EFFECTIVE MANAGEMENT OF LEARNER INTEGRITY IN ONLINE ASSESSMENTS: AN APPROACH USING KEYSTROKE DYNAMICS

H. A. A. K. M. Seneviratne

Registration No. 158054D

Degree of Doctor of Philosophy

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2026

EFFECTIVE MANAGEMENT OF LEARNER INTEGRITY IN ONLINE ASSESSMENTS: AN APPROACH USING KEYSTROKE DYNAMICS

H. A. A. K. M. Seneviratne

Registration No. 158054D

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Computer Science & Engineering

Faculty of Engineering

University of Moratuwa

Sri Lanka

June 2026

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 2026.06.10

The above candidate has carried out research for the PhD thesis under our supervision. We confirm that the declaration made above by the student is true and correct.

Name of Supervisors: Eng. Prof. Indika Perera

Eng. Prof. Shantha Fernando

Signatures of the Supervisors: Date: 2026-06-10

Date: 2026-06-10

DEDICATION

I dedicate this work to everyone in all parts of the world, who have contributed and are contributing, to promote, uphold, and maintain merit systems within academia.

ACKNOWLEDGEMENT

I would begin by expressing my deep gratitude to my supervisors Eng. Prof. Indika Perera and Eng. Prof. Shantha Fernando. Eng. Prof. Perera who served as my communicating supervisor, extended continued guidance and mentoring to me during my doctoral research. It marked a period of lasting impressions filled with inspiration, understanding, and practical wisdom, all of which turned out to be vital ingredients in developing a research mindset. Eng. Prof. Fernando provided valuable guidance and encouragement at the stage of formulating my research proposal and in refining the publications that were made connected with my research, as well as the thesis. I thank both of them wholeheartedly.

The valued critiques received from my distinguished panel of examiners comprising Eng. Prof. Dileeka Dias, Eng. Prof. Sanath Jayasena, and Eng. Prof. Indika Perera, at all the progress reviews which were extensive and thorough, helped me immensely in achieving each milestone leading to the completion of my research.

The research assistance received from Eng. Imesha Sudasingha in the initial development of a keystroke logger for exclusive use in my experiments is gratefully acknowledged. The insights provided by Prof. Nandana Rajatheva in making this logger lean and the assistance provided by Dr. Thushan Sivalingam in developing and testing a prototype is also gratefully remembered.

I wish to express my sincere appreciation of the pivotal role played by Eng. Devinda Abayagunawardana and Mr. Brian Dunsford, in engaging the services of volunteers to roleplay as subjects in my experiments, which enabled the collection of lifeline data for the research. Their role and commitment in keeping the volunteers adequately briefed while maintaining their anonymity is gratefully admired. On the same note, I extend my thanks to all those volunteers whose identities are not known to me, both individually and collectively, for devoting their time and effort to roleplay in my experiments.

Finally, I would like to thank my wife Manjula, my son Dinindu, my daughter Chinthani, and my daughter in law Panchalie who were towers of strength, for their optimistic companionship and understanding.

ABSTRACT

Educational institutions worldwide face a serious issue when conducting online assessments due to certain candidates cheating. Review of related work in the domains of learner integrity and technology assisted fraud prevention indicated gaps in the application of technology for maintaining learner integrity during online assessments. This thesis addressed these gaps by conceptualizing a system for the effective management of learner integrity during online assessments. The use of the behavioral biometric called 'keystroke dynamics' was thoroughly reviewed and how it could be adapted to authenticate candidates continuously and non-obtrusively while they take online assessments with no accessories deployed at the remote site was conceptualized. The thesis describes how the proof of this concept was established through experiments. This includes the setting up of a fictitious online testing institution and deploying a group of volunteers to roleplay as candidates. Tests were conducted online to simulate a structured initial registration to acquire biometrics for profile creation. Novel features extracted from the acquired biometrics were used to train anomaly detection models employing a variety of binary classifiers. This was followed by online tests that simulated genuine test taking as well as fraudulent test taking. These scenarios were played back offline while continuous authentication was exercised using the trained classifiers. Both conventional metrics and metrics applicable to continuous authentication that have been established recently were used for the performance assessment of the classifiers. Quantitative analysis of the test results indicate that the system is capable of detecting impersonation with high accuracy while other fraudulent acts are also detected with dependable accuracy. The system is also capable of the absolute and profile independent detection of clipboard transfer of text during online assessments. The thesis also discusses schemes for improving the overall detection capability of the system and suggests several directions to inspire future work.

Keywords: Learner Integrity, Continuous Authentication, Keystroke Dynamics.

TABLE OF CONTENTS

Declaration i
Dedication ii
Acknowledgement iii
Abstract iv
Table of Contents v
List of Figures xii
List of Tables xvi
List of Abbreviations xvii
List of Appendices xviii
CHAPTER 1 1
INTRODUCTION 1
1.1 Problem with online assessments 1
1.2 Attempts to prevent cheating at online assessments 2
1.3 Research Problem and the Research Questions 6
1.4 Research objectives 7
1.5 Outline of research work undertaken 7
1.6 Expected scientific and social outcome of the research 8
1.7 Organization of the thesis 8
CHAPTER 2 11
CRITICAL REVIEW OF RELATED WORK 11
2.1 Chapter description 11
2.2 Learner integrity 11
2.2.1 Brief history of academic dishonesty 12
2.2.2 Emergence of cheating within the sphere of online learning 13
2.2.3 Attempts to curb technology aided academic dishonesty 16
2.3 Overview of biometrics and biometric authentication 18
2.3.1 Physiological biometrics 19
2.3.2 Behavioural biometrics 20
2.3.2.1 Keystroke dynamics 21

2.3.2.2 Rhythm-Click Characteristics 24
2.4 Applications of Keystroke Dynamics 25
2.4.1 Principles and Techniques 25
2.4.2 Strengthening conventional authentication 27
2.4.3 Use as a Soft Biometric 28
2.4.4 Use of non-temporal features 29
2.4.5 Other applications of keystroke dynamics 30
2.5 Continuous Authentication (CA) 30
2.6 Performance assessment of biometric systems 32
2.7 Benchmark Datasets 33
2.8 Applicability to online authentication 34
2.9 Performance assessment of continuous authentication systems 34
2.10 Device Dependence 35
2.11 Challenges in applying authentication to Learner Integrity Management 36
2.12 Challenges posed by time variance and strategies for coping 37
2.13 Challenges posed by the availability of Large Language Models 39
2.14 Ethical considerations and legal implications 39
2.15 Conclusions drawn from the review of related work and their relevance to management of learner integrity 40
2.16 Recapitulation 44
CHAPTER 3 45
THE RESEARCH METHODOLOGY 45
3.1 Chapter description 45
3.2 Research methodology adopted 45
3.2.1 Approach to the research 45
3.2.2 Design of the research 45
3.2.3 Methods used for data collection 45
3.2.4 Strategy for collecting subjects 45
3.2.5 Methods used for the analysis of data gathered 46
3.2.6 Ethical and administrative considerations 46
3.2.7 Scope of the research and its limitations 46
3.3 Distinctive challenges associated with online assessments 47

3.4 Suitability of Keystroke Dynamics as the biometric 47
3.5 Use of Lockdown Browsers 48
3.6 The authentication system architecture and workflow 48
3.7 Description of the features to be extracted 53
3.7.1 The 'Dwell Time Vector' 54
3.7.2 The 'Hot Sequence Vector' 54
3.7.3 The 'Quick Sequence Vector' 55
3.7.4 The 'Far Sequence Vector' 55
3.7.5 The 'Pause Duration Ensemble' 56
3.7.6 Extraction of non-temporal features 59
3.8 Creation of the Candidate's Profile 59
3.9 Continuous Authentication during subsequent tests 60
3.10 Method of anomaly detection 60
3.11 Training the classifiers 66
3.12 Conformity to the specifications 68
3.13 Legal implications and ethical considerations 69
3.14 Recapitulation 70
CHAPTER 4 71
THE EXPERIMENTS 71
4.1 Chapter description 71
4.2 The experimental protocol 71
4.3 The data acquisition phase 72
4.3.1 Description of the tests 72
4.3.2 Selection of volunteers for roleplay 77
4.4 The profile building phase 79
4.4.1 Extraction of the 'Dwell Time Vector' 79
4.4.2 Extraction of the 'Flight Time Related Features' 80
4.4.3 Extraction of non-temporal features 84
4.5 The detection model development phase 85
4.5.1 Model development for dwell time related features 86
4.5.1.1 Generation of training data for 'genuine test taking' 86
4.5.1.2 Generation of training data for 'fraudulent test taking' 87

4.5.1.3 Building the composite dataset 88
4.5.1.4 Selection of roleplayers for simulation of 'fraudulent test taking' 88
4.5.1.5 Choosing a classifier model 89
4.5.2 Model development for the remaining features 89
4.5.3 Selecting the best classifier 90
4.5.4 Consistency with the concept 91
4.6 The analysis phase 92
4.6.1 Repeatability of experiments 92
4.6.2 Simulation of impersonation 93
4.6.3 Simulation of looking up and real time collaboration 95
4.6.4 Use of non-temporal features 96
4.7 Supplementary experiments 96
4.7.1 Protocol analysis of the client server communication during the acquisition of keystroke data 96
4.7.1.1 Examining the nature of the data transfer 97
4.7.1.2 Examining the effect of the volume of text typed 97
4.7.1.3 Testing of a lean keystroke logging system 98
4.7.2 Examination of clipboard transfers during online test taking 101
4.8 Choice of techniques and methods 103
4.9 Resources used in the experiments 103
4.10 Recapitulation 105
CHAPTER 5 107
INTERPRETATION OF RESULTS 107
5.1 Chapter description 107
5.2 Interpretation of the validity of the profiles created 107
5.2.1 Validity of the experimental subjects 107
5.2.2 Validity of the mock test 107
5.2.3 Validity of the features extracted 108
5.2.4 Validity of the training datasets 109
5.2.5 Validity of selection of the best classifiers 111
5.2.6 Overall validity of profiles created 114
5.3 Interpretation of results of the simulation of genuine test taking 114

5.3.1 Accuracy of genuine detections 114
5.3.2 False Rejection Worst Interval (FRWI) for genuine test taking 115
5.3.3 Validity of non-temporal features 115
5.3.4 Overall interpretation of results of the simulation of genuine test taking 116
5.4 Interpretation of results of the simulation of impersonation 116
5.4.1 Accuracy of imposter detection..... 120
5.4.2 False Acceptance Worst Interval (FAWI) during impersonation 121
5.4.3 Time to First Imposter Detection (T2FD) during impersonation..... 122
5.4.4 Unusability of non-temporal metrics in imposter detection..... 123
5.4.5 Overall interpretation of results of the simulation of impersonated test taking..... 123
5.5 Interpretation of results of the simulation of lookup..... 124
5.5.1 Accuracy of detecting lookup 124
5.5.2 Time related metrics during lookup 126
5.5.3 Overall interpretation of results of simulation of test taking while looking up..... 127
5.6 Interpretation of results of the simulation of collaboration..... 127
5.6.1 Accuracy of detecting collaboration 127
5.6.2 Time related metrics during collaboration 129
5.6.3 Overall interpretation of results of simulation of test taking while collaborating..... 130
5.7 Comparison of performance when detecting different fraudulent acts 131
5.8 Interpretation of results of the Protocol Analysis experiments 131
5.9 Interpretation of Results of the Clipboard Transfer Experiment..... 137
5.10 Key findings 140
5.11 Recapitulation..... 141
CHAPTER 6 143
DISCUSSION 143
6.1 Chapter description 143
6.2 Performance benchmark comparison 143
6.3 Establishment of Proof of the Concept 145
6.4 Revisiting the research questions 146

6.5 The drawbacks in the experiments 150
6.5.1 Low accuracy in detecting fraudulent acts other than impersonation ... 150
6.5.2 Non-usability of the biometrics of certain subjects 151
6.5.3 Inadequate replication of human factors during tests in the experiments 153
6.6 Suggestions for improving the authentication system 154
6.6.1 Alternative classifiers for detection of fraudulent acts other than impersonation 154
6.6.2 Use of machine learning and Artificial Intelligence 154
6.6.3 Improvements to the profile creation process 155
6.6.4 Improvements to the biometrics acquisition process 155
6.6.5 Improvements to the client – server communication during online tests and the processing of keystroke logs 156
6.7 Identification of points of failure 157
6.8 Key takeaways from the discussion 158
6.9 Recapitulation 158
CHAPTER 7 159
CONCLUSION 159
7.1 Chapter Description 159
7.2 Overall conclusion made through the research 159
7.3 Contributions of the research 160
7.4 Directions for future work 160
7.4.1 Additional features of non-obtrusive nature 160
7.4.2 Improved classifiers 162
7.4.3 Profile updating 163
7.4.4 Mitigation strategies 163
7.4.5 Population analysis 164
7.4.6 Device dependence 164
7.4.7 Adaptation of the system to other languages 165
7.5 Recapitulation 165
REFERENCES 167
BIBLIOGRAPHY 177
APPENDIX - A 179

ALGORITHMS USED IN THE EXPERIMENTS 179
Algorithm – 1: Extraction of dwell time vector from biometrics: 179
Algorithm – 2: Extraction of pause duration ensemble from biometrics: 182
Algorithm – 3: Extraction of flight time vectors (generic) from biometrics: .. 184
Algorithm – 4: Generation of training data for genuine test taking: 190
Algorithm – 5: Generation of training data for fraudulent test taking: 191
Algorithm – 6: Simulation of genuine and impersonated test taking: 192
Algorithm – 7: Simulation of fraudulent test taking: 193
Algorithm – 8: Selection of anomalous data sources for a roleplayer: 194
Algorithm – 9: Generation of training dataset for a roleplayer for a given feature: 195
Algorithm – 10: Generic algorithm to train and test a classifier: 197
Algorithm – 11: Selection of best classifier for a vector based feature: 198
Algorithm – 12: Selection of best classifier for an ensemble based feature: ... 200
APPENDIX - B 203
INDEX OF SOURCE LISTINGS 203
APPENDIX - C 221
RELATED PUBLICATIONS 221

LIST OF FIGURES

Figure #	Description	Page
2.1	Growth trend of research on 'keystroke dynamics'	23
2.2	Definitions of timing measurements on keystroke events	27
2.3	Variation of assessment metrics with sharpness of authentication system	33
2.4	Application of the 'Fraud Diamond' to the scenario of online assessments.	42
2.5	'Learner Integrity Managed' Online Testing System	43
3.1	Workflow of the proposed authentication system	49
3.2	The architecture of the proposed authentication system	50
3.3	First stage of offline profile creation – Reference vector extraction	51
3.4	Second stage of offline profile creation – Training data generation	52
3.5	Final stage of offline profile creation – Classifier training	53
3.6	Measurement of epochs for key events	53
3.7	Allocation of keys to each finger in each hand, for touch typing.	56
3.8	Algorithm to extract 'Dwell Time Vector'	57
3.9	Distinct clustering of pause durations of two persons	58
3.10	Algorithm to extract 'Pause Duration Ensemble'	58
3.11	3D/2D analogy for the formation of the vector pair μ_s and $\widehat{\mu_s}$ and the angle of departure θ.	61
3.12	Generation of a stream of θ values while the candidate types	62
3.13	Anomaly detection using the angle of departure	63
3.14	Different scenarios of occurrence for the 'Angle of Departure'	64
3.15	Anomaly detection using the mean value of pause durations	66
3.16	Algorithm for creating training data for binary classifier for dwell time related features	67
4.1	The phases of the experiments	71
4.2	A sample screen capture showing the response panel for a typical question in the online test interface	72
4.3	Contents of an example file storing the key event data	74

Figure #	Description	Page
4.4	Simulation of lookup during an online test	76
4.5	Simulation of real time collaboration during an online test	77
4.6	Relative frequencies of the highest occurring 20 sequences of two letters in English text	80
4.7	Structure of the 3-D array used to store hot sequences	82
4.8	Structure of 2-D array for storing the parameters of 'hot sequences'	82
4.9	Structure of the data frame for storing 'Quick Sequences'	83
4.10	Algorithm to generate datapoints for genuine test taking	86
4.11	Algorithm to generate datapoints for fraudulent test taking	87
4.12	Algorithm for simulation of continuous authentication amidst impersonation	93
4.13	Algorithm for simulation of continuous authentication amidst looking up	95
4.14	The test character sequence typed into the answer box of question 1 of Test 01, just before submission	97
4.15	Protocol Analyzer screen with filter 'HTTP' selected, showing the data exchange while the test sequence was typed	98
4.16	5-byte data structure used to encode a single key event	100
4.17	Algorithm for lean coding of a single key event	101
5.1	Frequency distribution of the size of keystroke logfile across the population of 38 roleplayers	108
5.2	Frequency distribution of the size of 'Pause Duration Ensemble' file across the population of 38 roleplayers	109
5.3	Scatter plots of training datasets based on 'Angle of Departure' between Dwell Time Vectors	110
5.4	Scatter plots of training datasets based on 'Angle of Departure' between Hot Sequence Flight Time Vectors	110
5.5	Scatter plots of training datasets based on 'Pause Duration Ensembles'	111
5.6	Frequency distribution of FRR (%) across the population of roleplayers	114
5.7	Frequency distribution of FRWI across the population of roleplayers	115
5.8	Frequency distribution of FAR (best performance) across the population of roleplayers	120

Figure #	Description	Page
5.9	Frequency distribution of average FAR across the population of roleplayers	121
5.10	Frequency distribution of average worst case False Acceptance Interval (in minutes) when roleplayers impersonate each other	122
5.11	Frequency distribution of the T2FD (minutes) when roleplayers impersonate each other	123
5.12	Frequency Distribution of FAR across 34 roleplayers when they take online test while looking up	125
5.13	Frequency Distribution of FAWI while looking up, across 20 roleplayers whose test taking exhibit an anomaly detection accuracy of more than 40%	126
5.14	Frequency Distribution of T2FD while looking up, across 20 roleplayers whose test taking exhibit an anomaly detection accuracy of more than 40%	127
5.15	Frequency distribution of FAR across 26 roleplayers taking test while collaborating	129
5.16	Frequency Distribution of T2FD across 24 roleplayers whose test taking exhibits a non-zero anomaly detection accuracy	129
5.17	Frequency Distribution of FAWI across 24 roleplayers whose test taking exhibits a non-zero anomaly detection accuracy	130
5.18	Contents of the JSON file holding the keystroke events transmitted from the browser client.	132
5.19	Protocol analyzer screen showing the loopback traffic while the test sequence was typed	133
5.20	Anatomy of the first HTTP datagram that carried the keystroke data of the test sequence to the server	134
5.21	The contents of the byte file resulting from the modified keystroke logging	135
5.22	Data on the wire when the test sequence ‘1234ABCD’ was typed	136
5.23	Variation of HTTP payload size with the number of keys typed for both types of keystroke logging	136
5.24	Response panel of the modified Test 01 to examine text transfer via the clipboard	138
5.25	The responses captured separately on the JSON file and the answer file.	139
6.1	Algorithm used by the back end script for logging keystroke data	152

Figure #	Description	Page
6.2	Algorithm for converting ‘byte file’ to a JSON log	156

LIST OF TABLES

Table #	Description	Page
1.1	Countermeasures for fraudulent practices during online assessments	03
1.2	Effectiveness of technical countermeasures	04
3.1	Comparison of proposed system with other commonly used biometric authentication systems	69
4.1	Completion of roleplays by the volunteers	79
4.2	6-bit code to encode the distinct keys on a QWERTY keyboard	99
5.1	Test completion by the roleplayers and usability of biometrics	107
5.2	List of classifiers for a single feature	112
5.3	Number of roleplayers for whom each classifier has been selected as the best	113
5.4	Comparison of non-temporal features extracted from Tests 01 and 02	116
5.5	Best performance metrics of simulated impersonations by 37 imposters	117
5.6	Average values of performance metrics of simulated impersonations by 37 imposters	118
5.7	The variance of average test session parameters across the population of 38 roleplayers, when impersonated by 37 roleplayers	119
5.8	Performance metrics of test taking by 34 roleplayers while looking up	124
5.9	Performance metrics of test taking by 26 roleplayers while collaborating	128
5.10	Comparison of detection performance for different fraudulent acts	131
5.11	Observations made upon different scenarios of text input	137
6.1	Performance comparison with a benchmark system	145
6.2	Evaluation of Authentication Methods deployable during online tests	147

LIST OF ABBREVIATIONS

Abbreviation	Expansion / Description
2FA	Two-Factor Authentication
AI	Artificial Intelligence
ANGA	Average Number of Genuine Actions
ANIA	Average Number of Imposter Actions
ASCII	American Standard Code for Information Interchange
.CSV	Comma-Separated Values (extension for files containing tabular data)
CA	Continuous Authentication
CBAS	Continuous Biometric Authentication System
CBT	Computer Based Testing
DB	Benchmark Dataset (*abbreviated in French*)
DT	Dwell Time
ECMA	European Computer Manufacturers Association
ECG	Electrocardiogram
EER	Equal Error Rate
ENISA	European Network and Information Security Agency
FAR	False Acceptance Rate
FAWI	False Acceptance Worst Interval
FRR	False Rejection Rate
FRWI	False Rejection Worst Interval
FT	Flight Time
FTP	File Transfer Protocol
GenAI	Generative AI
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
LLM	Large Language Model
LMS	Learning Management System
LR	Logistic Regression
LTS	Long Term Support
MOOC	Massive Open Online Course
NB	Naïve Bayes
NLP	Natural Language Processing
PHP	PHP : Hypertext Preprocessor (*recursive acronym*)
RF	Random Forest
SFTP	Secure FTP
SVM	Support Vector Machine
T2FD	Time to First Detection
TCR	Time to Correct Reject
URL	Uniform Resource Locator
UTC	Universal Time Coordinated (or Coordinated Universal Time)
WPM	Words per Minute
XAI	Explainable AI

LIST OF APPENDICES

Appendix #	Description	Page
Appendix - A	Algorithms Used in the Experiments	179
Appendix - B	Index of Source Listings	203
Appendix - C	Related Publications	221