Citation for published version:
Abrar Ullah, Hannah Xiao, Trevor Barker, and Mariana Lilley, 'Evaluating security and usability of profile based challenge questions authentication in online examinations', *Journal of Internet Services and Applications*, Vol. 5 (2): March 2014.

DOI:
https://doi.org/10.1186/1869-0238-5-2

Journal of Internet Services and Applications
a SpringerOpen Journal

RESEARCH Open Access

Evaluating security and usability of profile based challenge questions authentication in online examinations

Abrar Ullah*, Hannan Xiao, Trevor Barker and Mariana Lilley

Abstract

Student authentication in online learning environments is an increasingly challenging issue due to the inherent absence of physical interaction with online users and potential security threats to online examinations. This study is part of ongoing research on student authentication in online examinations evaluating the potential benefits of using challenge questions. The authors developed a Profile Based Authentication Framework (PBAF), which utilises challenge questions for students' authentication in online examinations. This paper examines the findings of an empirical study in which 23 participants used the PBAF including an abuse case security analysis of the PBAF approach. The overall usability analysis suggests that the PBAF is efficient, effective and usable. However, specific questions need replacement with suitable alternatives due to usability challenges. The results of the current research study suggest that memorability, clarity of questions, syntactic variation and question relevance can cause usability issues leading to authentication failure. A configurable traffic light system was designed and implemented to improve the usability of challenge questions. The security analysis indicates that the PBAF is resistant to informed guessing in general, however, specific questions were identified with security issues. The security analysis identifies challenge questions with potential risks of informed guessing by friends and colleagues. The study was performed with a small number of participants in a simulation online course and the results need to be verified in a real educational context on a larger sample size.

Keywords: Security; Usability; Online learning; Online examination; E-learning; MOODLE; Challenge questions; Authentication

1. Introduction

This study investigates student authentication in online learning and examinations. Student identification in online learning is largely reliant upon remote authentication mechanisms. The absence of face-to-face identification can make online learning and high stakes examinations vulnerable to a number of authentication threats and therefore, the security of online learning environments is highly important. Online learning offers a number of advantages including availability, reliability, flexibility and reusability [1,2]. Besides the anticipated benefits of online learning, it has some limitations including the security of online examinations as one of the major concerns.

In typical online environments, examination is an integral part of the learning process. In online examinations, face-to-face invigilation is often replaced with authentication systems and therefore, security becomes a critical factor with regard to their credibility. Secure authentication is particularly relevant to the success of high stakes online examinations. Effective authentication approaches are important to ensure secure, reliable and usable student authentication mechanisms in an online learning and examinations context. The implementation of a reliable and secure approach to students' authentication is vital to ensure trust of the stakeholders in the assessment process. It has been an active research area and a number of authentication techniques have been implemented in order to ensure secure online examinations. A diverse set of authentication techniques have been developed in earlier research work, which verify online

* Correspondence: abrar.ullah@gmail.com
School of Computer Science, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

users' identities based on knowledge or "*What one knows*" [3], possession of objects or *"What one has"* [4] and biometrics or *"What one is"* [5].

In our earlier study [6], we developed the Profile Based Authentication (PBAF) approach for student authentication in online examinations and presented a usability analysis of using challenge questions as a second factor authentication. The results of this study have been presented [7]. In them, we discussed the impact of the clarity and memorability of questions on effectiveness of the PBAF method. The study [7] also analysed participants' feedback through an online survey to determine various usability attributes as well as user satisfaction.

The current paper further explores the strengths and weaknesses of the PBAF method in terms of usability, security and the effect of question design on the overall authentication process. In addition to the above, this paper presents a detailed analysis of the security of the PBAF method in a follow-up guessing authentication attack to risk assess and mitigate any threat. Participants of the follow-up abuse case scenario were selected from the original users group, who participated in the previous phases of the study. The guessing attack was performed to analyse the resilience of challenge questions to informed guessing by friends and colleagues. The findings also contributed to the design and implementation of a traffic light system in the PBAF.

The structure of the paper is organised into 5 sections. The paper starts with an introduction to online learning, examination and authentication challenges in Introduction. The work background and literature review is presented in Background and related work. The research methodology including empirical design, participant recruitment and empirical implementation phases are presented in Study design and methodology. The results, analysis and findings of empirical investigations are discussed in Results. The concluding remarks including work summary and future directions are presented in Conclusion.

2. Background and related work

The online examination is an important feature and critical asset of online learning [8]. A number of previous studies have acknowledged that student authentication in online examinations faces many security threats. Unethical conduct has been growing in online learning due to un-controlled environment in online examinations as a result of use of technology and the Internet [9,10]. Agulla [9] suggests that it can be a real challenge to verify the identity of an individual in an online environment without any physical interaction. Colwell and Jenks [11] argue that online examinations are more vulnerable to academic dishonesty than traditional face-to-face examinations. A large number of authentication techniques have therefore been developed, which can be implemented to enhance the security of online examinations.

The traditional authentication techniques are classified into three categories:

- Knowledge Based Authentication (KBA) e.g. login-identifier and password, passphrase, challenge questions
- Object Based Authentication (OBA) e.g. smart cards, ID cards
- Characteristics Based Authentication (CBA) or Biometrics e.g. fingerprint, audio or voice recognition, signature recognition and face recognition.

The above authentication techniques have their strengths and weaknesses in terms of cost, usability and security, when applied to online learning environments [6]. KBA are the most prevalent, cost effective and widely accepted approaches [12]. However, KBA approaches can be vulnerable to security attacks including collusion, guessing, lost credentials, dictionary attacks and brute-force attacks [3]. The OBA approaches are widely used in banking, transports, hotels and parking areas, with a potential for use in online learning [13]. The OBA features may be useful to resist adversaries' attacks. However, the authentication objects can be shared, lost or stolen for use in authentication attacks. The OBA features require special purpose input devices, which incurs additional cost. The use of special purpose input devices may limit the implementation of OBA in online learning environments. The CBA approaches free individuals from remembering passwords and carrying cards. An individual's physical or behavioural characteristics are a key to the identification and therefore, CBA (biometrics) are seen as the most reliable authentication features [14]. The CBA features also require special purpose input devices for recording and authentication, which incurs additional cost. The special purpose input devices may limit the scope of CBA implementation in a wider Internet context. The CBA approaches have been reported with algorithm challenges like False Accept Rate (FAR), False Reject Rate (FRR), Equal Error Rate (ERR), Failure to Enrol Rate (FER) and Failure to Capture Rate (FCR) [15].

In light of the above discussion, it is desirable to develop an authentication feature, which is secure, cost effective and accessible to a large online population using standard input devices. The authors designed and developed the PBAF method, which implements challenge questions coupled with login-identifier and password features for authentication purposes. The PBAF approach is chosen for a number of reasons. Primarily, the PBAF integrates learning and the examination process, whereby answers to profile questions collected in the learning process are utilised to authenticate students in the examination process. Unlike biometrics and object-based methods, the PBAF, being a knowledge-based method, can be implemented to cover a large online population using standard input devices. The design, development, implementation and maintenance of the PBAF method can be cost effective. In our previous work, we:

- implemented the PBAF method in an online learning environment, to authenticate students, firstly at a course access level and secondly at examination access level [6].
- organised an empirical study to research the usability of the PBAF method in terms of memorability of questions, clarity of questions, syntactic variation and implementation of a traffic light system [7].
- performed an in-depth analysis of the design of questions and their impact on the usability attributes. The study reported an analysis of completion time of the profile questions and the results of a post study survey to present participants' feedback on layout and usability [16].

The challenge questions are a key to the PBAF approach and are designed to be reliable and unique as they pertain to information known to individual users. It is widely seen as a credential recovery technique [17]. Challenge questions are also employed for customer verification in online and telephone banking [18]. In a recent study, Just and Aspinall [19] reviewed the use of challenge questions as a second factor authentication in 10 UK banks, which indicated that the method was reliable and used for the security of monetary transactions in financial institutions.

Besides the anticipated benefits, challenge questions have some limitations. Some studies have reported usability and security issues related to the use of challenge questions in credential recovery [17,20]. In [17], it is also argued that the collection of sensitive information about users can raise privacy and ethical issues. The usability of any authentication approach is highly important for reliability and security. It is recognized that the memorability of challenge questions and lack of clarity may cause security and usability issues [7,21].

From the above discussion, it is evident that challenge questions can be useful as a second factor authentication. However, to achieve effective authentication using the PBAF method in online examinations, usability and security issues need to be investigated.

2.1 Profile based authentication

The PBAF is a multi-factor knowledge based authentication approach, which utilises login-identifier and password and challenge questions. It integrates the learning and examination processes, whereby answers to profile questions collected during learning activities are utilised for authentication in the examination process.

Using the PBAF method, students are provided with a unique login-identifier and password for logging into the learning environment. After successful login, students are required to answer profile questions in order to gain access to learning resources. The profile questions are used to collect answers in order to build and update individuals' profiles. The profile is a student's description in the form of questions and answers. It is anticipated that learning is a recurrent activity and the students' profiles are consolidated in multiple visits. The secondary authentication process is triggered when students request to access an online examination. They are then required to provide matching answers to a set of challenge questions randomly selected from their profiles. The PBAF being a knowledge-based method can be implemented to cover a large online population and may provide adequate security against many authentication attacks. The PBAF was implemented on a Modular Object Oriented Dynamic Learning Environment (MOODLE) Learning Management System (LMS) for the purpose of this empirical study. MOODLE is a free source environment with a modular and extendable structure. A brief description of how the PBAF approach to student authentication works can be found below:

- *PBAF Setup:* The PBAF provides a configurable web interface. This is used to add pre-designed questions to the library for use as profile and challenge questions. The number of profile and challenge questions requested at learning and authentication phases are configurable items in this interface.
- *Profile Questions:* Profile questions are presented to students in order to build their profiles. Each profile question is presented to each individual student once. The profile questions are a subset of pre-designed questions added in the PBAF setup. Students are required to supply answers to these questions on each visit to obtain access to learning resources.
- *Challenge Questions:* The PBAF generates and presents random challenge questions when access to online examination is requested. The student registers n profile questions, and is presented with $t \leq n$ challenge questions upon authentication [7,22]. To an individual student, $\mathrm{r} = t$ challenge questions must be answered correctly in order to access online examination. However, if an error tolerant traffic light system is implemented, it is sufficient to answer $r \leq t$ challenge questions correctly in order to access online examination. The challenge

questions are randomized using a random floating-point value v in the range *0 < = v < 1.0*, which is generated by MySQL database [23]. The students' answers to challenge questions are authenticated and a *timestamp* is stored with individual questions in their respective profiles to exclude questions presented within the past 24 hours.

- *Traffic Light System:* To relax the authentication constraints for enhanced usability, a traffic light system is embedded in the PBAF. The traffic light system authenticates users based on the number of correct answers to challenge questions. A three scale classification is adopted to authenticate users, which are red, amber and green. Users in the red classification are locked out and denied access to examination. Users in the amber classification are presented more challenge questions to re-authenticate and users in the green classification are granted access to examination.
- *Authentication:* The authentication algorithm implements string-to-string comparisons to match the answers with the stored information. In earlier studies, researchers used a combination of algorithms for comparative analysis. In their work Schechter et al. [20] implemented an equality algorithm for string-to-string comparison, substring algorithms, and distance algorithms were also used. In another study, Just and Apsinall [24] proposed guidelines for designing usable and secure challenge questions which recommended removing white spaces, punctuation and capitalization for enhanced usability. The PBAF method implements the equality algorithm for exact match without the pre-processing of answers. The equality algorithm was chosen for better security and to use the results as a benchmark, which could be compared with those from revised algorithms to be investigated in future stages of this research. The nature of this algorithm means that students are allowed to access online examinations only if they provide exact answers to their challenge questions. The PBAF method implements randomization of questions during multiple attempts and poses questions which were not previously presented in the last 24 hours, in order to be effective against security threats including brute-force guessing attacks [25]. A specific number of incorrect answers to challenge questions locks out the user from further attempts and requires administrator intervention to unlock the account.

3. Study design and methodology

The aim of this study was to analyse the usability and security of the PBAF method in the context of online examinations. A set of 20 questions was compiled to cover the academic, personal, contact, favourite and date themes. The experiment was performed in an online environment and the empirical design and methodology was approved by the University of Hertfordshire's research ethics committee. The study was conducted to test the following hypotheses:

- The PBAF meets standard usability criteria of efficiency and effectiveness.
- The traffic light system enhances the usability of PBAF method by relaxing authentication constraints.
- The PBAF is secure against informed guessing attacks by friends and colleagues.

The above hypotheses were framed to analyse the usability attributes, which were informed by research work in the domain of usability and software quality [26,27]. Bevan [28] states that usability and quality complement each other and that usability is quality in use. As in [27], the quality factors include efficiency, effectiveness, satisfaction, accessibility, productivity, safety and international-ability. In a similar vein, Nielsen [29] defines usability as a property with multiple dimensions each consisting of different components. He also suggests that the different factors can conflict with each other. Nielsen defined a number of usability factors including learnability, efficiency, memorability, errors, and satisfaction. Learnability defines, how well a new user can use the system, while the efficient use of the system by an expert is expressed by efficiency. Effectiveness is the degree of accuracy and completeness with which the user achieves a specified task in a certain context [20]. If a system is used occasionally the factor memorability is used, which dictates effectiveness. Satisfaction is a qualitative attribute which largely depends upon users' feedback based on the effective and efficient use of the artefact. The authors evaluate applicable usability attributes in the context of online learning and examinations, which include efficiency, effectiveness, satisfaction and memorability of questions. In previous studies, the authors evaluated user satisfaction [16] and memorability [7] attributes, while this work analyses the efficiency and effectiveness of challenge questions used in the PBAF.

Previous research suggests that challenge questions can be vulnerable to guessing attacks by friends and colleagues [20,25]. Just and Aspinall [22] describe guessing in three categories, which are "Blind guessing", "Focused Guessing" and "Observation". In blind guessing, the attacker performs a brute-force attack without considering the question. In focused guessing, the attacker may still use a brute-force technique, however, the search space is cut down by considering the question type. In observation, the attacker performs an informed guess about both the user and question. Schechter [20] performed guessing attacks by acquaintances and statistical guessing in the context of credential recovery to evaluate security of challenge questions. We organised an informed guessing (observation) abuse case scenario in the context of online learning and examinations using the PBAF method. This study does not cover blind and focused guessing. The abuse case was performed to assess risks and mitigate any security threat using the method defined by ISO 31000 [30].

3.1 Participants recruitment

The participants were recruited from a pool of local and international undergraduate and postgraduate full time students from the UK and overseas universities. All the participants were informed and provided with study design and guidance notes explaining the aims and objectives of this research. Guidance notes were emailed to all participants to describe the registration procedure, access dates for learning, and the examination. Of the total 30 potential participants, 23 consented to participate in the experiment. In a follow-up abuse case scenario, we circulated a list of 10 participants requesting them to identify their colleagues and friends from the first cohort, who participated in the learning and examination phases of the study. A total of 6 participants consented to take part in the abuse case scenario. The participants recruited for the abuse case scenario were required to impersonate their friends and colleagues and attack the online examination for security analysis.

3.2 Questions design

The questions for this empirical study were compiled into five different themes i.e. academic, personal, favourite, contact, and date themes. The question design in the academic and contact themes was based on the University of Hertfordshire undergraduate admission form to minimize any privacy concerns. Questions in the personal and favourite themes were inspired from the corporate email service providers i.e. Google, Microsoft, AOL and Yahoo [20]. Usability, privacy and security were considered when designing the questions. The findings from PBAF adopting these questions will be used as a benchmark, which can be compared and optimised in the future stages of this research.

3.3 Empirical study phases

Our experiment was organised into five phases; setup phase, online registration phase, online learning phase, online examination phase and security test phase. The empirical activities shown in Figure 1 were performed remotely over the Internet in a simulated environment on MOODLE LMS. The PBAF was developed in PHP server side scripting language and integrated with the LMS deployed on a test server for the purpose of this empirical study. A simulation online learning course was created on a remote server and a mock-up online examination added to the course. The online course and examination were designed only to achieve the research objectives and was not an actual University course. Participants were required to answer the profile and challenge questions to authenticate their online examinations. The experiment was performed in the phases described below. Some initial configurations were performed in the initial setup phase before the study commenced.

Initial Setup Phase: An initial setup was required to set out values of the configurable variables. A set of 20 questions designed for the study was uploaded to the PBAF. The number of profile questions presented during the learning process is configurable and was set to 3. The number of challenge questions presented during the examination process is configurable and was set to 3. The following traffic light configuration was defined:

1. *Criteria 1-Red*: If the number of matched answers to the challenge questions is classified *red*, the participant is locked out and access to online examination is denied. The value of the red classification was set to 0.
2. *Criteria 2-Amber*: If the number of matched answers to the challenge questions is classified *amber*, the participant is presented with more challenge questions to authenticate iteratively. The value of the amber classification was set to 1.
3. *Criteria 3-Green*: If the number of matched answers to the challenge questions is classified *green*, the participant is authenticated and access to online examination is granted. The value of the green classification was set to 2.

Online Registration Phase: The experiment was started from the online registration phase as shown in Figure 1. The participants completed the registration and created their login-identifier and password. The login-identifier and password provides the primary authentication to access the simulation online course.

Online Learning Phase: The participants were required to access the LMS and visit the simulation online course accessed for a period of one month with a minimum of three days between each visit. As learning is a recurrent process, therefore, participants were required to visit the online course on multiple dates. The following steps were performed in the online learning phase.

- The Participants accessed the online course using their login-identifiers and passwords created in the registration phase.

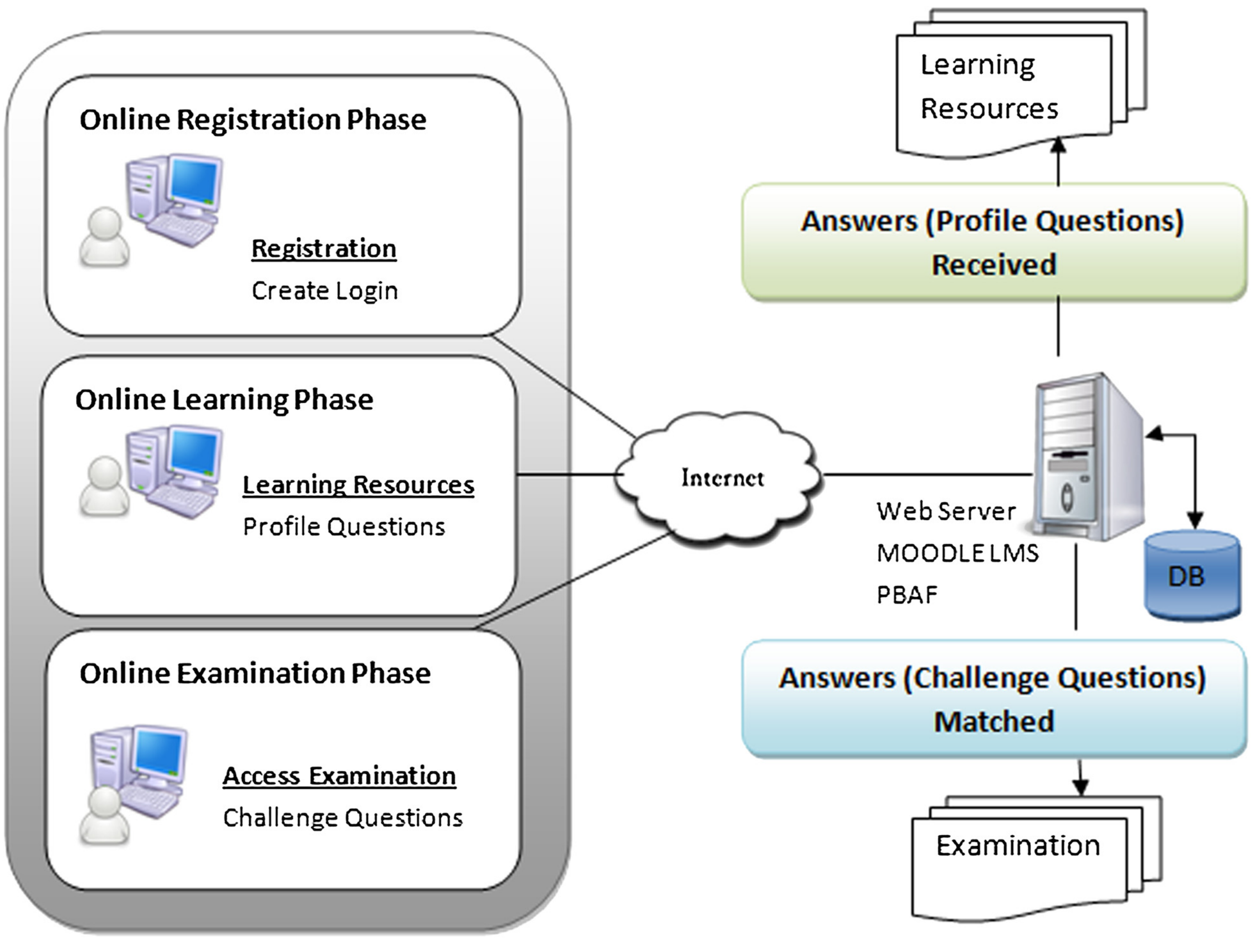

Figure 1 Empirical study phases.

- On each visit the participants were redirected to answer 3 profile questions in order to access the online course. For the purpose of the study reported here, the number of questions was set to 3 in the initial setup phase. This would allow the authors to collect sufficient data for the preliminary analysis, without causing fatigue to the participants.
- The profile questions and their answers were stored in the database to build and consolidate individual participant's profiles.

Online Examination Phase: On completion of the online learning phase, the participants were notified by email to access the online examination. There was an intervening period of 30 days between the participants' first access to learning and the online examination phases. The following steps were performed in the online examination phase.

- The participants accessed the online course using their login-identifier and password created in the registration phase.
- When the participants visited the online examination they were redirected to answer 3 challenge questions selected randomly from their profiles, in order to assess their access status. The challenge questions presented in the past 24 hours were excluded to mitigate brute-force, blind and focused guessing attacks.
- Authentication was performed using the *equality* algorithm for string-to-string comparison. The traffic light system was disabled in the participants' first visit to the online examination for comparative analysis of data with and without the traffic light system. The participants were granted access to the examination, when answers to all their 3 challenge questions matched the stored credentials. In the subsequent visits to online examinations, the traffic

light system was enabled as shown in Figure 2, and described below:

a) If the *number* of matched answers to the challenge questions is classified as *red,* deny access and block the participant's account.
b) If the *number* of matched answers to challenge questions is classified as *amber,* present more challenge questions and repeat the authentication. The amber classification is repeated until the status is changed or all the challenge questions in the individual's profile are exhausted. Those participants exhausting all their challenge questions are locked out.
c) If the *number* of matched answers to the challenge questions is classified as *green,* grant access to the examination.

Security Test Phase: We conducted a follow-up study for security assessment. An abuse case scenario was performed to risk assess the PBAF approach against guessing attacks. Research studies [20,25] suggest that challenge questions can be vulnerable to blind, focused and informed guessing attacks by adversaries, acquaintances, friends and colleagues. To evaluate the resilience of challenge questions to informed guessing attack by friends and colleagues, we performed an abuse case scenario involving pairs of friends and colleagues from the existing participants. As explained previously, this study does not cover statistical, blind and focused guessing. The use case presents a scenario, where an individual obtains the login-identifier and password of a friend or colleague, gains access to the online environment and performs *informed guessing* to answer challenge questions during authentication. The following steps were taken to perform the abuse case scenario:

- We required the participants to identify their friends and colleagues from the first cohort participating in the previous phases of the study. Of the first cohort of 23 participants, a group of 6 volunteered to take part in the abuse case scenario and notified their friends.
- We paired the participants with their friends and colleagues so each individual can cross attack a friend's account.
- Fictitious passwords were created for participants in the abuse case scenario. The login-identifiers and passwords of friends and colleagues were amended for privacy reasons and shared with the designated participants to enable them to impersonate as their colleagues.
- The participants visited the course using their friends' login-identifier and password.
- The participants visited the online examination on behalf of their friends and were presented with 3 random challenge questions. Answers to the challenge questions were submitted using informed guesses. The authentication feedback was not revealed to the participants and stored in the database for security analysis.
- The traffic light system was enabled using the criteria outlined in the online examination phase. Using the traffic light system, the participants meeting the criteria in red classification were locked out. The participants meeting the amber classification criteria were recurrently presented with more questions until the status was changed or

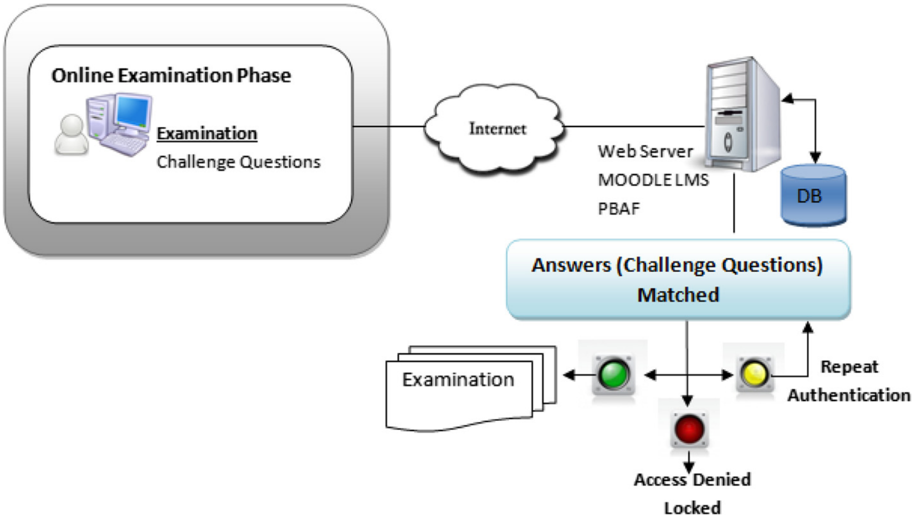

Figure 2 Traffic light system in online examination phase.

all the challenge questions in the respective profile were exhausted. The participants meeting the red classification criteria were locked out.

4. Results

Of all the invitees, 23 participated in the initial registration and 18 took part in the various phases of the empirical study by providing answers to 274 profile questions. A total of 13 participants answered 66 challenge questions in the online examination phase of the experiment and completed the authentication.

A group of 6 students participated in a follow-up security test phase and submitted answers to 24 challenge questions, guessing on behalf of their colleagues.

The usability and security analysis are discussed below.

4.1 Usability analysis

The usability results presented here are extracted from the data taken from the participants' interactions with the online learning and examination phases discussed in Study design and methodology. We have analysed the usability of questions in the online examination and traffic light authentication phases. In the online examination phase, participants managed to submit 38 (58%) matched answers, whereas, 28 (42%) unmatched due to various usability issues. The efficiency and effectiveness of questions in the context of online learning and examinations are evaluated in the discussion below.

4.1.1 Efficiency

Efficiency is a usability metric defined by ISO, which can be evaluated by measuring the completion time of each task and sub-tasks separately [27]. A system is considered efficient, if users are able to complete tasks in a reasonable time.

The efficiency was analysed from data collected during participants' answers to profile questions in the learning phase. To examine the efficiency of questions in the PBAF method, the "completion time" and "answer length" of answers to profile questions were measured. The mean score and standard deviation of completion time and answer length was computed and presented in Table 1. The correlation analysis of the two variables was measured to analyse the efficiency of profile questions used in this study. A Pearson Correlation was computed to examine the relationship between the "completion time" and the "answer length". Table 2 shows the Pearson r =0.152; p value 0.011 ($p < 0.05$) indicates a significant correlation between the two variables where $n = 274$. The small value of $r = 0.152$ suggests that there were other intervening variables affecting the completion time, however, these are not covered in this study. The potential factors that can impact the completion time include typing speed, question relevance to the individual, personal break, Internet connection speed and any privacy concerns. The efficiency of questions in various themes is discussed below.

Table 1 Usability analysis: efficiency

Question themes	Completion time (seconds)		Answer length (characters)	
Academic questions	**Mean**	**SD**	**Mean**	**SD**
Find out about this course	14.14	7.98	7.0	6.11
Student number	14.55	8.52	3.0	2.9
Name of last school attended	14.60	6.67	14.86	9.38
Grades in highest qualification	15.14	6.29	2.0	2.47
Year of highest qualification	15.20	7.16	4.0	0
Month started the current course	15.61	8.06	5.0	2.03
Year started the current course	16.18	8.98	4.29	1.07
Highest qualification	16.93	6.80	9.40	8.47
Personal questions				
Father's surname	13.55	8.76	4.71	1.26
Country of birth	13.78	7.25	7.20	1.37
Best friend's surname	14.47	6.95	5.79	2.57
Dream job as a child	18.03	8.65	9.85	5.24
Favourite questions				
Hero of your childhood	14.70	5.94	11.71	5.31
Tutor	15.06	8.13	8	3.48
Module on this course	18.34	9.8	7.5	5
Contact questions				
Home Tel no with country code	15.73	8.78	10.60	3
Home address town	16.83	9.36	15	13.75
House name or number	17.18	7.8	19.58	18.55
Mobile number with country code	17.43	8.98	11.69	1.43
Date questions				
Date of birth	16.42	6.75	6.36	3.91

Academic Questions: The relevance of questions is important to inform the efficiency of the PBAF approach. The participants responded to pertinent academic questions, with an efficient completion time. As an example, the completion time of answers to profile questions "*Where did you find out about this course*", "*student number*" and "*Last school attended*" was the

Table 2 Pearson correlation

		Answer length	Completion time
Answer length	Pearson correlation	1	.152*
	Sig. (2-tailed)		.011
	N	274	274
Completion time	Pearson correlation	.152*	1
	Sig. (2-tailed)	.011	
	N	274	274

*. Correlation is significant at the 0.05 level (2-tailed).

shortest in the academic theme with a mean completion time of 14.14, 14.55 and 14.60 seconds, which indicates that the relevance of questions is an important factor leading to increased efficiency.

Questions with answer hints can also contribute to enhanced efficiency. The findings indicate that embedded answer hints in questions were treated as an answer choice by participants, which enhanced efficiency. As an example, the profile question "*Where did you find out about this course*", shows a high degree of efficiency, because it was presented with an answer hint i.e. "Friend, Internet" to help participants understand the context of the question. Although the completion time of the question was efficient, 78% of the answers were identical and selected from the answer hint "Friend, Internet", which can be usable, but may lead to security risks.

The use of abbreviations in answers can affect the usability of challenge questions. It was noted that in spite of efficient completion time of 14.60 seconds, the length of answers to question "*Name of last school attended*", was the largest in the academic theme. To account for the length, further exploration of answers revealed that 44% of answers were abbreviations and 56% were full school names. Long school names resulted in increased answer length.

Question clarity is important for the efficient of responses. The completion time may increase for vague and unclear questions irrespective of their answer length. The completion time of answers to the profile question "*Grades in highest qualification*", was recorded in 15.14 seconds. The completion time was higher for an average answer length of 2 characters. The question does not explicitly specify grade type, which resulted in variations in answers. The detailed sorting of answers revealed that participant submitted different grade types (letters, percentage and description). The answers contained 64% letters "e.g. A, A*, A+", 22% percentage type and 14% descriptive texts.

Question context and relevance to individuals is highly important for the usability of the PBAF method. The profile question "*Month started current course*" was completed in 15.61 seconds. The detailed analysis of answers revealed that participants in the empirical study were originally enrolled on different courses at their respective institutions and questions in the context of the empirical simulation course needed further clarity. The participants were not particularly aware of "*current course*" in the context of a simulation course and the question vagueness contributed to delay in response time. Of the total answers to this question requesting "*month*" information, 50% were incorrect. A similar response was noted to profile question "*year started current course*" with a mean completion time of 16.18 seconds. The detailed exploration of answers revealed a 28% *"incorrect year"* or unrealistic answers. The increased completion time can be attributed to the relevance and clarity issues reported above with respect to "*current course*".

Questions with long anticipated answers can affect the usability. As an example, name of the institution or employers can be long and descriptive. The completion time of profile question "*highest qualification*" was 16.93 seconds, which is the largest in the academic theme with increased answer length.

Personal Questions: Personal questions are believed to be usable and widely used by the corporate email providers e.g. AOL, Yahoo, Google and Microsoft [20]. Our results indicate that the completion time of personal questions was efficient. The completion time of answers to profile questions "*Father's surname*", "*country of birth*" and "*Best friend's surname*" was 13.55, 13.78 and 14.47 seconds and the answer length was 4.71, 7.20 and 5.79 characters. The average completion time of the questions indicate slight variation with positive efficiency.

The personal questions requesting subjective information from the past resulted in a high completion time. As an example, the profile question "*Dream job as child*" resulted in higher completion time and answer length as 18.03 seconds and 9.85 characters.

In conclusion, the mean time incurred on all questions in the personal theme was 14.89 seconds, which is an efficient completion time in the online setting.

The results clearly indicate that better clarity and readability of questions in the personal theme was one of the factors resulting in enhanced efficiency.

Favourite Questions: Favourite questions have been widely used for credential recovery [20]. The favourite questions collect subjective information, which may change over time and circumstances, however, popular favourite questions can be usable. As an example, the completion time of profile questions "*Hero of childhood*" and "*Tutor*" was 14.70 and 15.06 seconds, which indicates positive efficiency.

As discussed earlier, the question's context and relevance is highly important for better usability. As an example, the completion time of the answer to profile question favourite "*Module on this course*" was 18.03 seconds. The "*module on this course*" in question was not relevant in the context of a simulation course and lacks clarity. The analysis of data revealed that 47% of answers contained unrealistic patterns like "NA, Nil, and Unknown".

A large number of questions requested subjective information; however, the overall efficiency of profile questions in the favourite theme was positive.

Contact Questions: The questions requesting contact information were created in a more generic way, to cover addresses for a wide range of participants in different geographic locations. However, this created clarity issues. The completion times of answers to profile questions "*Telephone number including country code*" and "*Address town"* was 15.73, 16.83 seconds respectively and answer length was 10.60, 15 characters. Detailed analysis of answers to "*Address town*" revealed that 33% of all answers contained full address and 67% were address town or city name, which indicates lack of clarity.

The completion time of answers to the profile question "*House name or number*" was 17.18 seconds with the largest answer length 19.58 characters. Analysis of the answers revealed that the generalization of question created ambiguity and answer lengths contained large variations. Participants' answers contained 42% full home address, 25% house number, 17% home phone number, 8% house name and 8% of city name, which shows rapid answers shift.

From the above discussion, a pattern can be noticed in answers to questions in the contact theme with increase in completion time and answer length, which may also affect the effectiveness during authentication process.

Date Questions: The date information is often presented and stored in varied formats. Without specifying a format, collection of date information can invite syntactic variation, which can affect the usability. The completion time of answers to profile question "*Date of birth*" was 16.42 seconds. The further analysis of participants' answers revealed that open and varied "date" format was used in answers with the use of special characters "/", "-" and descriptive "month name e.g. October". Using a standard date format can enhance the efficiency of date type questions.

Summary of Efficiency: In summary, the completion time reflects the efficiency and participants' understanding of questions and their ability to answer realistically. Questions with design flaws require extra thinking and time to respond and therefore it may result in distraction and have implications for the overall efficiency of the PBAF method. The shortcomings in question design may affect the efficiency of the PBAF and also reflect on usability during online examination, which is discussed below. Profile questions with an answer hint resulted in efficient completion time; however, this approach can create security risks.

The results reported here in terms of efficiency suggest that the question design should consider clarity, relevance and students' anticipation to conveniently answer the questions. Questions inviting long answers, as in the contact theme, may incur extra completion time and result in low efficiency.

For the reasons covered in this section, the efficiency hypothesis of the PBAF was supported for selective questions used in this study. However, it would be interesting to further investigate the efficiency of the PBAF method and revise questions with enhanced clarity in a real online course.

4.1.2 Effectiveness

Effectiveness may be considered to be the degree of accuracy of responses. Effectiveness, in the context of PBAF questions evaluation was taken to mean that participants were able to submit a maximum number of matched answers effectively with low error rate.

Effectiveness was analysed on data collected from participants' answers to challenge questions during the online examination. To examine effectiveness and accuracy, participants' answers to challenge questions were analysed into 5 common themes as academic, personal, contact, favourites and date. We used the equality algorithm in the empirical study. However, results were compiled to analyse the effectiveness if a more relaxed algorithm was implemented. The results of a relaxed algorithm were derived from the data collected in the online examination disregarding capitalisation, whitespaces and minor spelling errors using a combination of substring and distance algorithm as described in an earlier study [20]. Table 3 shows the crosstab analysis of data using the equality and relaxed algorithms under columns 3 to 6 headings. Data in columns 5 and 6 presented in bold-face show an increase in effectiveness when results were computed using a relaxed algorithm. The answers were submitted by all participants during authentication before access to the online examination was granted or denied. Since the challenge questions were posed randomly, therefore, the sample distribution was not uniform. The effectiveness of challenge questions using the equality and relaxed algorithms is discussed below.

Academic Questions: The relevance of questions can be important to recall answers and inform the effectiveness of the PBAF approach. It was hoped that questions with an answer hint would be easy to recall during authentication. However, the challenge question "*Find about this course*" received 2 (67%) matched answers during authentication. The analysis of answers revealed that one question failed to match as a result of syntactic variation.

Question context and relevance to individuals is important in reproducing the exact answers during authentication. The challenge question "*Month started current course*" received 2(100%) unmatched answers. As reported in the efficiency results, the text *"current course"* in the question is not relevant in the context of a simulation course, which led to usability issues.

Questions reported with clarity issues in the efficiency analysis, resulted in low effectiveness. One of the most obvious consequences of the question clarity can result in recall and syntactic variation in authentication during the online examination phase.

Table 3 Usability analysis: effectiveness

Question themes	Effectiveness and accuracy				
		Equality algorithm		Relaxed algorithm[1]	
Academic questions	**N[2]**	**Matched**	**Unmatched**	**Matched**	**Unmatched**
Student number	1	1(100%)	0(0%)	1(100%)	0(0%)
Year started the current course	3	3(100%)	0(0%)	3(100%)	0(0%)
Year of highest qualification	4	3(75%)	1(25%)	3(75%)	1(25%)
Highest qualification	4	3(75%)	1(25%)	**4(100%)**	**0(0%)**
Find out about this course	3	2(67%)	1(33%)	2(67%)	1(33%)
Name of last school attended	5	3(60%)	2(40%)	**4(80%)**	**1(20%)**
Grades in highest qualification	2	0(0%)	2(100%)	0(0%)	2(100%)
Month started the current course	1	0(0%)	1(100%)	**1(100%)**	**0(0%)**
Total		*15(65%)*	*8(35%)*	*18(78%)*	*5(22%)*
Personal questions					
Best friend's surname	6	6(100%)	0(0%)	6(100%)	0(0%)
Country of birth	4	4(100%)	0(0%)	4(100%)	0(0%)
Father's surname	3	2(67%)	1(33%)	**3(100%)**	**0(0%)**
Dream job as a child	2	1(50%)	1(50%)	**2(100%)**	**0(0%)**
Total		*13(87%)*	*2(13%)*	*15(100%)*	*0(0%)*
Favourite questions					
Tutor	6	1(17%)	5(83%)	**5(83%)**	**1(17%)**
Hero of your childhood?	3	3(100%)	0(0%)	3(100%)	0(0%)
Module on this course?	3	0(0%)	3(100%)	0(0%)	3(100%)
Total		*4(33%)*	*8(67%)*	*8(67%)*	*4(33%)*
Contact questions					
Home Tel no with country code	2	1(50%)	1(50%)	1(50%)	1(50%)
Home address town	4	1(25%)	3(75%)	**2(50%)**	**2(50%)**
House name or number	4	0(0%)	4(100%)	**1(25%)**	**3(75%)**
Mobile number including country code	1	0(0%)	1(100%)	0(0%)	1(100%)
Total		*2(18%)*	*9(82%)*	*4(36%)*	*7(64%)*
Date questions					
Date of birth?	5	4(80%)	1(20%)	5(100%)	0(0%)
Grand total	**66**	**38(58%)**	**28(42%)**	**50(76%)**	**16(24%)**

[1] Disregard capitalization, whitespace and minor spelling errors.
[2] Number of challenge questions.
Data in bold-face show an increase in effectiveness when results were computed using a relaxed algorithm.

Using the equality algorithm, the challenge questions in the academic theme received 15(65%) matched answers and 8(35%) unmatched answers, which shows acceptable effectiveness. However, there is a potential to further improve the usability by addressing the issues reported.

A more relaxed algorithm would increase the effectiveness of questions in the academic theme by 13%. Manual sorting of the data revealed that 3 answers were penalized for capitalization, spelling mistakes and spacing, which would benefit from using the relaxed algorithm. The implementation of the relaxed algorithm would decrease the error rate and increase the effectiveness to 18(75%).

Personal Questions: Personal questions are believed to be more memorable and therefore, widely used for credential recovery [20]. The challenge questions in the personal theme are reported with enhanced effectiveness in the online examination phase. The challenge questions "*Best friend's surname*" and "*Country of birth*" received 6 (100%) and 4 (100%) matched answers during authentication, which shows a high degree of effectiveness.

Syntactic variation including capitalization, spacing, spellings, writing syntax, can affect the usability of open text answers to challenge questions. The answers were lexicographically correct, nevertheless, the string to string match failed using the equality algorithm.

Using the equality algorithm, the challenge questions in the personal theme received 13(87%) matched and 2 (13%) unmatched answers, which indicates a high degree of effectiveness with a large number of accurate answers during authentication.

A more relaxed algorithm would increase the effectiveness of questions in the personal theme by 13%. Manual sorting of the data revealed that 2 answers were penalized for capitalization and spacing, which would benefit from using the relaxed algorithm. The implementation of the relaxed algorithm would decrease the error rate and increase the effectiveness to 15(100%).

Favourite Questions: The challenge questions in the favourite theme are a subset of personal questions, which pertains to individual's favourites. Popular challenge questions can be easy to recall. As an example, the popular challenge question "*Hero of childhood*" received 3(100%) matched answers during authentication, which indicates a high degree of effectiveness. It was reported with positive efficiency and submitted in the shortest completion time in the favourite theme during online learning.

Syntactic variation can increase the usability challenges. The challenge question "*Tutor*" received 5(83%)

unmatched answers and resulted in low effectiveness. The analysis revealed that 80% of answers were lexicographically correct; however the equality algorithm did not produce an exact match.

The challenge question "*module on this course*" was also reported with 3(100%) unmatched answers. The analysis revealed a complete shift in the answer pattern largely because of relevance and clarity issues reported in the efficiency analysis. The results clearly indicate a knock-on effect of unclear questions.

Using the equality algorithm, the challenge questions in the favourite theme received 4(33%) matched and 8(67%) unmatched answers, which indicates low effectiveness.

A more relaxed algorithm would increase the effectiveness of questions in the favourite theme by 32%. Manual sorting of the data revealed that 2 answers were penalized for capitalization, which would benefit from using the relaxed algorithm. The implementation of the relaxed algorithm would decrease the error rate and increase the effectiveness to 8(66%).

Contact Questions: The challenge questions in the contact theme were generalized for wider implementation. However, the generalization of questions created ambiguity, which resulted in poor usability.

The ambiguous questions reported in the efficiency analysis, had a knock-on effect and resulted in low effectiveness. The challenge question "*Address town*" received 1 (25%) matched answers. In a similar vein, the challenge questions "*House name or number*" received 4 (100%) unmatched answers, which indicates very low effectiveness. The variation in answers reported in the efficiency analysis increased the degree of difficulty for participants to produce the exact answers during the authentication phase.

Using the equality algorithm, the challenge questions in the contact theme were reported with poor effectiveness and received 9 (83%) unmatched answers, which indicates a sharp decrease in effectiveness. Questions in the contact theme were also reported with poor efficiency in the preceding Section.

A more relaxed algorithm would increase the effectiveness of questions in the contact theme by 18%. Manual sorting of the data revealed that 2 answers were penalized for spelling mistakes, which would benefit from using the relaxed algorithm. The implementation of the relaxed algorithm would decrease the error rate and increase the effectiveness to 4 (36%).

Date Questions: The challenge question "*Date of birth*" received 4 (80%) matched results during authentication. Syntactic variation in the date format was reported in the efficiency analysis. The *"Date of birth"* question received a single unmatched answer as a result of syntactic variation in the date format. The date was submitted in different formats such as "dd/mm/yyyy", "dd-mm-yyyy" and "day, month, year".

Using the equality algorithm, the challenge questions in the date theme indicate a high degree of effectiveness and no change was observed in the findings, if a more relaxed algorithm was implemented.

Summary of Effectiveness: In summary, the results that emerged from data analysis indicate a high number of matched answers for academic, personal and date themes. The questions with better relevance and clarity were reported with a high degree of effectiveness. The questions reported with low clarity, ambiguity and format issues had a knock-on effect during authentication and resulted in poor effectiveness. The participants failed to submit matched answers to a large number of questions in the 'favourite' and 'contact' themes using the equality algorithm implemented in empirical trail. The effectiveness of questions in the context of this study would further increase by 18%, if a more relaxed algorithm was implemented to compensate for capitalisation, spacing and spelling mistakes. The overall effectiveness will increase from 38 (58%) to 50 (76%), which is a large increase.

It was observed that questions with objective information remained efficient and effective during the learning and (authentication) examination phases. Also, responses to subjective answers were frequently changing during the learning and examinations phases resulted in failed authentication.

Concluding this section, we can say that question design needs particular consideration to address clarity, ambiguity and relevance to target users.

4.2 Traffic light system analysis

To address the usability challenges posed by the question design, we developed and implemented a traffic light system shown in Figure 2 and based on the criteria outlined in Study design and methodology. The data presented in Table 3, was collected from the PBAF implementation, with and without the traffic light system. The findings revealed that, before using the traffic light system, 23% of the participants submitted exact answers to all their 3 challenge questions and authenticated successfully. Of the total answers submitted, 38% participants provided exact answers to 2 out of 3 and 31% to 1 out of 3 challenge questions. However, 8% of participants provided no matching answers to challenge questions in the online examination phase. The reasons for unmatched answers are discussed in the preceding section. Before the traffic light system, the PBAF locked out participants who failed to submit exact answers to all of their 3 challenge questions. The participants, who provided exact answers to 1 or 2 of their 3 challenge questions, formed 69% (i.e. 31% + 38%) of the total unsuccessful attempts largely because of usability issues reported earlier.

Given the results of the online examination phase and in order to minimize the usability issues, we set up the traffic light system as shown in Figure 2. The system employed a three scale criteria outlined in the study methodology. The classification is setup to analyse PBAF performance by relaxing the constraints for compensating the usability issues. This may create a usability and security trade off, which needs further experimentation.

The results revealed that implementation of a traffic light system improved authentication success rate and minimized the impact of usability issues. A summary of data 'before' and 'after' the traffic light implementation is presented in Table 4. Overall, authentication success rate for participants increased from 23% to 92% (61% + 31%).

The traffic light system can provide an enabling environment to reduce the usability challenges and enhance the performance of the PBAF method. However, we are aware that, with the implementation of such a traffic light system, *security analysis* of the PBAF is warranted on a larger sample size.

4.3 Security analysis

The security analysis presented here, is extracted from the data taken from the participants' interactions with the security test phase described in Study design and methodology. We have analysed the security of questions against informed guessing attacks. The security test phase does not cover blind and focused guessing. An abuse case scenario was performed to evaluate the security of questions used in this study.

4.3.1 Guessing by friends and colleagues

The analysis collected from the abuse case scenario is presented in Tables 5 and 6. A total of 6 participants made 9 attempts to guess the challenge questions on behalf of their friends and colleagues. The participants were allowed to perform multiple attempts if the traffic light system criteria were met.

Table 5 shows analysis of abuse case scenario in terms of participants' attempts and traffic light results using the equality algorithm. Of the 6 participants, 3 (50%) failed to guess matched answers to any of their challenge questions on the 1st attempt and were classified *red*. The remaining 3 (50%) participants guessed matched answers to 1 out of 3 challenge questions and were classified *amber*. Of the 3 participants' classified *amber*, 1 dropped out of the process and the remaining 2 completed the abuse case scenario.

Table 4 Traffic light system

Authentication before traffic light system				
Attempt	**0/3 Matched**	**1/3 Matched**	**2/3 Matched**	**3/ 3 Matched**
1	1(8%)	4(31%)	5(38%)	3(23%)
Authentication after traffic light system				
	Red	**Amber**		**Green**
	0/3 Matched	***1/3 Matched***		***2-3/3 Matched***
1	1(8%)	4(31%)		8(61%)
2	0(0%)	2(12%)		3(19%)
3	0(0%)	0(0%)		2(12%)

Table 5 Security abuse case scenario and traffic light

Participants	Attempt	Matched	Unmatched	Authentication
P1	1st	0	3	Failed (Red)
P2	1st	0	3	Failed (Red)
P3	1st	0	3	Failed (Red)
P4	1st	1	2	Repeat (Amber)
P5	1st	1	2	Repeat (Amber)
P6	1st	1	2	Repeat (Amber)
P4	2nd	0	3	Failed (Red)
P5	2nd	0	3	Failed (Red)

In the second attempt, 2 participants were presented with more challenge questions for authentication and failed to guess exact answers to any of these. They were classified *red* and locked out.

Table 6 shows the crosstab analysis of abuse case scenario using the equality and relaxed algorithms under columns 3, 4, 5 and 6 headings. Data presented in boldface in column 5 and 6 shows any changes to security level, when results were computed using the relaxed algorithm. The participants were presented 24 challenge questions randomly on behalf of their friends and colleagues. Using the equality algorithm, answers to 3 (13%) were successfully guessed by participants, whereas 21 (88%) of the answers failed to match their respective profile answers. A more relaxed algorithm would increase the number of matched answers to 5 (21%) at the cost of increasing security risk.

To conclude this section, informed guessing by friends and colleagues was not highly successful and participants could not authenticate. However, questions in the public, friends and colleague domain were vulnerable to guessing. The abuse case scenario is discussed below to examine challenge questions in the individual themes.

Academic Questions: The participants submitted a total of 13 answers to challenge questions in the academic theme. The participants successfully guessed one answer in the academic theme.

It was anticipated that academic information would be vulnerable to guessing by friends and colleagues. However, participants' answers to a large number of the challenge questions failed to match.

Although, it was likely that challenge questions "*Month started current course*" and "*Year started current course*" could be guessed by individuals on the same course,

Table 6 Security analysis

Question themes	**Security abuse case**				
	N	**Equality algorithm**		**Relaxed algorithm**	
Academic questions		**Matched**	**Unmatched**	**Matched**	**Unmatched**
Student number	1	0(0%)	1(100%)	0(0%)	1(100%)
Year started the current course	3	0(0%)	3(100%)	0(0%)	3(100%)
Year of highest qualification	1	1(100%)	0(0%)	1(100%)	0(0%)
Highest qualification	2	0(0%)	2(100%)	0(0%)	2(100%)
Find out about this course	0	*NA	*NA	*NA	*NA
Name of last school attended	2	0(0%)	2(100%)	0(0%)	2(100%)
Grades in highest qualification	2	0(0%)	2(100%)	0(0%)	2(100%)
Month started the current course	2	0(0%)	2(100%)	0(0%)	2(100%)
Total		*1(8%)*	*12(92%)*	*1(8%)*	*12(92%)*
Personal questions					
Best friend's surname	1	0(0%)	1(100%)	0(0%)	1(100%)
Country of birth	2	1(50%)	1(50%)	**2(100%)**	**0(0%)**
Father's surname	1	0(0%)	1(100%)	**1(100%)**	**0(0%)**
Dream job as a child	0	*NA	*NA	*NA	*NA
Total		*1(25%)*	*3(75%)*	*3(75%)*	*1(25%)*
Favourite questions					
Tutor	1	0(0%)	1(100%)	0(0%)	1(100%)
Hero of your childhood?	0	*NA	*NA	*NA	*NA
Module on this course?	1	0(0%)	1(100%)	0(0%)	1(100%)
Total		*0(0%)*	*2(100%)*	*0(0%)*	*2(100%)*
Contact questions					
Home tel no with country code	1	0(0%)	1(100%)	0(0%)	1(100%)
Home address town	1	0(0%)	1(100%)	0(0%)	1(100%)
House name or number	1	0(0%)	1(100%)	0(0%)	1(100%)
Mobile number including country code	1	1(100%)	0(0%)	1(100%)	0(0%)
Total		*1(25%)*	*3(75%)*	*1(25%)*	*3(75%)*
Date questions					
Date of birth?	1	0(0%)	1(100%)	0(0%)	1(100%)
Grand total	**24**	**3(13%)**	**21(88%)**	**5(21%)**	**19(79%)**

*NA: Questions not presented due to randomization.
Data in bold-face show an increase in correct answers during abuse case when results were computed using a relaxed algorithm.

however, due to the clarity of questions reported earlier, participants failed to produce matching answers to these questions in all the 5 guesses.

The analysis of data using a more relaxed algorithm shows no change in the findings. However, the detailed exploration of the answers to challenge questions in the academic theme indicates security vulnerabilities and close guess possibilities by participants. A review of the academic questions is recommended to mitigate any risks.

Personal Questions: Participants submitted a total of 4 answers to challenge questions in the personal theme. It was anticipated that answers to personal questions would be by guessed by friends and colleagues. Schechter et al. [20] argue that the personal information can be found on the social media websites. Of all the personal challenge questions posed during the abuse case scenario, participants managed to guess matched answer to one question.

Personal information such as country of birth and place of birth can be vulnerable to informed guessing. The use of questions in the public domain can be vulnerable to guessing. It may not be true for all, but traditionally people use a common family and surname. Jobling [31] indicates that from five thousand years ago, fathers have passed their surname to children. The analysis of answers to profile question *"Father's surname"* in the learning phase revealed that, 64% of participants had a common surname as their fathers' and can be vulnerable to guessing attack.

A more relaxed algorithm would increase the security vulnerabilities of questions in the personal theme by 50% i.e. (75%-25%). Manual sorting of the data revealed that 2 answers failed to match during the security attack due to capitalization and spacing. The implementation of the relaxed algorithm shows decrease in security and increase in the number of matched answers from 1 (25%) to 3 (75%).

Favourite Questions: Participants submitted a total of 2 answers to challenge questions in the favourite theme. Questions in the favourite theme are widely used for credential recovery by email providers and banks. Although, an earlier empirical study [20] indicates that favourite questions are vulnerable to guessing, however, our findings indicate that questions in the favourite theme were resistant to an informed guessing attack.

The analysis of data in the favourite theme shows no change to the results, when a more relaxed algorithm was implemented.

Contact Questions: Participants submitted a total of 4 answers to challenge questions in the contact theme. Questions in the contact theme are likely to be known to friends and colleagues. Of all the challenge questions in the contact theme posed during the abuse case scenario, participants guessed matched answer to one question.

The challenge questions requesting phone or mobile numbers can be easily guessed by friends. It is likely that the contact numbers for friends and colleagues are stored in the phone or email address book and can be used for a guessing attack.

The analysis of data in contact theme shows no change to the results, when a more relaxed algorithm was implemented.

Date Questions: Participants submitted a single answer to challenge questions in the date theme. Although, "*date of birth*" is likely to be known by friends and colleagues, however, participants failed to guess a matched answer.

The analysis of data in date theme shows no change to the results, when a more relaxed algorithm was implemented.

Summary of Security Abuse Case: In summary, personal and academic questions are likely to be known to friends and colleagues. The challenge questions in the personal theme received one matched answer using an informed guessing attack. The questions in the personal theme were reported with positive efficiency, however, answer to personal questions can be guessed by friends and colleagues using the equality algorithm. Questions in the contact and academic themes can also be prone to guessing attacks by friends and family with one question each being successfully guessed by friends and colleagues. Although, the use of a relaxed algorithm may enhance the usability of challenge questions, however, it can also increase the security risks.

As a consequence of guessable and weak challenge questions and traffic light system, attackers may break security of the PBAF to reach their target.

5. Conclusion

The PBAF technique is a multi-factor knowledge based system, which uses challenge questions as repeat authentication in addition to login-identifier and password for student authentication in the online examination context.

In this study, the PBAF approach implemented text based academic, personal, favourite, contact and date questions for student authentication. The findings from the empirical study reported here suggest that challenge questions based authentication in online examinations can be an effective feature to resist adversaries' attacks, however, usability and security issues were reported in selective questions when used in the PBAF.

The usability metrics efficiency and effectiveness were evaluated. A large number of questions were reported with efficient completion time. The questions reported with clarity, ambiguity, relevance and format issues resulted in low efficiency and failed authentication, which also affected the effectiveness of the PBAF method. The results that emerged from data analysis using the equality algorithm indicate a high number of matched answers during authentication for academic, personal and date themes. The participants failed to submit matched answers to a large number of questions in the favourite and contact themes. The majority of the questions reported with the clarity issues resulted in failed authentication. The implementation of a more relaxed algorithm to compensate for capitalisation, spelling mistakes and spacing, would further improve the usability attributes. Question design has a measurable effect on the overall usability and security of the PBAF approach, which needs particular consideration to address clarity, ambiguity, relevance, subjective, and objective information. The subjective answers were frequently changing with time and a shift in answers patterns was observed.

The findings of the study suggest that participants may not provide 100% exact answers to all their 3 challenge questions set out for this work, largely because of the usability challenges such as syntactic variation and memorability issues. The implementation of a traffic light system improved authentication outcome from 23% to 92%, by enabling multiple chances. However, during the abuse case scenario, the traffic light algorithm granted 2 out of 6 attackers a second chance to answer more challenge questions in order to re-authenticate. Nevertheless, the participants guessed correct answers on behalf of their friends and colleagues, largely because of poor question design.

The security abuse case analysis revealed that questions related to friends, colleagues and common public knowledge can be a security risk. Some questions such as "*year of starting current course*" or "*father's surname*" can be intelligently guessed which may pose security threats. The overall results show a potential of using the PBAF authentication for online examination. However, secure and usable implementation of the PBAF method largely depends upon the quality of question design.

While the initial results are promising, further research is necessary to analyse question design and privacy. Furthermore, the number of participants in this study was small and more analysis is warranted on a larger sample size. There is a need to re-visit the design of questions to balance the trade-off between usability and security keeping in view the study results. The multiple attempts in the traffic light system may encourage the attacker to repeat the attack pattern. To prevent the attacker from repeating the attack pattern, a password change could be enforced in the future, if the student is locked out due to attacker activities. Virzi's empirical study [32] on the number of subjects for usability identification indicates that as few as 5 users can identify 80% of the usability issues. However, a number of conclusions cannot be drawn reliably for challenge questions in this security analysis due to a small number of participants and therefore, it is imperative to verify the security results in a real educational context on a larger sample size.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AU, HX and ML proposed the PBAF. AU designed, developed and implemented the PBAF in an online simulation course. AU also provided implementation guidance, put the layout of experimental validation and performance evaluation, and drafted the manuscript. HX, TB and ML carried out the structural and technical changes in the manuscript. TB helped and suggested statistical evaluation and recommended language modifications. All authors read and approved the final manuscript.

Acknowledgements

A special thank you goes to those who contributed to this paper: Bruce Christianson, Professor of Informatics, School of Computer Science University of Hertfordshire for his feedback on the research methodology and Paul Kirk Business Manager, IM & T, School of Postgraduate Medical and Dental Education Cardiff University for his help with the language review.

Received: 29 October 2013 Accepted: 25 February 2014
Published: 4 March 2014

References

1. Strother JB (2002) An assessment of the effectiveness of e-learning in corporate training programs. Int Rev Res Open Dist Learn 3(1):2, Article 3.1
2. Ruiz JG, Mintzer MJ, Leipzig RM (2006) The impact of e-learning in medical education. Acad Med 81(3):207
3. Huiping J (2010) Strong Password Authentication Protocols. In: 4th International Conference on Distance Learning and Education (ICDLE). IEEE, San Juan, Puerto Rico
4. Deo V, Seidensticker RB, Simon DR (1998) U.S. Patent No. 5,721,781. U.S. Patent and Trademark Office, Washington, DC
5. Moini A, Madni AM (2009) Leveraging biometrics for user authentication in online learning: a systems perspective. IEEE Syst J 3(4):469–476
6. Ullah A, Xiao H, Lilley M (2012) Profile Based Student Authentication in Online Examination. In: International Conference on Information Society (i-Society 2012). IEEE, London, UK
7. Ullah A, Xiao H, Lilley M, Barker T (2012) Usability of Profile Based Student Authentication and Traffic Light System in Online Examination. In: The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012). IEEE, London
8. Karaman S (2011) Examining the effects of flexible online exams on students' engagement in e-learning. Educ Res Rev 6(3):259–264
9. Agulla EG, Rifón LA, Castro JLA, Mateo CG (2008) Is My Student at the Other Side? Applying Biometric Web Authentication to E-Learning Environments. In: Eighth IEEE International Conference on Advanced Learning Technologies. IEEE, Santander, Cantabria
10. Harmon OR, Lambrinos J, Buffolino J (2010) Assessment design and cheating risk in online instruction. Online J Dist Learn Admin 13(3), Retrieved on Feb. 03, 2013 from http://www.westga.edu/~distance/ojdla/Fall133/harmon_lambrinos_buffolino13.html
11. Colwell JL, Jenks CF (2005) Student Ethics in Online Courses. In: 35th Annual Conference Frontiers in Education (FIE '05). IEEE, IA, USA
12. Chen Y, Liginlal D (2008) A maximum entropy approach to feature selection in knowledge-based authentication. Decis Support Syst 46(1):388–398
13. Bruns R, Dunkel J, Von Helden J (2003) Secure Smart Card-Based Access To An eLearning Portal. Proceedings of the 5th International Conference on Enterprise Information Systems (ICEIS), Angers, France
14. Gil C, Castro M, Wyne M (2010) Identification in Web Evaluation in Learning Management System by Fingerprint Identification System. In: Frontiers in Education Conference (FIE). IEEE, WA, USA
15. Sahoo SK, Choubisa T (2012) Multimodal biometric person authentication: a review. IETE Tech Rev 29(1):54
16. Ullah A, Xiao H, Lilley M, Barker T (2012) Using challenge questions for student authentication in online examination. Int J Infonom (IJI) 5(3/4):9
17. Just M, Aspinall D (2009) Challenging Challenge Questions. In: Socio-Economic Strand. Oxford University, UK
18. Rabkin A (2008) Personal Knowledge Questions for Fallback Authentication: Security Questions in the Era of Facebook. In: In SOUPS 2008: Proceedings of the 4th Symposium on Usable Privacy and Security, vol 23. ACM, New York, NY, USA
19. Just M, Aspinall D (2012) On the Security and Usability of Dual Credential Authentication in UK Online Banking. In: Internet Technology And Secured Transactions, 2012 International Conferece. IEEE, London, UK
20. Schechter S, Brush AJB, Egelman S (2009) It's No Secret. Measuring the Security and Reliability of Authentication via. In: 30th IEEE Symposium on Security and Privacy. IEEE, CA, USA
21. Griffith V, Jakobsson M (2005) Messin'with Texas Deriving Mother's Maiden Names Using Public Records. In: Third International Conference, ACNS. Springer, NY, USA
22. Just M, Aspinall D (2009) Personal Choice and Challenge Questions: A Security and Usability Assessment. In: Proceedings of the 5th Symposium on Usable Privacy and Security. ACM, CA, USA
23. (2012) Mysql. MySQL Reference Manaual 12.6.2. Mathematical Functions. MySQL -The worlds most popular opensource database., [cited 2012 15/10/2012]; 5.0:[MySQL 5.0 Reference Manual]. Available from: https://dev.mysql.com/doc/refman/5.0/en/mathematical-functions.html#function_rand
24. Just M (2004) Designing and evaluating challenge-question systems Security & Privacy. IEEE 2(5):32–39
25. Just M, Aspinall D (2009) Choosing Better Challenge Questions. In: Symposium on Usable Privacy and Security (SOUPS). ACM, CA USA
26. Standardization I. O. F (1998) Ergonomic Requirements for Office Work with Visual Dispaly Terminals, Part 11: Guidance on Usability. ISO 9241-11, Geneva
27. Seffah A, Kececi N, Donyaee M (2001) QUIM: A Framework for Quantifying Usability Metrics in Software Quality Models. In: Quality Software, 2001 Proceedings Second Asia-Pacific Conference. IEEE, Hong, Kong
28. Bevan N (2001) International standards for HCI and usability. Int J Human-Comp Stud 55(4):533–552
29. Nielsen J, Hackos JT (1993) Usability Engineering. Academic press, San Diego
30. Purdy G (2010) ISO 31000: 2009—setting a new standard for risk management. Risk Anal 30(6):881–886
31. Jobling MA (2001) In the name of the father: surnames and genetics. TRENDS Genet 17(6):353–357
32. Virzi RA (1992) Refining the test phase of usability evaluation: how many subjects is enough? Hum Fact: J Hum Fact Ergonom Soc 34(4):457–468

doi:10.1186/1869-0238-5-2
Cite this article as: Ullah *et al.*: **Evaluating security and usability of profile based challenge questions authentication in online examinations.** *Journal of Internet Services and Applications* 2014 **5**:2.