Journal of the Association for Information Systems

Volume 10 | Issue 2 Article 2

2-17-2009

A Behavioral Analysis of Passphrase Design and Effectiveness

Mark Keith
Arizona State University, mark.keith@asu.edu

Benjamin Shao
Arizona State University, benjamin.shao@asu.edu

Paul Steinbart
Arizona State University, paul.steinbart@asu.edu

Recommended Citation
Keith, Mark; Shao, Benjamin; and Steinbart, Paul (2009) "A Behavioral Analysis of Passphrase Design and Effectiveness," *Journal of the Association for Information Systems*, 10(2), .
DOI: 10.17705/1jais.00184
Available at: https://aisel.aisnet.org/jais/vol10/iss2/2

Journal of the Association for Information Systems
JAIS

Research Article

A Behavioral Analysis of Passphrase Design and Effectiveness*

Mark Keith
Arizona State University
mark.keith@asu.edu

Benjamin Shao
Arizona State University
benjamin.shao@asu.edu

Paul Steinbart
Arizona State University
paul.steinbart@asu.edu

Abstract

Although the use of multiple methods of user authentication for IT system increases security, passwords are often the only credential required for access. Consequently, the challenge is to discover ways to improve password strength without impairing usability. Longer pass "phrases" have received increased attention as a solution to this challenge because they are potentially more resistant to attacks yet are easy to remember. Recent evidence, however, suggests that passphrases increase the likelihood of typographical errors resulting in login failures and negative user perceptions. This paper presents experimental results that demonstrate well-designed passphrases do not increase login failures and, thereby, generate positive user perceptions. Implications are drawn to help IT managers develop effective IT security policies in utilizing passphrases to improve authentication and to assist researchers in identifying avenues for future research.

Keywords and phrases: *Passwords, passphrases, authentication, security, memory, usability, user behavior.*

* Izak Benbasat was the accepting senior editor. Jesper Johansson, Wei Yue, and Kenneth Walsh were the reviewers.
This was submitted on May 25, 2007 and went through two revisions.

Volume 10, Issue 2, pp. 63-89, February 2009

A Behavioral Analysis of Passphrase Design and Effectiveness

1. Introduction

IT security could be enhanced by using multiple methods to authenticate users, such as combining "something you know" (e.g., a password) with "something you have" (e.g., a smartcard or token) and "something you are" (e.g., a biometric characteristic). Although the use of biometrics and smartcards is growing, passwords are still the most common, and sometimes the only authentication mechanism used by many organizations (Whitman, 2003). Therefore, it is important to find ways to improve password effectiveness.

To be effective as an authentication mechanism, passwords must simultaneously satisfy two conflicting requirements: they must be difficult to compromise, yet easy to remember. This challenge underscores the importance of considering user behavior when developing security solutions. If users are allowed to create their own passwords, they tend to use common words, names, dates, or other personal information that can be easily remembered (Brown et al., 2004). Such passwords, however, can also be easily guessed by attackers with knowledge about their intended victim. In addition, user-generated passwords are often easy to compromise by various types of permutation attacks that use pre-compiled dictionaries or word-lists (Narayanan and Shmatikov, 2005). If organizations attempt to mitigate this threat either by developing strict guidelines for creating passwords or by assigning users system-generated passwords that are comprised of random characters, users can find such complex passwords difficult to remember (Yan et al., 2004). Therefore, they often write the password down and store it insecurely (e.g., by attaching it to the monitor or placing it under the keyboard). Even if users store the written-down password securely (e.g., in their wallets), doing so changes it from being "something you know" to "something you have," thereby reducing the number of different authentication factors being used.

Switching from passwords to pass*phrases* may provide a solution to this security vs. "memorability" predicament. Passphrases are long passwords created from multiple words to form a phrase (e.g., "I love to eat chocolate chip cookies"). The idea of using passphrases is not new (cf Porter, 1982), but until recently most systems have restricted password length to eight characters, making the use of longer passphrases infeasible. Now, many online authentication schemes and newer versions of both Windows and Mac operating systems support the use of much longer passwords, and security professionals recommend replacing passwords with passphrases (Burnett, 2005; Center for Internet Security, 2004, Section 2.2.2.4; Johansson and Riley, 2005, p. 345; Skoudis and Liston, 2006, p. 402). Passphrases are attractive because there is evidence that they are as easy to remember as user-generated passwords (Keith et al., 2007; Yan et al., 2004), yet are potentially more resistant to compromise than system-generated passwords (Burnett, 2005; Johansson, 2004c). Passphrases also increase the effectiveness of behavioral biometric tools like keystroke analysis in distinguishing between different people entering the same data (Huston, 2006). Thus, passphrases may not only be inherently more secure than passwords, but may also improve the joint level of security when used as part of a multi-factor authentication scheme.

Recent research, however, reveals a potential drawback to passphrases: increased typographic errors (Keith et al., 2007). Such errors not only result in more login failures but also negatively affect such user perceptions as ease-of-use. User perceptions are important because they can influence willingness to comply with the organization's security policies. If users decide not to comply with security policies that are perceived as being onerous and counterproductive, the overall level of security declines. Therefore, it is important to investigate whether switching from passwords to passphrases merely replaces one trade-off between security and usability with another.

This study reports the results of a 12-week field experiment that investigated whether passphrases can be constructed in a manner that makes them both more effective and easier to use than strong passwords. The remainder of this paper is organized as follows. The next section provides the theoretical foundation for our research design and develops our hypotheses. Then we describe our research methodology and the experimental task, and present our results. We discuss the implications of those results, including the limitations of the study, before we conclude this paper.

2. Literature Review and Hypotheses

Prior research on passwords has primarily focused on technical issues, particularly how the characteristics of the "something you know" credential affect its strength and ease of recall (e.g., Morris and Thompson, 1979; Pond et al., 2000; Wiedenbeck et al., 2005, Zviran and Haga, 1990). Password strength refers to its resistance to both guessing and "cracking" attacks. Guessing attacks are online: The perpetrator tries repeatedly to log in to the target account by trying various passwords. The likelihood of an attacker successfully guessing a user's password can be significantly reduced by enforcing policies about minimum password length, required frequency of changing, and the maximum number of attempted logins permitted before the account is locked. Proper policies, combined with periodic examination of logs to identify excessive failed attempts to log in to specific accounts, should make guessing attacks unlikely to succeed (Johansson and Riley, 2005, p. 327). On the other hand, password cracking involves stealing a copy of the encrypted or hashed password file, or capturing the challenge-response sequence, and attempting to create strings that match the captured credentials (Johansson, 2004a).[1] Precompiled hash files exist for words in almost every language; therefore, "cracking" any password that is contained in such a list is a relatively trivial task. Such dictionary attacks can be thwarted, however, by "salting" passwords with additional random text prior to hashing. Salted passwords and passwords comprised of random sequences of characters can only be cracked by brute force enumeration of every possible combination of allowable characters. Password strength, in terms of resistance to such brute force enumeration attacks, is a function of the size of the search space, which can be calculated by the formula n^L, where n represents the size of the allowable character set that is used to create the password (or passphrase) and L represents its length. Thus, increasing the length L of a password/passphrase exponentially increases the size of the search space. If the distribution of passwords or passphrases within the potential search space is uniform, attackers would need to enumerate, on average, one-half of the possibilities to successfully guess the login credential.

This n^L formula suggests that longer passphrases should be much stronger than fixed-length system-generated random passwords. The formula, however, assumes that each character is randomly chosen with equal probability. This assumption is not likely to hold for user-generated passwords and passphrases. For example, because passphrases consist of words, there are certain patterns of letter sequences (e.g., in English the letter q is almost certain to be followed by the letter u, sequences of three or more consecutive vowels are highly unlikely, etc.). As such, *entropy* is a better measure of password strength than simple length and character set, because it also takes into account the probability with which each individual character is chosen (Johansson, 2004b). Entropy is a measure of the randomness of a password; the more random the sequence of characters, the higher the entropy of a password and the more resistant it is to cracking. There is no universal agreement on how to calculate the entropy of user-generated passwords and passphrases. Nevertheless, even conservative calculations indicate that well-designed passphrases are likely to have higher entropy (i.e., be more resistant to brute-force guessing attacks) than eight-character system-generated passwords that represent truly random sequences of symbols[2] (Johansson, 2004b; Johansson, 2004c).

The n^L formula may also have to be adjusted for passphrases to reflect the fact that the relevant unit

[1] Theft of the password hash file means that the attacker has already compromised at least one machine. Cracking the password file enables the attacker to continue to access the system in a manner that is difficult to track (i.e., by logging in as a legitimate user). In addition, because users often use the same credential on multiple systems, cracking passwords on one system may enable attackers to successfully jump to other targets.

[2] For example, a common measure of entropy is log (base 2) of character set size (Johansson, 2004b). Passphrases consisting of 26 alphabetic characters plus the following 13 special characters enclosed in brackets [.,;:?"'()-!$%] and the space bar total 40 potential characters resulting in 5.3 bits of entropy per character (log(base2) of 40). Hence, a three-word (approx. 16 characters) passphrase consisting of the above 40 characters has about 84.8 bits of entropy (16 characters x 5.3 bits). In comparison, a completely random password has 6.6 bits of entropy per character (log(base2) of 95 possible characters). Therefore, an eight-character random password has only 52.8 bits of entropy (8 characters x 6.6 bits) compared to 84.8 for a passphrase.

of analysis is words, rather than characters.[3] Thus for passphrases, the character set size *n* is the size of the average user's vocabulary and *L* is the number of words used in the passphrase. Estimates of the size of an average adult's vocabulary vary dramatically, ranging from a few thousand to more than 50,000 words (Crystal, 2003; Wren, 2003). As was the case with entropy, however, even the most conservative estimates of vocabulary size indicate an advantage for passphrases over random system-generated passwords that contain a mix of alphanumeric and special characters. For example, a five-word passphrase drawn from a conservative vocabulary set of 3,000 words is more resistant to brute-force enumeration than an eight-character random system-generated password, and an eight-word passphrase is as strong as a 14-character random system-generated password.

However, technology and behavior are inseparable in information systems (Hevner et al., 2004). Password security is, thus, as much a behavioral issue as a technical one. Users adapt their behaviors to requirements imposed by a system, and they also attempt to modify or alter aspects of the system in order to make it easier to use (DeSanctis and Poole, 1994). The resulting "technology in practice" (Orlikowsky, 2000) may be quite different from what the system designer intended and may lead to counterintuitive results (Gray and Durcikova, 2005-6). Consider the situation where organizations permit users to create their own passwords and provide explicit guidance for developing strong passwords. Such guidance typically includes requirements to use multiple types of characters, to not use words from the dictionary, and to exceed some minimum length. If organizations do not monitor or enforce compliance with those guidelines, users tend to ignore them and create weak passwords (Zviran and Haga, 1999). If organizations employ password-checking programs (Ruffo and Bergadeno, 2005) to monitor and enforce compliance with password guidelines, users may resort to behaviors that satisfy the "letter" but not the "spirit" of those rules. For example, users may create passwords like !QAZxsw2 or pa$$W0rd. Although both examples appear to satisfy typical complexity requirements (i.e., they are case-sensitive and contain both numbers and special characters), each may be included in some precompiled dictionaries (the former because it consists of a predictable pattern based on the keyboard layout and the latter because it is a regular word, albeit with special characters replacing normal letters). Another undesirable user behavior is the propensity to use the same password on multiple systems (Ives et al., 2004; Johansson & Riley, 2005). This creates what is referred to as a "security dependency" or "domino effect" in which compromising a weak system provides attackers with authentication credentials for access to other more secure systems.

Besides the act of *generating* a password, *using* the password during regular logins also takes on an adaptive structure. Password authentication systems are designed to restrict access to only those people who correctly enter a username and a password. The underlying assumption is that legitimate authorized users can successfully meet these requirements. Users, however, often make typographical errors or may even forget their passwords altogether. Many systems are designed to lock users out after a certain number of failed attempts (usually three to five). Such lockouts not only result in extra time and cost spent resolving these problems, but are also likely to cause users to alter their behaviors to avoid future problems. For example, if users have difficulty remembering a complex password, they may write it down. Alternatively, to avoid typing errors, users may store their password or passphrase electronically so that they can simply copy and paste it when authenticating themselves to a system. Both responses transform the authentication credential from "something you know" to "something you have," with concomitant unfavorable changes in the effective level of security provided.

The preceding examples illustrate how user behaviors can cause the actual level of security provided by a specific authentication credential to be much lower than the analysis of its technical specifications would predict. They also underscore the need to consider user experience with and perceptions about authentication mechanisms. If users have negative perceptions about a system, they are less likely to use it voluntarily; or if use is mandatory, users are likely to circumvent or modify features that are perceived as too burdensome (Ives et al., 1983; Mahmood et al., 2000; McKeen et al., 1994). Thus, the difficulty of use (in the form of login failure rate, for example) is likely to be

[3] We thank an anonymous reviewer for calling attention to the need for this adjustment in unit of analysis to words.

directly associated with user attempts to modify the implementation of an authentication mechanism. Consequently, when deciding whether to replace passwords with passphrases, system designers and security administrators should consider the relative propensity of each technique to cause login failures.

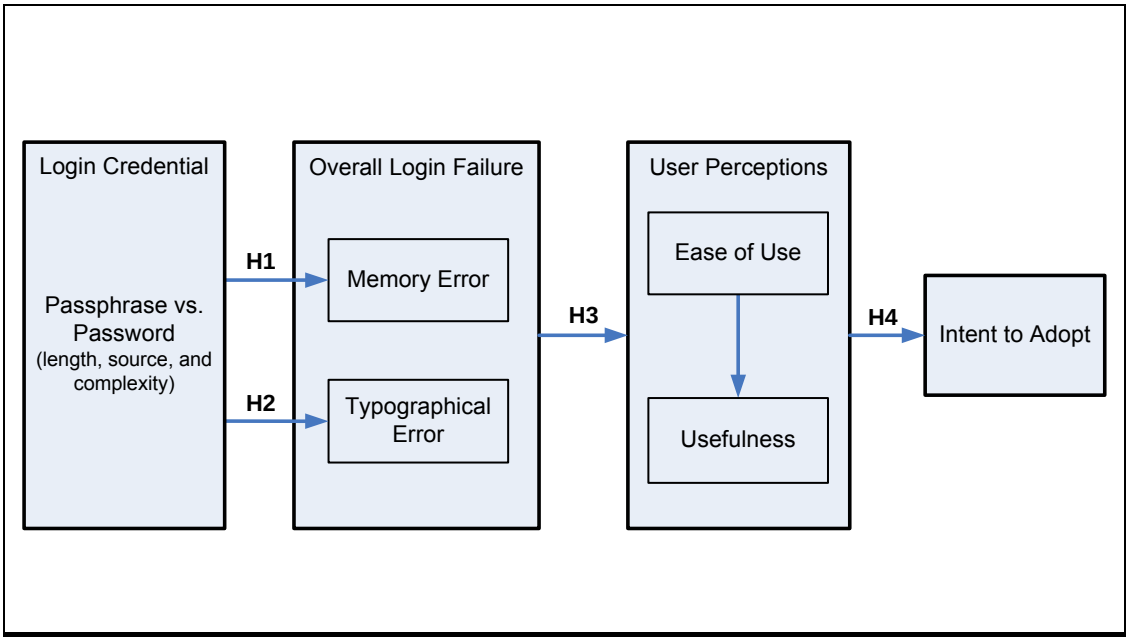

Figure 1. Theoretical model of login failure, user perceptions, and intentions

Login failures by legitimate authorized users can be due to either forgetting the login credential or making typographical errors when entering it. Therefore, the relative effectiveness of passwords and passphrases depends, in part, on how easy it is to both remember and successfully enter each type of credential. Figure 1 displays the theoretical relationships between password characteristics and login failure as well as between login failure and the subsequent user perceptions and intentions. The following sections discuss these relationships and constructs in greater detail.

2.1. Memory-related Issues with Authentication Credentials

In order to truly be an authentication credential that is "something you know," passwords or passphrases must be memorized and not written down. One important factor affecting memorization is the ability to retain the to-be-learned information in short-term (or "working") memory long enough to transfer it to long-term memory (Anderson, 2005; Driscoll, 2005, p. 86-87). This transfer depends upon the relationship between the amount of information to be learned and the storage capacity of short-term memory. Initially, short-term memory was estimated as being capable of holding between five to nine "chunks" of information (Miller, 1956), but more recent studies suggest that the effective optimal size is actually three to five "chunks" of information (Cowan, 2001; Doumont, 2002). However, this is not as stringent a limitation as it may seem because a chunk of information need not be limited to a single character or number, but may instead be any *meaningful* set of related items (Simon, 1974, p. 482). Thus, both a single word like "baseball" and a phrase like "roses are red, violets are blue" may each be treated as one chunk of information, even though the former consists of eight and the latter of 31 characters. Consistent with this chunking theory, prior research has indeed found that passphrases are as easy to remember as user-generated passwords (Keith et al., 2007; Zviran and Haga, 1993).

Because of the ease of cracking passwords based on common words, most organizations require user-generated passwords to consist of a mixture of alphanumeric and special characters. To facilitate memorization, users often create passwords like "$3Cur!ty" that appear to be random

sequences of symbols but in reality are still based on words that can be easily remembered (e.g., by substituting the characters $, 3, and ! for the letters s, e, and i in the word "security"). Such memory-enabling artifices, however, are also known to attackers and can be used to simplify what would otherwise be brute-force guessing. Consequently, some organizations mandate the use of system-generated passwords that consist of truly random sequences of characters. Although such passwords (e.g., "8Wk$nP3!") may be harder to guess or crack, they are also much more difficult to memorize because each character will be treated as a separate chunk of information.

Users must both memorize their authentication credentials and be able to recall them correctly when attempting to access the system. People can better remember meaningful words and non-words if they generated those items themselves as opposed to merely read them (Cameron et al., 2005; Jacoby, 1978; Johns and Swanson, 1988; Slamecka and Graf, 1978). This so-called *generation* effect suggests that it should be easier for users to recall self-generated passwords or passphrases than system-generated passwords.

There is reason to suspect, however, that self-generated passwords and passphrases may themselves differ in ease of recall. People often attempt to use the same or slightly modified passwords on different systems in order to reduce the number of different passwords they need to remember (Ives et al., 2004). For example, if different systems impose different requirements (e.g., length, use of special characters, etc.), users may simply modify a password they already use on a system with less stringent requirements (e.g., replacing the letter "l" with the number "1"). Yet, a robust finding in memory research, referred to as the "phonological similarity effect" (PSE) (Baddeley, 1966; Conrad, 1964; Lian et al., 2001), suggests that sets of phonologically similar words are more difficult to distinguish and recall than dissimilar words. Thus, although it may be easier to *generate* new passwords that are similar to other passwords, doing so likely makes it more difficult for users to *recall* the correct password later when attempting to authenticate to a particular system. Currently, few systems require the use of passphrases. As a consequence, passphrases should be quite dissimilar to passwords used to access other systems and, therefore, should be easier to recall.

The preceding discussion leads to the following hypotheses:

H1a: Users of passphrases will experience fewer login failures due to memory errors than will users of either self-generated passwords or system-generated random passwords.

H1b: Users of self-generated passwords will experience fewer login failures due to memory errors than will users of system-generated random passwords.

2.2. Typing-related Issues with Authentication Credentials

People may correctly remember their password or passphrase, but still make a typographical error when entering it. This problem is aggravated by the design of human-computer login interface, which masks the characters being entered to prevent "over-the-shoulder" password discovery. As a result, users do not have the ability to "recognize" that they have made a mistake and to correct it immediately (as they can do when using a word processing program). On the contrary, they only receive feedback (in the form of a login failure) after they finish entering an invalid password.

Typing is a process that involves four stages (Salthouse, 1986): 1) the *input* stage where text is converted in memory into chunks, 2) the *parsing* stage where the chunks are decomposed into strings of characters, 3) the *translation* stage where characters are converted in memory to specific finger movements, and 4) the *execution* stage where the key presses are carried out. Although errors can occur at any one of these stages (Salthouse, 1986), the latter two are most likely to produce "typos" when entering login credentials. The most common problems include *substitution* errors, which occur by accidentally striking a nearby key instead of the correct one; *temporal* errors, which involve transposing correct letters; and *execution* errors that arise from entering too few or too many keystrokes (Logan, 1999).

The probability of making a typing error increases with the amount of material being entered. Thus,

one might expect that passphrases, because of their increased length, should result in more typing errors than short passwords. Indeed, Keith et al. (2007) report that user-created passphrases were more than twice as long as user-generated passwords and resulted in significantly more login failures due to typing errors.

Length, however, is only one of many factors that can cause typing errors. It has been argued that because of the many potential error factors affecting each key press, "the challenge is more one of explaining accuracy rather than errors" (Logan, 1999, p. 1769). One such explanation is offered by Rieger (2004), who argues that typists can develop "automatic activation" for common words and phrases—meaning that actions can become so learned that the translation stage (that of converting letters into keystrokes) can require almost no effort. In other words, typing patterns that have become well-learned will result in higher speeds and greater accuracy due to the strengthening of cognitive information processing pathways[4] during training (Cohen et al., 1990; John, 1996; Rumelhart and Norman, 1982). This argument has significant implications for the design of passphrases and passwords. Nowadays, most users have considerable word processing experience. Therefore, familiar sequences of characters that consist of commonly typed words and that use typical spelling and punctuation should be typed much more quickly and accurately than unusual sequences of characters that include numbers and special characters. Consequently, security practitioners have argued that it is easier to type passphrases than complex passwords that consist of not only alphanumeric but also special characters (cf Skoudis and Liston, 2006, p. 402). For example, users should make fewer typing errors when entering the word "scooter" than when typing "Sc00ter" or "$c00t3r" because the former conforms to normal spelling conventions used when typing a text document. We will refer to typing passwords and passphrases consistent with the contents of normal text documents as being in "word processing mode" (WPM).

Users are likely to enter into WPM when typing a passphrase because the credential is a sentence or sentence fragment. Consequently, passphrases that are similar in structure to what would be typed in a word processing document should be easier to type correctly than those that violate normal word processing rules, because the former represent a well-learned skill. For example, when numbers are included in a passphrase, they are less likely to result in typing errors if they are used as dates or ages than if they are appended to or mixed into the words that comprise the passphrase (e.g., passphrases like "In 1492 Columbus sailed the ocean blue" or "Bill Jones is 23" should result in fewer typing errors than passphrases like "Th1s is my s3cr3t passphrase" or "Once upon a time3"). The use of spaces to separate words is also more consistent with WPM and, therefore, should result in fewer typing errors than attempting to enter the same phrase either without any spaces or using the underscore character to separate words.

Lack of conformity to WPM may explain Keith et al.'s (2007) finding that passphrases increased the number of login failures due to typing errors. In their study, users who were assigned to the passphrase condition were required to create credentials that were case sensitive and that contained at least one non-letter. Our examination of the credentials created by their subjects reveals that most participants did not create phrases consistent with WPM. Instead, they either did not use spaces or used a special character, such as the underscore, to separate the words in the phrase. Therefore, Keith et al.'s (2007) finding that passphrases result in more typing errors may not represent an inherent problem with passphrases but, rather, may simply reflect problems associated with creating passphrases that are inconsistent with WPM. If people create passphrases that follow normal word processing conventions, their well-developed word processing skill should offset the increased length and may even make such passphrases less prone to typing errors than user-generated passwords comprised of a non-WPM sequence of alphanumeric characters. This leads to the following hypotheses:

H2a: Users who create passphrases that conform to word processing mode (WPM) will

[4] The term "processing pathways" refers to the pattern of cognitive activities that comprise human actions. For example, finger movements occur "via a set of connected modules that form a pathway" (Cohen et al., 1990, p. 335) through the brain. The speed and accuracy at which the finger moves depends on the "strength" of the pathway.

experience fewer typing errors than will users of either system- or self-generated passwords.

H2b: Users who create passphrases that conform to WPM will experience fewer typing errors than will users who create passphrases that are not consistent with WPM.

H2c: Users of system-generated random passwords will experience more typing errors than will users who generate their own passwords.

2.3. User Perceptions of Alternative Authentication Credentials

In general, people prefer to minimize effort when using systems, provided then achieve an acceptable level of accuracy (Todd and Benbasat, 1994; 1999; 2000). At first glance, it would appear that it takes more effort to use passphrases than either user-generated or system-generated passwords because they are longer and require more keystrokes to enter. However, the preceding sections have presented reasons to believe that passphrases may be easier to remember and less prone to typing errors than either user-generated or system-generated passwords. Thus, passphrases may actually require less effort to use *successfully* than either user-generated or system-generated passwords.

Moreover, according to the technology acceptance model (Davis, 1989; Adam et al., 1992; Venkatesh and Davis, 2000; Wixom and Todd, 2005), if users cannot successfully log in either because they make a typographic error or because they forget their password, they are likely to form negative perceptions about the "ease of use" and "usefulness" of that credential, and be less satisfied overall. Both typing and memory errors are expected to create similar negative perceptions, because the design of the login interface only provides feedback that the login attempt was unsuccessful, but does not indicate the cause for the login failure. Indeed, the two types of errors can become intertwined and hard for users to distinguish. If users experience repeated login failures due to typing errors, they may question whether they remember the correct password for that system. If they then enter a different password, they will experience additional login failures, thereby further exacerbating their perceived ease of use and usefulness about the login credential being used. Hence, the *overall* login failure rate, rather than the cause of the login failure, is likely to determine user perceptions. The more frequently users encounter problems when using a credential to attempt to login, the more negative their perceptions are likely to become. As a result of negative perceptions, they will have lower intentions to voluntarily adopt that credential for future use. This leads to the following hypotheses:

H3: Overall login failure rates will be inversely related to perceptions about the ease-of-use and usefulness of login credentials.

H4: User perceptions about the ease-of-use and usefulness of a login credential increase their intentions to voluntarily adopt this type of credential for future use.

Table 1 summarizes the variables of interest and how they relate to our hypotheses.

3. Methodology

3.1. Participants and Task

We conducted a longitudinal field study to test our hypotheses. Undergraduate students from a large public university who were enrolled in an elective course on web development offered by the school of business participated in the experiment. Course instructors used a class website to provide resources and materials for web design. The site contained a restricted section where assignments, materials, and tutorials could be accessed only via username and password. Before classes began, a database was generated containing each student's university-assigned username. Next, each username was randomly assigned to one of three groups for password generation requirements: 1) Standard, user-generated passwords at least eight characters long and containing one or more non-letters, 2) Random, system-generated passwords that were eight characters long and created from the 95-character base FIPS standard (1985), and 3) User-generated passphrases at least 16 characters long. Hereafter, we refer to these as the *standard*, *random*, and *passphrase* groups, respectively.

Table 1. Theory, Variables, and Hypotheses about Password and Passphrase Use

Theory	Explanation	Variable(s)	Hypotheses
"Chunking" (Cowan, 2001; Miller, 1956)	The ability to "chunk" information makes related letters and words easier to remember. Therefore, passphrases and user-generated passwords that can be represented as one chunk are easier to remember than system-generated passwords, because each random character is likely to be treated as a separate chunk.	Memory-based login failures	H1a: Users of passphrases will experience fewer login failures due to memory errors than will users of system-generated random passwords. H1b: Users of self-generated passwords will experience fewer login failures due to memory errors than will users of system-generated random passwords.
Phonological Similarity Effect (Baddeley, 1966; Conrad, 1964; Lian et al., 2001)	Multiple "orthogonal" words are easier to remember at once than highly similar words. Therefore, multiple passwords and passphrases are easier to remember if they are more different from others in memory. Because passphrases are seldom used at the present time, they should differ markedly from any passwords previously used and, therefore, should be easier to remember than self-generated passwords.	Memory-based login failures	H1a: Users of passphrases will experience fewer login failures due to memory errors than will users of self-generated passwords.
Skilled Typing (Rumelhart and Norman, 1982; Salthouse, 1986)	Those characters that are less commonly typed in word processing (e.g., numbers and non-alphanumeric characters) are less automatic in terms of motor skills. As more of these unusual characters are used, passwords and passphrases become more difficult to type correctly.	Typo-graphical-based login failures	H2a: Users who create passphrases that conform to word processing mode (WPM) will experience fewer typing errors than will users of either system- or self-generated passwords. H2b: Users who create passphrases that conform to WPM will experience fewer typing errors than will users who create passphrases that are not consistent with WPM. H2c: Users of system-generated random passwords will experience more typing errors than will users who generate their own passwords.
User Perceptions and Technology Acceptance (Davis, 1989; Venkatesh and Davis, 2000; Wixom and Todd, 2005)	User experiences with login credentials determine their subsequent perceptions. In the text-based authentication context, these measures are based upon the user's rate of login failure, which is a function of the rates of both memory- and typographically-based errors.	Perceived usefulness, perceived ease-of-use, intent to adopt	H3: Overall login failure rates will be inversely related to perceptions about the ease-of-use and usefulness of login credentials. H4: User perceptions about the ease-of-use and usefulness of a login credential increase their intentions to voluntarily adopt this type of credential for future use.

In the first class, each participant accessed the course's website registration page and entered his or her university username. Depending on the group to which they were randomly assigned, participants then received one of three different sets of instructions. The interfaces for each group were identical except for the instructions about the nature of the login credential that would be used to access the materials on the course website. Participants assigned to the standard group were asked to generate a password that was at least eight characters long and contained at least one non-letter. Participants assigned to the random group were asked to provide an email address where they would have their password sent to them. Participants assigned to the passphrase group were asked to generate a password based on a three- to five-word phrase at least 16 characters long.[5] Website functionality was built in to ensure that the standard and passphrase participants met length and character requirements. In addition, participants assigned to the passphrase and standard password groups saw a pop-up window that reminded them not to create a login credential that was similar to one they used on any other system.

The course included eight homework assignments that could only be obtained by logging into the website. These assignments were given at a rate of one per week, leaving six weeks with no login requirement toward the semester's end. During those last six weeks, participants could continue to log in to the website to retrieve missed assignments, download lecture slides, and obtain web development resources. Every time a participant attempted to log in, the website application recorded the username, login credential entered, timestamp, and outcome (success or failure). Participants who forgot their passwords had to personally contact the instructor, who then verbally provided them with their password. If participants needed a password reminder outside of class hours, they were told to call the instructor at a specified phone number or send an email with a phone number at which the instructor would call them back.

3.2. Measures

Login Failure and Error

The overall login failure rate, failures due to memory errors, and failures due to typing errors were calculated for each individual. *Rates* of failures were used rather than *totals* in order to normalize the varying numbers of individuals' login attempts (i.e., participants who have more login failures will naturally have higher login totals). Distinguishing between memory-errors and typing-errors was accomplished using a hybrid of objective measures and subjective judgments as described below.

Keith et al. (2007) used a formula known as the "Levenshtein" (L) distance (Levenshtein, 1966) to distinguish typo- from memory-related errors. The resulting L-score is a measure of the difference in the characters between two strings. For example, entering "paswword" instead of "password" results in an L-score of 1. The algorithm also accounts for difference in string lengths, so that when compared to the correct word "password" both "password," and "passwords" have an L-score of 1.

Incorrectly typing one character results in an L-score of 1. Thus, it is tempting to classify any login failure with an L-score of 1 as a typographical error, and any login failure with an L-score greater than 1 as a memory problem. Although objective, such a rule is inadequate, because it does not take into account the *context* of the login failures. For example, if a user makes only one keystroke error when entering his password, then it seems likely that he has simply made a typographical error. On the other hand, if the user makes the *same* one-character mistake several times in a short period, then it is more likely that he has, in fact, forgotten the exact spelling of the password. Nonetheless, the L-scores would be 1 for each such login failure. In addition, the L-score formula does not take into account the position of the keystroke error or its context in a word. For example, assume that "hairball" is *user1*'s password and that *user1* experiences two separate login failures when entering "hairball\" and "hairballs" instead of the correct password. The L-scores for both mistakes would be 1;

[5] Because many participants might never have used passphrases before, two examples were provided in the instructions. The first example was "Ilovetosnowboard" and the second was "I love to snowboard". The system ensured that participants could not simply copy and paste either of the examples provided. In addition, the "remember password" option was disabled during the course of our experiment.

consequently, use of the rule that an L-score of 1 represents a typing error would categorize both mistakes as typographical errors. The two errors, however, are actually quite different. The "\" character is just above the enter key and, therefore, may conceivably have been pressed by accident. In contrast, the letter "s" is not next to either the letter "l" or the enter key and, consequently, more likely reflects a memory error (i.e., the user forgot whether the password was singular or plural).

Large L-scores are also potentially problematic. A large L-score can represent a memory error, in which the user entered the wrong credential for that system. On the other hand, it could also result from accidentally pressing the "Caps Lock" key, instead of the shift key when typing the first character of the login credential.

Examination of the sequence of login attempts, however, can often provide clues about the cause of a login failure. For example, entering a different password after experiencing a login failure suggests that the preceding login failure was due to a memory error (i.e., the user forgot which password to use on this system). Similarly, responding to a failed login attempt by re-entering the same password but typing it correctly is evidence that the previous failure represented a typing error.

The preceding examples demonstrate why it is necessary to use subjective judgment in combination with an objective measure like an L-score in order to classify the cause of a login failure. Therefore, we asked two judges (neither of whom was aware of the hypotheses being tested) to use both L-scores and information about the login context to determine the cause of login failures. These judges were given a list of every login failure, including the user identification number, that user's correct password or passphrase, incorrect credential entered, L-score, and a timestamp so that they could take into account both the context and sequences of login attempts. In addition, we explained the L-scoring technique to them.

The judges individually categorized each login failure and recorded the reasons for their decisions. The two judges then met and compared their individual assessments. They agreed on 92.3 percent of the 351 login failures. The Kappa coefficient is 0.775 ($p < 0.001$), indicating that their agreement rate was significantly greater than chance. The judges discussed the cases where they disagreed with each other and reached agreement on how to categorize every such login failure. We used the consensus classification identified for each login failure as the dependent variable.

User Perceptions and Intentions

We measured user perceptions and intentions by administering a survey on the last day of class. Each survey item was based on a Likert-type scale ranging from 1 to 7 (1 = strongly disagree, 7 = strongly agree). Items measuring *ease-of-use* and *usefulness* were based on the dimensions used by Adams et al. (1992), with modifications for the password context. We also included single item to measure intent to adopt (see Appendix A for the list of survey items).

4. Experimental Results

Of the 58 undergraduate students who began the experiment, five (two from each password group and one from the passphrase group) dropped the class during the first five weeks. In addition, we removed one subject from the passphrase group who did not follow the generation guidelines to generate a three of five word phrase from the final analysis, resulting in a sample size of 52 individuals. This resulted in 18 participants in the standard password group, and 17 participants each in the random and passphrase treatments.

Over the course of the semester, the participants attempted 1,540 logins (of which 351 were unsuccessful) to the class website, resulting in an average of almost 30 attempts per individual. Although a minimum of only eight successful attempts were required to retrieve the eight homework assignments, the additional login attempts were a result of both login failures and a desire to access the additional web development resources available on the site.

4.1. Demographics and Descriptive Data

Thirty-eight of the 52 participants were male. Forty-two participants were native United States Citizens. Thirty participants were students from the school of business where the web development course was offered, and the remainder came from the schools of fine arts, liberal arts and sciences, and engineering. Of those participants who chose to report their academic standing, there was one freshman, two sophomores, 19 juniors, 23 seniors, and four graduate students. The average participant age was nearly 24, with a range from 17 to 43.

Table 2 contains descriptive information concerning the authentication credentials for these 52 participants. The passphrase group's credential averaged 18.2 characters in length (and 3.63 words) compared with 9.7 and 8 characters for the standard and random password groups, respectively. The average character base is determined by the presence of lowercase letters, uppercase letters, numbers, and non-letters in the password. For example, a password consisting of only lowercase letters has a character base of 26; one containing both uppercase and lowercase letters has a character base of 52; one that is case sensitive and includes numbers has a character base of 62; and one that used all keyboard characters on a standard QWERTY keyboard has a character base of 95.[6]

Table 2. Password and Passphrase Descriptive Data

Group	Group size	Ave. length	S.D. of length	Ave. char. base	# of possible combinations
1. Standard Password	18	9.7	1.52	37.1	2.4E+15
2. Random Password	17	8.0	0.00	88.0	3.6E+15
3. Passphrase	17	18.2	1.69	31.0	9.9E+26
	Ave. # of lowercase letters	Ave. # of uppercase letters	Ave. # of numbers	Ave. # of spaces	Ave. # of non-alphanumeric characters excluding spaces
1. Standard Password	6.9	0.0	2.7	0.0	0.2
2. Random Password	2.3	2.9	0.8	0.0	2.0
3. Passphrase	16.3	0.2	0.6	2.0	0.0

Table 2 also shows that user-generated passwords were more likely to contain numbers than special characters. Passphrases, however, seldom included numbers or any special characters other than spaces. We ran a popular password-cracking program called "Ophcrack" for two hours against a Windows Vista password hash dump file of the login credentials used in this experiment to evaluate their relative strength. We were able to crack 16 of the 18 user-generated passwords, one of the system-generated random passwords, but none of the passphrases.[7]

As a manipulation check, the user perceptions survey administered at the conclusion of the experiment included an item asking participants to indicate how similar this password or passphrase was to any previous or current passwords they have used. Based on a Likert scale from 1-7 (1 = very similar, 7 = very dissimilar), the standard, random, and passphrase groups reported means of 4.50,

[6] It can be argued that the mere presence of non-alphanumeric characters does not indicate a "true" 95-character base because the 10 characters associated with the number keys on a keyboard are used far more often than the other 23 non-alphanumeric characters. While this may be true, the only two user-generated passwords in Table 2 that we classified as using the full 95-character set both used non-alphanumeric characters not associated with the number keys.

[7] As one anonymous reviewer noted, the failure to crack any passphrases may be an artifact of the design of most current password cracking programs, which are limited to brute-force guessing of credentials that are less than 15 characters in length.

6.57, and 6.63 respectively, indicating that each group considered their passwords to be different from those they used in other settings.

4.2.. Login Performance

We performed an analysis of the differences in login performance between groups using one-way ANOVAs with credential type as the independent variable and the participant's login failure rates due to memory and typographical errors as dependent variables. We used T-tests to compare the hypothesized group differences[8] whereas comparisons that were not hypothesized.

Overall (combining memory and typographical errors), the passphrase group experienced the lowest login failure rate (10.98 percent), followed by the standard (20.32 percent) and random (30.15 percent) groups. Perhaps most interesting, four of the passphrase users never experienced a single login failure, compared to only two of the standard password users and one of the random password users.

The judges classified 299 of the 351 login failures as reflecting memory errors and 52 as representing typographical errors (see Table 3 for group details). We performed multivariate analysis of variance (MANOVA) using Wilks's lambda criterion to check for mean differences among the three groups across memory and typographical error rates (See Appendix B for details). Results of the MANOVA indicated that significant differences exist (F = 4.059, p = 0.004). Therefore, we performed independent ANOVAs for each dependent variable (see the results in Appendix C).

Table 3. Descriptive Data of Login Attempts

Group		Total Attempts	Successful Attempts	Memory Errors	Typo-graphical Errors
1. Standard Password	Mean	34.17	25.94	7.56	0.67
	Std. Dev.	13.23	8.62	9.50	0.91
2. Random Password	Mean	29.71	20.00	8.12	1.59
	Std. Dev.	11.85	8.14	6.57	1.18
3. Passphrase	Mean	24.71	22.47	1.47	0.76
	Std. Dev.	12.68	11.73	1.33	1.15

Memory Errors

Figure 3 shows that the passphrase group experienced significantly fewer memory errors (7.80 percent) than either the standard group (18.22 percent, p = .017) or the random group (24.78 percent, p < .001). Thus, H1a is supported. However, the standard and random groups did not differ significantly in the number of login failures due to memory errors (p = .124). Thus, H1b is not supported.

Typographical Errors

Hypothesis H2a predicts that users who created WPM-consistent passphrases would make fewer typing errors than would users of either self-generated or system-generated random passwords. Four participants created passphrases that were not consistent with WPM. Therefore, we tested H2a by excluding those four participants from the analysis. Figure 4 shows that the typographical error rate for the 13 passphrase users who created WPM-consistent passphrases was 2.34 percent, which was significantly lower than the typing error rate for the random group (5.36 percent, t = 1.927, p = .032). The difference in typing error rates between the passphrase and standard password groups, however, was not significant (means of 2.34 percent vs. 2.11% percent t = 0.185, p = .854). Thus, H2a is only partially supported.

[8] Levene's test was used to measure the homogeneity of variances and t-test adjustments. Only one pairwise comparison did not pass the Levene test for the homogeneity of variances (p > .05); hence the corresponding t-test for that comparison was performed assuming unequal variances.

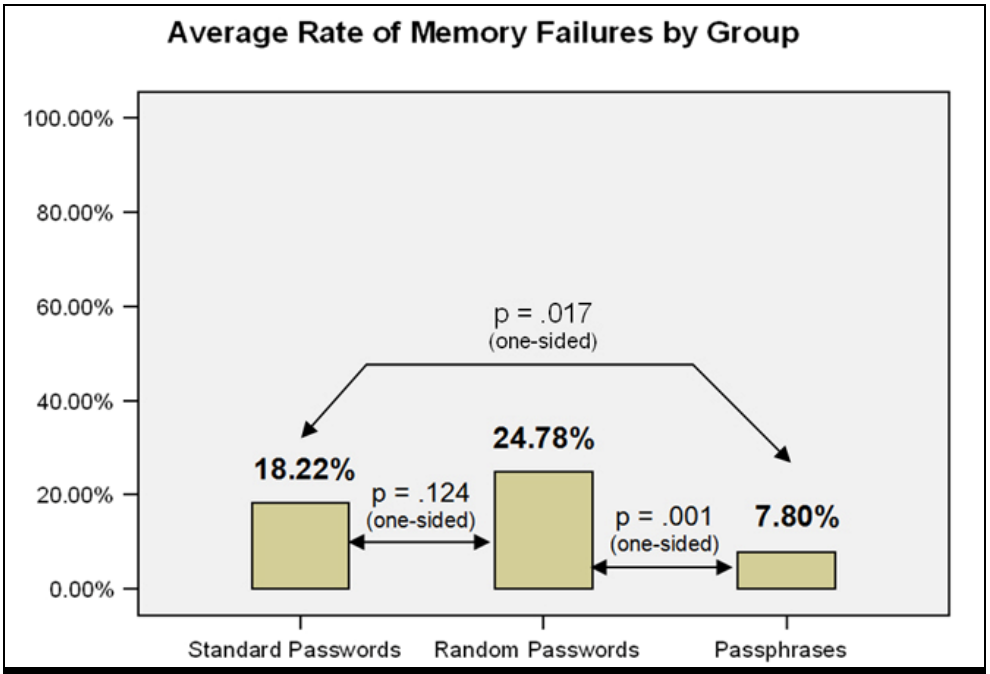

Figure 3: Average Rate of Memory Errors by Group

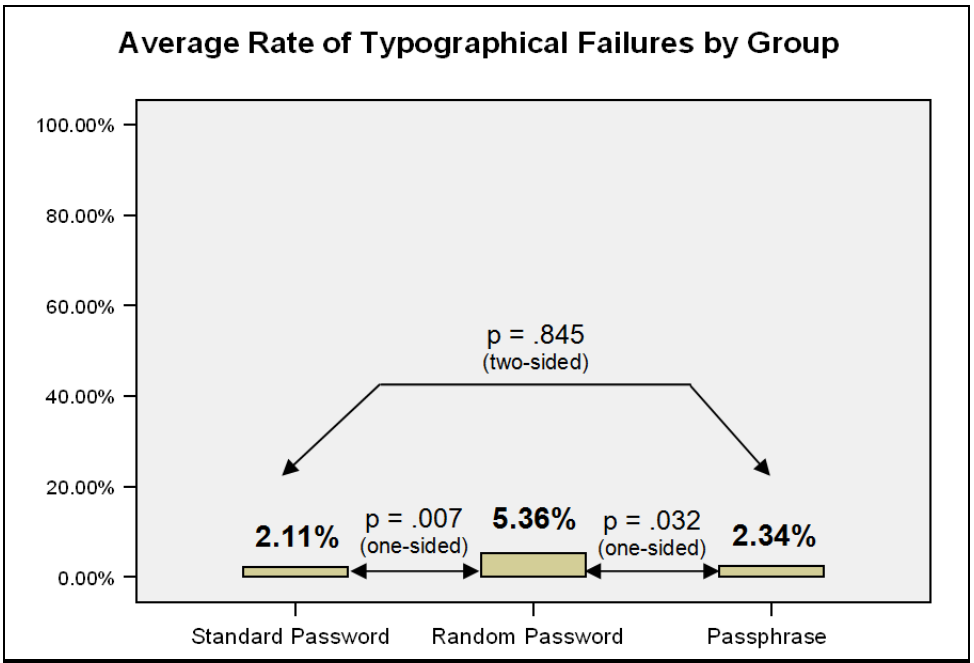

Figure 4: Average Rate of Typographical Errors by Group

Note: This graph and corresponding ANOVA does not include passphrase users who did not conform to WPM

H2b predicts that the degree to which passphrases are WPM-consistent affects the likelihood of making a typographical error. Sample size constraints prevented us from creating two passphrase groups in this study. Therefore, to test H2b, we obtained the data from the passphrase group in Keith et al.'s (2007) experiment. Their passphrase group received similar instructions with the exception that their subjects were required to include at least one non-alphabetic character, one upper-case

letter, and one lower-case letter. The two passphrase groups are of the same size (n = 17) and created passphrases of similar length (means of 18.06 characters in that study vs. 18.12 characters in this study, t = 0.065, p = 0.949).

To test H2b, we grouped the 34 individual passphrases created in both studies into those that conformed to WPM (e.g., use of spaces, no alternative characters, etc.) and those that did not. This classification resulted in 18 participants in the WPM group (13 from the current study and five from Keith et al. (2007)) and 16 in the non-WPM group (four from the current study and 12 from Keith et al. (2007)). The same judges who coded the cause of login errors for participants in this study followed the same procedures to classify the cause of login errors in the data from Keith et al. (2007). One-tailed t-tests reveal that the typing error rate was lower for passphrases that were consistent with WPM (3.19% vs. 8.29%, p = 0.019). Thus, H2b is supported.[9] Figure 4 also shows that H2c is supported: users of system-generated random passwords made significantly more typographical errors than did users of self-generated passwords (5.36 percent vs. 2.11 percent, p =.007).

Collectively, the results of testing Hypotheses H2a-c indicate that users of WPM-consistent credentials made fewer typographical errors than did users of credentials that were not WPM-consistent. To further examine this issue, we performed a regression analysis of the rate of typographical errors as a function of the types of characters included in the login credential. The analysis contained in Appendix D shows that the only factor significantly associated with the rate of typographical errors was the inclusion of non-alphanumeric characters (other than spaces). This provides additional evidence in support of the WPM effect. It also explains the results of testing Hypotheses H2a-c: as shown in Table 2, system-generated random passwords included an average of two such non-alphanumeric characters, but self-generated passwords (passphrases) seldom (never) used any such characters.

4.3. User Perceptions

At the end of the project, participants completed the perceptions survey (shown in Appendix A) during the final class period. Although the items were based on validated instruments, they were adjusted for the specific context of passphrase usage. The three items measuring ease-of-use formed a reliable construct (Cronbach's alpha = 0.90). However, the item concerning credential "effectiveness" was removed in order to achieve a reliable, four-item measure of usefulness (Cronbach's alpha = 0.92). Figure 5 displays the differences between groups on all three perception measures: ease-of-use, usefulness, and intent to adopt.

Hypothesis H3 predicts that overall login failure rates will affect user perceptions of usefulness and ease-of-use. To test these hypotheses, we formulated a partial least squares (PLS) model using SmartPLS (Ringle et al., 2005). Figure 6 displays the PLS model, which confirms H3 that the rate of overall login failures is inversely related to both ease-of-use ($\beta = -0.29$, $p < 0.01$) and usefulness ($\beta = -0.14$, $p < 0.05$).

Hypothesis H4 predicts that user perceptions will affect their intent to voluntarily adopt a particular type of login credential for use on other systems. In addition, prior research on technology acceptance has demonstrated that perceived usefulness also mediates the relationship between perceived ease-of-use and the intention to adopt (Venkatesh and Davis, 1996). The PLS model confirmed that perceived ease-of-use is a significant indicator of perceived usefulness ($\beta = 0.77$, $p < 0.001$) and that perceived usefulness is positively related to a participant's intention to adopt his or her login credential ($\beta = 0.58$, $p < 0.001$). However, the coefficient for the relationship between perceived ease-of-use and intention to adopt was not significant. This result is not surprising, because prior research has often demonstrated that usefulness is the stronger predictor of a user's intention to adopt because of the mediating effect of usefulness[10] (Taylor and Todd, 1995; Venkatesh and Davis, 1996; 2000).

[9] There was no difference in the rates of memory errors associated with the two types of passphrases (13.44 percent for the WPM vs. 10.10 percent for the non-WPM groups, p = 0.547). This is consistent with arguments that any amount of meaningfully related information can be represented as a single chunk (Simon, 1974).

[10] However, it is possible that this study simply did not have a large enough sample size to detect the small effect of perceived ease of use on intent to adopt.

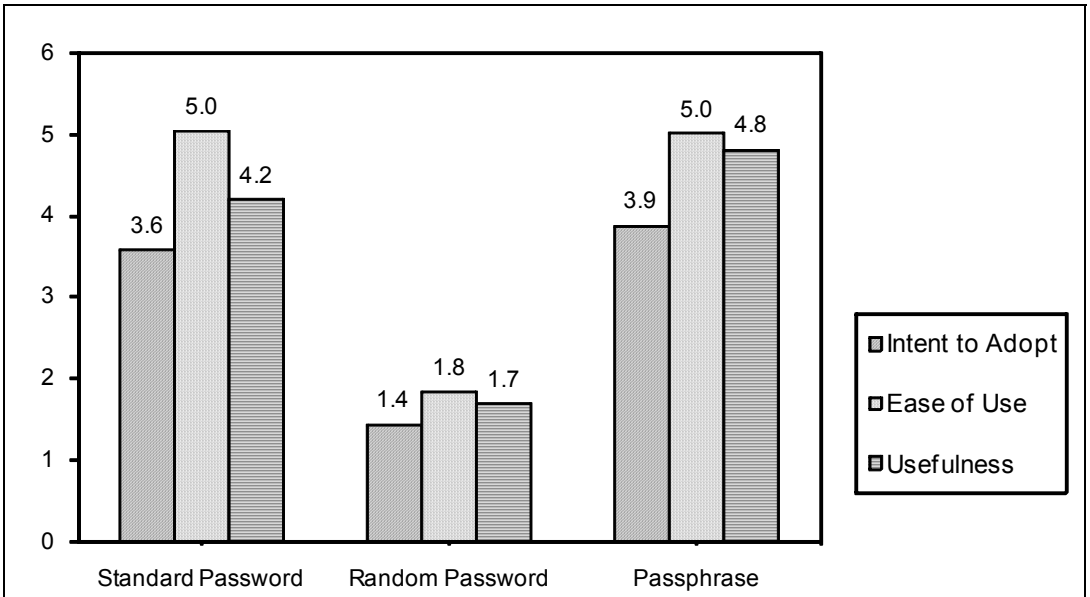

Figure 5: Average User Perceptions by Group

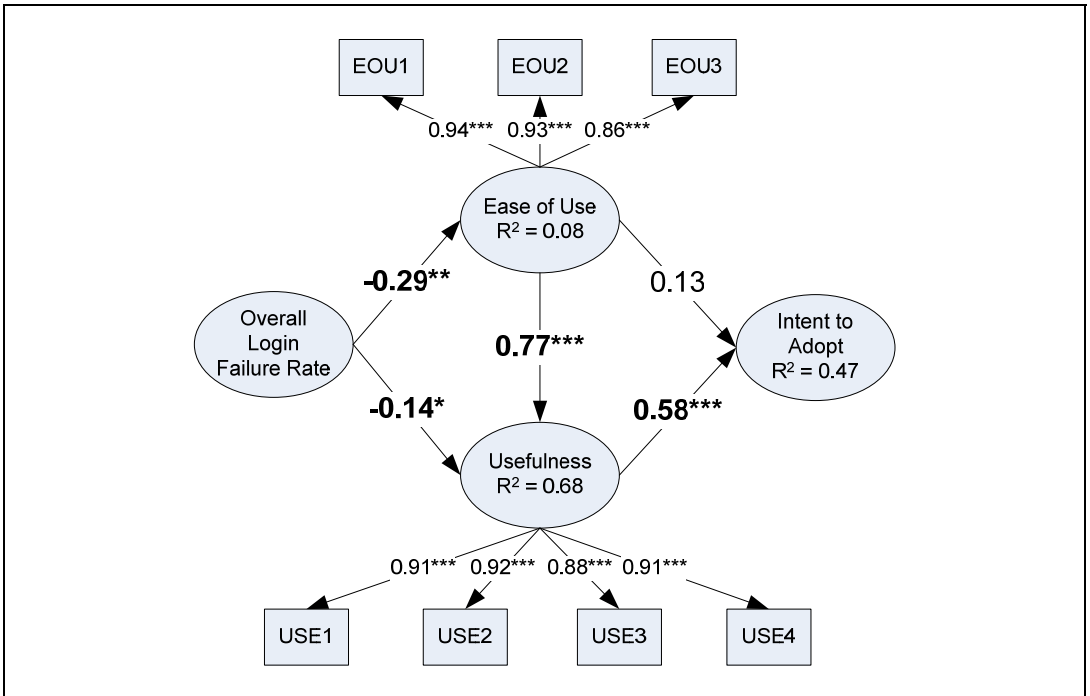

Figure 6: PLS Model Testing H3 and H4

5. Discussion

Table 5 summarizes our findings. As predicted, passphrases are easier to remember (i.e., resulted in fewer login failures due to memory errors) than either user-generated standard passwords or system-generated random passwords. Contrary to prior research (cf Yan et al., 2004), we did not find that user-generated passwords were easier to remember than system-generated random passwords, but they did result in fewer login failures due to typographical errors. The hypotheses concerning the

effect of WPM on typographical errors were mostly supported. Users of system-generated random passwords, which contained not only alphanumeric but also other special characters, experienced more login failures due to typographic errors than did users of standard passwords or users of WPM-consistent passphrases. In addition, users of WPM-consistent passphrases experienced fewer login failures due to typographical errors than did users of passphrases that were not WPM-consistent. There was, however, no difference in frequency of typographical errors between passphrases and standard passwords. As anticipated, user perceptions were more favorable when they did not encounter problems logging in. The relative level of participants' perceptions was directly related to their overall login failure rates, with passphrase users having the most favorable perceptions and random password users reporting the least favorable perceptions. Finally, user perceptions about usefulness were positively related to their expressed intention to voluntarily adopt that type of credential to authenticate themselves to other systems.

Table 5. Summary of Research Findings

Behavioral Effects	
H1a: Users of passphrases will experience fewer login failures due to memory errors than will users of either self-generated passwords or system-generated random passwords.	Supported
H1b: Users of self-generated passwords will experience fewer login failures due to memory errors than will users of system-generated random passwords.	Not Supported
H2a: Users who create passphrases that conform to word processing mode (WPM) will experience fewer typing errors than will users of either system- or self-generated passwords.	Partially Supported
H2b: Users who create passphrases that conform to WPM will experience fewer typing errors than will users who create passphrases that are not consistent with WPM.	Supported
H2c: Users of system-generated random passwords will experience more typing errors than will users who generate their own passwords.	Supported
Psychological Effects	
H3: Overall login failure rates will be inversely related to perceptions about the ease-of-use and usefulness of login credentials.	Supported
H4: User perceptions about the ease-of-use and usefulness of a login credential increase their intentions to voluntarily adopt that type of credential for use in authenticating to other systems.	Supported

5.1. Limitations

Before discussing the implications of our findings, it is important to consider the limitations of this study. First, we conducted a controlled experiment with college students as participants. Controlled experiments increase internal validity, but raise questions about external validity. Thus, it is possible that our participants' behavior may not generalize to that of employees in a work setting. One difference concerns frequency of logins. Employees must log in at least once each day, and sometimes more often (e.g., if password-protected screen savers are used or if policies require logging out prior to extended breaks). In contrast, participants in our study only needed to log in periodically, perhaps as infrequently as once per week. However, participants in our study were required to use their authentication credentials in a meaningful setting to perform a task for which there were non-trivial rewards (i.e., they could only complete the assigned homework by successfully logging in to the course website). Moreover, we measured *actual* login behavior, rather than tested responses to hypothetical situations.

A second limitation is that the length of our experiment was too short to study the effects of mandatory password changes across different types of authentication credentials. Therefore, our results may only be representative of the *initial* use of passphrases as an authentication credential. However, this study did utilize a long enough time period to reliably measure whether users can remember and correctly enter long passphrases, and hence it represents a significant step forward from the short-

term and cross-sectional tests used in most prior password research. Nevertheless, future research is needed to determine how difficult it is to generate new passphrases, and how a mandatory requirement to periodically change passphrases affects subsequent login failure rates and user perceptions.

5.2. Implications

The findings of this study have a number of important implications for practice and research.

Practice

In recent years security practitioners have begun advocating the use of passphrases over passwords. One argument in favor of passphrases is that they are easier to remember (cf Johansson and Riley, 2005, p. 338). Our finding that passphrase users experienced fewer login failures due to memory errors supports that argument. This suggests that switching to passphrases may improve security by reducing users' inclination to write them down, thus ensuring that the credential remains "something they know." Our finding that passphrase users made significantly fewer typographical errors than did users of system-generated random passwords also supports practitioners' arguments that it should be easier to correctly type passphrases than complex passwords that include special non-alphanumeric characters other than spaces (cf Skoudis and Liston, 2006, p. 402).

Our results also suggest that passphrase design is important. Specifically, users of WPM-consistent passphrases made fewer typographical errors than did users who created passphrases that were not consistent with WPM. This finding suggests that IT managers may wish to strongly encourage users to create passphrases that are WPM-consistent so as to help reduce login failures due to typographical errors. The ease of typing WPM-consistent passphrases has additional implications for IT managers who desire to increase the resistance of passphrases to brute-force attacks by including numbers and special characters. Our findings suggest that this intent will be easier to achieve and more successful if users choose symbols normally used in word processing (e.g., commas, periods, quotes, exclamation points, question marks, etc.). In addition, numbers are more likely to be correctly entered if they represent dates, quantities, or ages. Thus, our results suggest that it is better to use passphrases verbatim than to create a password consisting of the first letter of each word in the phrase because, in most cases, the resulting password will not conform to standard spelling rules.

A potential problem with adopting passphrase policies is that users may tend to generate passphrases consisting of only a small number of words. If we accept conservative estimates that the average person's vocabulary consists of approximately 3,000 words, then the potential search space for passphrases that consist of four words is actually smaller and, therefore, easier to "guess" through brute-force enumeration, than the search space for eight-character alphanumeric, case-sensitive passwords. On the other hand, five-word passphrases drawn from a 3,000 word vocabulary have a search space that is larger than that for system-generated random passwords that contain not only alphanumeric but also special characters ($3000^5 > 95^8$), and an eight-word passphrase constructed from such a vocabulary has a search space comparable to that of a 14-character system-generated random password ($3000^8 \sim 95^{14}$). Thus, IT managers contemplating the use of passphrases may wish to establish a minimum requirement that passphrases consist of at least five words.

Research

Our results suggest a number of promising avenues for further research. One important topic concerns identifying the reasons why passphrases are easier to remember than user-generated passwords. The concepts of chunking and phonological similarity provide competing explanations of this phenomenon. On the one hand, a passphrase, regardless of length, may be easy to memorize because it can be represented as one chunk. In contrast, a user-generated password that conforms to complexity requirements may be represented as multiple chunks: one for the base word, and one for each substitution of a number or special character for a letter. Thus, the password *$3cur!ty* may consist of four chunks (i.e., security, $, 3, and !). On the other hand, psychology research has found that recall accuracy is inversely related to the degree to which the target item is phonologically similar to other items stored in memory (Baddeley, 1966, Conrad, 1964; Lian et al., 2001). Thus, our finding that passphrases are easier to remember may merely reflect their novelty and may disappear over

time as they become more widely used. Only future research can disentangle these alternative explanations.

We did not find a significant difference in typing errors between passphrases and user-generated passwords, even though the latter were not WPM-consistent when they included at least one non-alphabetic character. However, users of self-generated passwords rated their passwords as being more similar to the credentials they used on other systems than did passphrase users (4.50 vs. 6.63 on a 7-point Likert scale, p=0.002). This means that users of self-generated passwords in our study had prior practice in typing passwords that were not WPM-consistent. Practice improves the performance of cognitive, perceptual, and motor tasks, assuming that individuals are motivated and receive immediate feedback (Ericsson et al., 1993). Users in our study were motivated to successfully log in as they were trying to access course materials, and they received immediate feedback concerning that endeavor. Thus, with practice, users may become proficient in typing short passwords that are not WPM-consistent. Nevertheless, while typing error rates in both this study and Keith et al. (2007) declined over time during the course of the experiment, the groups that initially experienced the highest typing error rates formed the most unfavorable perceptions. Thus, further research investigating the effects of password and passphrase composition on typing errors is warranted.

Our finding that WPM-consistent passphrases were less prone to typing errors than non-compliant passphrases has implications for the potential strength of passphrases. Current password cracking programs focus on manipulating individual characters in the credential, and those character manipulation techniques are not optimized for cracking passphrases. Information security like national defense, however, is an ongoing process of continual innovation both offensively and defensively. As passphrases become more widely used, brute-force methods for attacking them are also likely to evolve to focus on manipulating words rather than individual characters, thus significantly weakening the potential strength of passphrases. Obvious countermeasures include deliberately misspelling words in the phrase or replacing some individual letters with numbers or other special characters. Our results suggest, however, that either strategy is likely to increase login failures due to typing errors. Therefore, additional research identifying ways to increase passphrase strength without sacrificing usability is needed.

6. Conclusion

Passphrases, like all authentication credentials, are artifacts. It is important for IS researchers to carefully examine and rigorously demonstrate the utility of such artifacts (Hevner et al., 2004). This paper does so by showing how technical features (e.g., password composition requirements) and reactive behaviors (e.g., memory and typing errors) are so intertwined that focusing on one side of the issue but ignoring the other leads to incomplete, or even erroneous, conclusions. Our results suggest that passphrases do enhance the usability of including "something you know" as an authentication credential. Nevertheless, additional questions remain to be answered by future research.

Acknowledgements

The authors thank three anonymous reviewers, the Senior Editor Dr. Izak Benbasat, and participants at the University of Arizona workshop for their constructive comments on the paper. Any errors that remain are the sole responsibility of the authors.

References

Adams, D. A., Nelson, R. R., and Todd, P. A. "Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication," *MIS Quarterly* (16:2), 1992, pp. 227-247

Anderson, J. R. *Cognitive Psychology and its Implications 6th Edition*, New York, NY, Worth Publishers, 2005

Baddeley, A. D. "The Influence of Acoustic and Semantic Similarity on Long-Term Memory for Word Sequences," *Quarterly Journal of Experimental Psychology* (18:4), 1966, pp. 302-309

Brown, A. S., Bracken, E., Zoccoli, S., and Douglas, K. "Generating and Remembering Passwords,"

Applied Cognitive Psychology (18:6), 2004, pp. 641-651
Burnett, M. *Perfect Passwords: Selection, Protection, and Authentication*, ebrary.com, Syngress Publishing, 2005
Cameron, K. A., Haarmann, H. J., Grafman, J., and Ruchkin, D. S. "Long-Term Memory is the Representational Basis for Semantic Verbal Short-Term Memory," *Psychophysiology* (42:6), 2005, pp. 643-653
Center for Internet Security, *Windows XP Professional Operating System Legacy, Enterprise, and Specialized Security Benchmark Consensus Baseline Security Settings, Version 1.3*, The Center for Internet Security, www.cisecurity.org, 2004
Cohen, J. D., McClelland, J. L., and Dunbar, K. "On the Automatic Control of Processes: A Parallel Distributed Processing Account of the Stroop Effect," *Psychological Review* (97:3), 1990, pp. 332-361
Conrad, R. "Acoustic Confusions in Immediate Memory," *British Journal of Psychology* (55), 1964, pp. 75-84
Cowan, N. "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity," *Behavioral and Brain Sciences* (24:1), 2001, pp. 87-185
Crystal, D., *Cambridge Encyclopedia of the English Language*, Cambridge, MA, Cambridge University Press, 2003
Davis, F. "Perceived Usefulness, Perceived Ease of User, and User Acceptance of Technology," *MIS Quarterly* (13:3), 1989, pp. 319-339
DeSanctis, G. and Poole, M. S. "Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory," *Organization Science* (5:2), 1994, pp. 121-147
Doumont, J. "Magical Numbers: The Seven-Plus-or-Minus-Two Myth," *IEEE Transactions on Professional Communication* (45:2), 2002, pp. 123-127
Driscoll, M. P. *Psychology of Learning for Instruction 3rd Edition*, Boston, MA, Pearson, 2005
Ericsson, K. A., Krampe, R. T., and Tesch-Romer, C. "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review* (100:3), 1993, pp. 363-406
Federal Information Processing Standards (FIPS) Publication 112, "Password Usage," 1985 (http://www.itl.nist.gov/fipspubs/fip112.htm) last accessed December 8, 2008
Gray, P. H. and Durcikova, A. "The role of knowledge repositories in technical support environments: Speed versus learning in user performance," *Journal of Management Information Systems* (22:3), 2005-6, 159-190
Hevner, A. R., March, S. T., Park, J., and Ram, S. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), 2004, pp. 75-105
Huston, B. "Review of BioPassword Internet Edition," *Information Security*, 2006, pp. 65
Ives, B., Olson, M. H., and Baroudi, J. J. "The Measurement of User Information Satisfaction," *Communications of the ACM* (26:10), 1983, pp. 785-793
Ives, B., Walsh, K. R., and Schneider, H. "The Domino Effect of Password Reuse," *Communications of the ACM* (47:4), 2004, pp. 75-78
Jacoby, L. L. "On interpreting the effects of repetition: Solving a problem versus remembering a solution," *Journal of Verbal Learning and Verbal Behavior* (17:6), 1978, pp. 649-667
Johansson, J. "The Great Debates: Pass Phrases vs. Passwords. Part 1 of 3," Microsoft TechNet, (October 1, 2004a), retrieved 27 March 2007, http://www.microsoft.com/technet/community/columns/secmgmt/sm1004.mspx
Johansson, J. "The Great Debates: Pass Phrases vs. Passwords. Part 2 of 3," Microsoft TechNet, (November 1, 2004b) retrieved 27 March 2007, http://www.microsoft.com/technet/community/columns/secmgmt/sm1104.mspx
Johansson, J. "The Great Debates: Pass Phrases vs. Passwords. Part 3 of 3," Microsoft TechNet, (December 1, 2004c) retrieved 27 March 2007, http://www.microsoft.com/technet/community/columns/secmgmt/sm1204.mspx
Johansson, J. M. and Riley, S. *Protect Your Windows Network: From Perimeter to Data*, Upper Saddle River, New Jersey, Addison-Wesley, 2005
John, B. E. "TYPIST: A Theory of Performance in Skilled Typing," *Human-Computer Interaction* (11:4), 1996, pp. 321-355
Johns, E. E., and Swanson, L. G. "The Generation Effect with Nonwords," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (14:1), 1988, pp. 180-190

Keith, M., Shao, B., and Steinbart, P. "The Usability of Passphrases for Authentication: An Empirical Field Study," *International Journal of Human-Computer Studies* (65:1), 2007, pp. 17-28

Levenshtein, V. I. "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics-Doklady* (10), 1966, pp. 707-710

Lian, A., Karlsen, P. J., and Winsvold, B. "A Re-Evaluation of the Phonological Similarity Effect in Adults' Short-Term Memory of Words and Nonwords," *Memory* (9:4-6), 2001, pp. 281-299

Logan, F. A. "Errors in Copy Typewriting," *Journal of Experimental Psychology: Human Perception and Performance* (25:6), 1999, pp. 1760-1773

Mahmood, M. A., Burn, J. M., Gemoets, L. A., and Jaquez, C. "Variables Affecting Information Technology End-User Satisfaction: A Meta-Analysis of the Empirical Literature," *International Journal of Human-Computer Studies* (52:4), 2000, pp. 751-771

McKeen, J. D., Guimaraes, T., and Wetherbe, J. C. "The Relationship between User Participation and User Satisfaction: An Investigation of Four Contingency Factors," *MIS Quarterly* (18:4), 1994, pp. 427-451

Miller, G. A. "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," *Psychological Review* (63:2), 1956, pp. 81-97

Morris, R. and Thompson, K. "Password Security: A Case History," *Communications of the ACM* (22:11), 1979, pp. 594-597

Narayanan, A. and Shmatikov, V. "Fast Dictionary Attacks on Passwords Using Time-Space Tradeoff," *Proceedings of the 12th ACM Conference on Computer and Communications Security CCS '05,* November 7-11, 2005, Alexandria, Virginia, pp. 364-372

Orlikowsky, W. J. "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations," *Organization Science* (11:4), 2000, pp. 404-428

Pond, R., Podd, J., Bunnell, J., and Henderson, R. "Word Association Computer Passwords: The Effect of Formulation Techniques on Recall and Guessing Rates," *Computers & Security* (19:7), 2000, pp. 645-656

Porter, S. N. "A Password Extension for Improved Human Factors," *Computers & Security* (1:1), 1982, pp. 54-56

Rieger, M. "Automatic Keypress Activation in Skilled Typing," *Journal of Experimental Psychology: Human Perception and Performance* (30:3), 2004, pp. 555-565

Ringle, C. M., Wende, S., and Will, A, *SmartPLS*, Hamburg, Germany: University of Hanburg, 2005.

Ruffo, G. and Bergadeno, R. "EnFilter: A Password Enforcement and Filter Tool Based on Pattern Recognition Techniques," *Lecture Notes in Computer Science* (3617), 2005, pp. 75-82

Rumelhart, D. E. and Norman, D. A. "Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance," *Cognitive Science* (6:1), 1982, pp. 1-36

Salthouse, T. A. "Perceptual, Cognitive, and Motoric Aspects of Transcription Typing," *Psychological Bulletin* (99:3), 1986, pp. 303-319

Simon, H. "How Big is a Chunk?" *Science* (183: 4124), 1974, pp. 482-488

Skoudis, E. and Liston, T. *Counter Hack Reloaded, 2nd Edition*, Upper Saddle River, New York, Prentice-Hall, 2006

Slamecka, N. J., and Graf, P. "The generation effect: Delineation of a phenomenon," *Journal of Experimental Psychology: Human Learning and Memory* (4:6), 1978, pp. 592-604

Taylor, S. and Todd, P. A., "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research* (6:2), 1995, pp. 144-176

Todd, P., and Benbasat, I. "The Influence of Decision Aids on Choice Strategies: An Experimental Analysis of the Role of Cognitive Effort," *Organizational Behavior and Human Decision Processes* (60:1), 1994, pp. 36-74

Todd, P., and Benbasat, I. "Evaluating the Impact of DSS, Cognitive Effort, and Incentives on Strategy Selection," *Information Systems Research* (10:4), 1999, pp. 356-374

Todd, P., and Benbasat, I. "Inducing Compensatory Information Processing Through Decision Aids That Facilitate Effort Reduction: An Experimental Assessment," *Journal of Behavioral Decision Making* (13:1), 2000, pp. 91-106

Venkatesh, V. and Davis, F. D., "A Model of the Antecedents of Perceived Ease of Use: Development and Test," *Decision Sciences* (27:3), 1996, pp. 451-481

Venkatesh, V. and Davis, F. D., "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46:2), 2000, pp. 186-204.

Whitman, M. E. "Enemy at the Gate: Threats to Information Security," *Communications of the ACM* (46:8), 2003, pp. 91-95

Wiedenbeck, S., Waters, J., Birget, J., Brodskiy, A., and Memon, N. "PassPoints: Design and Longitudinal Evaluation of a Graphical Password System," *International Journal of Human-Computer Studies* (63:1-2), 2005, pp. 102-127

Wixom, B. H. and Todd, P. A. "A Theoretical Integration of User Satisfaction and Technology Acceptance," *Information Systems Research* (16:1), 2005, pp. 85-102

Wren, S. "Vocabulary," (http://www.balancedreading.com/vocabulary.html) last accessed on December 9, 2008

Yan, J., Blackwell, A., Anderson, R., and Grant, A. "Password Memorability and Security: Empirical Results," *IEEE Security & Privacy* (2:5), 2004, pp. 25-31

Zviran, M. and Haga, J. W. "Cognitive Passwords: The Key to Easy Access Control," *Computers & Security* (9:8), 1990, pp. 723-736

Zviran, M. and Haga, J. W. "A comparison of password techniques for multilevel authentication mechanisms," *The Computer Journal* (36:3), 1993, pp. 227-237

Zviran, M. and Haga, J. W. "Password Security: An Empirical Study," *Journal of Management Information Systems* (15:4), 1999, pp. 161-185

Appendix A: Survey Items

Likert Scale 1 = Strongly Disagree to 7 = Strongly Agree

Similarity	1. The password or passphrase I used for the [experiment][11] is quite similar to others I have used in the past or are currently using.
Ease of Use	2. It would be easy for me to remember passwords like the one I used for [the experiment] as a login credential on other restricted sites that I regularly access.
	3. It would be easy for me to correctly enter passwords like the one I created for [the experiment] as a login credential on other restricted sites that I regularly access.
	4. It would be easy for me to become skillful at generating passwords in the format I used for [the experiment] (i.e. length, character set, and composition) as a login credential on other restricted sites that I regularly access.
Usefulness	5. I can enter passwords during routine logins more quickly by using the format of my [experiment] password (i.e. length and character set, and composition).
	6. The password I created for [the experiment] would be very difficult for an unauthorized user to crack or discover.
	7. Using a password format like the one I used for [the experiment] would help reduce my need for password reminders or changes.
	8. I would not need to write down or store my passwords as much if I always used a password format (i.e. length and character set, and composition) like the one I used for [the experiment]
	9. I could remember more passwords at once if I always used a password format (i.e. length, character set, composition) like the one I used for [the experiment]
Intent to Adopt	10. I intend to use the same password format I used for [the experiment] (i.e. length, character set, and composition) in the future to create passwords to login to other systems.

[11] In the survey instrument the course name appeared in place of the phrase [the experiment]

Appendix B: MANOVA Results on the Effect of Credential Type on Memory and Typographical Error Rates

Multivariate Tests

Effect		Value	F	Hyp. df	Error df	Sig.	Partial Eta Squared	Power
Intercept	Pillai's Trace	.683	51.605	2.000	48.000	.000	.683	1.000
	Wilks' Lambda	.317	51.605	2.000	48.000	.000	.683	1.000
	Hotelling's Trace	2.150	51.605	2.000	48.000	.000	.683	1.000
	Roy's Largest Root	2.150	51.605	2.000	48.000	.000	.683	1.000
CredType	Pillai's Trace	.282	4.026	4.000	98.000	.005	.141	.899
	Wilks' Lambda	.732	4.059	4.000	96.000	.004	.145	.901
	Hotelling's Trace	.348	4.087	4.000	94.000	.004	.148	.903
	Roy's Largest Root	.280	6.855	2.000	49.000	.002	.219	.905

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Power
Corrected Model	Rate of Memory Errors	.250	2	.125	5.868	.005	.193	.854
	Rate of Typo Errors	.010	2	.005	2.756	.073	.101	.519
Intercept	Rate of Memory Errors	1.490	1	1.490	70.062	.000	.588	1.000
	Rate of Typo Errors	.066	1	.066	37.791	.000	.435	1.000
Credential Type	Rate of Memory Errors	.250	2	.125	5.868	.005	.193	.854
	Rate of Typo Errors	.010	2	.005	2.756	.073	.101	.519
Error	Rate of Memory Errors	1.042	49	.021				
	Rate of Typo Errors	.085	49	.002				
Total	Rate of Memory Errors	2.787	52					
	Rate of Typo Errors	.159	52					
Corrected Total	Rate of Memory Errors	1.292	51					
	Rate of Typo Errors	.095	51					

Appendix C: ANOVAs Comparing Credential Types

One-way ANOVA testing the effects of credential type on memory errors

	Sum of squares	df	Mean square	*F*	Significance	Partial Eta^2	Power (α = .05)
Between Groups	0.250	2	0.125	5.868	0.005	0.193	0.854
Within Groups	1.042	49	0.021				
Total	1.292	51					

Cell means for memory errors by credential type

Credential Type	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Standard	.182	.034	.113	.251
Random	.248	.035	.177	.319
Passphrase	.078	.035	.007	.149

One-way ANOVA testing the effects of credential type on typographical errors

	Sum of squares	df	Mean square	*F*	Significance	Partial Eta^2	Power (α = .05)
Between Groups	0.011	2	0.005	3.816	0.029	0.145	0.664
Within Groups	0.065	45	0.001				
Total	0.076	47					

Cell means for typographical errors by credential type

Credential Type	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Standard	.021	.010	.001	.041
Random	.054	.010	.033	.074
Passphrase	.032	.010	.012	.052

Appendix D: Regression Models Predicting Typographical Errors

Linear regression was used to further understand the effects of password length and specific character types on typographical errors. The results indicate that password and passphrase *length* did not explain any of the variance in typo errors in this study. Neither did the amount of uppercase letters or numbers. However, the inclusion of non-alphanumeric characters (e.g., #, @, and &) did have a significant effect on the rate of typo errors ($p < .001$). None of the passphrases used these characters, but they did appear in all but one of the system-generated passwords and also in two of the user-created passwords. The resulting credential is not WPM-consistent, which may account for the fact that the random group had the highest rate of typing errors, as predicted by the rationale underlying H2a and H2b.

Regression Model Estimates for Typographical Errors	
Independent Variable	Model
Password Length	0.001
	(0.764)
Rate of 'Other Characters'	0.168***
	(3.419)
Rate of Uppercase Letters	-0.026
	(-0.580)
Rate of Numbers	-0.001
	(-0.033)
N	52
R-squared	0.256
F-statistic	4.043

Notes: t-statistics are in parentheses below coefficients.* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

About the Authors

Mark Keith is a Clinical Assistant Professor of Information Systems in the W. P. Carey School of Business at Arizona State University. His research interests include organizational impacts of IS, service-oriented systems, software project coordination, IS security and usability, and information retrieval. His research has appeared in such journals as *INFORMS Decision Analysis*, *International Journal of Human-Computer Studies*, and *Journal of the Association for Information Systems*.

Benjamin B. M. Shao is an Associate Professor of Information Systems in the W. P. Carey School of Business at Arizona State University. His current research interests are in IT impacts, IS security, distributed systems, and software project management. His research has appeared in such journals as *Communications of the ACM*, *Decision Support Systems*, *European Journal of Operational Research*, *IEEE Transactions*, *Information & Management*, and *Journal of the Association for Information Systems*. He also serves on the editorial board for *Journal of the Association for Information Systems*, *Information Technology & Management*, and an *MIS Quarterly* Special Issue.

Paul John Steinbart is a Professor of Information Systems in the W. P. Carey School of Business at Arizona State University. His research interests include information security, privacy, and IT governance. His research has appeared in numerous journals, including *MIS Quarterly*, *The Accounting Review*, and *Decision Sciences*. He is currently the editor of the *Journal of Information Systems* (published by the Information Systems section of the American Accounting Association) and is also the co-author of the leading Accounting Information Systems textbook.

Journal of the Association for Information Systems

ISSN: 1536-9323