Citation for published version:

Abrar Ullah, Hannan Xiao and Trevor Barker, ‘A study into the usability and security implications of text and image based challenge questions in the context of online examination’, *Education and Information Technologies*, 2018.

DOI:

https://doi.org/10.1007/s10639-018-9758-7

Educ Inf Technol
https://doi.org/10.1007/s10639-018-9758-7

A study into the usability and security implications of text and image based challenge questions in the context of online examination

Abrar Ullah[1] · Hannan Xiao[2] · Trevor Barker[2]

Received: 28 August 2017 / Accepted: 7 June 2018

Abstract Online examinations are an integral component of online learning environments and research studies have identified academic dishonesty as a critical threat to the credibility of such examinations. Academic dishonesty exists in many forms. Collusion is seen as a major security threat, wherein a student invites a third party for help or to impersonate him or her in an online examination. This work aims to investigate the authentication of students using text-based and image-based challenge questions. The study reported in this paper involved 70 online participants from nine countries completing a five week online course and simulating an abuse case scenario. The results of a usability analysis suggested that i) image-based questions are more usable than text-based questions ($p < 0.01$) and ii) using a more flexible data entry method increased the usability of text-based questions ($p < 0.01$). An impersonation abuse scenario was simulated to test the influence of sharing with different database sizes. The findings revealed that iii) an increase in the number of questions shared for impersonation increased the success of an impersonation attack and the results showed a significant linear trend ($p < 0.01$). However, the number of correct answers decreased when the attacker had to memorize and answer the questions in an invigilated online examination or their response to questions was timed. The study also revealed that iv)

✉ Abrar Ullah
aaaullah@cardiffmet.ac.uk

Hannan Xiao
h.xiao@herts.ac.uk

Trevor Barker
t.1.barker@hert.ac.uk

[1] Department of Computing and Information Systems, Cardiff Metropolitan University, Western Avenue, Cardiff CF5 2YB, UK

[2] School of Computer Science, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK

Springer

Published online: 19 June 2018

an increase in the size of challenge question database decreased the success of an impersonation attack ($p < 0.01$).

Keywords Computer security · Online examinations · Authentication · Collusion · Image based authentication · Challenge questions

1 Introduction

Educational institutions are increasingly moving toward the use of online learning systems for the delivery of courses. In typical online learning environments, students interact with learning resources and take examinations from remote locations which raise the security concerns of stakeholders. With the increasing demand for online learning, there is a rising concern about the integrity of the online examination process (Watson and Sottile 2010). Academic dishonesty is one of the major security threats which have been a widely researched area. It has been reported as a serious challenge, due to vulnerable authentication approaches and the difficulty of verifying the identity of remote students. Face-to-face invigilation can be expensive and logistically challenging in dispersed geographical locations. However, in high-stake examinations many educational institutions prefer invigilated examinations to the use of online examinations due to the difficulty in the authentication of a remote user with no face-to-face interaction (Moini and Madni 2009).

Academic dishonesty can take place in a number of different ways. However, the work presented in this paper investigates an impersonation abuse case, wherein students invite third party impersonators to take the test on their behalf. From this scenario, students take advantage of weak authentication mechanisms and the absence of physical verification.

This paper presents the findings of an empirical study conducted in an online course with remote international participants. The work focuses on research that aims to investigate the authentication of examinees via the use of a challenge questions approach (Ullah et al. 2012a). The authors developed a profile based method, which implements challenge questions and login-identifier and password. Besides the traditional text-based challenge questions, this study implemented multiple-choice image-based questions for the evaluation of usability and security. Using this method, a student profile is built and consolidated during the teaching and learning process. A subset of profile information is used for authenticating students in online examinations. This study aimed to:

- examine the usability of text-based and image-based questions in a real online learning course.
- examine the effect of sharing different numbers of challenge questions in an impersonation using varying database sizes.
- examine the effect of using memory, printed and electronic sources when answering challenge questions during impersonation.

2 Background and related work

2.1 Academic dishonesty and collusion

The threat level of collusion in online examinations is different from other online applications such as banking where implicit collusion is unlikely to happen (Rabkin 2008).

In his earlier work, Ercole et al. (2002) studied collusion in a comparative empirical study using multiple choice questions in face-to-face and online examinations. It is one of the major security threats (Laubscher et al. 2005) which challenges the validity of online examinations (Carter et al. 2003). It can be classified in the following categories based on its occurrence in different scenarios (Ullah et al. 2016):

- *Impersonation (Operated by a Third Party Impersonator)*: This type of attack happens when a student invites a third party helper to impersonate and take an online test on his or her behalf. Impersonation can happen in different ways described below:
  - *Email (Asynchronous):* A student shares access credentials with a third party impersonator via email asynchronously, when they are unable to interact during an online examination in real-time due to implementation of locking and monitoring mechanisms (Kitahara et al. 2011).
  - *Smart Phone (Real-time):* Students are authenticated using a dynamic mechanism e.g. code texted on a mobile phone in real-time. To circumvent this security, a student and a third party share access credential in real time via instant messaging e.g. Skype, Viber, WhatsApp, Phone, SMS (Church and De Oliveira 2013) etc.
  - *Remote Desktop Sharing:* In this case, a student logs in to an online test and shares his or her screen with an impersonator remotely.
- *Abetting (Operated by a Student Aided by a Third Party)*: A student takes an online test, while a third party helper shares answers. This type of attack can happen in the different ways described below:
  - *Same Location*: A student takes a test while a third party helps with solving the exam questions based in the same location (Rowe 2004).
  - *Remote Location*: A student takes an online test, while a third party helps with solving the exam questions from a remote location via different communication means (Wheeler et al. 2003; Hart and Friesner 2004).

Given that security measures such as "secure browser" can be implemented to mitigate instant messaging (on computer), Internet browser access and remote desktop sharing during an examination session (Kitahara et al. 2011), a student may still be able to share access credentials with a third party before an online test.

2.2 Authentication approaches

The conventional authentication approaches fall into three categories based on "what you know" e.g. password and secret information "what you have?" e.g. a smart card and "what you are" e.g. biometrics (Jin et al. 2004). These methods are driven by knowledge, objects and human characteristics. In the light of the literature review and the benefits and limitations of various authentication approaches, the following criteria were framed to evaluate an accessible, cost effective, secure and useable authentication feature in the context of online examination:

- *Accessibility*: To ensure that the method can be used and accessed by a wide range of online participants using standard input devices. This frees users from a need to have access to special purpose devices that can limit implementation. Advances in mobile technology increase demand for accessible authentication approaches.
- *Cost Effectiveness:* The need for a cost effective approach is essential and this factor relates to the cost of development, implementation and maintenance. Bailie and Jortberg (2009) state that cost is an important consideration for technical and academic professionals in designing identity verification.
- *Security:* It is important to ensure that a method provides adequate protection to online examinations against the identified threats.
- *Usability*: Security mechanisms can only offer the intended protection, if usable. It is important to ensure that a method is reliable in terms usability. It describes the ability of authentication mechanisms to meet usability standards. The common attributes defined by the International Organization for Standards (ISO) (Iso9241-11 1998) which contributes to the usability includes efficiency and effectiveness.

Following is a review of traditional authentication features in light of the above criteria.

2.2.1 Knowledge-based authentication

Knowledge-based features employ the method of verifying users by matching one or more secrets supplied by an individual against data associated with the same individual (Chen and Liginlal 2008). The login-identifier, password and challenge question methods are commonly known as knowledge-based features. This is a widely acceptable approach because of its simplicity, availability and accessibility on a wide range of platforms (Hayashi et al. 2011). It is a low cost and preferred authentication method implemented in a majority of secure systems due to simple administration requirements (Hafiz et al. 2008).

However, the method may not prevent collusion attacks as passwords and personal information can be easily shared.

2.3 Challenge question authentication

Challenge question authentication is a knowledge-based feature, which is widely seen as a credential recovery technique (Just and Aspinall 2009a; Schechter et al. 2009). This method has been used as a second factor feature and employed for customer verification in online and telephone banking (Rabkin 2008; Just and Aspinall 2012). It can be a cost effective and accessible approach to cover a large online population. However, the reliability of challenge questions is dependent upon the context of use and choosing usable and secure questions (Just and Aspinall 2009b; Ullah et al. 2012b). The following section presents an analysis of the previous studies on challenge questions approach.

2.3.1 Previous research

Table 1 shows usability and security analysis of the previous studies on challenge questions authentication. In their influential work, Haga and Zviran (1990; Zviran and Haga 1993) investigated the effectiveness of text-based questions with several user groups. Their findings revealed participants reproduced correct answers to 70% opinion based and 74% fact based questions. However, their earlier study (Zviran and Haga 1990) showed that participants who were close family or friends correctly guessed answers to 33–39%. Research studies conducted by Just and Aspinall (2009a, c) revealed that 75 and 82% answers were correctly produced by participants. In their study (Just and Aspinall 2009a), Just and Aspinall reported that, of the 117 challenge questions asked, 88 (75%) of the total answers were recalled exactly while 21 (18%) had different punctuation/capitalization (typically performed when registering answers), and 8 (7%) were completely different citing memorability issue in a span of 28 days. The authors identified that memorability issue was 8 (7%) which was higher than the password memorability of 4.28% reported by (Florencio and Herley 2007). Just and Aspinall performed analysis of security level against blind guessing, focused guessing and observation attacks. Blind guessing is a brute force attack where attacker has no knowledge of the questions. In focus guessing, attacker can read and understand the question in order to guess the answer from a relevant search space. Security findings of their study indicate low security level for 5 of their 60 questions. Schechter et al. (2009) investigated

Table 1 Challenge questions: previous studies

		Usability		Security
S. no	Study	Efficiency	Effectiveness	Guessing
1	Zviran and Haga (1990, 1993)	NA	70–74%	33–39%
2	Just and Aspinall (2009c)	NA	82%	8.3% Low
3	Just and Aspinall (2009a)	NA	75%	46% Low
4	Schetcher et al. (2009)	NA	80%	10–13%
5	Ullah et al. (2014a)	15.7 s	58–76%	12%/29%
6	Bailie and Jortberg (2009)	NA	92%	NA
7	Renaud and Just (2010)	NA	68%	NA
8	Renaud and Just (2010)	NA	77%	38%
9	Babic et al. (2009)	0 s	2.23 (1–3)	NA
10	Ullah et al. (2015)	0 s	68%	NA

text-based questions widely used by corporate email services including Microsoft, AOL, Google, and Yahoo. Schetcher found that participants were able to answer 80% of their questions correctly; however, family and friends of participants were able to guess 10% of answers. Also, 13% of answers could be guessed within five attempts, which includes guessing the most popular answers of other participants. Bailie and Jortberg (2009) collected participants information from the US consumer database and created 150–300 challenge questions for each participants. The findings showed 92% correct answers which is the highest rate in all previous studies for text based questions. The collection of personal information could be a challenging task which may as well pose privacy concerns for wider implementation world-wide. To address the memorability issue, Renaud and Just (2010) proposed associative picture based cues with multiple choice answers, which achieved a 13% increase in the memorability with 77% correct answers. However, the security analysis revealed that 38% of the times, answers were guessed by close friends. The picture-based questions were related with participants' day to day activities. In the guessing attack, participants were asked to attribute characteristics to a person and associate to a well-known location. Babic et al. (2009) implemented activity based questions for authentication. They utilized a memorability scale (3 being easy to recall the answer and 1 being the opposite) and scored 2.23 correct answers. Ullah et al. (2015) implemented activity based questions in an online examination with 68% accuracy. However, further investigation is required to understand the security of activity based questions.

In their previous work, the authors conducted an online course to analyse usability, security and privacy factors in an online examination context (Ullah et al. 2014a, 2012b, c). The overall findings of the study reported positive outcomes. However, the following usability and security issues were identified. These issues were identified in the text-based questions as discussed in the literature review above (Just and Aspinall 2009a, c; Schechter et al. 2009).

- Questions with clarity, relevance and ambiguity issues were less usable. This influenced efficiency, effectiveness and memorability.
- Weak question design could lead to successful guessing.

These findings were helpful in setting out a benchmark for future research including the study reported in this paper.

2.4 Multiple-choice image-based questions

Our previous study revealed usability and security issues associated with the use of text-based challenge questions due to question design (Ullah et al. 2014a). In response to the risks and usability issues identified, multiple-choice image questions were introduced for use in this study. The use of image authentication has been adopted for a number of online services. For example, the Bank of America utilizes a site key image combined with text-based challenge questions (Youll 2006). Renaud and Just (2010) reported enhanced usability while using association-based image questions. This method can be further classified into recall based, cue recall based and recognition based schemes. In this study, we implemented recall based and recognition based image authentication (Ullah et al. 2014b). The description and background of these questions is given below:

- *Recall Image-based Questions:* Recall is the ability to memorize items without help. Shephard (1967) indicates that humans are better at recalling images than words, which is driven by the "picture superiority effect". This system requires a user to recall and select their previously chosen images.
- *Recognition Image-based Questions:* These rely upon an individual's ability to judge whether he/she has seen or selected an image before. It has been used in various studies by asking users to select previously chosen images from a large subset with distraction images (Brostoff and Sasse 2000; Hayashi et al. 2011; Ullah et al. 2015).
- *Cued Recall Image-based Questions:* This approach relies upon an individual's ability to recall an image, however, it is aided with a cue to stimulate recollection of a previous selection (Hayashi et al. 2011). This approach can be implemented using text-based information stored by a user (Rabkin 2008) or automated cues created programmatically (Wiedenbeck et al. 2005).

3 Research methodology

In this study, we implemented text-based and image-based challenge questions. A profile based authentication approach was employed to implement challenge questions in the learning and examination processes (Ullah et al. 2012a) as shown in Fig. 1. It was achieved by developing and integrating the system in Modular Object Oriented Distributed Learning Environment (MOODLE). Using this approach, learners register answers to pre-defined questions in order to access learning resources. A learner's profile is built during the learning process. In order to access an assessment activity, the

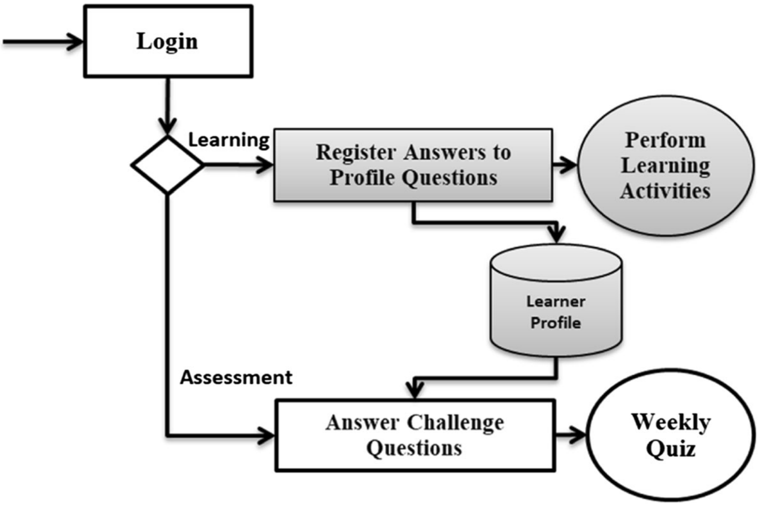

Login
Learning
Register Answers to Profile Questions
Perform Learning Activities
Learner Profile
Assessment
Answer Challenge Questions
Weekly Quiz

Fig. 1 Profile based authentication

learner is required to authenticate and provide correct answers to a random subset of challenge questions from his/her profile.

The study was organized in two phases described below.

3.1 Study phase – I usability evaluation

In the first phase, an online course was used involving remote students in order to evaluate the usability of text-based and image-based questions using a profile based authentication method as shown in Fig. 1. The following research methodology and process were adopted for usability evaluation.

3.1.1 Usability test approach (methodology)

Some earlier studies (Just and Aspinall 2009c; Just 2004) have reported usability as one of the major challenges in using challenge questions. A usability analysis is important in evaluating how effectively security measures can be implemented. The following usability attributes defined by the International Organization for Standards (ISO) (Iso9241-11 1998) and described in the Quality in Use Measurement Model (QUIM) (Seffah et al. 2001) were chosen:

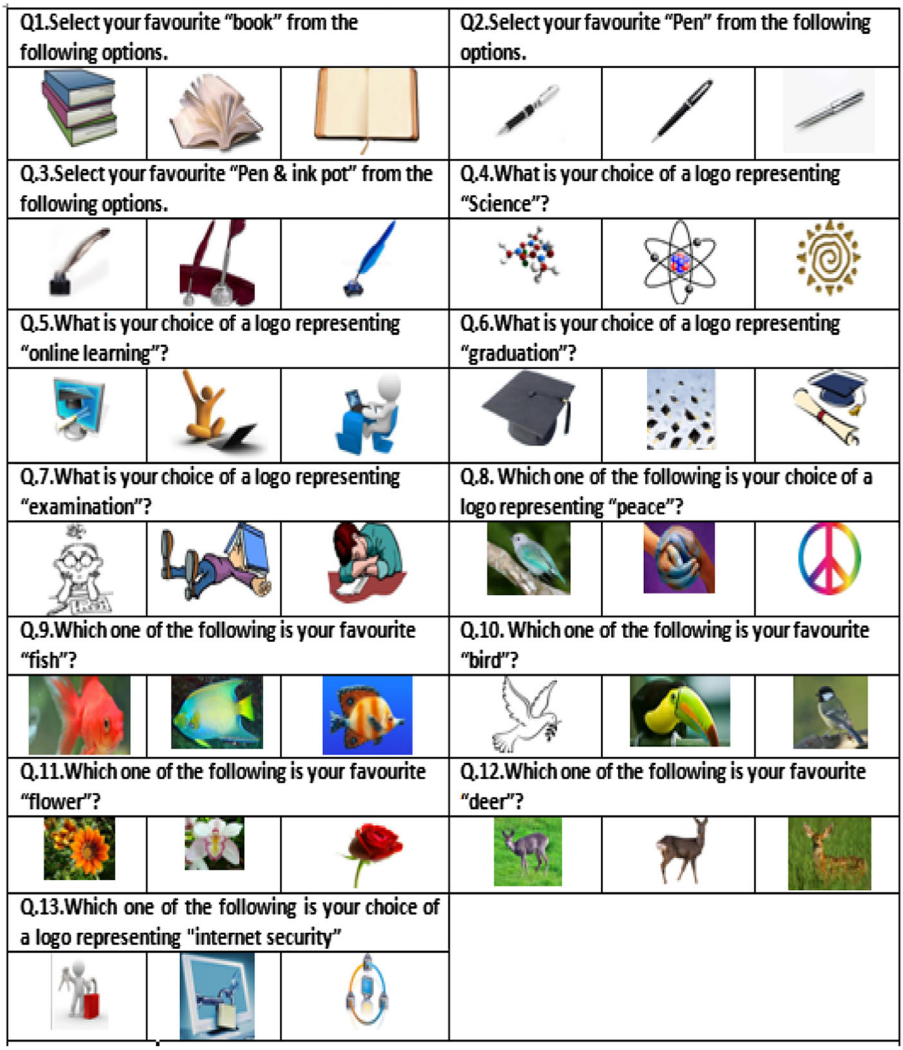
- *Efficiency:* It is a usability metric defined by ISO, which can be evaluated by measuring the completion time of each task and sub-tasks separately (Seffah et al. 2001). A system is considered efficient, if users are able to complete tasks in a reasonable time. In the context of this work, challenge questions completion time is measured to compute the efficiency of the proposed approach.
- *Effectiveness*: It is an important usability factor which indicates a degree of completeness with which users achieve a specified task in a certain context (Seffah et al. 2001). The effectiveness of questions was analyzed on the number of correct and incorrect answers to challenge questions in order to report the completion of authentication task and error rate.
- *Recall or Memorability:* An answer can be classified as memorable if it can be easily recalled (Just 2004). In the context of this work, memorability was evaluated based on the answers recalled during the authentication process. If a user's answer did not completely match with a previously registered answer, it is considered to be the result of recall or memorability failure.

3.1.2 Usability testing using an online course (process)

Following is the description of research process in order to evaluate usability attributes described in the research methodology above:

- *Questions Design:* Questions reported with usability and security issues in a previous study (Ullah et al. 2014a) were replaced with alternatives giving a careful design consideration to reduce ambiguity and clarity issues. Text-based and image-based questions were classified into five themes: academic, favourite, personal, date and image. Image based questions were further classified into recall and recognition questions as shown in Fig. 2. Text-based questions were based on the most common types implemented by leading email providers for credential recovery. For example AOL utilizes favourite and personal questions, Google uses a small set of personal questions and Microsoft and Yahoo implement a combination of personal and favourite questions (Schechter et al. 2009).

As shown in Fig. 2, recall image questions were presented as multiple-choice questions and students were required to choose an answer during the learning process, which was used for authentication during the examination. Recognition image questions were also presented as multiple-choice questions. However, a student was required to identify his/her previously chosen image, which was presented with multiple distraction images.



Q1.Select your favourite "book" from the following options.

Q2.Select your favourite "Pen" from the following options.

Q.3.Select your favourite "Pen & ink pot" from the following options.

Q.4.What is your choice of a logo representing "Science"?

Q.5.What is your choice of a logo representing "online learning"?

Q.6.What is your choice of a logo representing "graduation"?

Q.7.What is your choice of a logo representing "examination"?

Q.8. Which one of the following is your choice of a logo representing "peace"?

Q.9.Which one of the following is your favourite "fish"?

Q.10. Which one of the following is your favourite "bird"?

Q.11.Which one of the following is your favourite "flower"?

Q.12.Which one of the following is your favourite "deer"?

Q.13.Which one of the following is your choice of a logo representing "internet security"

Fig. 2 Image based questions

- *Course Setup:* An online course in PHP and MYSQL was setup and deployed in the MOODLE Learning Management System (LMS) on a remote web server. The course contents were released on a daily basis to engage participants and increase their interest and number of visits. A weekly online multiple-choice question (MCQ) quiz was set up as a summative online examination. Participants were recommended to invest 10 h weekly learning effort for 25 days in a span of 5 weeks.
- *Participants Recruitment:* An earlier study was conducted in a simulation environment. However, to understand the usability attributes in a real situation, an online course was organized and offered free of cost on the *University of Hertfordshire* online portal to attract students who were already enrolled on other distance learning courses. Participants were required to have basic programming knowledge in order to enroll. 70 students were recruited. The distribution of participants was not uniform across countries and cities, but there was a good representation from a diverse group of students from 9 countries. Of the 70 students, 50 (71%) students were from the United Kingdom. 11(16%) students from Pakistan, 2(4%) students from Malta and Nigeria, 1(1%) each from Ireland, Greece, India, Trinidad and Tobago, and Togo took part. One of the disadvantages of the above distribution is drawing conclusion using characteristics of the sample population. However, the study was directed to a specific user group involved in distance learning.
- *Student Registration:* Guidance notes and an enrolment key for registration were emailed to all participating students. Registration was a standard MOODLE sign up process, which was essential to create login credentials to access the course. Upon successful registration, students received their login-identifier and password.
- *Online Learning Weeks 1–5:* The course was presented over a period of 5 weeks. To collect data for the evaluation of usability and security, the transactional information including completion time of profile questions and challenge questions authentication results were stored in a database.
- *Examination Weeks 1–5 Quizzes:* The online course contained 5 quizzes released on a weekly basis towards the end of each week. Only students completing the quizzes were able to continue their study. Students were authenticated against their individual profiles recorded earlier.

3.2 Study phase – II collusion abuse case

In the second phase, a simulation study was conducted to evaluate collusion attacks when a text-based challenge questions approach is implemented. The following research methodology and process were adopted for security evaluation.

3.2.1 Risk based security assessment (methodology)

A risk based assessment model was adapted to plan, test and mitigate security risks. This model provides quantification of security level risks associated with various secure operations. This approach focuses on the test of features and functions of artefacts based on risks of their failure (Mcgraw 2004). An overview of the planning steps for this is given below.

- Identify Functions: The focus of challenge questions authentication in this study was the security of online examinations; therefore, weekly quizzes in an online course were identified as secure assets.
- Identify Risks: The ISO definition of risk is the "probability of occurrence of harm and its effect on objectives" (Purdy 2010). The security test in this study focused on the risk of collusion attacks when the challenge questions authentication is implemented.
- Identify Abuse Case: A collusion abuse case scenario was created and simulated using a web-based application in order to evaluate the security of the challenge questions approach, which is described later in this study.

3.2.2 Abuse case scenario simulation (process)

Phase –II of the study was organized to simulate an impersonation abuse case scenario after completion of the online course. Description of the simulation process is given below:

- *Designing Challenge Questions:* A total of 50 text-based challenge questions were created to simulate an impersonation abuse case scenario. A subset of these text-based questions was implemented in Phase –I of this research and the number of questions was increased for better results.
- *Online Simulation Databases:* An online challenge questions database and web-based application was setup to simulate impersonation. 50 challenge questions designed above were uploaded to the web based database application. The web based database application containing three different database sizes i.e. 20, 30, and 50, were hosted on a web server.
- *Participants Recruitment:* A total of 15 participants from four universities i.e. University of Hertfordshire, Southampton University, Cardiff University, University of South Wales and Institute of Management Sciences Pakistan volunteered to take part in the simulation abuse case tests. Although, this represents a small sample size, however, security testing is a specialist task and therefore, researchers involved in computer science were invited to collaborate. The participants were invited to help with simulating the impersonation attack and make a genuine effort to test the security of challenge questions approach.
- *Simulated Abuse Case Scenario:* The following collusion abuse case scenario was simulated sharing different number of questions and database sizes:

> "*A student is registered on an online course. The course uses challenge questions approach for authentication of students in online examinations. The student is due to write his final semester online test. He or she wants to boost his/her grades and recruit a third party to impersonate and take the test. However, to satisfy the challenge questions authentication, the student is required to share his/her challenge questions and answers with the third party helper in order to help with the impersonation. The third party helper would use the shared information to answer the randomly presented challenge questions for authentication*"

Given the above scenario, this study simulated the following sharing on database sizes containing 20, 30, and 50 questions. Different number of profile questions and answers were shared as shown in Table 2:

The simulation process is described below starting with a guessing attack with no answers shared using the database size 50:

1) A participant was asked to access the application and randomly guess answers to 50 challenge questions as shown in Table 2 and Fig. 3a.

To understand the influence of sharing using different database sizes, the following sharing conditions were tested for three database sizes 20, 30 and 50 respectively in a sequence shown in Table 2 and Fig. 3b. The researcher/moderator performed sharing with participants.

2) Starting from the database size "20" as shown in Table 2 and Fig. 3b, a total of "8" challenge questions and answers were shared with a participant via email to simulate impersonation.

Table 2 Collusion abuse case scenario: database size and questions shared

Database size (50)	
1) 0 or no sharing:	A student is unable to share any questions with a third party impersonator. In an attempt to impersonate and access the online examination, the third party helper uses random guessing to answer the challenge questions. This attack was simulated on the largest database size (50).
Database size (20)	
2) Share 8 questions	A student shares 8 questions and answers of his Database size (20) with a third party impersonator.
3) Share 12 questions	A student shares 12 questions and answers of his Database size (20) with a third party impersonator.
4) Share 20 questions	A student shares 20 questions and answers of his Database size (20) with a third party impersonator.
Database size (30)	
5) Share 12 questions	A student shares 12 questions and answers of his Database size (30) with a third party impersonator.
6) Share 18 questions	A student shares 18 questions and answers of his Database size (30) with a third party impersonator.
7) Share 30 questions	A student shares 30 questions and answers of his Database size (30) with a third party impersonator.
Database size (50)	
8) Share 20 questions	A student shares 20 questions and answers of his Database size (50) with a third party impersonator.
9) Share 30 questions	A student shares 30 questions and answers of his Database size (50) with a third party impersonator.
10) Share 50 questions	A student shares 50 questions and answers of his Database size (50) with a third party impersonator.

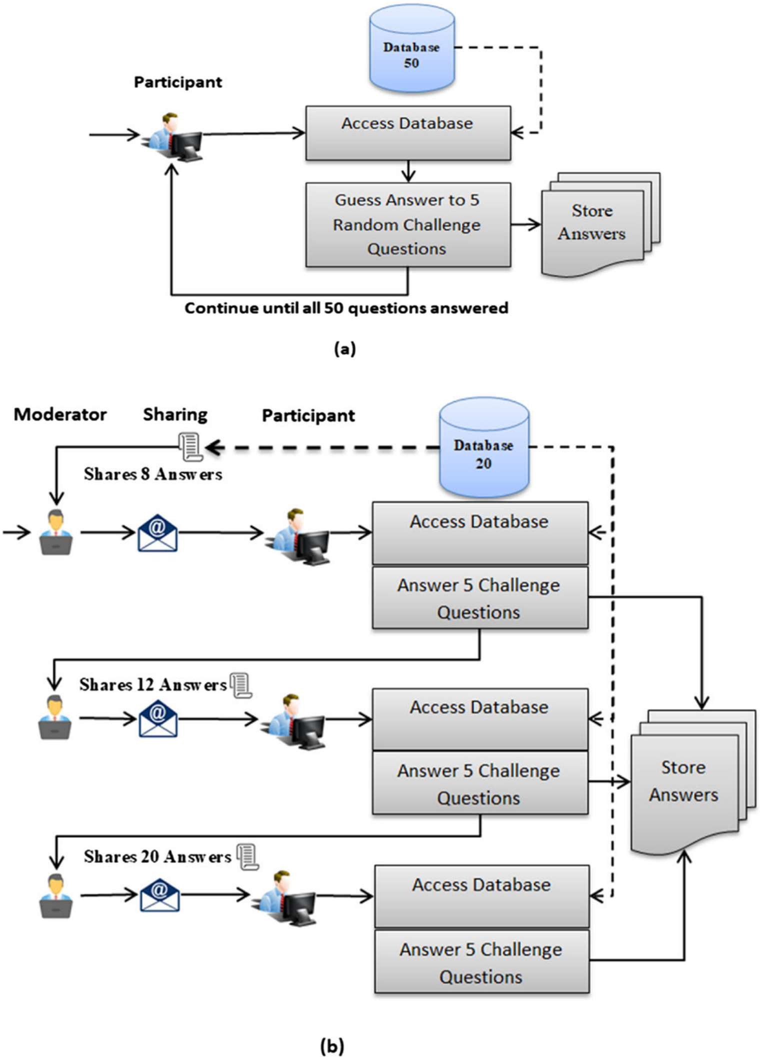


Fig. 3 Abuse case simulations

3) The participant accessed the database and answered "5" challenge questions randomly presented from "Database size (20)" using the shared questions and answers.
4) The number of shared questions was increased to "12" and "20" respectively.
5) The above steps were repeated for "Database size (30) and size (50)" and the number of questions shared.

Of the total 15 participants simulating the above scenarios, 10 participants answered the challenge questions using an *electronic or printed copy* shared via email.

A total of 5 participants volunteered and answered the challenge questions by memorising the answers from the shared email. They were not allowed to copy or see the shared email while answering the questions from memory. Data from the study was stored in the respective database for analysis.

4 Usability results and analysis

Seventy participants registered answers to 2315 profile questions in phase –I of the study during the course. The weekly quizzes were attempted by 48 participants who answered 1347 challenge questions. The usability analysis is discussed below.

4.1 Efficiency

We evaluated the efficiency of challenge questions by computing completion times by students on each visit to the course. The total number of profile questions collected was higher than the number of challenge questions posed for authentication. A student needed to access the course recurrently and the completion time of profile questions presented during the course would be expected to relate to the efficiency. Mean scores of the completion times are shown in Table 3.

A gradual decrease in the completion time of profile questions 74.87 s to 40.57 s is shown in Table 3, which indicates an increased efficiency with increased number of visits. In order to test the significance of any trend in the data presented in Table 3, a one-way ANOVA was performed with linear contrasts. A significant trend was confirmed for completion time in participants' multiple visits F (1,544) = 8.42, $p < 0.01$. A Pearson correlation was computed to assess the direction of the trend on subsequent visits ($r = -0.171$, $n = 558$, $p < 0.01$). The findings indicate a decrease in the completion time with an increasing number of visits.

Table 3 Mean completion times for profile questions

Visit No.	Completion time in seconds		
	Mean	SD	N=Visitors
1	74.87	59.48	70
2	62.28	61.77	60
3	53.22	63.52	54
4	43.26	47.92	50
5	32.07	15.13	44
6	45.18	41.37	40
7	43.05	38.15	38
8	44.42	41.98	38
9	46.11	34.20	35
10	47.32	38.84	34
11	37.93	23.43	29
12	43.50	30.18	24
13	42.50	67.65	23
14	40.57	31.08	19
	49.59	47.13	558

4.2 Effectiveness of text-based questions

To examine the effectiveness of challenge questions, an analysis of the correct answers to challenge questions during the quizzes was performed. We implemented an string-to-string comparison algorithm for authentication purposes (Schechter et al. 2009; Ullah et al. 2014a). Results in Table 4 show that, of the 890 text-based challenge questions randomly presented to students, 583 (66%) were answered correctly during the authentication process using an equality algorithm, which would increase to 74% if a more relaxed algorithm was implemented. The relaxed algorithm compensates for spelling mistakes and syntax variation. 307 (34%) answers were incorrect as a result of recall and syntactic variation.

As shown in Table 4, text-based challenge questions were classified into four themes. In order to test the significance of any differences in the means of correct responses to text-based challenge questions shown in Table 5, a one-way ANOVA test of significance was performed. The results of this analysis show that there was no significant difference in the means of correct answers between different themes ($p > 0.05$).

The "Academic" theme in Table 4 shows 64% correct response. Ambiguous academic questions resulted in failed authentication. The detailed sorting of incorrect answers revealed a complete shift in students' answers during learning and authentication. Some of these answers were semantically correct, however, not an exact match for the purpose of authentication. The incorrect answers registered by participants in the learning process resulted in failed authentication during all subsequent attempts in spite of a correct answer during the examination process.

Challenge questions in the "Date" theme shown in Table 4, received 72% correct answers during authentication. The syntax variation of the date format can be a usability challenge. Challenge question "Date of Birth" received 50% correct answers. Detailed sorting of answers revealed that a large number of answers were semantically correct with a variation in the answer format. As an example, answers "09/04/90" and "09/04/1990" submitted in learning and examinations phases were correct but penalized for failing string-to-string match during authentication; however, this could be prevented by enforcing validation rules or by using a different data entry method such as a calendar/date picker.

Table 4 Effectiveness of text-based challenge questions

Questions theme	Equality algorithm	Failure reason		Relaxed algorithm
	Correct/Incorrect N (%) N = number of answers	Syntactic variation	Recall	Correct/Incorrect N (%) N = number of answers
Text-based questions				
Academic	117(64%) / 67 (36%)	15 (22%)	52 (78%)	130 (71%) / 54 (29%)
Favourite	301(65%) / 162(35%)	31(19%)	131(81%)	321 (69%) / 142 (31%)
Personal	109 (66%) / 56 (34%)	17 (30%)	39 (70%)	128 (78%) / 37 (22%)
Date	56 (72%) / 22 (28%)	21 (96%)	1 (4%)	77 (99%) / 1 (1%)
Total	583(66%) / 307(34%)	84(27%)	223(73%)	656 (74%) / 234(26%)

Table 5 Effectiveness of image-based challenge questions

Question description	Type	Correct /Incorrect N (%) N = number of answers
Recall based image questions		
Pen	Object	15 (79%) / 4 (21%)
Book	Object	7 (70%) / 3 (30%)
Pen & Inkpot	Object	10 (63%) / 6 (38%)
Examination	Logo	15 (100%) / 0 (0%)
Science	Logo	18 (100%) / 0 (0%)
Online learning	Logo	16 (94%) / 1 (6%)
Graduation	Logo	24 (73%) / 9 (27%)
Internet security	Logo	10 (53%) / 9 (47%)
Peace	Logo	17 (89%) / 2 (11%)
Fish	Nature	20 (100%) / 0 (0%)
Flower	Nature	12 (86%) / 2 (14%)
Deer	Nature	20 (77%) / 6 (23%)
Bird	Nature	8 (62%) / 5 (38%)
Sub total		192(80%)/47(20%)
Recognition based image questions		
Select an image you've previously seen/chosen	Mixed	197 (90%) / 21 (10%)
Sub total		197 (90%) / 21 (10%)
Grand total		389 (85%) / 68 (15%)

The authentication of challenge questions used a string-to-string comparison method referred to as the equality algorithm. The algorithm penalized answers with syntactic variation and spelling mistakes, which would otherwise be considered correct if a more relaxed algorithm was used. The data in Table 4 columns 2 and 3 show results of authentication using an equality algorithm and column 5 shows results if a more relaxed algorithm was used. The data of the relaxed algorithm was compiled using a substring and distance algorithms (Schechter et al. 2009). The results of a relaxed algorithm compensates for syntactic variation such as date format, spelling mistakes and white spaces. A paired sample t-test showed a significant difference in the effectiveness between the equality ($M = 66.12$, $SD = 12.6$) and relaxed ($M = 75.35$, $SD = 14.09$) algorithms conditions $t\ (30) = -4.33$; $p < 0.01$.

4.3 Effectiveness of image-based questions

The effectiveness results of both "Recall" and "Recognition" image-based challenge questions are shown in Table 5.

The image questions are shown in Fig. 2. As discussed earlier, the "Recognition" image questions were derived non-intrusively in the background from students' answers to their multiple-choice image-based questions. A student's answer was used with a random subset of distraction images. These distraction images were not shown to participants previously. They were required to recognize their previously chosen image from a set of distraction images.

Of the total of 457 image-based challenge questions, 389 (85%) were answered correctly during authentication. The effectiveness result for the text-based questions described above was 66% using the equality algorithm and 74% using the relaxed algorithm. Implementation of multiple-choice questions addressed the issue of syntactic variation, capitalization, formatting and spelling mistakes, which increased the effectiveness. Results in Table 5 show that "Recall" and "Recognition" image-based questions received 192 (80%) and 197 (90%) correct answers respectively.

Following section presents a comparative analysis of text-based and image-based questions.

4.4 Comparison of effectiveness between text-based and image-based questions

The effectiveness of image-based questions was significantly better than the text-based challenge questions ($p < 0.01$). In order to test the significance of any differences in the means of correct answers between text and image questions shown in Tables 4 and 5, a one-way ANOVA test of significance was performed. The results of this analysis showed that there were significant differences in the means F (1, 42) = 13.5, $p < 0.01$.The use of image-based questions resulted in better effectiveness by minimizing usability problems such as syntactic variation, spacing, capitalization, spelling mistakes and memorability.

In order to test the significance of any differences in the means of correct answers between text and image questions shown in Tables 4 and 5 according to the *equality* and *relaxed* algorithms, a one-way ANOVA test of significance was performed. The results of this analysis showed that there were significant differences in the means F (5, 72) = 6.11, p < 0.01. Post hoc comparisons of the groupings yielded the following significant results.

Text-based (equality algorithm) x Image-based, mean difference (MD) = −14.33, Standard Error (SE) = 4.94, $p < 0.01$ Text-based (equality algorithm) x Text-based (Relaxed algorithm), MD = −9.2, SE = 3.39, $p < 0.01$. No other significant differences were found in the post hoc comparisons. The findings indicate that the use of image-based questions increased the effectiveness by addressing the issues related with syntax, spellings, spacing and formatting. However, the use of a relaxed algorithm also increased the effectiveness which compensated the stated issues. There was no significant difference in the effectiveness between image-based questions and text-based questions using a relaxed algorithm.

The implementation of image-based questions is encouraging and set a new direction for this research. In an earlier study, Renaud and Just (2010) reported a 13% increase in memorability while using association-based pictures in authentication. Multiple-choice image questions indicate more potential and increased answer recall.

5 Security results and analysis

Tables 6 and 7 show the security analysis of an impersonation abuse case performed by 15 participants using three different database sizes. Results of the 10 participants, who

Table 6 Impersonation abuse case scenario: answers copied for impersonation

Database size (50)		Database size (20)			Database size (30)			Database size (50)		
P#	0 *n = 50	8 n = 5	12 n = 5	20 n = 5	12 n = 5	18 n = 5	30 n = 5	20 n = 5	30 n = 5	50 n = 5
1	3(6%)	2(40%)	5(100%)	5(100%)	1(20%)	3(60%)	5(100%)	2(40%)	2(40%)	5(100%)
2	1(2%)	1(20%)	5(100%)	5(100%)	2(40%)	4(80%)	5(100%)	2(40%)	2(40%)	5(100%)
3	1(2%)	3(60%)	4(80%)	5(100%)	1(20%)	2(40%)	5(100%)	1(20%)	3(60%)	5(100%)
4	2(4%)	2(40%)	2(40%)	5(100%)	2(40%)	2(40%)	5(100%)	1(20%)	1(20%)	5(100%)
5	1(2%)	2(40%)	4(80%)	5(100%)	2(40%)	3(60%)	5(100%)	0(0%)	2(40%)	5(100%)
6	1(2%)	2(40%)	3(60%)	5(100%)	3(60%)	2(40%)	5(100%)	2(40%)	3(60%)	5(100%)
7	3(6%)	2(40%)	2(40%)	5(100%)	1(20%)	2(40%)	5(100%)	2(40%)	2(40%)	5(100%)
8	1(2%)	3(60%)	3(60%)	5(100%)	2(40%)	3(60%)	5(100%)	2(40%)	2(40%)	5(100%)
9	3(6%)	2(40%)	3(60%)	5(100%)	2(40%)	2(40%)	5(100%)	1(20%)	2(40%)	5(100%)
10	1(2%)	2(40%)	2(40%)	5(100%)	2(40%)	2(40%)	5(100%)	2(40%)	3(60%)	5(100%)
	17(3%)	**21(42%)**	**33(66%)**	**50(100%)**	**18(36%)**	**25(50%)**	**50(100%)**	**15(30%)**	**22(44%)**	**50(100%)**

*n = number of questions presented

Table 7 Impersonation abuse case scenario: answers memorized for impersonation

Database size (50)		Database size (20)			Database size (30)			Database size (50)		
P#	0 *n = 50	8 n = 5	12 n = 5	20 n = 5	12 n = 5	18 n = 5	30 n = 5	20 n = 5	30 n = 5	50 n = 5
1	1(2%)	2(40%)	3(60%)	2(40%)	2(40%)	0(0%)	4(80%)	1(20%)	1(20%)	3(60%)
2	1(2%)	2(40%)	2(40%)	1(20%)	1(20%)	1(20%)	2(40%)	2(40%)	2(40%)	1(20%)
3	0(0%)	1(20%)	2(40%)	3(60%)	1(20%)	3(60%)	2(40%)	2(40%)	3(60%)	3(60%)
4	2(4%)	2(40%)	1(20%)	1(20%)	2(40%)	1(20%)	2(40%)	2(40%)	2(40%)	2(40%)
5	3(6%)	2(40%)	3(60%)	2(40%)	1(20%)	4(80%)	2(40%)	2(40%)	3(60%)	3(60%)
	7(2.8%)	**9(36%)**	**11(44%)**	**9(36%)**	**7(28%)**	**9(36%)**	**12(48%)**	**9(36%)**	**11(44%)**	**12(48%)**

*n = number of questions presented

answered challenge questions from an electronic or printed copy, are presented in Table 6. Results of the 5 participants, who memorized the answers before responding the challenge questions, are presented in Table 7.

5.1 The effect of "number of questions shared" on impersonation

In order to test the significance of any trend in the data presented in Table 6 for different numbers of sharing in an impersonation attack using database size (20), size (30), and size (50), a one-way ANOVA was performed with linear contrasts. A linear trend was found for all sharing conditions on Database size (20), F (1, 36) = 293.8, $p < 0.01$, Database size (30), F (1, 36) = 507.6, $p < 0.01$, and Database size (50), F (1, 36) = 507.67, $p < 0.01$. A Pearson correlation was performed on data presented in Table 6 to test the direction of the trend for all sharing conditions on Database size (20), $r = 0.94$, $n = 40$, $p < 0.01$, Database size (30), r = 0.94, n = 40, $p < 0.01$ and Database size (50), $r = 0.93$, n = 40, $p < 0.01$.

The above results show that an increase in the number of shared questions has increased the number of correct answers in a collusion abuse case. Figure 4 shows a strong linear trend for all sharing conditions using all database sizes.

The findings revealed that an impersonation attack is more successful if a student is able to share a large number of questions with a third party impersonator. In the absence of monitoring or timing a user response, an impersonator can answer challenge questions copying from a printed or electronic source shared by a student in order to authenticate. In the abuse case simulation, challenge questions were randomized, however, the impersonator was able to search and copy the correct answers from the shared information. The findings revealed that the impersonator may circumvent the challenge questions approach, irrespective of the size of database, if an online examination is not monitored or students are not restricted to answer the questions in a limited time.

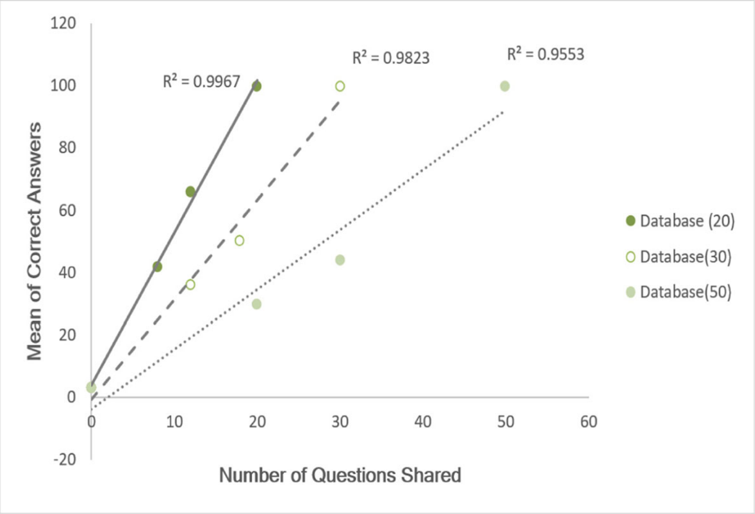


Fig. 4 Trend analysis sharing vs correct answers using database sizes

5.2 The effect of "database size" on impersonation attacks

This section provides an analysis on the "Database size" and how this affects the success of an impersonation attack. In order to test the significance of any trend in the data presented in Table 8 using database size (20), size (30) and size (50) for all sharing conditions, a one-way ANOVA was performed with linear contrasts. A trend was found for all database sizes (20), (30) and (50) F (1, 29) = 11.45, $p < 0.01$. A Pearson correlation was performed on data presented in Table 8 to test the direction of the trend on all database sizes for $r = -0.559$, $n = 30$, $p < 0.01$.

The above findings revealed that an impersonation attack was less successful with an increase in the database size. The trend line in Fig. 5 shows a decrease in the number of correct answers with an increase in the database size. Also, an increase in the database

Table 8 Impersonation using database sizes: answers copied for impersonation

P#	Database (20)	Database (30)	Database (50)
1	12(80%)	9(60%)	9(60%)
2	11(73%)	11(73%)	9(60%)
3	12(80%)	8(53%)	9(60%)
4	9(60%)	9(60%)	7(47%)
5	11(73%)	10(67%)	7(47%)
6	10(67%)	10(67%)	10(67%)
7	9(60%)	8(53%)	9(60%)
8	11(73%)	10(67%)	9(60%)
9	10(67%)	9(60%)	8(53%)
10	9(60%)	9(60%)	10(67%)
	104(69%)	**93(62%)**	**87(58%)**

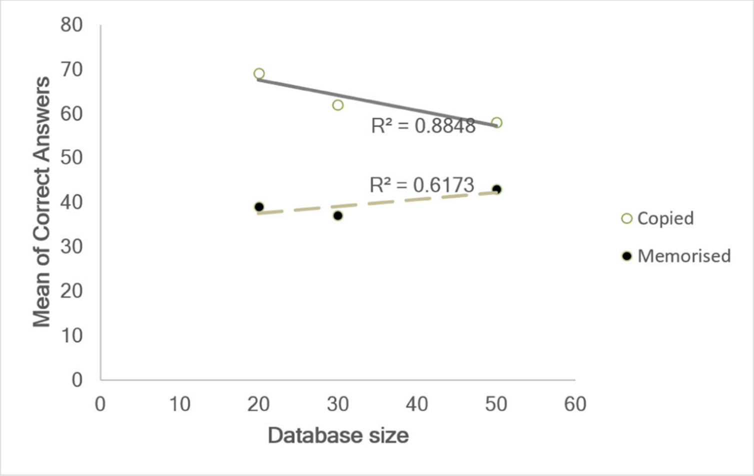


Fig. 5 Trend analysis: database sizes

size decreases the probability of randomly getting the same subset of questions shared by a student for impersonation. It is anticipated that an increase in the database size would make it harder for a student to share all answers with a third party impersonator.

If answers to challenge questions are timed or monitored, it would be expected to increase the difficulty for an impersonator to search for the correct answers from a shared source for larger database sizes. As shown in Fig. 5, the impersonation attack was less successful when participants had to memorise and answer the challenge questions. It is discussed in more detail below.

5.3 The effect of answering challenge questions from memory

In a practical situation, it is anticipated that students would answer challenge questions in a limited time. In the above discussion, participants were allowed to search the shared information in order to answer the questions with no time constraints. However, if answers to challenge questions are timed or the authentication process is monitored, an impersonator would be required to memorise the shared information. In order to test the significance of any trend in the data presented in Table 7 for four sharing conditions in an impersonation attack using Database size (20), size (30) and size (50), a one-way ANOVA was performed with linear contrasts. A linear trend was found for all sharing conditions on Database size (20), F (1, 16) = 17.8, $p < 0.01$, Database size (30), F (1, 16) = 13.5, $p < 0.01$, and Database size (50), F (1, 16) = 30.09, $p < 0.01$. A Pearson correlation was performed on data presented in Table 7 to test the direction of the trend for all sharing conditions on Database size (20), $r = 0.61$, $n = 20$, $p < 0.01$, Database size (30), $r = 0.66$, n = 20, $p < 0.01$ and Database size (50), $r = 0.75$, n = 20, $p < 0.01$.

The findings showed an increasing trend in correct answers with an increase in the number of shared answers for impersonation. However, the number of correct answers decreased when the impersonator answered the questions from memory. Figures 6, 7 and 8 show a difference in the correct answers for all sharing conditions using different

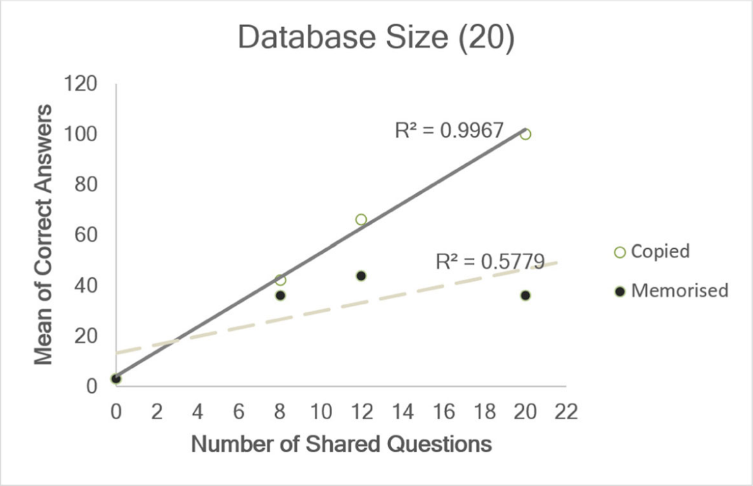


Fig. 6 Database size(20): electronic or printed source vs memory

database sizes and the way impersonator answered these questions. It shows that for all sharing conditions and database sizes, answers were less successful when attempted from memory. In order to test the significance of any differences in the means of correct answers between "Answers copied for impersonation" and "Answers Memorised for impersonation", a one-way ANOVA test of significance was performed on data shown in Tables 8 and 9. The results of this analysis showed that there were significant differences in the means for Database size (20) conditions F (1, 13) = 47.4; $p < 0.01$, Database size (30) conditions F (1, 13) =43.18; $p < 0.01$, and Database size (50) conditions F (1, 13) = 12.47; $p < 0.01$. This indicates that if answers to challenge questions are timed or an online examination process is monitored, it might discourage the impersonator from searching a printed or electronic source for answers.

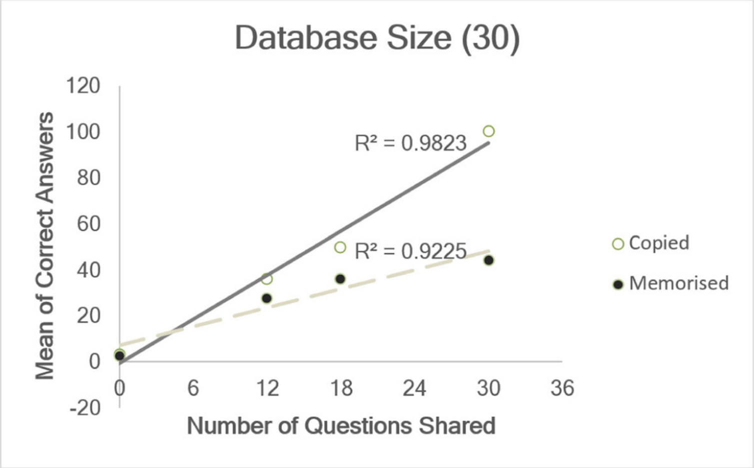


Fig. 7 Database size(30): electronic or printed source vs memory

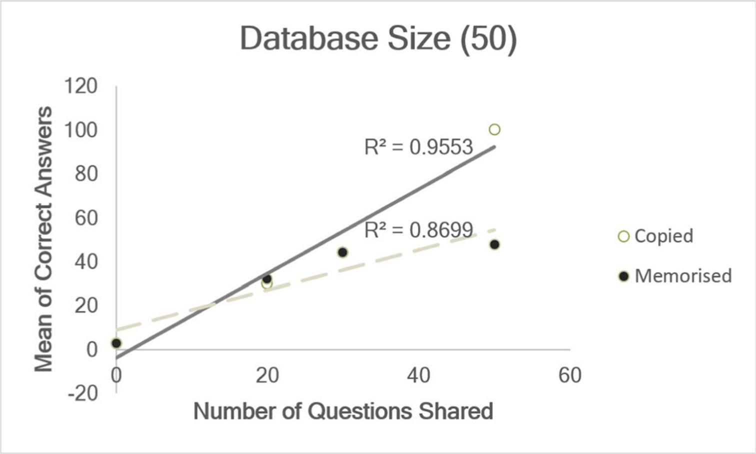


Fig. 8 Database size(50): electronic or printed source vs memory

6 Discussion

This work is part of an ongoing research to identify a secure and usable authentication approach in order to mitigate collusion in online examinations. This study uses a two phased approach to examine usability and security on two different user groups. The usability analysis of text-based questions is not significantly different than other similar studies discussed in Table 1. While the effectiveness of text-based questions increased with improved question design from our previous study (Ullah et al. 2014a), the issues contributing to incorrect answers such as syntactic variation, capitalization, incorrect spellings, memorability remained unchanged. In their study, Just and Aspinall (2009a) reported similar concerns for incorrect answers i.e. punctuation, capitalization and memorability. In order to address these issues, this study provides a comparative analysis between equality and relaxed algorithms. Our findings revealed that the relaxed algorithm improved the usability of text-based questions significantly ($p < 0.01$). This algorithm implemented an increased tolerance level and considered answers with spelling mistakes, capitalization and syntactic variation as correct. Bailie and Jortberg (2009) allowed two attempts to increase the tolerance level of challenge questions approach with 92% accuracy. However, their study lacks detail on authentication protocol. In their influential study, Schechter et al. (2009) implemented substring

Table 9 Impersonation using database sizes: answers memorized for impersonation

P#	Database (20)	Database (30)	Database (50)
1	7(47%)	6(40%)	5(33%)
2	5(33%)	4(27%)	5(33%)
3	6(40%)	6(40%)	8(53%)
4	4(27%)	5(33%)	6(40%)
5	7(47%)	7(47%)	8(53%)
Total	29(39%)	28(37%)	32(43%)

and distance algorithms which showed increase in correct answers. However, it also increased the success of guessing attack during security analysis.

Many studies (Just and Aspinall 2009a; Schechter et al. 2009; Ullah et al. 2014a; Renaud and Just 2010) reported memorability as one of the key issues with the use of text-based challenge questions. The results of usability analysis in this study revealed that memorability contributed to 74% of the total incorrect answers. The participants were unable to recall their answers correctly. Schechter et al. (2009) reported 57% incorrect answers citing memorability issue, however, 13% participants indicated that they registered bogus answers to their challenge questions. In order to address these issues, this study utilized multiple-choice image-based questions. The results revealed increased effectiveness compared to text-based questions. The findings showed that "Recall" and "Recognition" image-based questions received 192 (80%) and 197 (90%) correct answers respectively. There was a significant difference ($p < 0.01$) in the effectiveness between image questions and text-based questions. Furthermore, recognition based question were more useable as users had to recognized their previous image selection, which was presented with a set of random distractor images. In a similar study Renaud and Just (2010) achieved 13% increase in memorability with associative picture-based authentication. This study utilised image-based questions using three multiple-choice options, which implies 33% probability of a correct random guess. However, an increase in the number of multiple choice options will decrease the probability of random guessing. Although, the use of image-based questions enhances the usability, however, the above discussion shows a usability and security trade-off.

Previous research work on challenge questions focused largely on conventional threats i.e. guessing attacks. This study investigates the potential use of challenge questions to mitigate non-conventional impersonation threats. The security analysis based on the abuse case scenario revealed that the success of an impersonation attack was influenced by the number of answers shared with a third party impersonator. There was a significant linear trend ($p < 0.01$), when impersonators answered their challenge questions from a printed or electronic copy of the shared questions. The number of correct answers decreased, when impersonators memorized and answered the questions to simulate a scenario when an online examination is monitored or answers to challenge questions timed. The response time is identified an important authentication factor. The study also revealed that an increase in the database size decreases the number of correct answers during a collusion attack. This indicates that an increase in the profile size increases the resilience of the challenge questions approach against collusion attacks.

In a practical scenario, the success of impersonation attack will depend upon the ability of students to share maximum answers with third party impersonators. It can be inferred from the findings that such attacks may be more successful for smaller profiles. Since text-based questions are associated with individual's personal information, it may relatively easy for students to share such information with an impersonator particularly in a high stake examination. A decrease in sharing personal information can influence the impersonation. Image-based questions could be a potential alternative to discourage students from sharing; however, this needs further investigation.

7 Conclusion

This study presents a comparative analysis of usability and security between text-based and image-based challenge questions in the context of online examinations. Text-based questions were reported with common usability issues reported in previous research including spelling mistakes, capitalization, spacing, and memorability. The use of image-based questions increased usability results including efficiency and effectiveness. It can be a usable and secure concept to prevent conventional security threats. This may help universities and researchers to investigate the potential for using image-based challenge question to mitigate security threats including collusion attack.

The security analysis based on the abuse case scenario revealed that the success of an impersonation attack was influenced by the number of answers shared with a third party impersonator, database size and the response time during authentication process. Therefore, to mitigate impersonation, it is essential to implement question type which minimizes the ability of students to share their credentials. The findings are not sufficient to determine the student's ability of sharing personal information. However, it can be assumed that a student should be able to share personal information in a high-stake process. Further research is warranted to understand individual's ability of sharing information with an impersonator in an online examination context.

Future research will explore other question types which are usable but minimizes the risks of sharing in order to mitigate impersonation.

Acknowledgements A special thank you goes to those who contributed to this paper: Paul Kirk Business Manager, Jay Beavan, MARS Programmer, and Lyndon Ellison Web Programmer, School of Postgraduate Medical and Dental Education, Cardiff University for their help and support with the study.

Author's contribution AU, HX, and TB contributed equally in the design and preparation of this research work.

Compliance with ethical standards

Competing interests There are no competing interests.

References

Babic A., Xiong H., Yao D., Iftode L., editors. (2009). Building robust authentication systems with activity-based personal questions. Proceedings of the 2nd ACM workshop on Assurable and usable security configuration; ACM.

Bailie, J. L., & Jortberg, M. A. (2009). Online learner authentication: Verifying the identity of online users. *Bulletin-Board Postings, 547*, 17.

Brostoff, S., & Sasse, M. A. (2000). Are Passfaces more usable than passwords? A field trial investigation. In S. Mcdonald, Y. Waern, & G. Cockton (Eds.), *People and computers XIV—usability or else* (pp. 405–424). London: Springer.

Carter, J., Ala-Mutka, K., Fuller, U., Dick, M., English, J., Fone, W., & Sheard, J. (2003: ACM). How shall we assess this? *ACM SIGCSE Bulletin, 35*, 107.

Chen, Y., & Liginlal, D. (2008). A maximum entropy approach to feature selection in knowledge-based authentication. *Decision Support Systems., 46*(1), 388–398.

Church K., De Oliveira R., editors. (2013). What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS. Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services; ACM.

Ercole, A., Whittlestone, K., Melvin, D., & Rashbass, J. (2002). Collusion detection in multiple choice examinations. *Medical Education, 36*(2), 166–172.

Florencio D., Herley C., editors. (2007). A large-scale study of web password habits". Proceedings of the 16th international conference on World Wide Web; ACM.

Hafiz M. D., Abdullah A. H., Ithnin N., Mammi H. K., editors. (2008). Towards identifying usability and security features of graphical password in knowledge based authentication technique. Modeling & Simulation, 2008 AICMS 08 Second Asia International Conference on; 2008: IEEE.

Hart, M., & Friesner, T. (2004). Plagiarism and poor academic practice–a threat to the extension of e-learning in higher education? *Electronic Journal on e-Learning, 2*(1), 89–96.

Hayashi E., Hong J., Christin N., editors. (2011).Security through a different kind of obscurity: Evaluating distortion in graphical authentication schemes. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2011: ACM.

Iso9241-11. (1998). Ergonomic requirements for office work with visual dispaly terminals, Part 11: Guidance on usability. ISO 9241-11. Geneva1998.

Jin, A. T. B., Ling, D. N. C., & Goh, A. (2004). Biohashing: Two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition., 37*(11), 2245–2255.

Just, M. (2004). Designing and evaluating challenge-question systems. *Security & Privacy, IEEE., 2*(5), 32–39.

Just M., Aspinall D., editors. (2009a). *Challenging challenge questions. Socio-economic strand*. Oxford University UK.

Just M., Aspinall D., editors. (2009b). Choosing better challenge questions. Symposium on usable privacy and security (SOUPS); CA, USA: ACM.

Just M., Aspinall D., editors. (2009c). Personal choice and challenge questions: a security and usability assessment. Proceedings of the 5th Symposium on Usable Privacy and Security; CA,USA: ACM.

Just M., Aspinall D., editors. (2012). On the security and usability of dual credential authentication in UK online banking. Internet Technology and Secured Transactions, 2012 International Conferece For; IEEE.

Kitahara, R., Westfall, F., & Mankelwicz, J. (2011). New, multi-faceted hybrid approaches to ensuring academic integrity. *Journal of Academic and Business Ethics., 3*(1), 1–12.

Laubscher R., Olivier M. S., Venter H. S., Eloff J. H. P., Rabe D. J., editors. (2005). The role of key loggers in computer-based assessment forensics. Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries; 2005: South African Institute for Computer Scientists and Information Technologists.

Mcgraw, G. (2004). Software security. *Security & Privacy, IEEE., 2*(2), 80–83.

Moini, A., & Madni, A. M. (2009). Leveraging biometrics for user authentication in online learning: A systems perspective. *IEEE Systems Journal., 3*(4), 469–476.

Purdy, G. (2010). ISO 31000: 2009—setting a new standard for risk management. *Risk Analysis, 30*(6), 881–886.

Rabkin A., editor. (2008). Personal knowledge questions for fallback authentication: Security questions in the era of Facebook. In SOUPS: Proceedings of the 4th Symposium on Usable Privacy and Security; 2008; 23, New York: ACM.

Renaud K., Just M., editors. (2010). Pictures or questions?: examining user responses to association-based authentication. Proceedings of the 24th BCS Interaction Specialist Group Conference; British Computer Society.

Rowe, N. C. (2004). Cheating in online student assessment: Beyond plagiarism. *Online Journal of Distance Learning Administration., 7*(2).

Schechter S., Brush A. J. B., Egelman S., editors. (2009). It's No Secret. Measuring the Security and Reliability of Authentication via 'secret' questions. 30th IEEE Symposium on Security and Privacy; IEEE.

Seffah A., Kececi N., Donyaee M., editors. (2001).QUIM: A framework for quantifying usability metrics in software quality models. Quality Software, 2001 Proceedings Second Asia-Pacific Conference on; IEEE.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior., 6*(1), 156–163.

Ullah, A., Xiao, H., & Lilley, M. (2012a). *Profile based student authentication in online examination. International conference on information society*. London: IEEE.

Ullah A., Xiao H., Lilley M., Barker T., editors. (2012b). Usability of profile based student authentication and traffic light system in online examination. The 7th International Conference for Internet Technology and Secured Transactions (ICITST); London, UK: IEEE.

Ullah, A., Xiao, H., Lilley, M., & Barker, T. (2012c). Using challenge questions for student authentication in online examination. *International Journal for Infonomics (IJI), 5*(3/4), 9.

Ullah, A., Xiao, H., Barker, T., & Lilley, M. (2014a). Evaluating security and usability of profile based challenge questions authentication in online examinations. *Journal of Internet Services and Applications., 5*(1), 2.

Ullah A., Xiao H., Barker T., Lilley M., editors. (2014b) Graphical and text based challenge questions for secure and usable authentication in online examinations. The 9th International Conference for Internet Technology and Secured Transactions (ICITST); London, UK: IEEE.

Ullah, A., Xiao, H., & Barker, T. (2015). Usability of activity-based and image-based challenge questions in online student authentication. In T. Tryfonas & I. Askoxylakis (Eds.), *Human aspects of information security, privacy, and trust HAS 2015 lecture notes in computer science* (pp. 131–140). Cham: Springer.

Ullah A., Xiao H., Barker T., editors. (2016). A classification of threats to remote online examinations. Computing and Communication (IEMCON), 2016 International Conference and Workshop on; 2016: IEEE.

Watson, G., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? *Online Journal of Distance Learning Administration., 13*(1), n1.

Wheeler, D., Whittlestone, K., Smith, H., Gupta, A., & Menon, D. (2003). A web-based system for teaching, assessment and examination of the undergraduate peri-operative medicine curriculum. *Anaesthesia, 58*(11), 1079–1086.

Wiedenbeck S., Waters J., Birget J.-C., Brodskiy A., Memon N., editors. (2005). Authentication using graphical passwords: Effects of tolerance and image choice. Proceedings of the 2005 symposium on Usable privacy and security; ACM.

Youll J. (2006). Fraud vulnerabilities in sitekey security at bank of america." Available: www.cr-labs com/publications/SiteKey-20060718 pdf.

Zviran M., Haga W. J., editors. (1990). User authentication by cognitive passwords: An empirical assessment". Information Technology, 1990 'Next decade in information technology', Proceedings of the 5th Jerusalem Conference on (Cat No 90TH0326-9); IEEE.

Zviran, M., & Haga, W. J. (1993). A comparison of password techniques for multilevel authentication mechanisms. *The Computer Journal, 36*(3), 227–237.