$

based on equation (4.22), equation (4.23) and equation (4.24), we have

$$unhide(aenc(m,hide(K_X,r)),r^{-1}) = aenc(m,K_X) \tag{4.25}$$

Now since, this recovered message is encrypted with the public key of X, C cannot decrypt it with his/her private key. This proves that, **the message encrypted with the disguised public key, achieves enforced confidentiality of message.**

Equation (4.24) produces the message encrypted with the public key of X.

This encrypted message can be subsequently decrypted by only X, as X is in a possession of

corresponding private key.

Thus, it is possible to restore the original message back, in spite of using the disguised public key for encrypting the message.

In other words we can state that:

**The message encrypted with the disguised public key conserves the original message.** □

### 4.4.3 Properties

The proposed disguised public key cryptosystem (refer Theorem 4.4.1) satisfies the following security properties:

1. The producer knows nothing about the correspondence between $K_X$ and $s'$ i.e. the producer cannot trace the link between the owner of the public key $K_X$ and disguised public key, $s'$.

2. The intermediary cannot derive m from c. In other words, the intermediary cannot create original message m from c as c is the encrypted message produced using the public key of the consumer.

## 4.5 Modelling the Question Paper/Answer-script Delivery Protocols in Applied $\pi$ Calculus

We assume a human originator / sender represented as examination authority $C$ (and examinee $E$ for answer-script delivery protocol described in subsection 4.5.2), a human addressee represented as examinee $E$ (and examination authority for answer-script delivery protocol) and an online publicly readable system under examination authority's control - the pub. Both communicating partners rely on existing Public-Key Infrastructure (PKI).

Examinee and examination authority has a cryptographic key pair (private and public key), and both partners know the other's public keys (The exchange of cryptographic keys is not covered by our protocol.).

### 4.5.1 Question paper Delivery (QPDA) Protocol

A message sequence chart describing the protocol to deliver the question paper from examination authority to examinees is shown in Fig. 4.2.

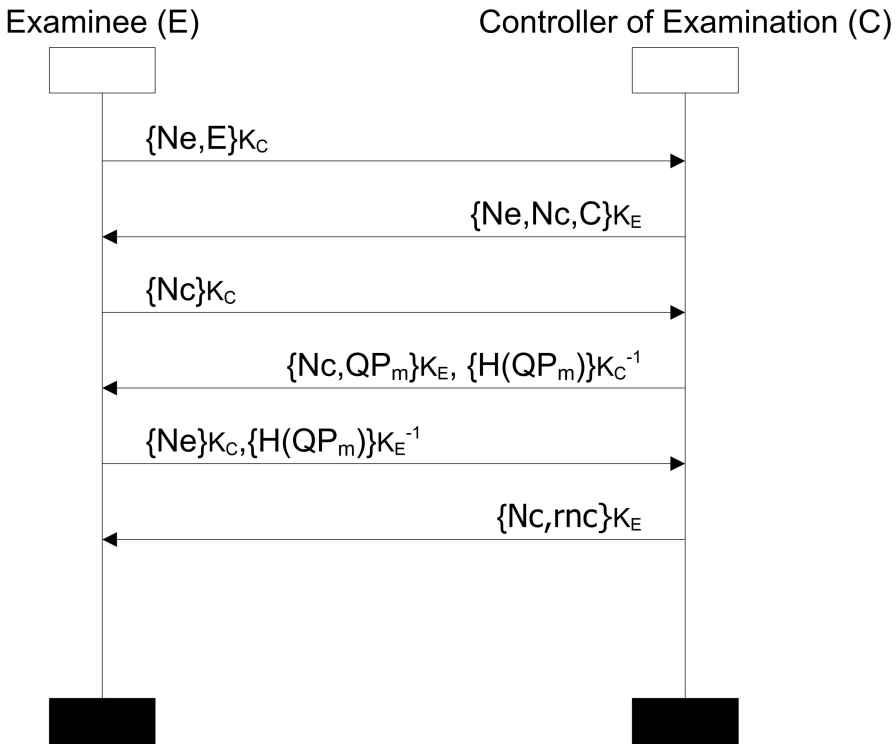



Fig. 4.2: Question Paper Delivery

The protocol steps and their rationale are illustrated in Protocol 1 (QPDA). For the sake of brevity, we focus on the main protocol and skip the first three steps used for authentication [NS78, Low96].

**Protocol 1** (QPDA) : Question Paper (QP) delivery to examinees.

**4:** After authentication session is confirmed(First 3 steps are ignored):

4.1: Examination authority generates a master copy of question paper, $QP_m$ using the pool of question bank available at its disposal.

4.2: Examination authority computes message digest of $QP_m$ using private key $K_C^{-1}$ of examination authority.

4.3: Examination authority encrypts $QP_m$ along with session key $N_C$ with public key $K_{E_i}$ of examinee.

4.4: Examination authority pairs messages created in step 4.2 and 4.3 and sends it to examinee.

**Message 4:** $C \rightarrow E_i : \{N_C, QP_m\}K_{E_i}, \{\mathscr{H}(QP_m)\}K_C^{-1}$

**Reason:** This is to build the evidence that examination authority generated the question paper $QP_m$.

**5:** When examinee receives message 4 from examination authority:

5.1: Examinee decrypts message 4 to read $QP_m$, $N_C$ and $\mathscr{H}(QP_m)$.

5.2: Examinee computes hash of $QP_m$ and compares it with message digest $\mathscr{H}(QP_m)$ received from examination authority.

5.3: Examinee acknowledges receipt of valid $QP_m$ to examination authority, if both hash values match.

**Message 5:** $E_i \rightarrow C : \{N_{E_i}, \{\{\mathscr{H}(QP)\}K_C^{-1}\}K_{E_i}^{-1}$

**Reason:** This step builds the undeniable evidence that examinee indeed received, question paper $QP_m$ sent by examination authority.

**6:** When examination authority receives message 5 from examinee:

6.1: Examination authority decrypts message 5 to read $\mathscr{H}(QP_m)$.

6.2: On receipt of original $\mathscr{H}(QP_m)$ from $E_i$, examination authority generates a random number $rn_C$ for random actual question paper $QP_i$ generation, based on $QP_m$.

6.3: Examination authority encrypts $rn_C$ using public key of examinee

6.4: Examination authority sends message created in step 6.3 to examinee.

**Message 6:** $C \rightarrow E_i : \{N_C, rn_C\}K_{E_i}$

**Reason:** This step is required to make the protocol fair to both the parties. Moreover, this step helps in delivering a unique question paper copy to each examinee and at the same time avoids the transmission of actual QP copy over the network.

Examinee generates the random copy of question paper, $QP_{E_i}$ from the master copy $QP_m$ based on $rn_C$ received from examination authority and $rn_A$. ($rn_E$ is a random no generated by examinee. It is assumed that a unique question paper $QP_{E_i}$ is generated by examinee using parameters $rn_E$ and $rn_C$ (The question paper generation does not form part of our protocol.).

### 4.5.2 Protocols for Delivery of Answer-scripts (ADAA) using Disguised Public Key

The examination authority maintain a collection of blinded public keys [Cha83] of all the examiners (X). We intend to use the disguised public keys of examiners to hide the identity of examiners from examinees ($E_i$) (refer Section 4.4.2). The dual objective of hiding answer-scripts from the examination authority and the identity of examinee from examiner is achieved with the help of the blinding technique. Another pertinent requirement of linking the question paper to answer-scripts unambiguously is achieved through dual signature.

#### Protocol I using Asymmetric Cryptosystem

Initially, we designed the answer-script delivery protocol using pure asymmetric cryptosystem. A message sequence chart describing the protocol to deliver the answer-scripts from examinees to the examination authority is shown in Figure 4.3. The detailed steps are elaborated in Protocol 2. For the sake of brevity, we focus on the main protocol and skip the initial protocol steps used for authentication [NS78, Low96].

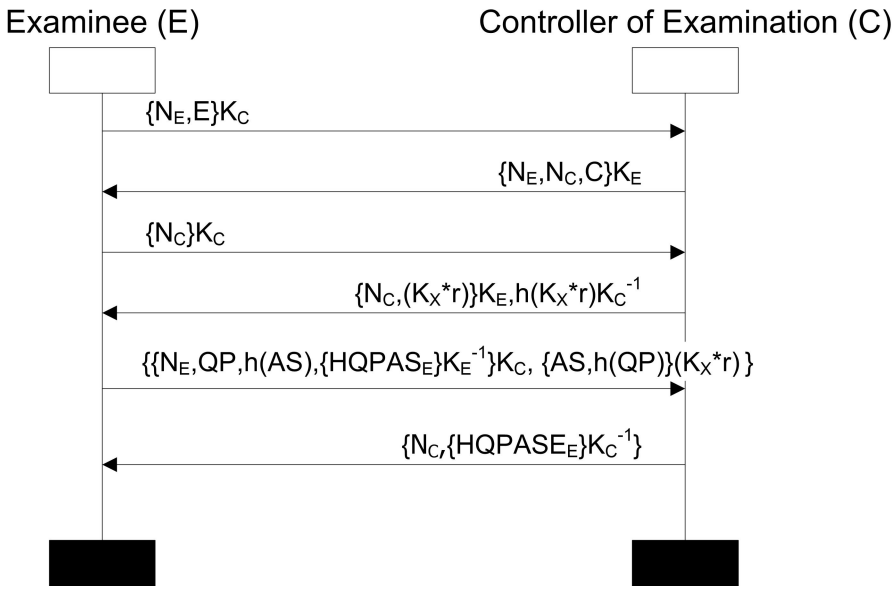



Fig. 4.3: Answer Script Delivery using Asymmetric Cryptosystem

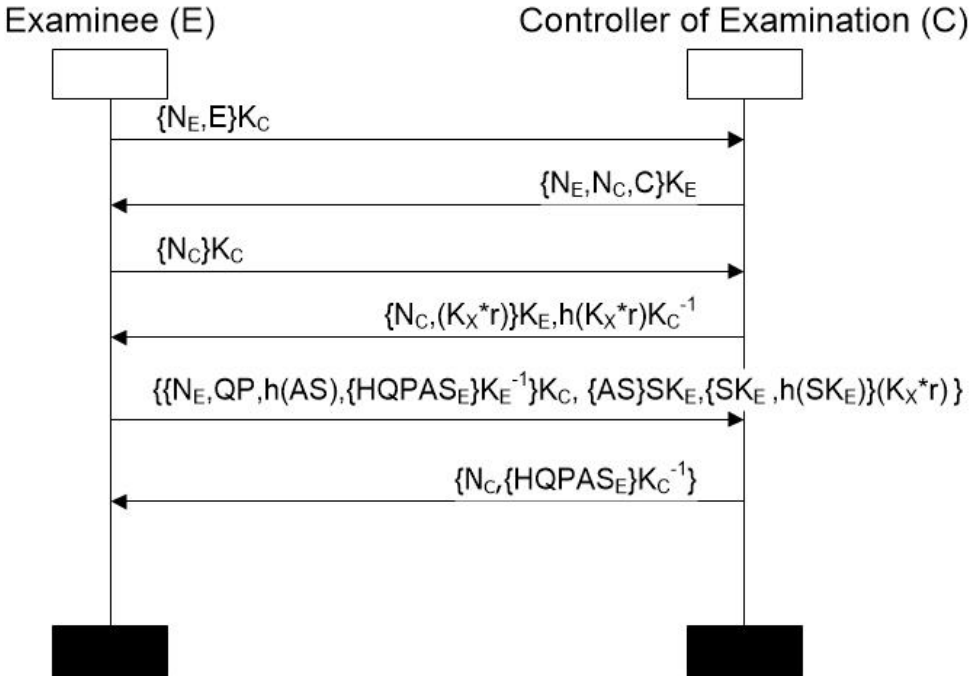



Fig. 4.4: Answer Script Delivery using Hybrid Cryptosystem

**Protocol 2** Answer-scripts (AS) Delivery using Asymmetric Cryptosystem for Anonymity.

**1:** After examinee ($E_i$) establishes authentication session with examination authority ($C$) (Authentication and key exchange steps are ignored):

**2:** Initially, examination authority disguises the public key of examiner (X) as follows:

2.1: First, examination authority selects the public key $K_X$ of examiner and chooses a random number (r) to disguise the public key $K_X$ as $(K_X * r)$.

2.2: Examination authority encrypts the disguised public key $(K_X * r)$ of examiner using the public key $K_{E_i}$ of examinee as $\{(K_X * r)\}K_{E_i}$.

2.3: Examination authority computes the message digest of $(K_X * r)$ and signs it using a private key $K_C^{-1}$.

2.4: Examination authority pairs the disguised public key and the message digest created in step 2.2 and 2.3 and sends it to examinee.

**Message 2:** $C \rightarrow E_i : \{N_C, (K_X * r)\}K_{E_i}, \{\mathscr{H}(K_X * r)\}K_C^{-1}$

**Reason:** Sending blind public key of examiner to examinee serves two crucial objectives. It aids in hiding the identity of examiner from examinee and assists in hiding the answer-scripts of examinee from examiner.

**3:** When examinee receives message 2 from examination authority,

3.1: Examinee decrypts message 2 to read $(K_X * r)$ and $\{\mathscr{H}(K_X * r)\}$.

3.2:Examinee computes hash of $(K_X * r)$ and compares it with the message digest $\{\mathscr{H}(K_X * r)\}$ received from examination authority.

3.3: If both hash values match, the protocol proceeds further.

3.4: Subsequently, examinee completes answer-script $AS_{E_i}$ and compute the message digest $\mathscr{H}(AS_{E_i})$ of $AS_{E_i}$. Examinee also computes the message digest $\mathscr{H}(QP_{E_i})$ of question paper $QP_{E_i}$ answered by it.

3.5: Examinee also computes a dual signature as defined in [OPT97, MS98].
Dual signature **HQPAS** is computed as $HQPAS_{E_i} = \mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$

3.6: Examinee pairs messages created in step 3.4 and 3.5 and sends it to examination authority as follows:

i. $\{QP_{E_i}, \mathscr{H}(AS_{E_i}), \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C$ using public key of examination authority.

ii. $\{AS_{E_i}, \mathscr{H}(QP_{E_i})\}(K_X * r)$ using the blinded public key of the examiner.

**Message 3:** $E_i \rightarrow C : \{\{N_{E_i}, QP_{E_i}, \mathscr{H}(AS_{E_i}), \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C, \{AS_{E_i}, \mathscr{H}(QP_{E_i})\}(K_X * r)\}$

**Reason:** By using the dual signature method, $QP_{E_i}$ and $AS_{E_i}$ can be linked together securely, while releasing only the necessary information to the relevant party.

**4:** When examination authority receives message 3 from examinee :

4.1: Examination authority decrypts first part of the message 3 to get $QP_{E_i} + \mathscr{H}(AS_{E_i}) + \{HQPAS_{E_i}\}K_{E_i}^{-1}$

4.2: Examination authority finds $\mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$.

4.3: Examination authority decrypts the digital signature from message 3 $\{HQPAS_{E_i}\}K_{E_i}^{-1}$ with the public signature key of examinee.

4.4: Compare the result of step 4.2 with 4.3 to verify integrity of the received message.

4.5: Examination authority acknowledges the receipt of $QP_{E_i}$ and $AS_{E_i}$ by sending signed dual signature.

**Message 4:** $C \rightarrow E_i : \{N_C, \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C^{-1}$

**Reason:** The acknowledgement builds the undeniable evidence that examination authority indeed received, the linked question paper $QP_{E_i}$ and $AS_{E_i}$ sent by examinee.

**Note:** Similar to above, during answer-script delivery examination authority provides $\{AS_{E_i}, \mathscr{H}(QP_{E_i})\}(K_X * r)$ and $\{HQPAS_{E_i}\}K_C^{-1}$ to examiners.

**Protocol II using Hybrid Cryptosystem**

Public-key cryptosystems often rely on complicated mathematical computations and are thus generally much more inefficient than comparable symmetric-key cryptosystems. A hybrid cryptosystem is one which combines the convenience of a public-key cryptosystem with the efficiency of a symmetric-key cryptosystem.

In this section, we describe answer-scripts delivery using hybrid cryptosystem. A message sequence chart describing the protocol to deliver the answer-scripts from examinees ($E_i$) to the examination authority (C) is shown in Figure 4.4 and detailed steps are elaborated in Protocol 3.

**Protocol 3** Answer-scripts (AS) Delivery using Hybrid Cryptosystem for Anonymity.

**1:** After examinee ($E_i$) establishes authentication session with examination authority ($C$) (Authentication and key exchange steps are ignored):

**2:** Initially, examination authority disguises the public key of examiner (X) as follows:

2.1: First, examination authority selects the public key $K_X$ of examiner and chooses a random number (r) to disguise the public key $K_X$ as $(K_X * r)$.

2.2: Examination authority encrypts the disguised public key $(K_X * r)$ of examiner using the public key $K_{E_i}$ of examinee as $\{(K_X * r)\}K_{E_i}$.

2.3: Examination authority computes the message digest of $(K_X * r)$ and signs it using a private key $K_C^{-1}$.

2.4: Examination authority pairs the disguised public key and the message digest created in step 2.2 and 2.3 and sends it to examinee.

**Message 2:** $C \rightarrow E_i : \{N_C, (K_X * r)\}K_{E_i}, \{\mathscr{H}(K_X * r)\}K_C^{-1}$

**Reason:** Sending blind public key of examiner to examinee serves two crucial objectives. It aids in hiding the identity of examiner from examinee and assists in hiding the answer-scripts of examinee from examination authority.

**3:** When examinee receives message 2 from examination authority,

3.1: Examinee decrypts message 2 to read $(K_X * r)$ and $\{\mathscr{H}(K_X * r)\}$.

3.2: $E_i$ computes hash of $(K_X * r)$ and compares it with the message digest $\{\mathscr{H}(K_X * r)\}$ received from examination authority.

3.3: If both hash values match, the protocol proceeds further.

3.4: Subsequently, examinee completes answer-script $AS_{E_i}$ and compute the message digest $\mathscr{H}(AS_{E_i})$ of $AS_{E_i}$. $E_i$ also computes the message digest $\mathscr{H}(QP_{E_i})$ of question paper $QP_{E_i}$ answered by it.

3.5: Examinee generates a secret key $SK_{E_i}$ .

3.6: Examinee encrypts $AS_{E_i}$ using its secret key $SK_{E_i}$ and pairs the secret key $SK_{E_i}$ and $\mathscr{H}(SK_{E_i})$ using disguised public key of examiner and sends it to examination authority.

3.7: Examinee also computes a dual signature as defined in [OPT97, MS98].
Dual signature **HQPAS** is computed as $HQPAS_{E_i} = \mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$

3.8: Examinee pairs messages created in step 3.4, 3.5 and 3.6 sends it to examination authority as follows:

i. $\{QP_{E_i}, \mathscr{H}(AS_{E_i}), \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C$ .

ii. $\{AS_{E_i}\}SK_{E_i}$ and $\{SK_{E_i}, \mathscr{H}(SK_{E_i})\}(K_X * r)$ .

**Message 3:** $E_i \rightarrow C : \{\{N_{E_i}, QP_{E_i}, \mathscr{H}(AS_{E_i}), \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C, \{AS_{E_i}\}SK_{E_i}, \{SK_{E_i}, \mathscr{H}(SK_{E_i})\}(K_X * r)$

**Reason:** By using the dual signature method, $QP_{E_i}$ and $AS_{E_i}$ can be linked together securely, while releasing only the necessary information to the relevant party.

**4:** When examination authority receives message 3 from $E_i$ :

4.1: Examination authority decrypts first part of the message 3 to get
$QP_{E_i} + \mathscr{H}(AS_{E_i}) + \{HQPAS_{E_i}\}K_{E_i}^{-1}$

4.2: Examination authority finds $\mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$.

4.3:Examination authority decrypts the digital signature $\{HQPAS_{E_i}\}K_{E_i}^{-1}$ from message 3 with the public signature key of $E_i$.

4.4: Compare the result of step 4.2 with 4.3 to verify integrity of the received message.

4.5: Examination authority acknowledges the receipt of $QP_{E_i}$ and $AS_{E_i}$ by sending signed dual signature.

**Message 4:** $C \rightarrow E_i : \{N_C, \{HQPAS_{E_i}\}K_{E_i}^{-1}\}K_C^{-1}$

**Reason:** The acknowledgement builds the undeniable evidence that examination authority indeed received, the linked question paper $QP_{E_i}$ and $AS_{E_i}$ sent by ($E_i$).

The proposed protocol is useful in achieving anonymity and enforced confidentiality in situations where information is exchanged between two parties through an intermediary.

### 4.5.3 Modelling Examination Authority and Examiner processes in ProVerif

The behaviour of the examination authority (C) and examinee (E) is modelled using ProVerif (refer Chapter 3, Section 3.3.2, refer Process 1 and refer Process 2). The corresponding ProVerif code is depicted in Appendix B. The modelling of examination authority and examinee processes and subsequent formal analysis is carried on the basis of protocol 2.

#### The Examination Authority Process

Examination authority, after sending the unique question paper to the examinee, generates a fresh nonce and a fresh blind factor. Next, examination authority disguises the public key of the examiner and sends the signed disguised public key of examiner to examinee [DK16], expecting an encrypted question paper and the answer-script pair bound by dual signature [MS98]. The role of examination authority is to verify the question paper legitimacy, which is done by matching the dual signature with a combined hash of question paper sent and hash of answer-script received. If both the hash values match, it confirms that no alterations were carried in the question paper by any of adversaries.

**Process 1** : The Examination Authority

```
let C =
new Nb:nonce;
new rf:bkey;
new quespap:bitstring;
let hexkey= hide(pkEx,rf) in
let sbexkey=pkeytobitstring(hexkey) in
let sbkHash=sign(hash(sbexkey),ssecEA) in
let qpHash=sign(hash(quespap),ssecEA) in
let authQBk = encrypt(((((Nb,quespap),qpHash),bexkey),
sbkHash),pkST) in
out(ch,authQBk);
in(ch, studQPAS:bitstring);
let ((((Na:nonce,=quespap),asHash:bitstring),
dualsign:bitstring),
encansscr:bitstring) = decrypt(authQBk,skEA) in
let hqphas = (hash(quespap),asHash) in
if hash(hqphas)=checksign(dualsign,spubST) then
let sdualsign=sign(dualsign,ssecEA) in
out(ch,sdualsign);
```

**The Examinee Process**

Examinee (E) on receipt of a valid question paper and at the end of answering the answer-script, generates a fresh nonce for authentication purpose. Next, he uses the disguised public key of the examiner received from the examination authority (C) to encrypt the answer-script produced by it. Examinee also computes a dual signature of question paper and answer-script to bind the answer-script to the question paper securely. The question paper, the hash of the answer-script, the dual signature and the encrypted answer-script together with the hash of question paper with disguised public key is sent to examination authority for necessary action.

**Process 2** : The Examinee

```
let E =
new Na:nonce;
in(ch,authQBk:bitstring);
let ((((Nb:nonce,quespap:bitstring),qpHash:bitstring),
usExkey:pkey),sbkHash:bitstring) =
decrypt(authQBk,skST) in
let susExkey = pkeytobitstring(usExkey) in
if hash(quespap)= checksign(qpHash,spubEA)
&& hash(susExkey) = checksign(sbkHash, spubEA) then
let hqphas = (hash(quespap),hash(ansscr))
let dualsign = sign(hash(hqphas),ssecST) in
let encansscr = encrypt((ansscr,hash(quespap)),usExkey) in
let studQPAS = encrypt(((((Na,quespap),hash(ansscr)),
dualsign),encansscr),pkEA) in
out(ch,studQPAS);
in(ch,sdualsign:bitstring);
```

## 4.6 Formal Analysis

We analyse our ADAA protocol in ProVerif , a security protocol verifier that allows the automatic analysis of authentication and privacy properties. The input language of ProVerif is a variant of the applied $\pi$ calculus. In the next Section, first we define security properties desired during delivery of answer-script from examinee to the examination authority. Also, we use ProVerif tool and manual proofs to verify the correctness of the protocol.

### 4.6.1 Question paper and Answer-script Associativity

A protocol with examinee process $E(QP,AS,id)$ and examination authority process $C$ , safeguards $QP$ and $AS$ associativity, if process $C$ possesses non-repudiable evidence to distinguish between the received copy of $QP$ and $AS$ pair from the claimed copy.

**Definition 14.** ***(Question paper and Answers-script Associativity)****, An examination system offers question paper* $(QP)$ *and answer-script* $(AS)$ *associativity with unique examinee*

*identity (id), if it is possible to unambiguously distinguish when a examinee $E_1$ produce answer-script $AS_{E_2}$ corresponding to the received question paper $QP_{E_1}$ from the case where examinee claim of producing $AS_{E_2}$ corresponding to altogether different question paper $QP_{E_2}$ (refer Fig.* (4.5)*).*

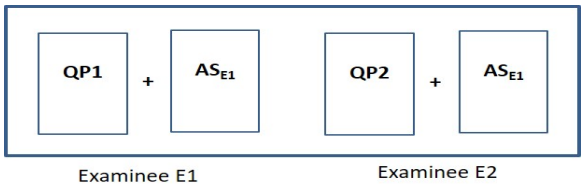



Fig. 4.5: Answer-script collusion/plagiarism

*The question paper and answer-script associativity is formally specified as follows:*

$$\begin{aligned} &\upsilon\tilde{n}.(E\{QP_{E_1}/x, AS_{E_2}/y, E_1/z\}|C) \\ &\not\approx_l \upsilon\tilde{n}.(E\{QP_{E_2}/x, AS_{E_2}/y, E_1/z\}|C) \end{aligned} \tag{4.26}$$

This association is required to build a reliable evidence for resolution of any dispute related to question paper/answer-script originality/correctness. We now show that the proposed ADAA protocol provides the associativity between a unique question paper and a answer-script, even when all but one examinee is dishonest.

**Theorem 4.6.1.** *ADAA protocol satisfies* associativity *between a given pair of question paper and answer-script.*

*Proof:* In order to prove Theorem 4.6.1, we need to show that in the context of ADAA protocol, it is possible to unambiguously distinguish when an examinee $E_1$ produces an answer-script $AS_{E_2}$ corresponding to the received question paper $QP_{E_1}$ from the case where an examinee claims of receiving an altogether different question paper $QP_{E_2}$.

Let us assume that examinee $E_1$ is dishonest. Initially $E_1$ produces answer-script $AS_{E_2}$ corresponding to the question paper $QP_{E_1}$ received by him. Let us also assume that $AS_{E_2}$

produced by $E_1$ is a plagiarized copy of the answer-script produced by neighbouring examinee $E_2$ corresponding to the question paper $QP_{E_2}$.

This action of the examinee would amount to getting poor grade to the examinee in the examination as it has produced totally unrelated answers.

If, the examinee $E_1$, claims that, he had submitted $AS_{E_2}$ corresponding to the question paper $QP_{E_2}$ and not $QP_{E_1}$, then the dispute handling authority needs concrete evidence to falsify the claim of the dishonest examinee. During the process of answer-script delivery, the following frames are generated.

$$\begin{aligned}
&\varphi_0 = \{pk(C)/v1\}|\{pk(E_i)/v2\} \mid \{pk(E_i)/v3\} \mid \{hexKey = hide(pk(E_i), rf)\} \mid \\
&\{enc(QP_{Ei}, E_i)|i = 1..n\}, \\
&\varphi_1 = \varphi_0|\{QP_{E1}/x, AS_{E2}/y\}, \\
&\varphi_2 = \{QP_{E2}/x, AS_{E2}/y\}, \\
&\varphi_k = \{\varphi_{k-1}\} \mid \{sign(hash(hQP_{E1}hAS_{E2}), ssecST)\} \mid \{hash(AS_{E2}) \mid hash(hQP_{E1}hAS_{E2}) \mid \\
&\{enc((AS_{E2}, hash(QP_{E1})), hexKey)\} \mid \{enc((AS_{E2}, hash(QP_{E1})), pk(E_i)\}, \\
&\varphi_\delta = \varphi_n|\{dec(QP_{E1}, C)|\{dec(AS_{E2}, E_i))\}
\end{aligned} \tag{4.27}$$

$\varphi_0$ corresponds to the initial knowledge of the communicating entities. It contains the public data and the public keys.

$\varphi_1$ corresponds to the question paper answer-script pair submitted by the dishonest examinee.

$\varphi_2$ corresponds to the claim of the dishonest examinee after the completion of the examination and the declaration of the result.

$\varphi_k$ corresponds to the knowledge of the examination authority/examiners after submission of the answer-script by $E_1$.

$\varphi_\delta$ corresponds to the final decryption of the received data.

Here, rf is a blind factor used for hiding the public key of the examiner from the examinees, ssecST is signing key of the examinee, hexkey is the hidden public key of examiner, $hQP_{E_1}hAS_{E_2}$ is concatenated hash of the question paper and answer script.

The dispute handling authority can use this non-repudiable evidence to prove and falsify any illicit claim of the examinee.

Considering the information available at the disposal of dispute handling authority as indicated in Eq. (4.27), we are able to show that, the claim of dishonest examinee of unfair evaluation due to the use of wrong question paper during evaluation can be proved incorrect as follows:

The dual signature $ds = hash(hQP_{E_1}hAS_{E_2})$ is signed by the examinee entity corresponding to the actual question paper ($QP_{E_1}$) received and the answer-script ($AS_{E_2}$) submitted by it.
The new claim of the examinee of receiving $QP_{E_2}$ and producing $AS_{E_2}$ is true, then, the dual signature of question paper/answer-script pair would have been:
$ds' = hash(hQP_{E_2}hAS_{E_2})$

$\exists\ QP_{E_2}$ s.t. $\mathscr{H}(QP_{E_1}) = \mathscr{H}(QP_{E_2})$ and $\exists\ ds = ds'$

It is unlikely that the two distinct question papers map to the same hash value since $QP_{E_1} \cap\ QP_{E_2} \neq \emptyset$.

Since $(ds = cds)\phi$ and $(ds' \neq cds)\phi 1$ , $\phi \not\approx_s \varphi 1$.
i.e., two frames $\phi$ and $\phi 1$ are statically not equivalent. This means that $\phi$ and $\phi 1$ are distinguishable to the dispute handling authority.

This holds true for any frame $\phi_i$ for $i > 0$.

Since, dispute handling authority is successful in distinguishing between original pair and altered pair, i.e, $P[QP_{E_1}/q1, ASE_2/a1] \not\approx P[QPE_2/q1, ASE_2/a1]$, we can conclude that ADAA

protocol ensures Unambiguous Associativity between given QP and AS pair.

### 4.6.2 Answer-script Secrecy

A protocol with the examinee process $E(QP,AS,id)$ and the examination authority process $C$ provides answer-script secrecy, if answer-scripts received by process $C$ are indistinguishable to it.

**Definition 15.** ***(Answer-script Secrecy)****, An examination system offers answer-script secrecy, if it is not possible for the examination authority to distinguish the answer-scripts received. This is formally specified by:*

$$\begin{aligned} &\upsilon\tilde{n}.(E\{AS_{E_1}/x, AS_{E_2}/y\}|C) \\ &\approx_l \upsilon\tilde{n}.(E\{AS_{E_2}/x, AS_{E_1}/y\}|C) \end{aligned} \tag{4.28}$$

Answer-script secrecy states that, the answer-scripts produced by the examinees need to remain secret from the examination authority. This is desired because the examination authority have no role to play in the evaluation of the answer-scripts.

**Lemma 4.6.1.** *ADAA protocol ensures the* secrecy of answer-scripts *from the examination authority.*

*Proof:* In order to prove lemma 4.6.1, we need to show that, it is not possible for the examination authority to distinguish the received answer-scripts from each other. Based on the equational theory and local knowledge of the examination authority (C) (refer Eq. (4.27)), we propose the following inference system.

$$\frac{C \qquad aenc((AS_{E_i}, hash(QP_i), dualsign), pk(E_i))}{\neg(AS_{E_i})}$$

The above inference system indicates that, the answer-scripts received by the examination authority are encrypted using the public key of the examiner. Since, the private key of the examiner is required to decrypt the answer-script produced by the examinee, it can be deduced that the secrecy of the answer-scripts is protected by the protocol from the examination authority, i.e., each received answer-script is observationally equivalent for the examination authority as indicated in Eq. (4.29).

$$P[(\{AS_{E_1}/x, AS_{E_2}/y\}] \approx [\{AS_{E_2}/x, AS_{E_1}/y\}] \tag{4.29}$$

Thus, we state that, the ADAA protocol provides secrecy of the answer-scripts from the examination authority.

### 4.6.3 Examinee Anonymity

A protocol with examination authority process $C(QP, AS, pseudo\_id)$ and examiner process $X$ provides examinee anonymity, if answer-scripts received by process $X$ are indistinguishable to it in terms of examinee identity.

**Definition 16.** ***(Examinee Anonymity)****, An examination system ensures examinee anonymity, if it is not possible for examiners to map the answer-scripts received by it to the actual examinee identity, i.e., examinee* $E1$ *producing answer-script* $AS_{E_1}$ *is indistinguishable from examinee* $E2$ *producing answer-script* $AS_{E_2}$*. This is formally specified by:*

$$\begin{aligned} &\upsilon\tilde{n}.(C\{\{AS_{E_1}, pid_{E_1}\}, \{AS_{E_2}, pid_{E_2}\}\}|X) \\ &\approx_l \upsilon\tilde{n}.(C\{\{AS_{E_1}, pid_{E_2}\}, \{AS_{E_2}, pid_{E_1}\}\}|X) \end{aligned} \tag{4.30}$$

An examination system with examinee anonymity ensures that, the examiner cannot infer the

author of the answer-scripts from the given answer-scripts. Examinee anonymity is required to prevent any attempt of the examinee and examiner from coercing with each other and trace the answer-script of the examinee based on the known examinee identities and the given answer-scripts.

**Lemma 4.6.2.** *ADAA protocol ensures* examinee anonymity *from the examiners.*

*Proof:* In order to prove lemma 4.6.2, we need to show that, it is not possible for the examiners to find the authors of the answer-scripts from its knowledge base. Based on the equational theory and local knowledge of the examiners (X) (refer Eq. (4.27)), we propose the following inference system.

$$\frac{X \qquad aenc((AS_{E_i}, pid_i), pk(X_i))}{(AS_{E_i}, pid_i)}$$

The examination authority, send the pseudo identity of the examinee ($pid_i$) to the examiners. The private key required to reveal the examinee identity back is known to only the examination authority. In other words, though examiners get the answer-scripts for evaluation, the examinee identity is not available to the examiners during evaluation, i.e., two given answer-scripts are observationally equivalent to the examiners in the absence of knowledge of actual examinee identity(refer Eq. (4.31)).

$$\begin{aligned} &P[AS_{E_1}/x, pid_{E_1}/y | AS_{E_2}/x, pid_{E_2}/y] \approx \\ &P[QP_{E_1}/x, pid_{E_2}/y | AS_{E_2}/x, pid_{E_1}/y] \end{aligned} \tag{4.31}$$

Thus, we state that, the ADAA protocol ensures *examinee anonymity* from the examiners.

### 4.6.4 Verifying Elements Generated for Dispute Handling

Any fair and reliable protocol needs to generate an irrefutable evidence for safeguarding the interests of all the communicating entities. The ADAA protocol generates series of elements during the run of protocol. The elements held by examinee and examination authority for effective dispute handling are listed in table 4.4.

Table 4.4: Verifying Elements held by each Stakeholder for Dispute Handling

| | Examinee ($E_i$) | Examination Authority (C) |
|---|---|---|
| Message | $QP_m$<br>$(K_{X_i} * r)$<br>$AS_{E_i}$ | $QP_m$<br>$(K_{X_i} * r)$<br>$\{AS_{E_i}, \mathscr{H}(QP_{E_i})\}(K_{X_i} * r)$ |
| Hash | $\mathscr{H}(QP_m)$<br>$\mathscr{H}(K_{X_i} * r)$<br>$\mathscr{H}(AS_{E_i})$<br>$HQPAS_{E_i} = \mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$ | $\mathscr{H}(QP_m)$<br>$\mathscr{H}(K_{X_i} * r)$<br>$\mathscr{H}(AS_{E_i})$<br>$HQPAS_{E_i} = \mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$ |
| Signature | $\{\mathscr{H}(QP_m)\}K_C^{-1}$<br>$\{\mathscr{H}(K_{X_i} * r)\}K_C^{-1}$<br>$\{HQPAS_{E_i}\}K_{E_i}^{-1}$ | $\{\mathscr{H}(QP_m)\}K_C^{-1}$<br>$\{\mathscr{H}(K_{X_i} * r)\}K_C^{-1}$<br>$\{HQPAS_{E_i}\}K_{E_i}^{-1}$ |

The verification process involves submission of a set of elements to a trusted third party for handling disputes. The compliance of the protocol to the security goals is verified using the verifying elements as listed in table 4.4.

**Confidentiality of Answer-scripts from all except Examiner**

An examinee submits $\{AS_{E_i}, \mathscr{H}(QP_{E_i})\}(K_{X_i} * r)\}$ containing encrypted answer-script $AS_{E_i}$ to examination authority. The answer-script is encrypted using disguised public key $(K_{X_i} * r)$ of examiner $X_i$. The answer-script encrypted using the public key of examiner $X_i$ can be decrypted using corresponding private key of the examiner $X_i$ only.

**Question paper and answer-script Binding**

The hash of answer-script and the corresponding question paper is linked together using dual

signature $HQPAS_{E_i} = \mathscr{H}[\mathscr{H}(QP_{E_i}) + \mathscr{H}(AS_{E_i})]$. Dual signature is computed by first finding hash value of question paper and answer-script individually.Then, a new hash is computed by combining the hash of question paper and answer-script.The signed dual signature $\{HQPAS_{E_i}\}K_{E_i}^{-1}$ is sent to the examination authority. This method binds the question paper and answer-script securely and prevents the examinee from claiming the receipt of altogether different question paper then the originally delivered to the examinee.

**Anonymity of Examiner and Examinee**

Examination authority provides disguised public key $(K_{X_i} * r)$ of examiner to examinee. Examinee needs to encrypt the answer-scripts present in his possession with the help of disguised public key. It is infeasible for examinees to compute the original public key of examiner based on the knowledge of all the public keys available in its repository.

Similarly, answer-scripts submitted by examinees are forwarded to examiners for evaluation after hiding the identity of examinees with the help of pseudonyms. In this way identity of examiners and examinees remain hidden from each other.

## 4.7 Summary

In summative examinations, the two crucial security requirements are anonymity and confidentiality. Anonymity is required to hide the identity of the examinee and the examiner from each other and confidentiality is necessary to maintain the secrecy of answer-scripts from the examination authority. In this Chapter, we described a dual purpose cryptographic scheme, namely 'disguised public key' to achieve anonymity and confidentiality in summative E-examinations. In our approach, the sender is provided with the disguised public key of the recipient to de-link the identity of the recipient from the sender. We, also provided a formal specification of the security protocol using applied $\pi$ calculus and verified the correctness of protocol properties using the ProVerif tool. We defined series of

associativity and anonymity properties to analyse the correctness of our proposed protocol. The defined associativity and anonymity properties are intended to link the question paper and answer-script of the examinee together without revealing unnecessary information to the other communicating entity. We used manual proofs and ProVerif tool to prove that our proposed protocol fully satisfies the properties of associativity and anonymity. The proposed mechanism is suitable in general, for achieving anonymity and confidentiality in applications where communication between the sender and the recipient is achieved through the intermediary third party.

# CHAPTER 5

# Computer-Assisted Evaluation using Rubrics

Institutions worldwide conduct public examinations to evaluate performance of examinees. Such examinations are often subjective in nature and need manual evaluation. The manual evaluation of subjective answer-scripts usually suffers from evaluation anomalies and the impact of 'Examiner variability' or 'Examiner subjectivity'. Examiner subjectivity/ variability mainly occurs due to the tendency of examiners to be careless, erratic, strict or liberal during the course of the evaluation. Most of the currently employed methods partly address the problem of evaluation errors/lapses and examiner subjectivity with the aid of extra checks such as re-checking, re-verification, re-evaluation, etc. We need a pragmatic and unified approach to ensure uniformity and error-free evaluation.

In this Chapter, we present a method of computer-assisted evaluation of subjective answer-scripts using rubrics. In which the main focus is reduction/elimination of errors as well as the degree of variability of the examiners, thereby, enhancing the quality of the evaluation of the answer-scripts. Section 5.2 considers the manual answer-script evaluation process to pinpoint various vulnerabilities associated with the manual evaluation of subjective answer-scripts. Some of the frequently used approaches for countering the evaluation anomalies, are discussed in section 5.3. Section 5.4 describes the summative examination model and specification of the proposed solution, i.e., Computer-Assisted Evaluation using Rubrics (CAER) for controlling the evaluation errors, intra and inter examiner variations in evaluation and for reduction of the lengthy evaluation cycle time.

Section 5.5 describes the research methodology used in comparing and contrasting the CAER solution with manual evaluation and finally section 5.6 validates the effectiveness of the CAER solution.

## 5.1 Introduction

Summative examinations form the crucial method for evaluation of examinees [Bou00, VTK05]. These summative examinations when conducted on a large scale are referred to as 'Public examinations'. Such examinations are often subjective in nature and need manual evaluation.

As Public examinations comprise of a large number of answer-books, they need a large number of examiners for evaluation of answer-scripts of each course paper. In the evaluation of subjective answer-scripts, each examiner applies his own yardstick to assess the answer-scripts. The independent evaluation scale results into large variation (wide difference in average marks and range of marks in a particular course) in allotment of marks [Bro12, OW15]. The causes of major variation in allotment of marks is due to a large quantum of answer-scripts, subjective evaluation and lack of uniform evaluation guidelines. Besides these, errors can creep in during evaluation of subjective answer-scripts. Such errors are normally introduced in the form of marks totalling errors or marks transfer errors.

The serious flaws in the current evaluation system are apparent from the significant changes in the examinees overall marks during subsequent verification/re-evaluation of answer-scripts. The current measures/approaches minimize the errors in evaluation and result compilation with additional efforts, such as, moderation, re-checking, re-verification and re-evaluation [Rea90, WCC11].

The methods that are employed in evaluation need to be consistent, fair and error-free for all the examinees. Each examinee needs to be evaluated with the same scale and criteria. A rubric is one such mechanism which offers a uniform and a consistent evaluation platform [RA10]. It includes criteria for rating important dimensions of performance, as well as the standards of attainment for those criteria. A rubric mechanism establishes assessment criteria and marking scheme for bringing marks variation under control [JS07].

In this Chapter, we consider some of the public examination systems to analyse and quantify the heterogeneity between and within examiners/graders during evaluation, coupled with, the errors committed during computation and presentation of the marks/grades and evaluation cycle time. We, also present a pragmatic computer-assisted evaluation technique using rubrics (CAER) for enabling examiners to improve the evaluation and the result compilation tasks. We achieve this objective by reducing evaluation anomalies, controlling examiner heterogeneity and reducing wastage of examination resources and time. Finally, an in-depth analysis of the manual evaluation and CAER solution is carried out to ascertain the effectiveness of the proposed CAER solution.

## 5.2 Iniquities of Subjective Answer-scripts Evaluation

It is essential to carry the entire answer-scripts evaluation process accurately and diligently to safeguard the interest of examinees, teachers and academic institution [CIM14, BBP13]. The presence of a large number of answer-scripts for evaluation, normally gives rise to inconsistency, errors and negligence irrespective of whether it is a single examiner or multi-examiner evaluation. The key issues plaguing the evaluation of subjective answer-scripts are, negligent evaluation, intra and inter examiner heterogeneity, lengthy evaluation cycle time and enormous wastage of examination resources.

In this section, we discuss the manual answer-script evaluation process in conventional examinations in order to understand the vulnerabilities associated with the manual evaluation.

### 5.2.1 Manual Answer-scripts Evaluation Process

Answer-scripts evaluation in the manual form involves a variety of human intensive tasks [Bou00, VTK05] as listed below:

- Examiner first read the contents of answer contained in the answer-script.
- Examiner makes a judgement about the marks to be assigned to the answer based on the answer content.
- Marks are recorded in the margin of the answer-book. In some cases marks are recorded on altogether separate marks input form.
- Calculate the subtotal of marks for each main question.
- Transfer the total of each main question on front page of answer-book.
- Calculate and record the grand total of marks on front page of answer-book.
- Transfer the grand total of marks obtained by each examinee on the statement of marks.

### 5.2.2 Evaluation Anomalies

Examiners commit a variety of errors/lapses during evaluation of subjective answer-scripts. The frequency of errors increases, especially while evaluating a large number of answer-scripts. Some of the vulnerable points leading to errors are:

**Incomplete Evaluation**

Examiner inadvertently ignore some part of the answer-script during evaluation, resulting into partial evaluation of answer-script. This incomplete evaluation especially occurs when examinee answer the question paper in a random sequence.

**Erratic Evaluation**

Examiner allots the marks randomly. This may occur due to time pressure, fatigue or due to lack of standard evaluation guidelines. Sometimes examiner fail to assign the best marks when optional questions are involved.

**Totalling Errors**

Evaluation of subjective answer-scripts is a highly tedious task. In addition to that, performing calculations at the end of every main question repeatedly for long hours creates considerable amount of fatigue. The repeated exposure to numbers and calculations for long hours seldom introduces errors in calculation of question-wise total and/or grand total.

**Transferring/Recording Errors**

The act of transferring wrong marks from inside of the answer-book to the main page of the answer-book or from front page of answer-book to the course statement of marks is also quite common.

**Data Entry Errors**

Errors can also occur during reading and entry of marks from the course statement of marks to a computerized system for the final compilation of results.

### 5.2.3 Examiner Heterogeneity

Variability in evaluation, specifically, takes place due to various factors, such as, examiners who has failed to understand the content they are evaluating [BYC04] or examiners deviate from the criteria decided for evaluation [SC06] or examiners interpret answers altogether differently [WPJ00] or attach importance to different aspects of the answer [OW15] or examiners have altogether different expectations of standards [GPZ08] for the answer they are evaluating.

Some of the scenarios in which a wide range of variation in evaluation is observed is illustrated next.

#### Intra-examiner variation

In a single examiner evaluation of a large number of answer-scripts, evaluation is likely to take several days. When evaluation spans for several days, it is unlikely that the examiner would remember the quality of answers of previously assessed answer-books. During the process of evaluation, the standard of evaluation seldom remains constant. This is due to factors such as a large number of answer-scripts for evaluation, perceived pattern in the answer-script content, order of evaluation, time of the day, fatigue, time constraints, etc. This brings about variation in allotment of marks by the same examiner and is referred to as "intra-examiner variation".

#### Inter-examiner variation

The advent of a vast number of answer-scripts of a specific subject, necessitates multiple examiners for carrying evaluation. In such a situation, subjectivity of the respective examiner creeps into the evaluation. Inevitably, there arises a variance between the range of marks and the average marks awarded by each examiner. This variation in allotment of marks in a multi-examiner evaluation is called as "inter-examiner variation".

#### Hawk-Dove effect

The evaluation is also affected by 'Hawk-Dove effect' [MMT06]. Herein, an examiner who is strict is prone to assign less marks even to a well-written answer-script. While, an examiner who is liberal tend to allot more marks even to an average answer-script. This would result in unfair differences in marks allotment. Such an evaluation creates unfair differences in marks allotment and ultimately to incorrect ranking of examinees.

### 5.2.4 Wastage of Resources and Time

In conventional examination, there is enormous wastage of precious human and consumable resources as well as lengthy cycle time leading to delay in declaration of results.

#### Wastage of Resources

Consider an institution who conduct two main examinations annually for about 1500 examinees in altogether 7 course papers. For this, on an average 02 page question paper and 20 page answer booklet is required for every examination per examinee. This translates into 4,62,000 paper sheets, which is an enormous amount of wastage of paper. Institutions also need separate facilities for storing paper-based answer sheet. Fetching out an old answer sheet is also a very daunting task. Secondly, huge amount of human resources are required to perform all the examination related tasks such as printing, sealing, distribution, collection, logistics and operations, verification, re-verification and maintenance.

#### Lengthy Cycle Time

In order to ascertain error-free evaluation it becomes essential to perform re-checking/re-verification of every activity carried during evaluation. This involves

re-checking of entire evaluation, total marks and checking the entered marks into the computer system. The additional verification process does not add any value to the examination system; on the contrary consume additional resources and add delay in completing the result compilation tasks.

## 5.3 Approaches for Countering Evaluation Anomalies

The answer-scripts evaluation process described in section 5.2.1 suffers from several vulnerabilities that can badly jeopardise the fairness and reliability of examination. Therefore, there is a need for effective subjective answer-scripts evaluation practices to control the evaluation iniquities.

In this section, some of the frequently used approaches for countering the evaluation anomalies, are discussed.

### 5.3.1 Moderation of Answer-scripts

Moderation procedure is devised to ensure equitable treatment to all the candidates and to judge them on merit by reducing the 'examiner variability' to the extent possible [Blo09]. In moderation, the meeting of each team of subject examiners is arranged to thoroughly discuss the appropriate answers and marking scheme. A sample valuation of answer-scripts are carried by the examiners concerned and this evaluation is reviewed by the head examiner. The head examiner again carries evaluation of randomly selected answer-scripts to verify deviations, if any, in the evaluation of selected answer-scripts. In spite of the presence of agreed moderation norms, many examiners tend to deviate from the expected or agreed norms. This occurs, when their propensity for strictness or liberality or erraticism or carelessness during the course of the evaluation, overrides caution.

Many academic institutions address intra/inter examiner variation in evaluation with the help of moderation of assessed answer-scripts. In this process, one subject expert acts as a moderator. The moderator selects some answer-scripts at random, evaluated by each examiner. He evaluates them independently. Examiner 'X' needs to evaluate all the answer-scripts again, if major variations are observed. However, it is observed that having rigorous moderation procedures adds little to accuracy and reliability in evaluation, on the contrary, they delay the assessment and final grading [Blo09].

### 5.3.2 Scaling

Scaling techniques are used in many standard public and competitive examinations for controlling inter examiner variation in evaluation. The entire basis for applying scaling to marks allotted by different examiners in the same subject/course paper is under the assumption that:

1. The answer scripts sent to each examiner for valuation are drawn randomly.
2. Each batch possesses equal abilities.

When the distribution of abilities in each batch is approximately equal, the mean marks and standard deviation of the scaled marks of each batch will be identical [HMG76]. However, in reality, it is unlikely to get a batch of examinees with equal abilities. Therefore, application of scaling techniques to batch of examinees mechanically yield unreliable outcome. The scaling technique, also does not address or rectify the effect of strictness or liberality of the examiner. Scaling may, to a limited extent, be successful in eliminating the general variation which exists between examiners. However, it does not resolve the problem of examiner variability due to 'Hawk-Dove effect', i.e., the effect of strictness or liberality of the examiner [GC88].

### 5.3.3 Rubrics based Evaluation

The computer-assisted grading system using rubrics has revealed that it solves the examiner variation as it clearly identifies specific criteria to be assessed to achieve objectivity in the assessment [AK09]. There are various tools for computer assisted evaluation using rubrics such as for semi-automatic grading of programming courses [AR06] , grading of descriptive type examinations [RSS09], essay grading [WDASG11], standard summative examination answer-scripts grading [DKW14] along with software's such as Moodle (https://moodle.org/) and Blackboard (http://www.blackboard.com) for evaluation of essays, assignments and descriptive questions.

### 5.3.4 In-house verification

In-house verification is performed to detect and correct evaluation errors and lapses. In this method some percentage or the entire lot of evaluated answer-scripts are verified again for detecting and correcting the errors in evaluation.

In in-house verification independent verifiers, verify all the evaluated answer-books. Each verifier is assigned a certain number of evaluated answer-books based on the total number of verifiers available and the total number of answer-books. Each verifier needs to verify:

1. The total marks entered on main answer-book and course statement of marks.
2. Grand total calculation.
3. Sub-question total calculation.
4. Carrying of sub-question total on front of the answer-book.
5. Incomplete evaluation.

6. Marks assigned to optional questions.

At the end of the verification process, each verifier prepares the verification report indicating the different types of evaluation anomalies as discussed in section 5.2.2.

However, in-house verification is not suitable to address the lapses or intra and inter examiner variation in the evaluation [DKW14].

### 5.3.5 Personal verification

The personal verification approach provides an opportunity for the examinees concerned to independently verify the evaluated answer-scripts to detect any errors or lapses in evaluation. In personal verification the examinees concerned are permitted to personally check whether any answer is unassessed or to check any computational/ recording errors. However, personal verification does not address examiner subjectivity and variation [DKW14].

### 5.3.6 Re-evaluation

In re-evaluation the desired answer-scripts are evaluated again by an independent examiner. The earlier evaluations (marks) are blocked by sticking paper on them. This is to avoid any bias in the fresh evaluation by the new examiner. This method is used to verify the degree of variation between the two evaluations before considering the best evaluation. However, re-evaluation also is not suitable to control examiner subjectivity and lapses [DKW14].

## 5.4 Computer-Assisted Evaluation using Rubrics (CAER)

The proposed CAER solution, comprises of modules for defining the structure of question paper, performing the evaluation, compiling the results and providing feedback to the

examinees. The CAER solution allows the examiner to mark the examinee answers as per predefined rubric criteria's. The completion of computer-assisted marking automatically generates question-wise marks along with the grand total for each examinee. CAER eliminates human interventions and thus, provides an error-free environment in evaluation and result compilation [DK18].

### 5.4.1 Assumptions

The current study is on the pretext of the following assumptions:

**Assumption 1**

The examiners evaluate subjective answer-scripts through a computer-aided system comprising of predefined digital evaluation template.

**Assumption 2**

Evaluation involving large number of answer-scripts needing multiple examiners for each subject/course paper.

**Assumption 3**

Around 10% of the answer-scripts corresponding to each subject/course paper are independently evaluated by all the subject examiners.

### 5.4.2 Architecture and Implementation of CAER

The CAER is a web based application developed using HTML, CSS, JavaScript, jQuery and PHP (for the front-end interface) and MYSQL (for the backend) and served through a web server, APACHE. CAER comprises of three main modules: Pre-Evaluation, Evaluation and Post-Evaluation. The entire architecture of CAER along with the work-flow is shown in Fig. 5.1

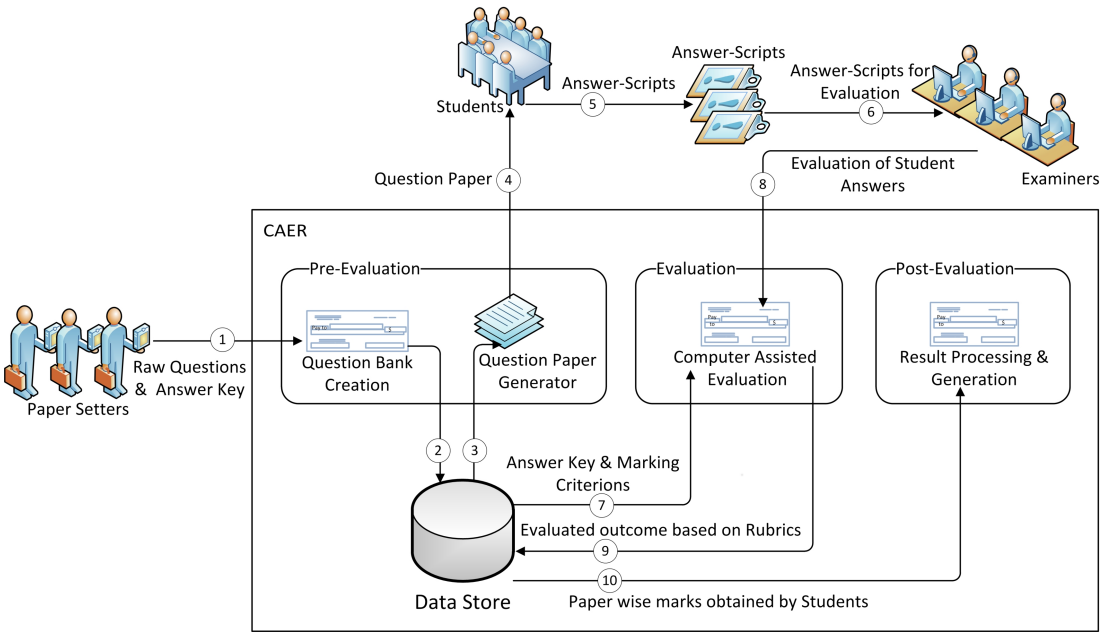



Fig. 5.1: CAER Architecture

### Pre-evaluation Module

The pre-evaluation module of CAER is designed to define the question paper format/structure along with actual question paper. Normally, the structure of a question paper is non-linear in nature involving multiple questions and sub-questions with options. A typical question paper may consist of the following format:

1. 'M' number of the main questions. All 'M' questions could be compulsory or examinees need to attempt any 'N' questions out of 'M'. Each question carries same weightage. Some other variations are also possible.

2. Usually each main question comprises of 'n' number of sub-questions. In this all 'n' questions could be compulsory or examinees need to attempt any 'p' questions from 'n' questions. Each question carries same weightage.

3. Choice between two questions, i.e., A or B. Each question carries equal weightage.

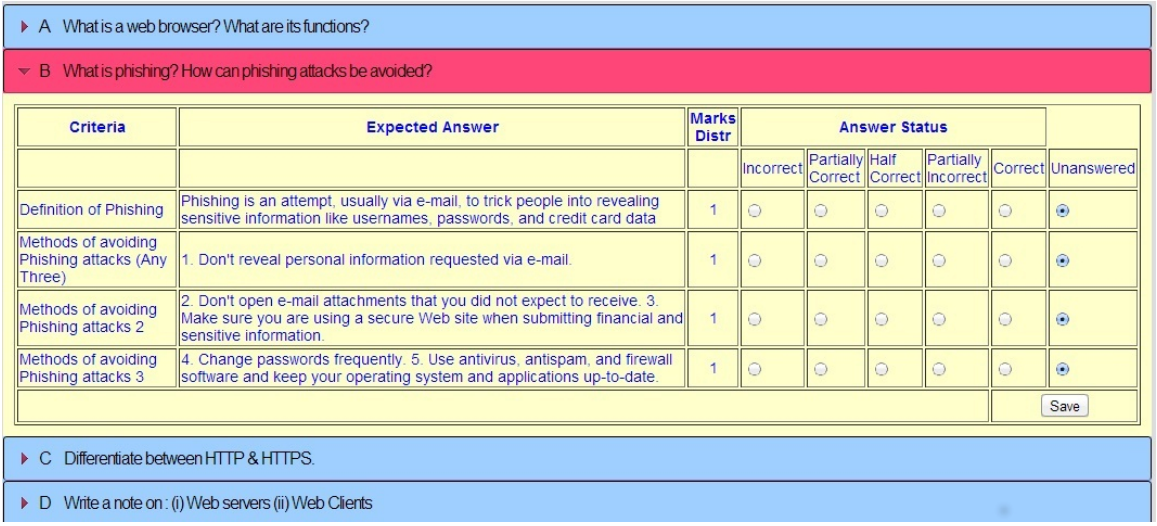



Fig. 5.2: Rubrics in CAER

Pre-evaluation module is also used to define the evaluation rubric for each question contained in the question paper. We considered **"Compliance of examinee answer to evaluation criteria"** as the evaluation rubric. The series of assessment criteria were defined for each question along with the marks weightage. The marking/evaluation criteria for each question and the corresponding marks are decided by the panel of subject experts. The parameters/answer status fully incorrect, fully correct, partially correct, partially incorrect, half correct and unanswered were set for marking each answer criterion.

The CAER solution was designed to assist examiners to calibrate evaluation of each answer within the framework as shown in Fig. 5.2. However, the examinees' creativity or flexibility is essentially not curtailed. The examiner does not assign any marks/grades during evaluation, he only identifies and marks with the mouse click the degree to which the required criterion/parameter is covered in the answer provided by the examinee.

**Evaluation Module**

Evaluation is the main module of CAER. This module is designed to mark the examinee answer-scripts based on the criteria defined in the pre-evaluation module.

Examiners pick up and read the examinee's handwritten answer-scripts and judge each answer against the predefined digital evaluation template comprising of criteria for each question vis-a-vis marking scheme. Degree to which the examinee answer corresponds to the expected answer criteria is marked on computer screen with a mouse click rather than the examiners' traditional red pen. The marking options available corresponding to each criterion are fully incorrect, fully correct, partially correct, partially incorrect, half correct and unanswered. The screenshot shown in Fig. 5.2 demonstrates the evaluation interface available to the examiners for answer-scripts evaluation.

The evaluation module records marks obtained by each examinee in each criteria of the question and then processes it further. For instance, if a particular answer criterion carries 2 marks and corresponding answer status is half correct than the system would automatically assign 1 mark to that answer criterion. Similarly, based on the answer status of the examinee proportionate marks is assigned automatically to each answer criterion. Thus, at the end of each answer-script evaluation, a digital copy of evaluated answer-book along with the marks obtained by each examinee, is available for further action.

**Post-evaluation Module**

The marks obtained by each examinee are already available in a digital format in the system based on on-screen evaluation carried by the examiner corresponding to the hand-written answer-scripts. The post-evaluation module of CAER handles processing of marks obtained and generation of the results. CAER solution totally eliminates the marks entry phase and associated data entry errors. The digital copy of each examinee's compliance with

predefined answer criteria is also available. This information if required can be made available to examinees for their reference and feedback.

### 5.4.3 Salient Features of CAER

Some of the salient features of the CAER are elimination of marking and calculation errors, reduction in examiner variation, elimination of marks entry and transparency.

#### Elimination of Marking and Calculation Errors

Evaluation of answer-scripts needs human intelligence to understand the quality of the answer presented by the examinee. The recording, totalling and comparing of marks can be done accurately through an automated process. The CAER handles with accuracy the tasks where human intervention is not required (refer section 5.6.1), thus eliminating the human errors in marks recording and calculations (see Fig. 5.3).

#### Reduction in Examiner Variation

The use of CAER provides examiners with well-defined evaluation criteria. When examiners evaluate answer-scripts within a defined framework, the scope for examiner variation is reduced considerably (refer section 5.6.2).

#### Elimination of Marks Entry

The system is capable of recording marks obtained by the examinee based on the criteria defined. At the end of the evaluation, CAER system is ready with course-wise marks obtained by each examinee, eliminating the marks entry task altogether.

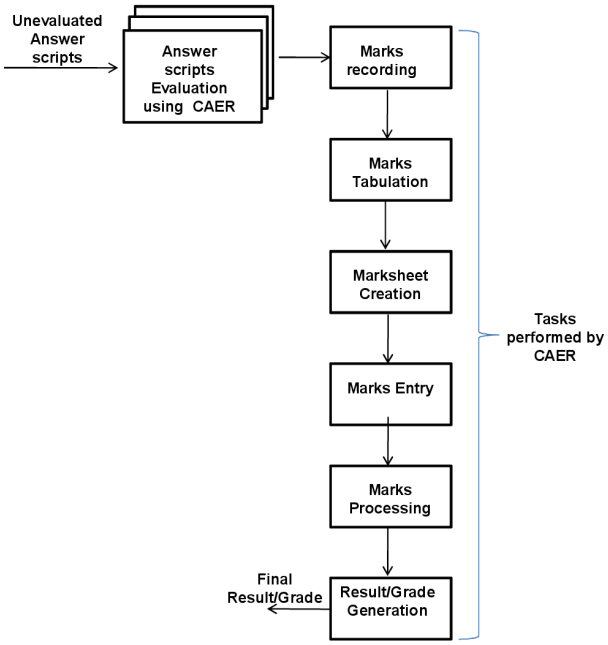



Fig. 5.3: Features of CAER

**Transparency**

In conventional evaluation, it is impractical to show the evaluated answer-scripts with feedback to all the examinees. In CAER the evaluation criteria and marking scheme is predefined. The degree to which an examinees' answers fits the predefined criteria is recorded in the system. This information can be easily provided to examinees for their reference together with any discrepancies observed during the evaluation.

## 5.5 Measurement of Efficacies of Evaluation

Educational institutions conduct Semester End Examinations (SEE) at the end of every semester to evaluate the performance of the examinees. The main corpus used in this study is based on the evaluation of answer-scripts from these examinations.

Initially, we identified the various types of lapses/errors committed by examiners during the

manual evaluation of subjective answer-scripts (refer section 5.2). We also explored the impact of CAER in controlling examiner variation and errors in evaluation. We resorted to data from In-house verification, personal verification, re-evaluation and intra/inter examiner evaluation as described in the following sections to measure the efficiency of CAER over manual evaluation.

### 5.5.1 Data from In-house Verification

We obtained the archived in-house verification (refer section 5.3.4) data of three years of about 5200 answer-scripts of each year from multiple higher educational institutions.

This instrument is used to quantify the lapses/errors that are committed by examiners during evaluation of answer-scripts.

### 5.5.2 Data from Personal Verification and Re-evaluation

All those examinees with grievances over the evaluation opt for personal verification/re-evaluation of answer-scripts (refer section 5.3.5 and 5.3.6).

We obtained the data from personal verification and re-evaluation of three years of about 120 cases opting for personal verification/re-evaluation in each year. We explored the errors/lapses and examiner variation in evaluation as discussed in section 5.2.2 and 5.2.3 along with the degree to which marks change during the personal verification/re-evaluation. This instrument is used to quantify the lapses/errors (refer section 5.2.2) or gross variation in the evaluation (refer section 5.2.3) of answer-scripts.

### 5.5.3 Data from Inter-Examiner Evaluation

We randomly selected about 80 answer-scripts pertaining to a particular course paper for evaluation. The corresponding question paper had altogether 25 questions with each question carrying a maximum of 4 marks and the total weightage of 80 marks. We utilized the services of three different examiners with similar experience and background for carrying out the evaluation. The evaluation was carried out by each examiner independently and no aspect of the evaluation was revealed to the other examiners to avoid any bias. Each examiner evaluated the same answer-scripts with manual evaluation as well as using CAER solution. This instrument was used to assess the inter-examiner variation (refer section 5.2.3) in manual and CAER.

### 5.5.4 Data from Intra-Examiner Evaluation

We randomly selected about 80 answer-scripts pertaining to a particular course paper for evaluation. The corresponding question paper had altogether 25 questions with each question carrying a maximum of 4 marks and the total weightage of 80 marks. The selected answer-scripts were evaluated on two different occasions by the same examiner using manual evaluation as well as CAER. We performed similar exercise with altogether three different examiners. We performed all these experiments without the knowledge of examiner concerned to avoid any bias in the evaluation.

This instrument is used to understand the intra-examiner variation (ref section 5.2.3) in evaluation in manual evaluation as well as CAER.

### 5.5.5 Evaluation Time

We identified the activities involved in subjective answer-scripts evaluation and result generation and recorded the time taken to complete each of the following activities:

1. Evaluation of answer-scripts
2. Marks entry on answer-book
3. Question-wise totalling and sub-totalling, grand totalling
4. Preparation of paper-wise statement of marks
5. Verification of evaluation
6. Marks entry into the computer system
7. Checking the entered marks

We calculated the total time required to evaluate each answers-script in manual evaluation and CAER. This instrument is used to measure the cycle time (refer section 5.2.4) in manual evaluation as well as CAER.

## 5.6 Results and Discussions

In this study, we analysed the different types of errors committed by examiners during evaluation of subjective answer-scripts based on the data obtained from in-house verification and verification/re-evaluation process of examination. This study also verified the inter and intra examiner variation in evaluation using the manual evaluation as well as CAER. We also compared the entire evaluation cycle time in manual evaluation and CAER. The following discussion will focus on the application, appropriateness and usefulness of the CAER

solution over manual evaluation from the examinees' perspective. The results obtained are discussed in the following sub-sections.

### 5.6.1 Performance of CAER in Reduction of Evaluation Anomalies

The main data forming the basis to measure the performance of CAER in reduction of evaluation anomalies are archived data of in-house verification (refer section 5.5.1), data of personal verification and re-evaluation (refer section 5.5.2) of answer-scripts.

The data from in-house verification served as an excellent tool to uncover the types of errors committed by examiners during evaluation of answer-scripts. The significant finding of this study was that, on an average 2% of the evaluated answer-books suffer from evaluation anomalies as discussed in section 5.2.2. Only subsets of such errors come to light in the absence of any in-house verification process and that too if the examinee concerned opts for verification/re-evaluation of answer-books. On an average 2% of the evaluated answer-scripts come for personal verification.

The analysis of the personal verification/re-evaluation of data indicated that at least 60% to 85% of the referred cases result into change in the marks. In this scenario, we can deduce that on an average at least 15% to 40% of the cases of evaluation anomalies go unnoticed due to absence of proper evaluation anomalies detection mechanism/tool.

The evaluation anomalies discussed in section 5.2.2 are automatically reduced/eliminated with the aid of CAER. The complete answer key in the form of well-defined rubrics, evaluation template along with marking scheme is predefined in CAER. CAER solution altogether got rid of manual marks assignment, totalling and marks entry. Examiner has to only mark extent to which examinee answer complied with expected answer. CAER solution relieved examiners from the task of marks assignment and calculations of sub-totals and the total and transferring marks from one document to another including any marks entry in the

computer system for result compilation.

The evaluation of answer-scripts with CAER provided examiners with evaluation template bearing all the questions contained in the question paper. If examiner inadvertently ignored evaluation of any answer, it will get detected due to absence of any marking in the evaluation template. Also, random sequence of answering the question paper will not put additional burden on the examiner as examiners role in CAER is to only perform focussed evaluation.

The presence of well-defined rubrics in CAER acted as a framework within which examiner had to carry evaluation. This provided limited scope for examiners to carry erratic evaluation. In addition to this, CAER system took care of assigning best marks when optional components were involved.

The CAER evaluation generated the digital copy of evaluated answer-scripts. This copy has potential to provide a reference/feedback mechanism for examinees to understand the defined criteria's for correct answers and marks weightage to each criterion vis-a-vis examinee performance in each defined criteria.

Thus, CAER solution provided significant improvement over conventional evaluation system by elimination of manual process of marking, totalling and data entry and associated data entry errors.

### 5.6.2 Performance of CAER in Reduction of Examiner Heterogeneity

This involved analysing the data from intra (refer 5.5.3) and inter examiner (refer 5.5.4) evaluation using manual and CAER solution.

The analysis of data from inter-examiner evaluation using manual approach indicated the evaluation variation to range from 0 to 75% whereas evaluation variation ranged from 0 to 25% in CAER (refer Fig. 5.4 and Fig. 5.5).

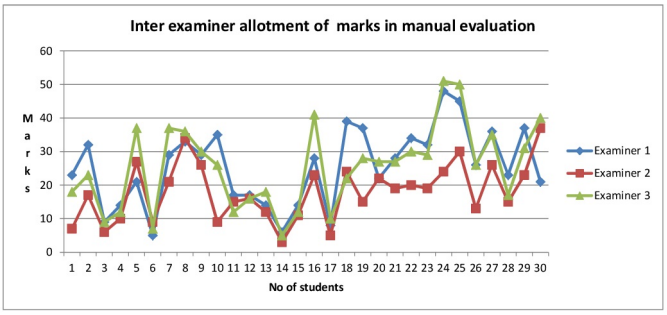



Fig. 5.4: Inter Examiner Variation in Manual Evaluation

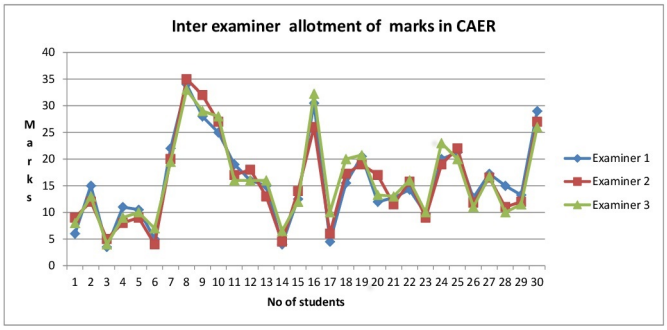



Fig. 5.5: Inter Examiner Variation in CAER Evaluation

A single measures, two-way mixed model, type absolute intra-class correlation coefficient (ICC) was used to calculate ICCs to measure the consistency of examiners evaluations. [KL16] classified the ICC as follows: values < 0.50 are considered poor consistency, between 0.50 to 0.75 are considered moderate consistency, between 0.75 to 0.90 is considered good consistency and value of ICC > 0.90 is considered excellent consistency. The values of ICCs for inter-examiner evaluation in manual approach are presented in table 5.1. It is observed that the 3 examiners exhibited moderate agreement among themselves in terms of how they carried the evaluation of answer-scripts as apparant from the single measures correlation coefficient of 0.636.

Table 5.1: Intraclass Correlation Coefficient (ICC) result indicating significant variation in Inter-Examiner evaluation in Manual approach

| | Intraclass Correlation[b] | 95 (%) Confidence Interval | | F Test with True Value 0 | | |
|---|---|---|---|---|---|---|
| | | Lower Bound | Upper Bound | Value | df1 | df2 |
| Single Measures | 0.636[a] | 0.350 | 0.812 | 9.141 | 29 | 58 |
| Average Measures | 0.840[c] | 0.618 | 0.928 | 9.141 | 29 | 58 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

The values of ICCs for inter-examiner evaluation in CAER are defined in table 5.2. It is observed that the correlation coefficient for single measures is 0.958 which suggest excellent reliability in evaluation of answer-scripts by each examiner. The results of manual and CAER evaluation provide sufficient evidence that the evaluation carried by using CAER is much consistent than the manual approach.

Table 5.2: Intraclass Correlation Coefficient (ICC) result indicating insignificant variation in Inter-Examiner evaluation in CAER

| | Intraclass Correlation[b] | 95 (%) Confidence Interval | | F Test with True Value 0 | | |
|---|---|---|---|---|---|---|
| | | Lower Bound | Upper Bound | Value | df1 | df2 |
| Single Measures | 0.958[a] | 0.925 | 0.978 | 68.435 | 29 | 58 |
| Average Measures | 0.986[c] | 0.974 | 0.993 | 68.435 | 29 | 58 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

A one way ANOVA was conducted to compare the inter-examiner variation in manual and CAER evaluation. The ANOVA result indicated that there was a statistically significant inter examiner variation in allotment of marks at the $p < 0.05$ in manual approach for three examiners (F (2, 87) = 4.735; sig. = 0.011) as shown in Table 5.3.

Table 5.3: ANOVA result indicating Inter-Examiner Variation in Manual and CAER Evaluation

| Method | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Manual | Between Groups | 1118.450 | 2 | 559.225 | 4.735 | .011 |
| | Within Groups | 10274.075 | 87 | 118.093 | | |
| CAER | Between Groups | 2.528 | 2 | 1.264 | 0.20 | .980 |
| | Within Groups | 5439.858 | 87 | 62.527 | | |

There was no statistically significant inter-examiner variation in allotment of marks at the $p < 0.05$ in CAER for three examiners (F (2, 87) = 0.020; sig. = 0.980) as shown in Table 5.3.

Post-hoc comparisons using the Tukey HSD test as shown in Table 5.4 indicated that there is a significant difference in evaluation between examiner 1 and examiner 2 (sig. = 0.022) & examiner 2 and examiner 3 (sig. = 0.028).

Table 5.4: Post Hoc Test using Tukey HSD for Determining Inter-Examiner Variation in Manual Evaluation

| (I) Examiner | (J) Examiner | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 7.600* | 2.806 | .022 | .91 | 14.29 |
| | 3 | .250 | 2.806 | .996 | -6.44 | 6.94 |
| 2 | 1 | 7.600* | 2.806 | .022 | -14.29 | .91 |
| | 3 | -7.350* | 2.806 | .028 | -14.04 | -.66 |
| 3 | 1 | -.250 | 2.806 | .996 | -6.94 | 6.44 |
| | 2 | 7.350* | 2. | .028 | .66 | 14.04 |

Note: *. The mean difference is significant at the 0.05 level.

The test indicated absence of statistically significant difference in evaluation between

examiner 1 and examiner 3 (sig. = 0.996).

Similarly, the range of intra-examiner variation in marks is from 0 to 50% in the manual approach as compared to 0 to 15% in CAER (Fig. 5.6 and Fig. 5.7).

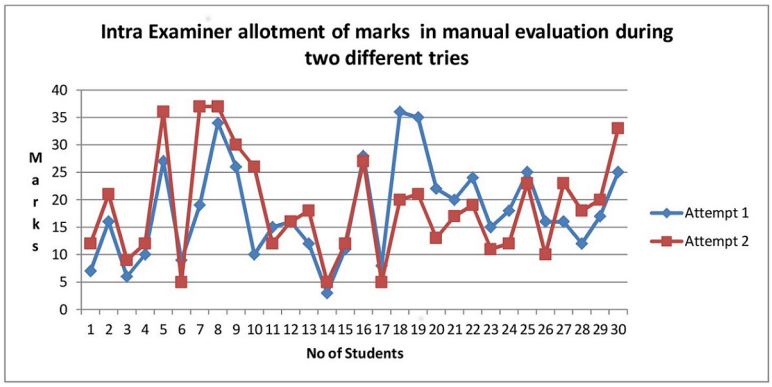



Fig. 5.6: Intra-Examiner Variation in Manual Evaluation

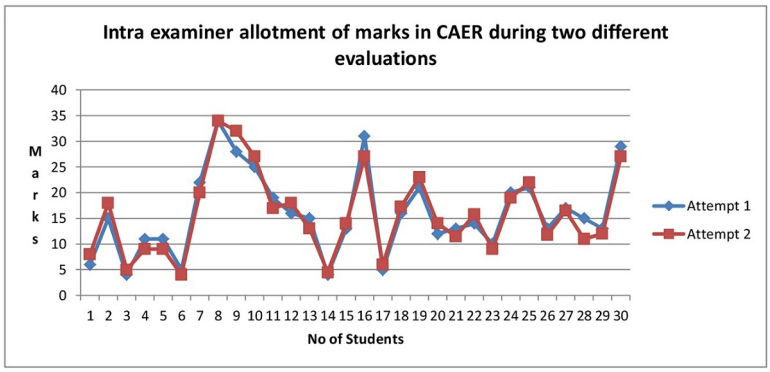



Fig. 5.7: Intra-Examiner Variation in CAER Evaluation

A paired-samples t-test was conducted to compare intra-examiner variation in marks allotment, with evaluation of same answer-scripts on two separate and independent occasions by the same examiner. There was a significant variation in marks allotment during each evaluation: first evaluation (M = 20.73, SD = 9.98) and second evaluation (M = 22.58, SD = 10.13) and t (60) = -2.74, p = 0.008 as shown in Table 5.5. These results suggest that the manual evaluation suffers from significant intra-examiner variation.

Table 5.5: Paired Samples Test Result indicating Intra-Examiner Variations in Manual and CAER Evaluation

| Method | Mean | Std. Deviation | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Manual Eval. Pair 1 Eval1 - Eval2 | -1.85246 | 5.27047 | -2.745 | 60 | .008 |
| CAER Pair 1 Eval1 - Eval2 | .10082 | 2.06578 | .381 | 60 | .704 |

When CAER solution is used, our results suggest that there is no statistically significant intra examiner variation in evaluation. The first evaluation indicated (M=22.85, SD=14.14) and second evaluation resulted into (M=22.75, SD=14.45) and t (60) =-.381, p = 0.74 as shown in Table 5.5. The result from Table 5.4 and Table 5.5 clearly indicates that the CAER solution effectively deals with inter and intra examiner variation in allotment of marks.

Further to this, there are a handful of studies confirming the intra and inter examiner variation in evaluation. This variation suggests that the technology can be used effectively in controlling examiner variation. According to [Sad09], some markers are characteristically generous, some are strict and others may be inconsistent.

### 5.6.3 Performance of CAER in Reduction of Evaluation Cycle time and Wastage of Resources

We recorded the average time required to complete evaluation and post-evaluation tasks per answer-script. We found that on an average in manual evaluation it takes altogether 560 seconds per answer-script to carry evaluation, entering marks, sub-totalling, totalling, recording marks, verification and entering marks and checking the entered marks into the computer system. In the manual examination system, the main tasks that add to waste in time are re-checking/re-verification processes involved in post-evaluation and result

processing. The additional verification process does not add any value to the system; on the contrary consume additional resources and add delay in completing the result compilation tasks. The time taken in percentage to complete each evaluation and result compilations task in manual approach is shown in Fig. 5.8

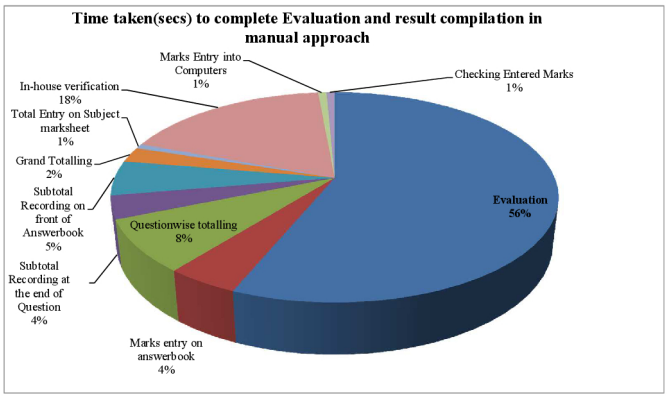



Fig. 5.8: Time in Manual Evaluation

On the other hand, it was observed that on an average it takes about 360 seconds per answer-script for evaluation and post-evaluation in CAER. In other words, CAER results in saving of about 250 seconds per answers-script. The human resources required to do the recording, totalling, preparing course statement of marks, verification of evaluated answer-books, marks entry, verification of entered marks are eliminated in CAER. This results in reduction of cost and subsequently a considerable amount of saving. The analysis of time consumed in evaluation and post-evaluations tasks in CAER along with percentage saving over manual approach reduces the time wastage by approximately 45% as summarized in Fig. 5.9

If we consider an average of 5220 answer-books for evaluation. CAER provides a saving of 250 seconds per answer-book. The total time saved in completing the evaluation/post-evaluation tasks is computed as follows:

Total time saved in days = 5220 * 250 seconds/60 * 60 * 24 days = 15 days Based on the above results, we can conclude that the result declaration process would take at least 15 days less, coupled with reduction in variations and errors.

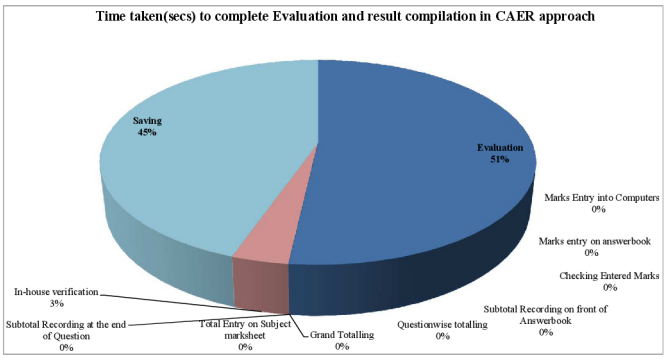



Fig. 5.9: Time saving in CAER

## 5.7 Summary

Summative subjective examinations are the predominant method adopted by institutions worldwide to evaluate examinees. However, the current examination systems suffer from a variety of lacunae at each stage of the examinations. The rise in the number of grievances received for verification and re-evaluation is an indicator that examinees have doubts about the current practices of evaluation. The current study undertaken reveals that these doubts are not completely unfounded. It is indeed true that evaluation suffers from the malaise of large-scale errors. The impact of 'examiner subjectivity' or 'examiner variability', also exists. The proposed CAER solution shows considerable promise in bringing improvement to the current practices of evaluation by controlling evaluation errors, examiner variability and wastage of time.

# CHAPTER 6

# E-Moderation for Detection and Correction of Evaluation Anomalies

In a typical public examination system, there is a huge quantum of answer-scripts. Therefore, it is necessary to distribute the answer-scripts among several examiners for evaluation. Each examiner has to evaluate a large number of answer-scripts pertaining to the subject/course paper. In such a system, variation in evaluation is bound to occur due to over-sight, differences in frame of reference/perception or simply due to the human error. Inevitably, therefore, even when the examiners with similar experience and background, evaluate large number of answer-scripts, there is an intra/inter examiner variation and errors/lapses in the evaluation.

Some of the currently used approaches such as moderation of answer-scripts, in-house verification, personal verification, re-evaluation of answer-scripts and scaling of marks (refer section 5.3 of Chapter 5), only provide cursory relief from anomalous and heterogeneous evaluation. This is apparent from the alarmingly increasing cases of verification/re-evaluation resulting into significant changes in the original marks. In this Chapter, we use machine learning techniques to classify each answer evaluation as negligent or normal and further predict the tuned marks to control the 'examiner heterogeneity' in the subjective answer-scripts evaluation. There are six sections included in this Chapter. Section 6.2, examines the application of machine learning techniques in education with a focus on examination and assessment. Section 6.3 describes the process of evaluation of answer-scripts along with the proposed E-moderation scheme. Section 6.4 provides the research methodology used in assessing the effectiveness of the proposed E-moderation

approach. Section 6.5 describes the two phases of E-moderation, namely, evaluation classifier and marks tuning. Section 6.6 validates the performance of the evaluation classifier and the marks predictor using a variety of evaluation metrics.

## 6.1 Introduction

Evaluation of subjective answer-scripts is a highly human intensive task. It needs focused and unbiased intervention of human resources such as examiners, moderators and verifiers. However, in the manual evaluation of a large number of answer-scripts, achieving consistency and uniformity with precision and perfection is a tall order.

In the public examination system, involving large number of examinees, answer-scripts pertaining to each subject/course paper are invariably distributed amongst several examiners for evaluation.

When several examiners are involved in evaluating the answer-scripts of a particular subject/course, then there is bound to be subjectivity. Moreover, each examiner's perception of what marks should be allotted to a particular answer is different. Additionally, some examiners tend to be liberal in marking whereas others are strict in evaluation. It implies that a liberal examiner will allot marks generously, giving rise to 'enhanced evaluation'. While, the strict examiner will assign marks frugally, resulting into 'reduced evaluation'. This phenomenon is termed as 'Hawk-Dove effect' [MMT06]. Such a situation can lead to a mediocre answer-script being assigned high marks and an excellent answer-script being assigned lower marks than the first. In other words, subjectivity of the examiners result in a wide difference between the marks assigned, whereby, examinees receiving 'reduced evaluation' inadvertently get detrimental ranking. Also, as the evaluation progresses, the same examiner can increase or decrease the standard of evaluation due to various factors such as improved understanding of the subject, the order of evaluation, time of the day, time

constraints, fatigue, etc. This leads to variation/negligence in allotment of marks, better known as 'intra examiner variation'. Apart from this, there are also instances of gross negligence/lapses in evaluation as apparent from the alarmingly increasing cases of personal verification/ re-evaluation converging into significant changes in the marks allotted during the initial evaluation. Therefore, there is a need to evolve a mechanism to ensure uniformity within the examiners so that the effect of 'examiner subjectivity' or 'examiner variability' is minimized.

This chapter is built on the knowledge garnered in Chapter 5. We performed the evaluation of the subjective answer-scripts pertaining to some of the public examination systems using evaluation system similar to CAER. We then used machine learning techniques to classify each evaluation as negligent or normal and further predicted tuned marks in an attempt to control the heterogeneity observed in the intra and inter examiner evaluation. Thus, the contribution of this chapter is twofold as listed below:

1. Use Support Vector Machine (SVM) classifier to classify the given evaluation as negligent or normal based on key evaluation parameters
2. Build the marks tuning system using Artificial Neural Network (ANN) to predict the normalized marks based on variation in evaluation

Marks tuning system proposed in this chapter is specifically designed for controlling examiner heterogeneity. The CAER solution proposed in Chapter 5 is capable of controlling evaluation anomalies as well as examiner heterogeneity. Having worked on both the methods, it appears that using CAER for controlling evaluation anomalies along with marks tuning for reducing examiner heterogeneity may yield excellent results.

## 6.2 Machine Learning Techniques in Assessment

Institutions conducting public examinations need to manage large collection of data pertaining to examinees and their academic performance. Very often, it becomes necessary to extract useful information from these huge chunks of data. In recent years, there is a transformation in the traditional processes of prediction leveraging the more sophisticated machine learning techniques to help reduce 'forecast errors' and eliminate unnecessary budgeting and planning.

### 6.2.1 Overview of Machine Learning Techniques

Machine learning techniques are commonly divided into three categories according to their purpose.

#### Supervised Learning

Supervised learning algorithms build a model by taking a known set of input data and known responses to the data (output/target) and train a model to generate predictions for the response to new data. The training process continues until the model achieves a desired level of accuracy on the training data. The learning acquired by the model in the training phase is checked on new data during the testing phase. Supervised algorithms can provide solutions to two broad types of problems, namely, classification (predicting discrete values) and regression (predicting continuous values).

1. **Classification**

   Classification algorithms are designed to predict the target class of discrete nature, based on input data. The target data is generally represented in a categorical form and represents a finite number of classes. Some of the machine learning algorithms for solving classification problems are decision trees, logistic regression, Naive Bayes, k

Nearest Neighbours (k-NN), Support Vector Classifier (SVC), etc. Some of the prominent applications of machine learning techniques in classification are:

- Filtering of emails as 'spam' or 'ham' using SVC [GC09, AB10].
- Prediction of tumour is 'malignant' or 'benign' [CW06, Aka09]
- Classification of images [FM04].
- Recognition of Protein Folds [DD01].

2. **Regression**

   Regression is a technique that uses a set of input data to predict the data of a continuous nature. Some of the machine learning algorithms for solving regression problems are Linear Regression, Regression Trees (e.g. Random Forest), Artificial Neural Network (ANN), Support Vector Regression (SVR), etc. Some of the applications of machine learning techniques in this area are:

   - Predict the future stock price based on current price and crucial market parameters [DFDSMD14].
   - Time-series forecasting. [QZR+14, Wei18].
   - Education and economic growth [BZ14].
   - Intelligent E-marketing campaigns [GCRAS15].

**Unsupervised Learning**

In unsupervised learning, the system is presented with unlabelled, uncategorized data. The unsupervised algorithms process the data without any prior training. Algorithms are left to their own formulations to discover and present the interesting structure in the data. Some of the unsupervised learning algorithms are K-means clustering, Neural Networks, Apriori algorithm for association rule learning problems.

Unsupervised learning problems can be further grouped into clustering and association problems.

1. **Clustering**

   Clustering is the task of dividing the population or data points into a number of groups in such a way that the data points in the same groups are similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a segregation of objects on the basis of similarity and dissimilarity between them.

2. **Association**

   Association analysis is the task of finding interesting relationships between various items or elements in large data sets. The hidden relationships are then expressed as a collection of association rules and frequent item sets. Frequent item sets are simply a collection of items that frequently occur together. Association rules suggest a strong relationship that exists between two items. The goal is to find associations of items that occur together more often than you would expect from a random sampling of all possibilities.

**Reinforcement Learning**

Reinforcement Learning is an area of machine learning which allows the machine or software agent to learn its behaviour based on feedback from the environment. This behaviour can be learnt once and for all, or it keeps on adapting as time goes by. The agent receives rewards by performing correctly and penalties for performing incorrectly. The agent learns without intervention from a human by maximizing its reward and minimizing its penalty.

As an agent, which could be a self-driving car or a program playing chess, interacts with its environment, it receives a reward state depending on how it performs, such as driving to

destination safely or winning a game. Conversely, the agent receives a penalty for performing incorrectly, such as going off the road or being checkmated.

### 6.2.2 Related Work in Assessment

We examined the literature to explore the use of machine learning techniques in the domain of education. Majority of studies have investigated issues related to prediction of performance of examinees, prediction of dropout and retention and improvement of assessment.

There have been studies using machine learning techniques such as support vector machines and neural networks for identifying the relationship between the examinee grades and past performance or demography as indicators of prediction of current performance [KPP04, HF13, DC13, RZPBK12, BP12]. Similarly, work carried by [CVNM07] used rule-based systems to predict examinee performance in an e-learning environment using fuzzy association rules. A multilayer perceptron topology is used for predicting the likely performance of a candidate being considered for admission into the university [OACO08].

Predicting dropouts is an important and challenging task for educational institutions and policymakers. Many researchers have explored the use machine learning techniques for predicting the likelihood of dropout/retention of students in a course of study relying mostly on academic performance, demographic, and financial data. In order to classify the dropout students, various approaches such as k-NN, Decision Tree (DT), Naive Bayes (NB) and Neural Networks (NN) have been successfully applied [YOT14, Del10, DGG$^+$12, Guo10].

In a study conducted by [LGM$^+$09], neural networks were used to cluster examinees into two groups based on the results of previously conducted tests. The groupings helped instructors address the specific needs of each group and adapt examinee training accordingly. There are some other studies where several classification algorithms have been applied for classifying students into groups such as passing or failing [HV06, SVM06]. In another work, learning

analytics is utilized to identify groups of students with similar patterns of performance and engagement. A tailored appraisal is provided to each group to help them master effectively the learning objectives of the course [KC17]. The work carried by [LAS+15], outlined an extensive framework that uses machine learning approaches to identify students who are at risk of not graduating high school on time. In addition to these studies, an extensive research deals with testing students' satisfaction level as well as construction of sophisticated measures of assessment [LLM12, WB13, BBE+10].

There is always a dilemma about how many variables one must include to achieve accurate prediction. According to [LGM+09, HF13], adding more predictor variables does not help in improving the average prediction accuracy for prediction of performance. However, the neural networks method leads to better prediction results compared to those of the normal regression analysis method [LGM+09]

Complimentary to the issue of improvement of assessment is creation and management of fair and error-free evaluation system. Research in this domain indicated that there is a dearth of work in solving the evaluation efficacies, more specifically the following issues need to be addressed:

1. Identification of evaluation as negligent or normal using key evaluation effecting parameters.

2. Adopting an uniform scale of valuation in multi-examiner evaluation for ascertaining uniformity and fairness in evaluation for each examinee.

## 6.3 Process of Evaluation with E-Moderation

The methods that are employed in evaluation need to be consistent, fair and error-free for all the examinees. Each examinee needs to be evaluated with the same scale and criteria. In

order to assess the quality of evaluation and establish uniformity in evaluation, we propose a solution in the form of E-moderation scheme comprising of two parts:

1. Classification of the evaluations carried by each examiner as either negligent or as normal.

2. Predicting the tuned marks, for each examiner in multi-examiner evaluation for controlling the intra and inter-examiner variation in evaluation.

The entire process of evaluation of answer-scripts along with the proposed E-moderation scheme is illustrated in Fig. 6.1. The E-moderation scheme defined in this chapter is on the

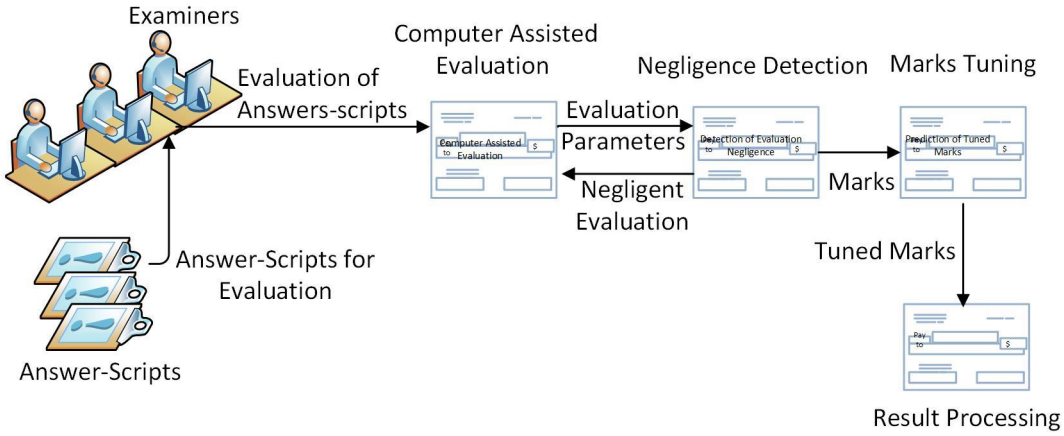



Fig. 6.1: Evaluation cycle in E-moderation scheme

identical assumptions as listed in section 5.4.1 of Chapter 5.

### 6.3.1 Detection of Negligent Evaluation

Initially, examiners are provided with computer-assisted evaluation system. This system helps in carrying evaluation of the subjective answer-scripts and recording the marks. The evaluation system is also designed to record other essential parameters such as evaluation

start and completion time, total time taken for evaluation of each answer, sequence of evaluation, etc.

We used machine learning techniques to segregate an evaluation as negligent or as normal depending on the features/patterns observed in the evaluation. We took into consideration, the actual time taken for evaluation and time required for evaluation based on examiner reading speed and comprehension accuracy. Similarly, we consider the actual marks assigned by the examiner for each answer and the maximum marks each answer deserves. Simultaneously, we also obtained the status of each evaluation as normal or negligent from the expert moderator to fine tune our prediction system. A given evaluation is considered as negligent if,

1. There is an apparent variation between the actual time taken for evaluation and examiner reading speed and comprehension accuracy.
2. There is an apparent variation in the marks assigned vis-a-vis number of words contained in the answer.
3. Expert moderator rating confirms that evaluation is negligent.

All the evaluations which fail in the above tests are classified as negligent and such evaluations are subject to corrective action. Then, a SVM classifier is used in order to classify the given evaluations as negligent or normal.

### 6.3.2 Prediction of Tuned Marks

When multiple examiners evaluate the answer-scripts related to a particular subject/course paper, each examiner tends to apply his own yardstick to assess the answer-scripts, resulting in inter examiner variation in evaluation.

The intra/inter examiner variation can be controlled by adjusting the marks assigned by respective subject examiners on one scale adopted by any one examiner. This is a regression

problem and we need to predict the tuned marks for each examiner to fit the evaluation of each subject examiner to the evaluation scale of the master subject examiner.

We apply ANN regressor on evaluations carried by different examiners pertaining to a particular course paper and predict the marks as if one examiner had evaluated all the answer-scripts. In this way, the entire evaluation is normalized onto one common scale to minimize the subjectivity of the examiner.

## 6.4 Design of E-moderation Model

The following are the goals of E-moderation model:

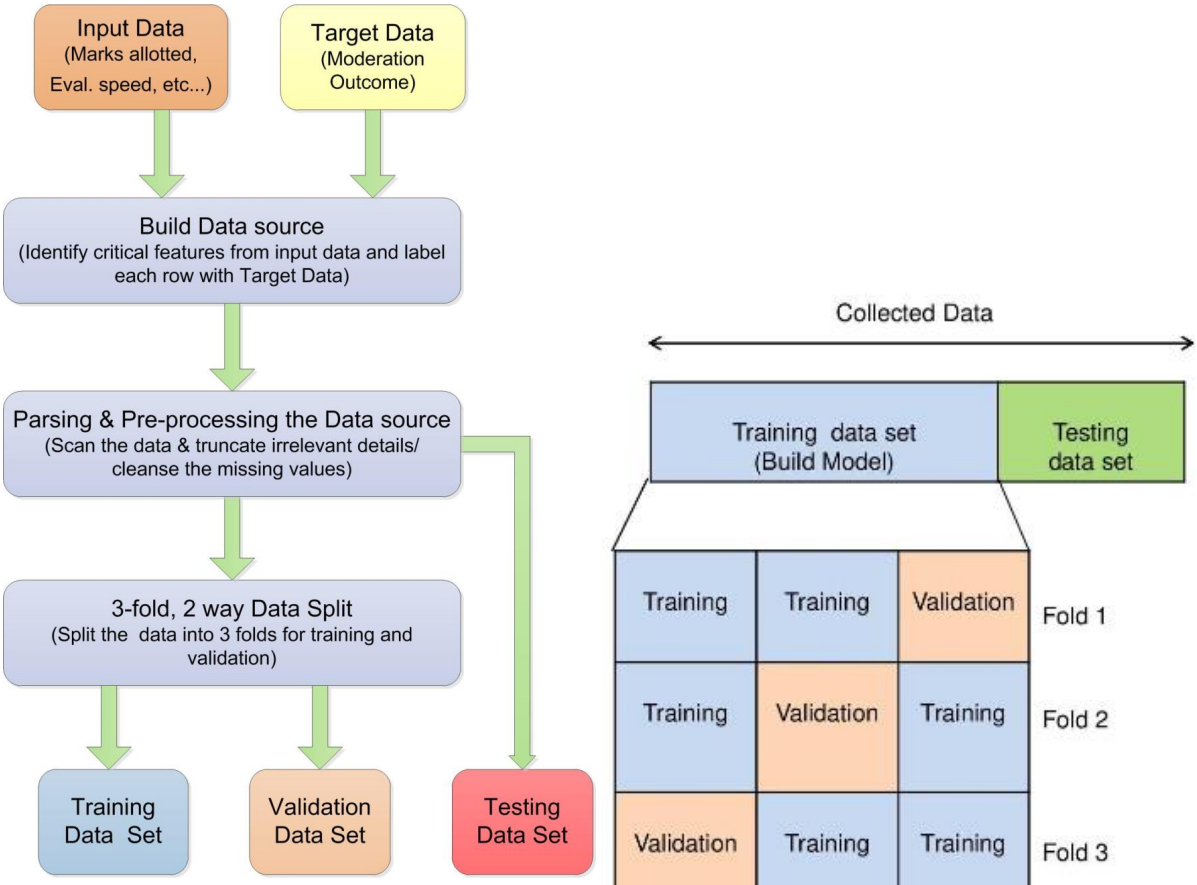
1. To analyse the evaluations carried by each examiner and classify it as negligent or normal.
2. To build a marks tuning system to map evaluation carried by each examiner on one common and normalized scale.

Thus, our E-moderation model comprises of two phases, i.e., detection of answer-scripts evaluation as normal or negligent and marks tuning. The whole design process of the proposed E-moderation model addressing the goals is explained in Fig. 6.2a.

### 6.4.1 Input Data

In the semester pattern of education, the Semester End Examinations (SEE) are conducted at the end of every semester to evaluate the performance of the examinees. The main corpus used in this study is based on the evaluation of answer-scripts from these examinations.

We designed custom answer-scripts evaluation recording system for aiding the examiners in evaluation and for recording the essential evaluation parameters. We used an online reading




(a) An Overview of General Steps Necessary for Performing E-moderation Tasks

(b) k-fold cross-validation (k=3)

Fig. 6.2: E-moderation model

and comprehension software (http://www.freereadingtest.com/ ) for keeping track of the actual reading / comprehension speed and accuracy of each examiner. Some percentage of evaluated answer-scripts were moderated by expert moderator.

We randomly selected about 100 answer-scripts pertaining to one course paper offered in undergraduate curriculum for evaluation. The corresponding question paper had altogether 25 subjective questions. Each question carried a maximum of 4 marks. The total weightage of the paper was 80 marks. The evaluation of answer-scripts was done by four examiners separately. The examinee identity was coded to prevent bias while evaluating. Similarly, to ensure independent evaluation, the marks awarded by each examiner was kept a secret from

other examiners. The entire evaluation produced a corpus of about 8000 evaluation records with roughly 2000 evaluations per examiner.

We identified the following critical features, considered to have direct or indirect influence in deciding whether evaluation carried is negligent or not.

- **Marks**

  The actual marks assigned by the examiner to the evaluated answer vis-a-vis the maximum marks designated to the question concerned.

- **Words**

  The total number of words contained in each answer corresponding to the minimum number of words expected in the answer. There is a relationship between the weightage of the marks allotted to a question and the quantity of content expected in the answer. e.g., a question which needs 6 to 8 lines for a suitable answer is answered in only 5 words. If in the evaluation, the answer is assigned above average marks, it creates suspicion. We need to formulate rules which help us in deciding whether evaluation is carried with negligence or not.

- **Evaluation Time**

  The total time taken by the examiner to evaluate each answer from the answer-script.

- **Reading and Comprehension**

  The normal reading speed of the examiner and reading speed of the examiner with a comprehension accuracy of at least 75% is considered. The amount of time taken for evaluation and actual reading speed of the examiner can be verified. This verification would ascertain, whether the examiner concerned, devoted fair time to evaluate the answer-script or not. e.g., if the reading speed of a particular examiner is 200 words per minute and the examiner spent only about 15 seconds. in evaluating an answer containing 130 words. Such an evaluation would definitely be unreliable.

- **Feedback/Rating**

  Moderators evaluate some percentage of the answer-scripts evaluated by each examiner 5.3.1. Moderator is a best person to assess the quality of evaluation of each examiner on a common scale. This information can be used in conjunction with other parameters to increase the correctness and efficiency of the evaluation model to some extent.

### 6.4.2 Target Data

A target vector is prepared for determining the E-moderation outcome using the critical features identified. We used categorical target labels for negligence detection as it is a binary classification problem. We used continuous data for marks tuning as it is a regression problem.

#### Target data for Detection of Negligent Evaluation

The evaluation is considered as negligent, if target=1 and normal, if target=0. The target prediction labels for training the model is based on relative marks ($\Delta m$) , relative evaluation time ($\Delta t$) and expert moderator's rating. The relative marks ($\Delta m$) is defined as the difference between the marks allotted by the examiner and the predicted marks. Thus, the relative marks ($\Delta m$) corresponding to each answer is defined as:

$$\Delta m = m_a - \frac{(w_a * m_{\max})}{w_{\min}} \tag{6.1}$$

s.t. $\Delta m = m_{\max}$ , if $(\Delta m) > m_{\max}$

Where $m_a$, represent the actual marks allotted by the examiner to the given answer $w_a$, represent the total number of words in a given answer

$m_{\max}$, represent maximum marks allotted to the question

$w_{\min}$, represent the minimum number of words expected in the answer concerned

Similarly, the relative time corresponding to each evaluation is defined as the difference between actual time taken for evaluation by the examiner and the expected time for evaluation based on examiner reading speed and comprehension accuracy. Thus, the relative time corresponding to each evaluation is defined as:

$$\Delta t = t_a - t_e \tag{6.2}$$

Where, $t_a$ represents actual time taken for evaluation and

$t_e$, represents expected time for evaluation.

In our proposed binary classifier for negligence detection, we defined target data for training purpose using the following inequalities

Negligent inclination, if $(\Delta m - m') > 0$ or $(\Delta t + t') < 0$

Normal evaluation, if $(\Delta m - m') \leq 0$ and $(\Delta t + t') \geq 0$

where m' and t' represent an acceptable level of marginal variation in marks allotted and evaluation time with respect to the corresponding computed values. i.e., if the total marks assigned is totally inconsistent with the content of the answers-script or if the total time taken for evaluation is much less than the actual time required for the evaluation, evaluation is considered as negligent. This labelling is cross checked with the expert moderator before finalizing the target labels for increasing the efficiency of the system.

**Target data for Marks Tuning**

The evaluation of answer-scripts pertaining to each subject/course paper are carried by multiple examiners. One examiner per subject/course paper is designated as a master evaluator. The target data for marks tuning stage during training process is the evaluation carried by the master evaluator. Our goal is to fit the marks assigned by the other

subject/course paper examiners onto the evaluations of master evaluator selected in the training process. In other words, we intend to uniformly tune the evaluations carried by each examiner into the evaluation of the master evaluator.

### 6.4.3 Parsing and Data Pre-processing

We selected those features that could contribute most to the prediction variable, filtering out irrelevant attributes from the data source. The following cases are considered during a pre-processing stage:

- Candidates do not attempt each and every question from the question paper. Corresponding to the un-attempted questions, the evaluation row is likely to remain blank. All such rows were truncated as missing data to prevent any undesired effect.

- The original data source was initially partitioned into 70% of training and 30% test data sets. This was performed before any other normalization/transformation step to prevent the contamination of the training data or leakage of test data into training data.

- The main features proposed for evaluation negligence detection are time and marks. Since both these features are measured on different scale, these features need to be rescaled. There are two methods for rescaling such features, namely, standardization and normalization. We preferred the standardization method for rescaling time and marks so that these data units follow the standard normal distribution with $\mu$=0 and $\delta$=1, where $\mu$ is the mean (average) and $\delta$ is the standard deviation from the mean. Standardization puts all features into similar range and no feature overshadows others due to heterogeneous scale.

- Multiple features are identified for predicting the negligence in evaluation. The Model was simplified by application of Applied Principal Component Analysis (PCA) that

transforms the multi-dimensional space into 2 dimensional linear combination.

- All those evaluations identified as negligent in the evaluation negligence detection phase are verified and necessary corrective steps are taken before feeding into the marks tuning phase.

### 6.4.4 k-fold Cross Validation

70% of partitioned train data set is further split into a train and a validation partition using 3-fold cross validation technique as shown in Fig. 6.2b for training and building the model. We used stratified sampling for splitting this dataset into 60% of the training and 40% for validation. The model is trained with all but one of the folds, and predictive performance measured on the part left out in the training process. The best model is selected based on the cross-validation error.

## 6.5 E-Moderation Phases

E-moderation model is comprised of two main phases, i.e., evaluation classifier and marks tuning. In the first phase, we identify and segregate the anomalous evaluation. We take corrective measures on all the anomalous evaluations and then submit it to the second phase for tuning the evaluations carried by each subject examiner onto one common scale.

### 6.5.1 Phase 1 - Evaluation Classifier

The objective of this phase is to classify the given evaluation as negligent or normal with the aid of input parameters.

Initially, we selected 60 answer-scripts for building the model. Each of the selected answer-script was evaluated by all the subject examiners concerned with the help of the computer-assisted evaluation system. Each examiner was blinded to evaluations carried by other examiners. The classification model was built and trained using evaluation parameters such as relative marks $\Delta m$ and relative evaluation time $\Delta t$ (refer Section 6.4.2 for details) as input and rating of expert moderator as a target. We followed the steps listed in algorithm 1 for training and validating the model.

**Algorithm 1** Detection of Negligent Evaluation

**Input:** Data Set for training and validation and testing(S)
**Output:** Evaluation classifier

1: Adopt k fold cross validation strategy (k=3) for splitting the input data set into 70% for the training and 30% for validation data set.
2: Obtain 3 subsets, each for training, $T = t_1, t_2, ..., t_k$ and validation $V = v_1, v_2, ..., v_k$
3: Initialize the SVM classifier, SVC with kernel function= RBF (Radial Basis Function), soft margin parameter (C)= 2 and selection parameter ($\Upsilon$) = 1. (Note: These parameter values are optimal values based on the grid search conducted using scikit-learn's GridSearchCV.)
4:
5: **for** $i = 1$ to $k$ **do**
6: Rescale $t_i$ and $v_i$ on a common scale using standardization technique.
7: Transform $t_i$ and $v_i$ into two dimensional space using PCA.
8: $m_i \leftarrow$ Build model by applying SVC to $t_i$.
9: $c_i \leftarrow$ Generate first level evaluation classifier by applying the model $m_i$ on validation data $v_i$.
10: **end for**
11: Generate the evaluation classifier c' by applying the best model $m_i$ obtained in step 8 on test data(S).

In the testing stage, we assigned the remaining 40 answer-scripts to each of the subject examiners. Each examiner evaluated all these answer-scripts with evaluation carried by each examiner ignorant about the other examiners, to avoid any bias in the evaluation. The evaluation classifier obtained in the training phase was applied to the evaluations carried by each examiner in the testing stage to classify the evaluation as negligent or normal. The accuracy of the evaluation classifier was verified by comparing the classified evaluation to

the classification carried by the expert moderator. In this phase, we applied the Python based Support Vector Machine (SVM) classifier using scikit-learn's SVM library. We used it to generate a classification model for classifying the evaluations as negligent or normal.

### 6.5.2 Phase 2 - Tuned Marks Predictor

The objective of this phase is to predict the tuned marks corresponding to the evaluations of each examiner. The input to this phase are the evaluations obtained from phase 1 with all the negligent evaluations detected and corrected.

One examiner per subject/course paper was selected as a master evaluator. The evaluation carried out by the master evaluator was moderated for correction of any evaluation anomalies. The prediction model was built and trained using the evaluations of each examiner as input and evaluation of the master examiner as a target. We followed the steps listed in algorithm 2 for training and validating the model.

In the testing stage, we assigned the remaining 40 answer-scripts to each of the subject examiner, including the master evaluator for evaluation. Each examiner evaluated all these answer-scripts, with evaluation carried by each examiner fully blinded from the other examiner to avoid any bias in the evaluation. The tuned marks predictor obtained in the training phase was applied to the evaluations carried by each examiner in the testing stage to obtain the tuned marks. The accuracy of the tuned marks predictor was verified by comparing the predicted tuned marks to the marks assigned by the master evaluator.

In this phase, we applied the python based Artificial Neural Network (ANN) learner using scikit-learn's neural_network library. We used it to generate a regression model for predicting the tuned marks.

**Algorithm 2** Prediction of Tuned Marks

**Input:** Data Set for training, validation and testing (S)
**Output:** Tuned Marks Predictor

1: Adopt k fold cross validation strategy (k=3) for splitting the input data set into 70% for the training and 30% for validation data set.
2: Obtain 3 subsets, each for training, $T = t_1, t_2, ..., t_k$ and validation $V = v_1, v_2, ..., v_k$
3: Initialize the MLPRegressor using perceptron architecture, with 10 hidden layers, an alpha value of 0.01, a stochastic gradient descent (sgd) solver, an adaptive learning rate and a logistic activation function. (Note: These parameter values are optimal values based on the grid search conducted using scikit-learn's GridSearchCV.)
4:
5: **for** $i = 1$ to $k$ **do**
6: Rescale $t_i$ and $v_i$ on a common scale using standardization technique.
7: Transform $t_i$ and $v_i$ into two dimensional space using PCA.
8: $m_i \leftarrow$ Build model by applying MLPRegressor to $t_i$.
9: $c_i \leftarrow$ Generate first level tuned marks predictor by applying model $m_i$ on validation data $v_i$.
10: **end for**
11: Generate the tuned marks predictor p' by applying the best model, $m_i$ obtained in step 8 on test data (S).

## 6.6 Results and Discussions

Evaluation of answer-scripts is the most crucial activity in high stake summative examinations. However, the intra and inter examiner variation in evaluation coupled with evaluation anomalies in the form of errors/negligence adversely affects the reliability and robustness of the summative examination. Identification of the anomalous evaluation and a tuning of the marks of all the examinees on a common scale in the event of intra/inter examiner variation is extremely necessary. The current study is undertaken to detect negligent evaluation and also predict the tuned marks to control intra/inter examiner variation in evaluation.

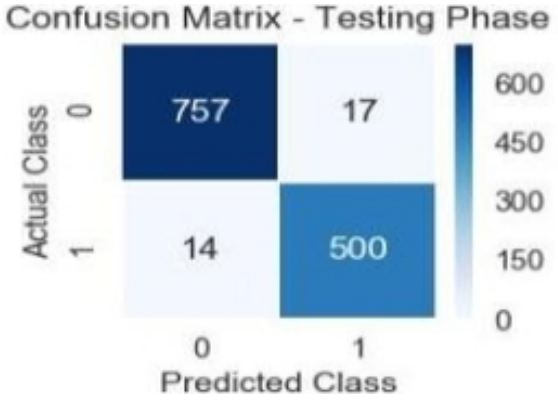
### 6.6.1 Performance of Evaluation Classifier

We verified the performance of the evaluation classifier using four verification metrics: confusion matrices/classification report, accuracy, ROC and AUC. The original dataset is split into 70% for training and 30% for testing. The 70% portion of partitioned train data set is further split into a train and a validation partition using 3-fold cross validation technique. The confusion matrices obtained in this process for each of the fold is illustrated in Fig. 6.3.


(a) Confusion matrix obtained in the training phase corresponding to fold 1.


(b) Confusion matrix obtained in the training phase corresponding to fold 2


(c) Confusion matrix obtained in the training phase corresponding to fold 3




(d) Confusion matrix obtained in the testing phase with the application of the best model of the training phase

Fig. 6.3: Heatmap Illustrating the Confusion Matrix for the Evaluation Classifier

In these confusion matrices, each column corresponds to the predicted class and each row

to the actual class. The positive class, i.e., negligent evaluation is labelled as 1 and normal evaluation is labelled as 0. The classifications that lie on the diagonal of the confusion matrix correspond to the True Negative (TN) and True Positive (TP) outcome. Higher values in the diagonal columns corresponding to the TN and TP as compared to the False Positive (FP) and False Negative (FN) columns, confirms the higher accuracy rate of the model.

We calculated the overall accuracy of the model using an equation:

$$Accuracy = \frac{TN + TP}{TN + FN + FP + TP} \tag{6.3}$$

We obtained accuracy of 95% with the simplest linear kernel function. The accuracy of the model further improved to 97% with the use of non-linear kernel (RBF kernel) with C=2 and $\gamma$= 1.0 (refer Fig. 6.3d). The consistent percentage of 97% for each of precision, recall and f1 produced by the evaluation classifier further confirms the reliability of the model (refer Table 6.1).

Table 6.1: Classification Report for the Evaluation Classifier

| Classification Label | Precision | Recall | f1-score | Support |
|---|---|---|---|---|
| 0 - Normal | 0.98 | 0.98 | 0.98 | 774 |
| 1 - Negligent | 0.97 | 0.97 | 0.97 | 514 |

Further, we obtained Receiver Operating Characteristic (ROC) curve (Graph of the true positive rate (Sensitivity= $\frac{TP}{TP+FN}$) v/s the false positive rate (Specificity= $\frac{TN}{TN+FP}$) for different cut-off points). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The high accuracy of the model can be confirmed from the position of the curve following the left hand border and then the top border of the ROC space (refer Fig. 6.4). A useless classifier is one that has its ROC curve exactly aligned with the diagonal. The ability of the model to correctly classify negligent and normal evaluation is further confirmed by AUC (Area Under Curve) value of 0.98 (refer Fig.

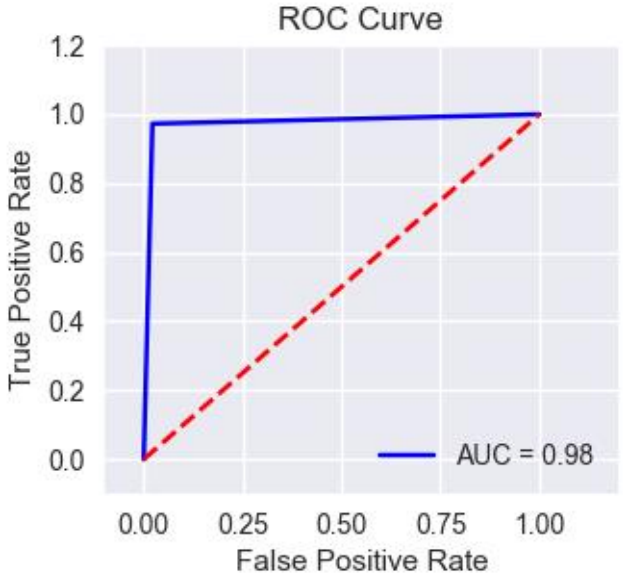



Fig. 6.4: ROC Curve for the evaluation classifier

6.4).

We investigated the problem of examiner heterogeneity in evaluation in Chapter 5 with the help of CAER. In this chapter we offered another perspective to the problem of examiner subjectivity with the help of machine learning techniques. In order to address this problem, first we attempted to remove the negligent data points from evaluation by classifying each evaluation as negligent or normal using SVM. Identification of diligent evaluation is a stepping stone in correcting the evaluation heterogeneity.

The results indicated that the major factors affecting the classification of evaluation, either as negligent or normal, were the actual time taken for evaluation and the time required for evaluation. Together with marks allotted by the examiner to each answer, length of answer and the expected length. For instance, if an examiner takes very little time for evaluation of an answer, it can be presumed that the examiner concerned is allotting the marks randomly without fully and properly reading the answer. Similarly, if the answer content contains few words and such an answer is allotted more than the deserved marks, then, the quality of the

evaluation is questionable.

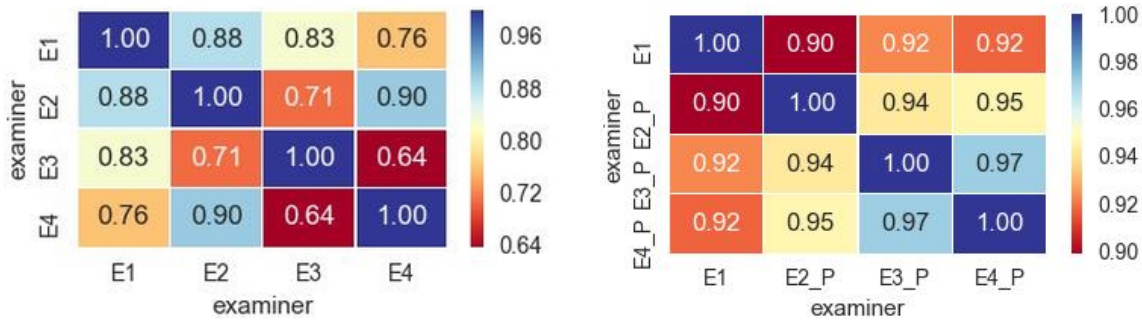
The classification matrix indicates that the value of false positives to be quite low (refer fig 6.3). This may be due to the presence of expert moderator rating as one of the factor in deciding whether the evaluation is negligent or normal. It appears that the classification model to some extent is biased towards expert moderator rating.

This study ignored many other critical factors such as answer-script content or key phrases, number of answer-scripts evaluated by the examiner, time of the day, time constraints, subject knowledge, etc., which directly/indirectly decide the quality of evaluation. Inclusion of these factors in the study can definitely improve the evaluation model considerably.

### 6.6.2 Performance of Tuned Marks Predictor

We verified the performance of the marks tuner using evaluation metrics, ANOVA, correlation heatmap and residual plot. The evaluation of the answer-scripts was done by four examiners separately, with evaluation carried by each examiner fully blinded from the other examiner to avoid any bias in the evaluation. First, we identified the degree of variation in the actual evaluation of each examiner with the help of ANOVA. An ANOVA test revealed statistically significant differences between the evaluation carried by each of the examiner at the $p < 0.05$ with F(3,176)=5.568; sig=0.001. Post-hoc comparisons using the Tukey HSD test also revealed that there is a significant difference in evaluation carried by each examiner at the 0.05 level. We also applied an ANOVA test on marks predicted by our ANN based marks tuner. The ANOVA test corresponding to the marks predicted by our ANN based marks tuner showed insignificant variation at the $p < 0.05$ with F(3,176)=0.609; sig=0.610). We also conducted post-hoc comparisons using the Tukey HSD test on marks predicted by ANN based marks tuner. The result also revealed an insignificant variation in the tuned marks predicted with the sig. value ranging from 0.86 to 0.99 at the 0.05 level.

We also relied upon heatmap of correlation coefficients to identify the degree of variation in the actual evaluation and the predicted marks for each of the examiner. As illustrated in Fig. 6.5a, most of the correlation coefficients are below 0.90, confirming the fact that there is a considerable difference in the evaluation carried by each examiner. Whereas, the correlation coefficients in Fig. 6.5b corresponding to the tuned marks of each examiner indicates correlation coefficients of above 0.90 confirming the fact that the tuned marks predictor is successful in predicting the tuned marks to one common scale, thus reducing the intra and inter examiner variation in evaluation.



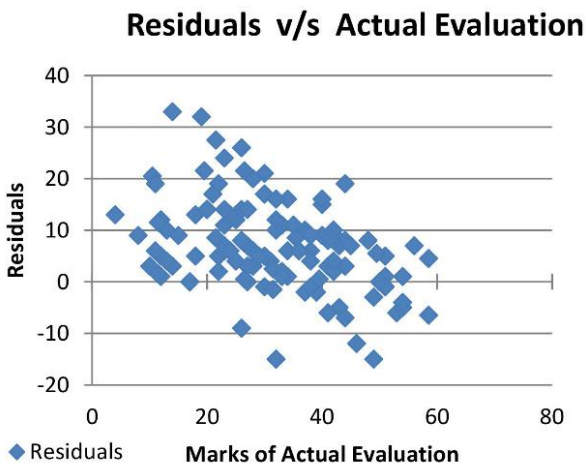

(a) Low correlation coefficients indicating high variation in evaluation for evaluation carried by four examiners (E1, E2, E3 and E4)

(b) High correlation coefficients indicating low variation for tuned marks predicted (E2_P, E3_P and E4_P) for three examiners (E2, E3 and E4) corresponding to the scale adopted by examiner E

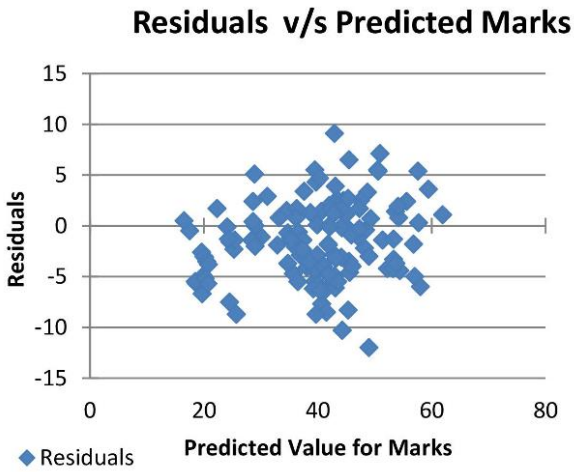
Fig. 6.5: Heatmap illustrating the correlation coefficients for the E-moderation model

A model is considered to fit the data well, if the difference between observed and predicted values is small. This is best illustrated in the form of a residual plot (refer Fig. 6.6). The plot in Fig. 6.6a shows the residuals (the difference between predicted point and its corresponding actual value) versus the consolidated marks assigned by different examiners for the same answer-scripts. We can notice that the difference in marks allotted by two examiners to the same answer-script is ranging from + 33 to -15. The reliability of the entire examination system is at stake with such a immense variation. The model validation graph in Fig. 6.6b shows the residuals versus the fitted/predicted values of marks obtained through a proposed

ANN model.




(a) Residuals v/s the consolidated marks assigned by different examiners for the same answer-scripts.




(b) Residuals v/s the fitted/predicted values of marks obtained through ANN Model

Fig. 6.6: Residual Plot Illustrating the Efficiency of E-moderation Model

The tuned marks predictor is designed to control the examiner heterogeneity in multi-examiner evaluation of answer-scripts. Before application of tuned marks predictor, we tried to remove all the evaluation perceived to be negligent in order to attain the reliable evaluation system. When multiple examiners are involved in evaluation of a particular subject/course paper attaining the evaluation of all examiners on a common scale can significantly increase the reliability of entire examination system. The tuned marks predictor proposed to attain this objective with the help of ANN Regressor.

The results suggest that the predictions made by the marks tuner using ANN model are quite consistent with the marks allotted by the master evaluator. In other words, with the inclusion of all the essential evaluation parameters, it is possible to accurately map the evaluation carried by one examiner onto the evaluation of another examiner with the prior knowledge of the scale adopted by those examiners. However, the scale chosen of one examiner for mapping onto the evaluation of other examiners needs to be free from any negligence or errors for increasing the accuracy of the model and reducing the bias.

The absence of any pattern and random distribution of points in Fig. 6.6b indicates that the model is unbiased and fitted the data considerably well. We also noticed that majority of the points lie within the range of 5, confirming that the intra/inter examiner variation in evaluation is brought under considerable control. Also, as this entire experiment was performed with limited set of answer-scripts and examiners evaluated answer-scripts in a controlled environment some sort of bias is likely to get introduced in the marks tuning system resulting into higher accuracy rate.

We tried different models for predicting the best outcome, though not a single model produced the best tuned marks for all the examiners involved. We found that different models produced excellent results for different examiners. This means that, there is still scope for improving the proposed model by combining multiple best models into one optimum model using ensembling techniques [ZWT02].

## 6.7 Summary

The evaluation of subjective answer-scripts suffers from large scale evaluation anomalies and the impact of 'examiner subjectivity' or 'examiner variability'. The currently adopted methods such as moderation and scaling only provide cursory relief from the menace of evaluation anomalies. The current study is undertaken to detect those evaluations that suffer from negligence and also predict the tuned marks in the event of intra and inter examiner variation in evaluation. We used a SVM classifier for classifying the evaluation as negligent or normal and an a ANN regressor for predicting the tuned marks. Findings from this research indicate that the given evaluations can be classified into negligent or normal with a great degree of accuracy. This study also provided evidence that we can predict the marks of one examiner based on the evaluation carried by another examiner with the proper training and validation of the model. Our proposed E-moderation model showed considerable

promise in controlling the intra and inter examiner variation as well as detecting the evaluation anomalies. Based on the insights gained through this study and experiments, we intend to improve the model further with the help of additional features and by incorporating ensembling techniques. It is also pertinent to observe the effect of CAER system proposed in Chapter 5 in conjunction with E-moderation system proposed in this chapter and to assess the effectiveness of combined approach for controlling evaluation anomalies as well as evaluation heterogeneity.

# CHAPTER 7

# Conclusion and Future Work

Public examinations are predominantly used to evaluate the performance and quality of examinees. As these examinations are high stake examinations, they suffer from a variety of threats and malpractices. An examination system with a large number of subjective answer-scripts pertaining to each course paper/subject invariably introduces anomalies and examiner subjectivity in manual evaluation. In this thesis, we investigated two specific assessment systems, namely, conventional assessment and E-assessment, to understand the source and type of threats along with security requirements (Chapter 2). An analysis of the conventional and electronic assessment indicated that both the systems are a way short of providing the optimum level of security and consistency in evaluation to the stakeholders concerned.

Public examinations are prone to a variety of security threats and malpractices. This is apparent from the ever-increasing cases of unfair means. It is necessary to control malpractices so that all examinees get a equal and fair footing. A variety of security approaches are currently being used for controlling malpractices in the conventional/electronic setup of examinations. Some of the most prominent approaches include:

1. Use of encryption techniques for preservation of secrecy of question papers/answer-scripts.

2. Use of hashing techniques to protect the integrity of question paper/answer-scripts.

3. Use of coding techniques and mixnet servers to establish anonymity of the examinees/examiners.

4. Use of Digital signatures to prevent denial of action of stakeholders concerned.

However, most of the work in summative examinations is focused on safeguarding the interest of only the examination authority, thus, creating a totally unfair system. Also, the majority of solutions applied security techniques in isolation resulting in a non-comprehensive security solution (Chapter 3).

## 7.1 Summary of Thesis Contributions

A fair and reliable examination platform needs to be built on the pillars of strong security mechanisms, well-defined layered protocols and a fair/homogenous evaluation system. The remainder of this section summarizes the specific technical contributions, we see emerging from our research work.

**Objective 1.** *Achieving anonymity of examinees and examiners in the exchange of answer-scripts.*

In public examinations, the two crucial security requirements are anonymity and confidentiality. Anonymity is essential to hide the identity of the examinee and the examiner from each other. Confidentiality is necessary to maintain the secrecy of the question paper before the commencement of the examination. Similarly, the answer-scripts need to be protected from all entities except the examiners.

Inorder to achieve anonymity between examiners and examinees in the process of answer-scripts delivery, we suggested a dual purpose cryptographic scheme, namely 'disguised public key' (Chapter 4). The said scheme is based on the concept of blind signature. We created the disguised public key for anonymizing the identity of examiners, using the following predicate:

$$unhide(aenc(m, hide(K_X, r)), r^{-1}) = aenc(m, K_X) \tag{7.1}$$

Where, $K_X$ represent the public key of the receiver (examiner),
r is the random number with corresponding inverse $r^{-1}$.

We also proposed the corresponding inverse predicate to recover the disguised public key at appropriate stage with the help of the following inverse predicate:

$$unhide(hide(K_X, r), r^{-1}) = K_X \tag{7.2}$$

The steps used in establishing anonymity and confidentiality using the disguised public key and corresponding inverse are summarized below:

1. Initially, the examination authority takes the public key of the examiner and disguises it using a random number as per equation 7.1. Then, the examination authority, encrypts the disguised public key using the public key of the examinee. The encrypted disguised public key is sent to the examinee.

2. The examinee decrypts the received message to get the disguised public key. The answer-script produced by the examinee is encrypted using a disguised public key. The examinee sends the encrypted answer-script to the examination authority.

3. The examination authority on receipt of the encrypted answer-scripts, applies equation 7.2 to unhide the public key of examiner. This produces an encrypted answer-script as if encrypted through the use of the public key of the examiner. Subsequently, the examination authority sends the encrypted answer-scripts to examiners.

In this approach, the sender (examinee) is provided with the disguised public key of the recipient (examiner) to de-link the identity of the recipient from the sender. The proposed

mechanism is suitable in general, for achieving anonymity and confidentiality in applications where communication between the sender and the recipient is achieved through an intermediary third party.

**Objective 2.** *Creating an inseparable association/bonding between a unique question paper and the answer-scripts exchanged between the examination authority and the examinees.*

In public examinations, the presence of a large number of examinees in each examination block provides ample opportunities to examinees to engage in collusion/plagiarism and cheating/copying. The malpractices such as question paper leakage and rampant collusion/plagiarism can be controlled to a great extent by generating a unique question paper, Just-In-Time (JIT). With the unique question paper, however, we require a mechanism to establish an unambiguous link between the examinee identity and the question paper. It is also necessary to associate the unique question paper received by the examinee to the corresponding answer-script produced by the examinee unambiguously. The established association needs to be strong enough to prevent both the sender and the receiver from denying their action in the future.

In Chapter 4, we presented a novel approach for linking the question paper and the answer-script associated with the examinee and revealing only the selective and essential part of the aggregated information to the examination authority (recipient). We presented a protocol based on the concept of digital signature, blind signature and dual signature for establishing an unbreakable association between the question paper and the answer-script submitted by the examinee.

We used applied $\pi$ calculus for providing a formal specification of the process of question paper/ answer-script delivery between examination authority and examinees. The series of associativity and anonymity properties emerging from the question paper/answer-script delivery protocols were validated using the ProVerif protocol verification tool and manual proofs.

The defined associativity properties ensured the strong link between the question paper and the answer-script pair without revealing unnecessary information to any of the communicating entities. We proved that the adversary is not able to obtain any significant information about the aggregated data (i.e., question paper cum answer-script) and break the association between question paper and answer-script without getting detected. The defined anonymity properties ensured that the identity of the examinee and examiners remained hidden from each other to avoid any bias in the evaluation.

**Objective 3.** *Measure the evaluation anomalies in a specific conventional examination system and develop a unified approach to ensure an error-free and a uniform evaluation.*

In the context of manual evaluation of subjective answer-scripts, we showed that evaluation in general suffers from large scale anomalies and heterogeneity (Chapter 5). The significant finding of the study is that, on an average 2% of the evaluated answer-books suffer from errors owing to totalling, recording and transferring. We also observed that over the years there is a rise in the number of verification and re-evaluation grievances, indicating that examinees have serious doubts about evaluation practices. The data pertaining to the analysis of personal verification and re-evaluation of answer-scripts indicated that on an average 85% of the referred cases result in a change in the marks either due to the errors in evaluation or examiner variability beyond the accepted standards.

We created a web based system, namely, Computer Assisted Evaluation using Rubrics (CAER) for performing the evaluation, compiling the results and providing feedback to the examinees. This system was designed to assist the examiner to mark the answer-scripts as per predefined rubric criteria. After the completion of computer assisted marking, CAER automatically generated question-wise marks along with the grand total for each examinee. Our approach significantly eliminated human interventions in the evaluation and in totalling. Thereby, providing a robust evaluation and result compilation environment.

We investigated the impact of CAER in controlling examiner variation and errors in

evaluation. This study revealed that the CAER approach brought considerable improvement to the current practices of evaluation by controlling evaluation errors, examiner variability and wastage of time.

**Objective 4.** *Devise an E-moderation scheme for classification of answer-script evaluation as anomalous or normal and prediction of normalized marks to control intra/inter examiner variation in evaluation.*

The manual evaluation of subjective answer-scripts suffer from the large scale evaluation anomalies and the impact of 'examiner subjectivity' or 'examiner variability' as shown in Chapter 5. The evaluation also suffers from 'Hawk-Dove effect', where some examiners are liberal in evaluation and some are strict in allotment of marks/grades. The currently adopted methods such as moderation and scaling only provide cursory relief from the problem of evaluation anomalies and heterogeneity.

We conducted a study to assess the evaluation pattern in multi-examiner evaluation using evaluation affecting parameters such as evaluation start and completion time, total time taken for evaluation of each answer, sequence of evaluation, etc. (Chapter 6). We used the machine learning technique, namely SVM classifier, to devise an evaluation classifier for classifying evaluation as negligent or normal by using critical parameters.

We analysed the performance of evaluation classifier using evaluation metrics, such as confidence matrix and ROC. The findings from this study demonstrated that the given evaluations can be classified into negligent or normal with a great degree of accuracy.

We also conducted a study to identify the heterogeneity in multi-examiner evaluation. The analysis of data with the help of ANOVA showed the existence of large scale intra/inter examiner variation in manual evaluation.

The intra/inter examiner variation can be controlled by adjusting the marks assigned by respective subject examiners on one scale adopted by any one examiner. We used ANN

regressor to devise marks predictor for predicting the normalized marks to fit evaluation of each subject examiner to the evaluation scale of any one of the subject examiner. We analysed the performance of marks predictor using metrics ANOVA and heatmap. The evidence obtained from analysis suggested considerable promise in controlling the intra/inter examiner variation as well as detecting the evaluation anomalies.

## 7.2 Direction for Future Work

The repeated incidents of malpractices and large scale evaluation anomalies in public examinations casts serious aspersions on the examination system. According to the information released by CBSE and an unofficial statistics, in every examination, there is a steep rise to more than 50% cases of malpractices every year. There is also increase in the grievances owing to evaluation anomalies. Institutions are seriously looking for more secure solutions to stem the anomalies that inadvertently affect our examination system. Many of these initiatives have opened up various avenues for conducting public examinations in a smooth manner and for devising more robust security measures.

Our study of some of the existing conventional/electronic examination systems revealed that the issue of full-fledged secure examination is still in its infancy. We foresee the likelihood of finding of new and more comprehensive approaches resulting from future and further research in this area. However, even the initial approaches that we proposed in this thesis raises many research questions.

In Chapter 4, we suggested cryptographic scheme for establishing anonymity between the examination authority, the examinee and the examiners during the exchange of the answer-scripts. We also devised security protocols for linking the question paper and the answer-script pertaining to each examinee. We discussed how the technique can be applied to the examination environment, but we do not have experimental results on the efficacy of doing

so. It would be intriguing to characterize the vulnerabilities associated with the suggested approach along with the limitations.

In this research, we are advocating the use of a unique question paper for controlling most of the public examination malpractices, but we have not identified the implications of it from a security perspective as well as from an academic viewpoint. Further research is needed to clarify the effect of unique question paper and associated security vulnerabilities in public examinations. There is little doubt that information yielded by such studies would enable institutions to deliver more effective services for all the stakeholders.

Another possibility on the basis of understanding of what makes an examination system fair and reliable is to design a new rule/constraint based system amenable to the needs of each stakeholder concerned. Another research direction is to experimentally test the scalability of conducting the subjective examination electronically. The additional future research direction is to characterize the entire examination process of setting up a question paper, conducting the examination, evaluation of answer-scripts and result declaration process. The emerging Blockchain technology can also be used for controlling the major intricacies of examination and for keeping track of critical activities in real time.

Much research also remains to be done on building and analysing a comprehensive security plan and framework for conducting summative E-examination. Additionally, we need security protocols for exchange of entire examination content securely amongst all the involved entities and build a seamless solution to control most of the examination malpractices.

Our research on intra/inter examiner variation in evaluation and evaluation anomalies (Chapter 5 and Chapter 6) suggests a number of future research directions. First, it would be desirable to create new mechanisms that do not share the major weaknesses that exist in manual evaluation. For example, we saw that in the multi-examiner and large quantum of

answer-script evaluation, there is a malady of both intra/inter examiner variation and it would certainly be interesting to characterize more generally how the E-moderation approaches proposed in this research, perform with existing approaches.

In Chapter 5, we saw how the computer-assisted evaluation using rubrics (CAER) can control evaluation anomalies. Another avenue for future research is to compare performance of the CAER approach to other existing models. Most significantly, while we have given a general template for controlling evaluation anomalies, much work needs to be done for testing the system, calibrating it and correcting the mis-grading to fully instantiate it.

Chapter 6, further extended the work of controlling evaluation anomalies, especially the examiner heterogeneity. An implicit assumption we made in devising an evaluation classifier is that the marking speed of examiner is fixed, whereas in reality examiners become trained to recognize common types of answers and become faster over time. It would be interesting to consider how this affects the negligence classification rules. As another research direction, we only characterized few evaluation effecting parameters for classifying each evaluation as normal or negligent, but there are certainly other instances/attributes that can make an evaluation bad or worse. Providing a more complete characterization that also classifies these instances/attributes would give us an even better understanding of the evaluation vulnerabilities.

Mechanisms for predicting the marks/grades on one normalized scale for all the examinees still leaves much to be desired. In part, this is primarily due to non-normal distribution of abilities of examinees in each batch and secondly due to really bad evaluation practices. Nevertheless, our results suggest that when an examinees' abilities possess normal distribution and the evaluation also is not randomly anomalous, then the examinees marks/grades can be mitigated by suitably training the machine for increasing the accuracy of prediction.

Difficulties for deciding the best prediction model do exist, even when restricting our attention to the controlled environment. This means that, there is still scope for improving the proposed model by combining the multiple best models into one optimum model by using the ensembling techniques. It is also important to note that further research is necessary to fit the marks/grades of all the examinees on a common scale to prevent any misjudgement/unfairness while comparing the abilities of each of the examinee.

Thus, it is evident that the use of technology provides promising ways of solving many of the lacunae associated with current examination system. The methods outlined in this thesis provide possible ways of ensuring fair and reliable examination system. While it remains to be seen how the proposed methods work in practical implementation. Considering the recurring and ever increasing incidents of anomalies and malpractices in public examination system, it appears that the current practices and security interventions are in infancy. Also, there are still various avenues to be explored for development, improvement and amalgamation of these methods into a comprehensive solution.

# APPENDIX A

# Disguised Public Key

## A.1 Illustration of working of Disguised Public Key

**Example A.1.1.** *Let n=187 (for modulus computation) and m= 2, the message created by producer (examinee). Let the public key/private of each entity defined in our protocol be as shown in Table A.1 Let the random factor held by intermediary (Examination authority) be*

Table A.1: Public/Private Key pair example

| | Public Key | Private Key |
|---|---|---|
| Producer(Examinee) | 19 | 59 |
| Consumer(Examiner) | 7 | 23 |

$r = 13$ *and its corresponding inverse,* $r^{-1} = 37$

The computational steps based on the defined public/private keys and random blind factors are illustrated in Table A.2:

Table A.2: Computational Steps

| Step | Computation | Result |
|---|---|---|
| Blinded key, $m' = (e * r)^{e'} mod\ n$ | $m' = (7 * 13)^{19} mod\ 187$ | 114 |
| Signed Key ,$s' = m'^{d'} mod\ n$ | $s' = 114^{59} mod\ 187$ | 91 |
| Encrypt the message using blinded key, $c' = m^{s'} mod\ n$ | $c' = 2^{91} mod\ 187$ | 178 |
| unblind the public key by removing blind factor, $c = c'^{r-1} mod\ n$ | $c = 178^{37} mod\ 187$ | 128 |
| Original Message, $m = c^{d} mod\ n$ | $m = 128^{23} mod\ 187$ | 2 |

# APPENDIX B

# ProVerif Code

## B.1 Modelling Answer-script Delivery with Associativity and Anonymity

```
(* Exam Authority is honest and can play an unbounded number
of instances *)
(* Unbounded number of dishonest examinees. Each examinee can play
 an unbounded number of instances *)

free ch: channel.                (* Public Channel *)
free chQ: channel [private].     (* Private Channel *)
free chA: channel [private].     (* Private Channel *)
free chX: channel [private].     (* Private Channel *)

type nonce.      (* Unique random session identification number *)
type pkey.       (* Public Key for asymmetric encryption *)
type skey.       (* Private Key for asymmetric encryption *)
type key.        (* Secret Key for symmetric encryption *)

type spkey.              (* signing public key *)
type sskey.              (* signing secret key *)

type bkey.               (* blinding key *)
type ukey.               (* unblinding key *)
```

```
type secret.
type public.
type authenticated.
type identity.  (* Examinee identity generated during registration *)
type pseudonym. (* Examinee Code generated before evaluation *)

free ans1:bitstring.
free ans2:bitstring.
(* Check whether attacker can find the question paper *)
query attacker(ans1).

fun pkeytobitstring(pkey): bitstring [data, typeConverter].

(* Public key encryption *)
fun pk(skey): pkey.
fun encrypt(bitstring, pkey): bitstring.
reduc forall x: bitstring, y: skey; decrypt(encrypt(x,pk(y)),y) = x.

(* Signatures *)
fun spk(sskey): spkey.
fun sign(bitstring, sskey): bitstring.
reduc forall m: bitstring, k: sskey; getmess(sign(m,k)) = m.
reduc forall m:bitstring, k:sskey; opensign(sign(m,k)) = m.

reduc forall m: bitstring, k: sskey; checksign(sign(m,k), spk(k)) = m.

(* Hash *)
fun hash(bitstring): bitstring.

(* Creation of Disguised public key based on the blind signature
scheme *)
```

```
fun bpk(bkey): ukey.
fun blind(pkey,bkey): pkey.
reduc forall r: bkey, m: bitstring, k: pkey;
unblind(encrypt(m,blind(k,r)),bpk(r)) = encrypt(m,k).



(* Tables *)
table stud_list(pkey, spkey).
table reg_students(pkey, spkey).
table reg_stud_code(spkey,pseudonym).
table examiners_list(pkey,bitstring).
table marks_list(pkey, bitstring).

free qp1,as1:bitstring.
free qp2,as2:bitstring.
free x,y: bitstring [private].



(* ---- Start of  Answer script delivery Protocol ------- *)

(* ------ Start of Exam Authority Role ---------- *)
let EA (pkEA:pkey,skEA:skey,ssec_EA:sskey,pkST:pkey,
pkEx:pkey,spub_ST:spkey) =
(* Registration of examinees *)
(* Receive question paper from the paper setters on a private channel *)
(* Deliver question paper to the examinees *)

new Nb:nonce;
new rf:bkey;     (* Random factor for blinding the public key *)
new ques_pap:bitstring;
```

```
(* Examination Conduct *)
(* For the sake of completeness, question paper delivery is shown here.
In reality, question paper is delivered separately *)
let bex_key= blind(pkEx,rf) in  (* Blind the public key of the examiner *)
let sbex_key=pkeytobitstring(bex_key) in
(* EA Calculate hash of blind public key used and sign it *)
let sbkHash=sign(hash(sbex_key),ssec_EA) in
(* EA Calculate hash of the question paper used *)
let qpHash=sign(hash(ques_pap),ssec_EA) in

(* EA encrypt the QP and blind key generated using public key of examinee.*)
let authQBk = encrypt(((((Nb,ques_pap),qpHash),bex_key),sbkHash),pkST) in
(* EA send QP and blind  public key to examinees *)
out(ch, authQBk);
(* EA Receive QP and AS pair from examinee *)
in(ch, studQPAS:bitstring);
(* EA authority decrypts QP AS pair *)
let ((((Na:nonce,=ques_pap),asHash:bitstring),dualsign:bitstring),
enc_ans_scr:bitstring) = decrypt(authQBk,skEA) in
(* Compute the dual hash based on the received QP and hash of AS *)
let hqphas = (hash(ques_pap),asHash) in
(* Verify the associativity of QP and AS  *)
if hash(hqphas)=checksign(dualsign,spub_ST) then
(* Compute the signature of the received dual signature *)
let sdualsign=sign(dualsign,ssec_EA) in
(* EA Send  acknowledgement to examinees *)
out(ch,sdualsign);

0.
(* ———— End of Exam Authority Role ———— *)
```

```
(* ————————Start of Examinee Role————————— *)
let ST(skST:skey,ssec_ST:sskey,pkEA:pkey,spub_EA:spkey,
ans_scr:bitstring) =
(* Registration of Examinees *)
(* Receive question papers from the EA *)
new Na:nonce;

(* For the sake of completeness, question paper receipt is shown here.
 In reality, question paper is received separately *)
(* Examinee receive blind public key from EA *)
in(ch,authQBk:bitstring);
(* Examinee  decrypts message received from EA *)
let ((((Nb:nonce,ques_pap:bitstring),qpHash:bitstring),usEx_key:pkey),
sbkHash:bitstring) = decrypt(authQBk,skST) in
let susEx_key = pkeytobitstring(usEx_key) in
(* Verify the Signature of the EA on  the received blind key *)
if hash(ques_pap)= checksign(qpHash,spub_EA) &&
hash(susEx_key) = checksign(sbkHash, spub_EA)  then
(* Combine hash of QP and AS together *)
let hqphas = (hash(ques_pap),hash(ans_scr)) in
(* Compute dual signature from hash of combined of h(QP) and h(AS) *)
let dualsign = sign(hash(hqphas),ssec_ST) in
(* Use  blinded key of the examiner  for encrypting the answer script *)
let enc_ans_scr = encrypt((ans_scr,hash(ques_pap)),usEx_key) in
(* Examinee encrypt the QP, blinded answer script using public key of EA.*)
out(ch,studQPAS);
let studQPAS = encrypt(((((Na,ques_pap),hash(ans_scr)),dualsign),
enc_ans_scr),pkEA) in
(* Send Examinee's QP and AS along  with dual signature *)
in(ch,sdualsign:bitstring);
0.
(* ———————— End of Examinee Role ———————— *)
```

```
(* ---------- Main Protocol ---------- *)
(* --------- Start process ----------*)
process
(* Public/Private keys of Examination Authority *)
new skEA: skey; let pkEA = pk(skEA) in out(ch, pkEA);
(* Public/Private signing keys of EA *)
new ssec_EA: sskey; let spub_EA = spk(ssec_EA) in out(ch, spub_EA);

(* Public/Private keys of Examiner *)
new skEx: skey; let pkEx = pk(skEx) in out(ch, pkEx);
(* Public/Private keys of examinee *)
new skST: skey; let pkST = pk(skST) in out(ch, pkST);
new ans_scr:bitstring;
(* Public/Private signing keys of examinee *)
new ssec_ST:sskey;let spub_ST=spk(ssec_ST) in out(ch, spub_ST);

(
(* Honest Exam Authority EA *)
(!EA(pkEA,skEA,ssec_EA,pkST,pkEx,spub_ST) )
(* Honest examinee *)        |
( new skC1: skey; let pkC1 = pk(skC1) in out (ch, pkC1);
new ssec_ST1:sskey;let spub_ST1=spk(ssec_ST1) in
out(ch, spub_ST1); !ST(skC1, ssec_ST1,pkEA, spub_EA,
 choice[ans1,ans2]) ) | ( new skC2: skey; let pkC2 = pk(skC2) in
 out (ch, pkC2); new ssec_ST2:sskey;let spub_ST2=spk(ssec_ST2) in
 out(ch, spub_ST2); !ST(skC2, ssec_ST2,pkEA, spub_EA, choice[ans2,ans1]) )

)
```

# Bibliography


[AASK08] Linda Anglin, Kenneth Anglin, Paul L Schumann, and John A Kaliski. Improving the efficiency and effectiveness of grading through the use of computer-assisted grading rubrics. *Decision Sciences Journal of Innovative Education*, 6(1):51–73, 2008. 28

[AB03] Martín Abadi and Bruno Blanchet. Computer-assisted verification of a protocol for certified email. In *International Static Analysis Symposium*, pages 316–335. Springer, 2003. 59

[AB10] Ola Amayri and Nizar Bouguila. A study of spam filtering using support vector machines. *Artificial Intelligence Review*, 34(1):73–108, 2010. 149

[ABB+05] Alessandro Armando, David Basin, Yohan Boichut, Yannick Chevalier, Luca Compagna, Jorge Cuéllar, P Hankes Drielsma, Pierre-Cyrille Héam, Olga Kouchnarenko, Jacopo Mantovani, et al. The avispa tool for the automated validation of internet security protocols and applications. In *International conference on computer aided verification*, pages 281–285. Springer, 2005. 58

[ABF07] Martín Abadi, Bruno Blanchet, and Cédric Fournet. Just fast keying in the pi calculus. *ACM Transactions on Information and System Security (TISSEC)*, 10(3):9, 2007. 59

[AC06] Judith A Arter and Jan Chappuis. *Creating and recognizing quality rubrics*. Prentice Hall, 2006. 28

[AF96] Masayuki Abe and Eiichiro Fujisaki. How to date blind signatures. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 244–251. Springer, 1996. 63

[AF01] Martín Abadi and Cédric Fournet. Mobile values, new names, and secure communication. *ACM SIGPLAN Notices*, 36(3):104–115, 2001. 58, 59

[AK09] Tuukka Ahoniemi and Ville Karavirta. Analyzing the use of a rubric-based grading tool. In *ACM SIGCSE Bulletin*, pages 333–337. ACM, 2009. 28, 125

[Aka09] Mehmet Fatih Akay. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert systems with applications*, 36(2):3240–3247, 2009. 149

[And00] Heidi Goodrich Andrade. Using rubrics to promote thinking and learning. *Educational leadership*, 57(5):13–19, 2000. 28, 29

[Apa10] Kikelomo Maria Apampa. *Presence verification for summative e-assessments*. PhD thesis, University of Southampton, 2010. 62

[AR06] Tuukka Ahoniemi and Tommi Reinikainen. Aloha-a grading tool for semi-automatic assessment of mass programming courses. In *Proceedings of the 6th Baltic Sea conference on Computing education research: Koli Calling 2006*, pages 139–140. ACM, 2006. 25, 28, 29, 125

[AWA10a] Kikelomo Maria Apampa, Gary Wills, and David Argles. An approach to presence verification in summative e-assessment security. In *Information Society (i-Society), 2010 International Conference on*, pages 647–651. IEEE, 2010. 62

[AWA10b] Kikelomo Maria Apampa, Gary Wills, and David Argles. User security issues in summative e-assessment security. *International Journal of Digital Society (IJDS)*, 1(2):1–13, 2010. 49

[AWA11] Kikelomo Maria Apampa, Gary Wills, and David Argles. Towards a blob-based presence verification system in summative e-assessments. *International Journal of e-Assessment*, 1(1), 2011. 62

[BAF08] Bruno Blanchet, Martín Abadi, and Cédric Fournet. Automated verification of selected equivalences for security protocols. *Journal of Logic and Algebraic Programming*, 75(1):3–51, 2008. 58, 60

[Bar17] Elaine Barker. Sp 800-67 rev. 2, recommendation for triple data encryption algorithm (tdea) block cipher. *NIST special publication*, 800:67, 2017. 41

[BBE+10] Michal Barla, Mária Bieliková, Anna Bou Ezzeddinne, Tomáš Kramár, Marián Šimko, and Oto Vozár. On the impact of adaptive test question selection for learning efficiency. *Computers & Education*, 55(2):846–857, 2010. 152

[BBP13] George A Brown, Joanna Bull, and Malcolm Pendlebury. *Assessing student learning in higher education*. Routledge, 2013. 118

[BF06] David Boud and Nancy Falchikov. Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4):399–413, 2006. 11

[BGD+08] Erik W Black, Joe Greaser, Kara Dawson, et al. Academic dishonesty in traditional and online classrooms: Does the media equation hold true? *Journal of asynchronous learning networks*, 2008. 49

[Bla01] B. Blanchet. An efficient cryptographic protocol verifier based on prolog rules. In *S. Schneider, editor, 14th IEEE Computer Security Foundations Workshop*, pages 82–96. IEEE Computer Society Press, 2001. 59

[Bla04] B. Blanchet. Automatic proof of strong secrecy for security protocols. In *IEEE Symposium on Security and Privacy*, pages 86–100. Oakland, California, 2004. 59

[Blo09] Sue Bloxham. Marking and moderation in the uk: false assumptions and wasted resources. *Assessment & Evaluation in Higher Education*, 34(2):209–220, 2009. 4, 123, 124

[Bou00] David Boud. Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2):151–167, 2000. 117, 119

[BP12] Brijesh Kumar Baradwaj and Saurabh Pal. Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*, 2012. 151

[Bro12] Val Brooks. Marking as judgement. *Research Papers in Education*, 27(1):63–80, 2012. 117

[BRW06] Gavin Busuttil-Reynaud and John Winkley. Jisc e-assessment glossary, 2006. 20, 22

[BS11] Mohini Bhardwaj and Amar Jeet Singh. Automated integrated examination system: A security concern. *Information Security Journal: A Global Perspective*, 20(3):156–162, 2011. 4

[BS18] Bruno Blanchet and Ben Smyth. Proverif: Automatic cryptographic protocol verifier user manual & tutorial. *Journal of Logic and Algebraic Programming*, 75(1):3–51, 2018. 60

[BW98] Paul Black and Dylan Wiliam. Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1):7–74, 1998. 12

[BYC04] David Baume, Mantz Yorke, and Martin Coffey. What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in Higher Education*, 29(4):451–477, 2004. 121

[BZ14] Nikos Benos and Stefania Zotou. Education and economic growth: A meta-regression analysis. *World Development*, 64:669–689, 2014. 149

[Cha83] David Chaum. Blind signatures for untraceable payments. In *Advances in cryptology*, pages 199–203. Springer, 1983. 8, 46, 87, 88, 99

[CIM14] Kevin Cox, Bradford W Imrie, and Allen Miller. *Student assessment in higher education: a handbook for assessing performance*. Routledge, 2014. 118

[Cop94] Don Coppersmith. The data encryption standard (des) and its strength against attacks. *IBM journal of research and development*, 38(3):243–250, 1994. 41

[CPVK16] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, and Vignesh Kalyanaraman. Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2:6–10, 2016. 48

[Cre14] Cas JF Cremers. Scyther user manual. *Department of Computer Science, University of Oxford: oxford, UK*, 2014. 58

[CRHJDJ06] Jordi Castella-Roca, Jordi Herrera-Joancomarti, and Aleix Dorca-Josa. A secure e-exam management system. In *The First International Conference on Availability, Reliability and Security, 2006. ARES 2006.*, pages 8–15. IEEE, 2006. 3, 61

[CVNM07] Félix Castro, Alfredo Vellido, Àngela Nebot, and Francisco Mugica. Applying data mining techniques to e-learning problems. *Evolution of teaching and learning paradigms in intelligent environment*, pages 183–221, 2007. 151

[CW06] Joseph A Cruz and David S Wishart. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2:59, 2006. 149

[DC13] Quang Hung Do and Jeng-Fung Chen. A neuro-fuzzy approach in the classification of students' academic performance. *Computational intelligence and neuroscience*, 2013:6, 2013. 151

[DD01] Chris HQ Ding and Inna Dubchak. Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics*, 17(4):349–358, 2001. 149

[Del10] Dursun Delen. A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4):498–506, 2010. 151

[DFDSMD14] Enric Junqué De Fortuny, Tom De Smedt, David Martens, and Walter Daelemans. Evaluating and understanding text-based stock price prediction models. *Information Processing & Management*, 50(2):426–441, 2014. 149

[DGG+12] Karel Dejaeger, Frank Goethals, Antonio Giangreco, Lapo Mola, and Bart Baesens. Gaining insight into student satisfaction using comprehensible data mining techniques. *European Journal of Operational Research*, 218(2):548–562, 2012. 151

[DGK+15] Jannik Dreier, Rosario Giustolisi, Ali Kassem, Pascal Lafourcade, and Gabriele Lenzini. A framework for analyzing verifiability in traditional and electronic exams. In *Information Security Practice and Experience*, pages 514–529. Springer, 2015. 62

[DH76] Whitfield Diffie and Martin E Hellman. New directions in cryptography. *Information Theory, IEEE Transactions on*, 22(6):644–654, 1976. 43

[DK15] Tassos Dimitriou and Ioannis Krontiris. Privacy-respecting auctions as incentive mechanisms in mobile crowd sensing. In *IFIP International Conference on Information Security Theory and Practice*, pages 20–35. Springer, 2015. 63

[DK16] Kissan Gauns Dessai and Venkatesh Kamat. A framework for analyzing associativity and anonymity in conventional and electronic summative examinations. In *Information Systems Security*, pages 303–323. Springer, 2016. 104

[DK18] Kissan Gauns Dessai and Venkatesh Kamat. omputer assisted evaluation using rubrics for reduction of errors and inter and intra examiner heterogeneity. *International Journal of Information and Communication Technology Education (IJICTE)*, 14(4):49–65, 2018. 127

[DKR09] Stéphanie Delaune, Steve Kremer, and Mark Ryan. Verifying privacy-type properties of electronic voting protocols. *Journal of Computer Security*, 17(4):435–487, 2009. 59

[DKW14] Kissan Dessai, Venkatesh Kamat, and Ramrao Wagh. Effective use of rubrics in computer assisted subjective answer-script evaluation. In *Proceedings of the 6th IEEE International Conference 2014 on Technology for Education*, pages 1083–1092, Kerala-India, 2014. IEEE. 2, 4, 125, 126

[Dow03] Steven M Downing. Validity: on the meaningful interpretation of assessment data. *Medical education*, 37(9):830–837, 2003. 27

[DR13] Joan Daemen and Vincent Rijmen. *The design of Rijndael: AES-the advanced encryption standard*. Springer Science & Business Media, 2013. 41

[DY83] Danny Dolev and Andrew C Yao. On the security of public key protocols. *Information Theory, IEEE Transactions on*, 29(2):198–208, 1983. 71, 73

[Eck03] Max A. Eckstein. *Combating academic fraud: Towards a culture of integrity*. International Institute for Educational Planning, 2003. 1, 2, 67

[ElG84] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. In *Advances in cryptology*, pages 10–18. Springer, 1984. 89

[Elt04] Lewis Elton. A challenge to established assessment practice. *Higher Education Quarterly*, 58(1):43–62, 2004. 11

[ET11] Ziba Eslami and Mehdi Talebi. A new untraceable off-line electronic cash system. *Electronic Commerce Research and Applications*, 10(1):59–66, 2011. 63

[FM04] Giles M Foody and Ajay Mathur. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(6):1335–1343, 2004. 149

[FWSC13] Chun-I Fan, Chien-Nan Wu, Wei-Zhe Sun, and Wei-Kuei Chen. Multi-recastable e-bidding game with dual-blindness. *Mathematical and Computer Modelling*, 58(1-2):68–78, 2013. 63

[GC88] FJ Good and MJ Cresswell. Placing candidates who take differentiated papers on a common grade scale. *Educational Research*, 30(3):177–189, 1988. 4, 124

[GC09] Thiago S Guzella and Walmir M Caminhas. A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, 36(7):10206–10222, 2009. 149

[GCRAS15] Jon Kepa Gerrikagoitia, Iñigo Castander, Fidel Rebón, and Aurkene Alzua-Sorzabal. New trends of intelligent e-marketing based on web mining for e-shops. *Procedia-Social and Behavioral Sciences*, 175:75–83, 2015. 149

[GLB13] Rosario Giustolisi, Gabriele Lenzini, and Giampaolo Bella. What security for electronic exams? In *International Conference on Risks and Security of Internet and Systems (CRiSIS), 2013*, pages 1–5. IEEE, 2013. 62

[GLR14] Rosario Giustolisi, Gabriele Lenzini, and Peter YA Ryan. Remark!: A secure protocol for remote exams. In *Cambridge International Workshop on Security Protocols*, pages 38–48. Springer, 2014. 3, 63

[Gol99] Dieter Gollman. Computer security. *John Wile & Sons, UK*, 1999. 40

[GPZ08] Peter Grainger, Ken Purnell, and Reyna Zipf. Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33(2):133–142, 2008. 121

[GTDPÁG06] A González-Tablas, A Diaz-Pabon, B Álvarez, and A Garnacho. Evaweb v2: Enhancing a web-based assessment system. In *Proceedings of the 4 th International Conference on Multimedia and Information Communication Technologies in Education, Sevilla, Spain*, pages 837–840, 2006. 61

[Guo10] William W Guo. Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction. *Expert Systems with Applications*, 37(4):3358–3365, 2010. 151

[Har05] Wynne Harlen. Teachers' summative practices and assessment for learning–tensions and synergies. *Curriculum Journal*, 16(2):207–223, 2005. 12

[HBG16] Ethan Heilman, Foteini Baldimtsi, and Sharon Goldberg. Blindly signed contracts: Anonymous on-blockchain and off-blockchain bitcoin transactions. In *International Conference on Financial Cryptography and Data Security*, pages 43–60. Springer, 2016. 63

[HF13] Shaobo Huang and Ning Fang. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, 61:133–145, 2013. 151, 152

[HM99] Jen Harvey and Nora Mogey. Pragmatic issues when integrating technology into the assessment of students. *S. Brown, P. Race, & J. Bull (Eds.), Computer-assisted assessment in higher education*, pages 7–20, 1999. 22

[HMG76] Arthur Edwin Harper, Vidya Sagar Misra, and Ramji Lal Gupta. *Research on examinations in India*. New Delhi: National Council of Educational Research and Training, 1976. 124

[Hol04] Gerard J Holzmann. *The SPIN model checker: Primer and reference manual*, volume 1003. Addison-Wesley Reading, 2004. 58

[HP10] Andrea Huszti and Attila Petho. A secure electronic exam system. *Publicationes Mathematicae Debrecen*, 77(3-4):299–312, 2010. 62

[HTLC10] Wen-Juan Hou, Jia-Hao Tsao, Sheng-Yang Li, and Li Chen. Automatic assessment of students' free-text answers with support vector machines. *Trends in Applied Intelligent Systems*, pages 235–243, 2010. 26, 28

[HV06] Wilhelmiina Hämäläinen and Mikko Vinni. Comparison of machine learning methods for intelligent tutoring systems. In *Intelligent tutoring systems*, pages 525–534. Springer, 2006. 151

[IKSA03] Subariah Ibrahim, Maznah Kamat, Mazleena Salleh, and Shah Rizan Abdul Aziz. Secure e-voting with blind signature. In *Telecommunication Technology, 2003. NCTT 2003 Proceedings. 4th National Conference on*, pages 193–197. IEEE, 2003. 62

[JS07] Anders Jonsson and Gunilla Svingby. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2):130–144, 2007. 118

[KC17] Hassan Khosravi and Kendra ML Cooper. Using learning analytics to investigate patterns of performance and engagement in large classes. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, pages 309–314. ACM, 2017. 152

[KKP03] Dharmendra Kanejiya, Arun Kumar, and Surendra Prasad. Automatic evaluation of students' answers using syntactically enhanced lsa. In *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing-Volume 2*, pages 53–60. Association for Computational Linguistics, 2003. 26, 28

[KL16] Terry K Koo and Mae Y Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016. 138

[Kni02] Peter T Knight. Summative assessment in higher education: practices in disarray. *Studies in Higher Education*, 27(3):275–286, 2002. 12, 27

[Kob91] Neal Koblitz. Cm-curves with good cryptographic properties. In *Advances in Cryptology - CRYPTO'91*, pages 279–287. Springer, 1991. 89

[KPP04] Sotiris Kotsiantis, Christos Pierrakeas, and Panagiotis Pintelas. Predicting students' performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5):411–426, 2004. 151

[KRS10] Steve Kremer, Mark Ryan, and Ben Smyth. Election verifiability in electronic voting protocols. In *European Symposium on Research in Computer Security*, pages 389–404. Springer, 2010. 59

[KS04] Tuomo Kakkonen and Erkki Sutinen. Automatic assessment of the content of essays based on course materials. In *Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on*, pages 126–130. IEEE, 2004. 26, 28

[Kuc10] Marcin Kucharczyk. Blind signatures in electronic voting systems. In *Computer Networks*, pages 349–358. Springer, 2010. 62

[LAS+15] Himabindu Lakkaraju, Everaldo Aguiar, Carl Shan, David Miller, Nasir Bhanpuri, Rayid Ghani, and Kecia L Addison. A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD*

*international conference on knowledge discovery and data mining*, pages 1909–1918. ACM, 2015. 152

[LCY+97] Kuang-Chih Lee, Keh-Ning Chang, Shiuh-Sheng Yu, Ing-Chau Chang, Chee-Wen Shia, Wen-Chin Chen, and Jau-Hsiung Huang. Design and implementation of important applications in a java-based multimedia digital classroom. *Consumer Electronics, IEEE Transactions on*, 43(3):264–270, 1997. 61

[LGM+09] Ioanna Lykourentzou, Ioannis Giannoukos, George Mpardis, Vassilis Nikolopoulos, and Vassili Loumos. Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology*, 60(2):372–380, 2009. 151, 152

[Lin08] Robert L Linn. *Measurement and assessment in teaching*. Pearson Education India, 2008. 12

[LLM12] Chee Kian Leong, Yew Haur Lee, and Wai Keong Mak. Mining sentiments in sms texts for teaching evaluation. *Expert Systems with Applications*, 39(3):2584–2589, 2012. 152

[Low96] Gavin Lowe. Breaking and fixing the needham-schroeder public-key protocol using fdr. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 147–166. Springer, 1996. 97, 99

[Mah11] VK Maheshwari. Malpractices in examinations–the termites destroying the educational set up, 2011. 1, 67

[MMT06] Ian Christopher McManus, J Mollon, and M Thompson. Assessment of examiner leniency and stringency ('hawk-dove effect') in the mrcp (uk) clinical examination (paces) using multi-facet rasch modelling. *BMC Medical Education*, 6(1):42, 2006. 6, 122, 146

[MO99] Chris Morgan and Meg O'reilly. *Assessing open and distance learners*. Psychology Press, 1999. 1, 11, 12

[MP12] Zohar Manna and Amir Pnueli. *The temporal logic of reactive and concurrent systems: Specification*. Springer Science & Business Media, 2012. 58

[MS98] Catherine Meadows and Paul Syverson. A formal specification of requirements for payment transactions in the set protocol. In *Financial Cryptography*, pages 122–140. Springer, 1998. 101, 103, 104

[MS13] Prerna Mahajan and Abhishek Sachdeva. A study of encryption algorithms aes, des and rsa for security. *Global Journal of Computer Science and Technology*, 2013. 43

[MSCB13] Simon Meier, Benedikt Schmidt, Cas Cremers, and David Basin. The tamarin prover for the symbolic analysis of security protocols. In *International Conference on Computer Aided Verification*, pages 696–701. Springer, 2013. 58

[Mur89] Tadao Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989. 58

[New78] Paul E Newton. Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2):149–170, 1978. 12, 27

[NS78] Roger M Needham and Michael D Schroeder. Using encryption for authentication in large networks of computers. *Communications of the ACM*, 21(12):993–99, 1978. 97, 99

[OACO08] VO Oladokun, AT Adebanjo, and OE Charles-Owaba. Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1):72–79, 2008. 151

[OPT97] Donal OMahony, Michael Peirce, and Hitesh Tewari. *Electronic Payment Systems*. Artech House Norwood, Dual Signature, 1997. 8, 101, 103

[OW15] Sally Roisin O'Hagan and Gillian Wigglesworth. Who's marking my essay? the assessment of non-native-speaker and native-speaker undergraduate essays in an

australian higher education context. *Studies in Higher Education*, 40(9):1729–1747, 2015. 117, 121

[PK01] Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability, and pseudonymity- a proposal for terminology. In *Designing privacy enhancing technologies*, pages 1–9. Springer, 2001. 49, 68

[PP02] Charles P Pfleeger and Shari Lawrence Pfleeger. *Security in computing*. Prentice Hall Professional Technical Reference, 2002. 40

[QZR+14] Xueheng Qiu, Le Zhang, Ye Ren, Ponnuthurai N Suganthan, and Gehan Amaratunga. Ensemble deep learning for regression and time series forecasting. In *Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on*, pages 1–6. IEEE, 2014. 149

[RA10] Y Malini Reddy and Heidi Andrade. A review of rubric use in higher education. *Assessment & evaluation in higher education*, 35(4):435–448, 2010. 28, 118

[Ram83] Arkalgud Ramaprasad. On the definition of feedback. *Systems Research and Behavioral Science*, 28(1):4–13, 1983. 12

[Rea90] James Reason. *Human error*. Cambridge university press, 1990. 117

[Rov00] Alfred P Rovai. Online and traditional assessments: what is the difference? *The Internet and Higher Education*, 3(3):141–151, 2000. 1, 12

[RSA78] Ronald L Rivest, Adi Shamir, and Len Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978. 46, 87, 89

[RSGL01] Peter Ryan, Steve A Schneider, Michael Goldsmith, and Gavin Lowe. *The modelling and analysis of security protocols: the csp approach*. Addison-Wesley Professional, 2001. 58

[RSS09] Archana Rane, Sandip Saha, and M Sasikumar. A tool for managing descriptive type examinations. In *International Conference on Management Technology for Educational Practices, July 2009*, 2009. 25, 28, 125

[RZPBK12] Vicente-Arturo Romero-Zaldivar, Abelardo Pardo, Daniel Burgos, and Carlos Delgado Kloos. Monitoring student progress using virtual appliances: A case study. *Computers & Education*, 58(4):1058–1067, 2012. 151

[SA92] J Michael Spivey and JR Abrial. *The Z notation*. Prentice Hall Hemel Hempstead, 1992. 58

[Sad89] D Royce Sadler. Formative assessment and the design of instructional systems. *Instructional science*, 18(2):119–144, 1989. 12

[Sad05] D Royce Sadler. Interpretations of criteria-based assessment and grading in higher education. *Assessment & evaluation in higher education*, 30(2):175–194, 2005. 11

[Sad09] D Royce Sadler. Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34(7):807–826, 2009. 142

[SB00] Alex Shafarenko and Dima Barsky. A secure examination system with multi-mode input on the world-wide web. In *null*, page 97. IEEE, 2000. 3, 61

[SC06] Erica Smith and Kennece Coombe. Quality and qualms in the marking of university assignments by sessional staff: An exploratory study. *Higher Education*, 51(1):45–69, 2006. 121

[Son99] Dawn Xiaodong Song. Athena: a new efficient automatic checker for security protocol analysis. In *Computer Security Foundations Workshop, 1999. Proceedings of the 12th IEEE*, pages 192–202. IEEE, 1999. 58

[Sta00] William Stallings. The set standard & e-commerce. *DOCTOR DOBBS JOURNAL*, 25(11):30–39, 2000. 47

[SVM06] Juan-Francisco Superby, JP Vandamme, and N Meskens. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Workshop on Educational Data Mining*, volume 32, page 234, 2006. 151

[TERS05] R Murray Thomas, Max A Eckstein, Ritva Reinikka, and Nathanael Smith. Combating academic fraud: Toward a culture of integrity, 2005. 1, 67

[TL13] Elise Trumbull and Andrea Lash. Understanding formative assessment: Insights from learning theory and measurement theory. *WestEd*, page 2, 2013. 12

[TVK05] Geoff Timmins, Keith Vernon, and Christine Kinealy. *Teaching and learning history*. Sage, 2005. 11

[Var14] Dale Varble. Reducing cheating opportunities in online test. *Atlantic Marketing Journal*, 3(3):9, 2014. 3

[VTK05] Keith Vernon, G Timmins, and C Kinealy. *Teaching and learning history in higher education*. Sage, 2005. 117, 119

[WB13] Marcelo Worsley and Paulo Blikstein. Towards the development of multimodal action based assessment. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 94–101. ACM, 2013. 152

[WCC11] Sarah Wiseman, Paul Cairns, and Anna Cox. A taxonomy of number entry error. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pages 187–196. British Computer Society, 2011. 117

[WDASG11] Andreas Weinberger, Heinz Dreher, M Al-Smadi, and Christian Guetl. Analytical assessment rubrics to facilitate semi-automated essay grading and feedback provision. In *Australian Technology Network Assessment Conference 2011*, pages 333–337. ACM, 2011. 125

[Wei05] Edgar R Weippl. *Security in e-learning*, volume 16. Springer Science & Business Media, 2005. 3, 61

[Wei18] Andreas S Weigend. *Time series prediction: forecasting the future and understanding the past*. Routledge, 2018. 149

[WPJ00] Frank Webster, David Pepper, and Alan Jenkins. Assessing the undergraduate dissertation. *Assessment & Evaluation in Higher Education*, 25(1):71–80, 2000. 121

[YOT14] Erman Yukselturk, Serhat Ozekes, and Yalın Kılıç Türel. Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and E-learning*, 17(1):118–133, 2014. 151

[Zim95] Philip R Zimmermann. *The official PGP user's guide*. MIT press, 1995. 43

[ZWT02] Zhi-Hua Zhou, Jianxin Wu, and Wei Tang. Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1-2):239–263, 2002. 171