

Quality Metrics for PDA-based M-Learning Information Systems

Ruti Gafni

The Open University of Israel, Raanana, Israel

rutiga@openu.ac.il

Abstract

PDA-based information systems are suitable for m-learning, enabling ubiquitous and idle-time utilization. However, the use of small devices and mobile systems requires special quality attention in order to get the desired benefits. This article examines the quality problems resulting from the characteristics of these systems, proposes some metrics to measure their quality and some practical implications for the designers and developers of PDA-based systems, including technical, implementation and content issues.

Keywords: personal digital assistant (PDA), m-learning, information systems usability, mobile, quality.

Introduction

Mobile learning (m-learning) enables learning independently of place and time, ubiquitous, through wireless networks and mobile devices, such as personal digital assistants (PDA), cellular phones, smart-phones, and mp-3 devices (Alexander, 2004). In this article the focus is on PDA-based m-learning information systems. Using PDA, the system can be activated online, while connected to the network, or offline, in case the information was previously loaded to the device memory and then used when disconnected from the network.

During the last years, the development of mobile information systems increased, since technology, devices, and infrastructures matured, including the emergence of third generation wireless networks (UMTS - Universal Mobile Telecommunications System, CDMA2000 1x - Code division multiple access, and GPRS - General Packet Radio Service). The use of mobile systems will further grow (Lau, 2006; Lee, 2003). This has led to development of diverse kinds of systems, including m-learning, which can be used by a variety of users for different tasks. Among others, students traveling by bus or train can read course summaries, perform online quizzes or assessments, and send them to the relevant instructor; company workers can use these systems for specific training, using any free time, wherever workers are situated; travelers can avoid carrying tourist travel books when visiting old towns and receive the desired information using their mobile devices.

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Publisher@InformingScience.org to request redistribution permission.

M-learning systems can create benefits for organizations and users, e.g., better idle time utilization, productivity en-

Editor: Alex Koohang

An earlier, shorter version of this paper was presented at the Chais conference 2009, in Raanana, Israel, and included in Y. Eshet-Alkalai, A. Caspi, S. Eden, N. Geri, & Y. Yair (Eds.), *Proceedings of the Chais conference on instructional technologies research 2009: Learning in the technological era*. Raanana: The Open University of Israel. http://www.openu.ac.il/research_center_eng/conferences.html

hancement, flexibility of processes and procedures, etc., which together emphasize competitive strategy, lower operation costs, and improve processes (Malladi & Agrawal, 2002). However, these benefits can be nullified by using sub-quality information systems. In order to use these systems successfully, they need to be of high quality (Terho, 2002). Mobile-wireless information systems face new kinds of problems, different from traditional and internet systems, which have been largely researched (Calero, Ruiz & Piattini, 2004; Covella & Olsina, 2006; Kan, 2002; Pandian, 2004). These problems include devices with small screens, keyboards, and memory, narrow bands, lack of coverage and diversity of users and devices. M-learning systems are affected by these special problems; thus, they require an enhanced quality definition. The purpose of this study is to find metrics to quantify the quality of such information systems in order to enable the evaluation, comparison, and analysis of m-learning information systems quality.

Quality Issues in PDA-based M-learning Systems

Although most benefits of mobility, as mentioned before, can be regarded as being based on technical quality, either of hardware, software, or the carrier network, mobile-wireless information systems face some unique problems originating from the mobile devices and from the characteristics of the network. Because m-learning systems are part of the mobile-wireless information systems, they suffer from the same problems.

The mobile devices cause some of these problems:

- The devices have technical limitations such as small memories (the amount of information stored on it cannot be large), short battery life (the continuous period of work cannot be long), and limited calculation and computation capabilities.
- There is a wide variety of devices, possessing different characteristics, and the application must be adaptable to all of them (Brady, Conlan, & Wade, 2004).
- The devices' use is uncomfortable, because of their size: tiny screens (which restrict the amount of data displayed), low resolution (pictures and video-clips are not seen in their optimally), and small keyboards difficult to operate.
- Security problems can arise when lost, due to possible unauthorized access to sensitive data.

These problems intensify when the m-learning system is targeted for a wide audience, where the users are faceless and there is neither user training nor implementation, but they are significant also in closed organizational information systems (firms, schools, universities, etc.).

When using the m-learning system online, networks cause other problems, including:

- Limited bandwidth, which influences directly on video-clips' display. This problem diminishes over the time because of the network generations' evolution (3G and up).
- Inconsistent connection stability and transfer delays, which interfere with learning continuity.
- Moreover, when users operate the system during mobility, the connection point to the network can change and obstacles can interfere, causing temporary disconnections, interruptions, or disturbances.
- Varied standards and protocols, some with high overhead, decrease the performance level; others with geographical restrictions restrict system use pervasiveness.
- Issues of security, privacy and, confidentiality (Di Pietro & Mancini, 2003; Herzberg, 2003) arise, which include risks of eavesdropping and the need for user identification and authentication, especially when conducting quizzes and exams or delivering private information.

- High costs of operation, especially when the users are low budget as students.

Using PDA-based m-learning systems, instead of smaller devices like cell phones, can be beneficial (Wood, 2003) because the user can get quick and easy access to data, the devices have a Qwerty style keyboard or a stylus that allows the user to interact with the system by touching the screen, some have memo pads and voice recorders that allow the user to record notes, and data can be gathered in the device and further sent to the remote system. Still, PDA devices do not solve the problems listed above. Therefore, these problems demand precise considerations, both in trying to avoid them, as much as possible, by specific and targeted design of the system and in measuring system quality, which is significantly influenced by all these features.

This article deals with measuring quality of PDA-based m-learning information systems with off-line activation, which enables learning, and not the quality of the learning itself. It covers the system quality level, identified by Kim and Ong (2005) as one of the three success factors of m-learning from the learners' perspective.

Relevant metrics are defined and validated using a PDA-based experiment.

Information Systems Quality Measuring

In order to measure the quality of PDA-based m-learning systems, there is a need to understand the significance of information systems quality measuring. Measuring creates a quantitative description which allows behavior comprehension and enables selection of tools and techniques to control and improve processes, products, and resources. Information systems quality cannot be measured only by software faults absence; it must be broader, including characteristics to cover all aspects, life-cycle phases, and viewpoints. Several hierarchical models for quality definition were proposed (Boehm, Brown, Kaspar, Lipow, & McCleod, 1978; McCall, Richards, & Walters, 1977). The most recent is ISO/IEC 9126 standard (ISO/IEC9126-1, 2001), which decomposes quality into seven characteristics, further divided into sub-characteristics (Figure 1). This standard defines, for each sub-characteristic, internal metrics to be measured without having to operate the system and external metrics to be measured while testing or executing the system. Gafni (2008) proposed to extend the metrics of the standard to cope with the problems of the mobile-wireless information systems.

Functionality	Reliability	Usability	Efficiency	Maintainability	Portability	Quality in Use
Suitability	Maturity	Understandability	Time Behavior	Analyzability	Adaptability	Effectiveness
Accuracy	Fault Tolerance	Learnability	Resource Utilization	Changeability	Installability	Productivity
Interoperability	Recoverability	Operability	Efficiency Compliance	Stability	Co-existence	Safety
Security	Reliability Compliance	Attractiveness		Testability	Replaceability	Satisfaction
Functionality Compliance		Usability Compliance		Maintainability Compliance	Portability Compliance	

Figure 1 – ISO/IEC 9126 quality characteristics and sub-characteristics

Measuring Quality in PDA-based M-learning Systems

PDA-based m-learning systems must be measured on the basis of traditional systems metrics, e.g., ease of maintainability, minimum complexity, lack of faults, and mean time between failures, which are the fundamental metrics for all information systems. Furthermore, when used online, they must be measured according to internet systems metrics, e.g., no broken links or ease of navigation. In addition, they need to be measured with special targeted mobile-wireless metrics (Spriestersbach & Springer, 2004).

The problems mentioned above raise some quality issues which affect the ISO/IEC 9126 quality characteristics and sub-characteristics and must be measured using new defined metrics.

Functionality includes the suitability, accuracy, interoperability, and security sub-characteristics. Suitability means that the functionality of the application fits the needs to fulfill a certain task without overwhelming the user. This is important in m-learning systems, especially because of the problems caused by the use of small devices. For this reason the information for these systems must be delivered in small and concise amounts (Parsons & Ryu, 2006; Venkatesh & Ramesh, 2002). Security is affected in different ways, for example, loss or theft of the mobile device or exposure of the messages in the network to unauthorized access.

Reliability includes the maturity, fault tolerance, recoverability, and reliability compliance sub-characteristics. During mobility, network problems, hiding obstacles, and hopping between antennas may disturb and interrupt communication. Thus, the system must be fault tolerant to maintain a specified level of performance, and, if connection is suspended, the system needs to re-establish the communication and recover the data directly affected, supporting recoverability.

Usability includes the understandability, learnability, operability, attractiveness, and usability compliance sub-characteristics. Usability is one of the most important characteristics when targeting systems to wide audiences, such as students, that need to operate an intuitive system, without direct training and support. Mobile users may not be able to concentrate on the system use, so the application should not be complicated, and the input must be easy to insert, intuitive, and simplified by using location aware functions (Terrenghi, Kronen, & Valle, 2005). The operability sub-characteristic is affected mainly by the mobile device attributes: screen, keyboard, and numeric pad, which restrict output and input interaction possibilities. Output capabilities of devices are determined by their screen, which limits the amount of information for simultaneous display. Applications need to consider this limitation, for example, by distributing information across multiple pages or adapting the content of the application. But, on the other hand, the operation of leafing through multiple pages in a mobile device decreases the operability. The input restrictions are due to size constraints of mobile devices and due to keyboard capabilities. This must be taken into consideration by applications that need to limit data input to minimal required data, pre-defined options, and automatic filled-in fields.

Moreover, the noisy surroundings (when operating the system while on a street or public environments) may distract the user and cause input errors, inaccuracy, and slowness. The ability to reach the relevant data “any place any time” enlarges system attractiveness.

All these reasons show that m-learning applications should minimize and simplify the required interaction between the user and the system.

Efficiency includes time behavior and resource utilization sub-characteristics. The high costs of wireless network usage, which currently have an economic impact greater than wired networks, mean that applications must be focused and the usage time reduced, otherwise users will avoid expensive systems. Moreover, users are accustomed to rapid systems, so slowness is unacceptable, especially when users are “on the move.” Mobile devices include small memory and low processing resources, so applications must be aware of these restrictions and optimize resource

utilization. These contribute to application's suitability as well as time and resource based efficiency.

Maintainability includes analyzability, changeability, stability, and testability. This set of sub-characteristics reflects mainly the technical stakeholders' viewpoint, such as the developers and maintenance people (Hordijk & Wieringa, 2005). Thus, the mobile-wireless features do not have significant influence, additional to those of the traditional systems features, upon them, and learners are not affected by the level of quality of the maintainability characteristic. If the application to be used for m-learning is the same used for e-learning, the architecture must be adaptive and rely on reusability of components (Goh & Kinshuk, 2006) so the level of maintenance quality will not lessen.

Portability includes adaptability, installability, co-existence, and replaceability. Portability is very important in m-learning systems that are targeted to wide audiences because of the large diversity of devices held by users and the lack of control over these devices, e.g., configuration devices and applications run by them. The application must adapt itself to the device features, both during installation and operation, and according to the user's preferences (Goh & Kinshuk, 2006). The application installation on the device must be invisible to laypersons. The application operation in the mobile device needs to exist harmoniously, sharing common resources, with other applications installed in the same device, especially co-existing with the telephone facilities when used in mobile or smart phones.

Quality in use is the user's view of quality: "The capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use" (ISO/IEC 9126-4, 2004). Since mobile applications may be used while driving or walking, tasks which need the user attention, the application manipulation must be simple and intuitive, enforcing safety.

Kim and Ong (2005) examined the success factors of m-learning from the learners' perspective, identifying three main dimensions of user satisfaction: system quality, content quality, and service quality. The system quality deals with the technical level, which is covered by the above characteristics, especially reliability and usability. Content quality deals with the semantic level, meaning the information quality and output accuracy and reliability. Service quality includes the installation assistance, product knowledge of the body supplying the service, training, support, and online help.

Parsons and Ryu (2006) divided the quality aspects of m-learning systems to three major issues: technical aspects, content related aspects, and context aspects. They proposed two new design issues that affect quality of m-learning: (a) the user role and collaboration and (b) extending the quality in use, considering the context of use. They suggested three new metrics to measure the proposed issues: (1) metaphor, which bridges from abstract and elusive vision to a concrete and complex situation, (2) functions to support interactivity with other members and tutors, regardless of location, and (3) learning content, which highlights whether an m-learning system can provide content that is both optimized for mobile delivery and justifies delivery through the mobile channel.

Quality Metrics for PDA-based M-learning Systems

The definition of new metrics to quantify the quality of m-learning systems is based on Gafni's (2008) methodology, which enables the definition of objective metrics, and consists of several phases:

- (1) Detection of quality problems and risks that outcome from the architectures and protocols of mobile-wireless information systems (Asunmaa et al., 2002; Green, 2003; Huber, 2004; Tarasewich, 2003; Varshney & Vetter, 2002; Vaughan-Nichols, 2004).
- (2) Choice of ISO/IEC 9126 quality characteristics affected by mobility and wirelessness.
- (3) Identification of objects (clustered into four entities: device, application, architecture, and user), which were decomposed into measurable attributes (ISO/IEC 15939, 2002) and assigned a unit measure and a scale, according to its meaning (Kitchenham, Pfleeger, & Fenton 1995).
- (4) Definition of metrics, methods for the measuring process with specific formulas, (ISO/IEC 15939, 2002) to allow objective measurement of mobile-wireless information systems quality.
- (5) Metrics theoretical and empirical validation.

The metrics defined are suitable especially for text-based m-learning systems, but they can be used with any kind of system, such as systems including video, audio, and other. Metrics to quantify the quality of the usage of multimedia were not developed in this paper. These can be developed using the same methodology.

Table 1 maps the defined and evaluated metrics for measuring the quality characteristics of a PDA-based m-learning system according to the different kind of problems which they measure.

Following, some of these metrics are defined.

For each metric defined, the expected output value was determined. In order to receive a concept of the total quality of the information system, there is a need to summarize the metrics results. Therefore, the expected values were defined in the range between 0 and 1, and in addition, the meaning of a higher result was consistently defined as a better quality. In some of the metrics this is the natural behavior. In others, there was a need to perform a transformation of the result. One

possible transformation is to calculate $X = \frac{1}{X'}$. Sometimes this transformation causes the results

to be very small and concentrated in the lower band of the range. Another type of transformation can be done by defining a maximum acceptance value, which can be defined according to the specifications of the information system. In this case, the transformation will be in the form

$$\text{of } X = \frac{\text{max} - X'}{\text{max}}.$$

Table 1 - PDA-based m-learning system metrics mapped according to mobile-wireless problems

Source of problems:	Network					Device					Mobility				
Type of problem	Narrow band	Connection stability	Security	Diverse standards	High costs	Variety of devices	Low memory	Tiny screens	Limited computational abilities	Small keyboard	Loss and information security	User adapted information	Security and privacy	Disturbances and disconnections	Users trust
Response time to get information from cache	X		X		X		X		X					X	X
Device memory cleanup after transaction							X								
Display load								X							
Messages conciseness	X				X		X	X							
Ease of input entering										X					
Ease of output use	X	X	X		X									X	X
Ease of use - displays per task	X	X			X			X	X					X	X
Secure messages and information on device											X	X	X		X
Use of user profile	X				X	X						X			
Installations success						X									
Ease of installation						X									

Definition of “Response time to get information from cache” Metric

Purpose of the metric: Measures the response time to get information stored in the device’s memory. The efficiency will be greater when the response time is smaller.

Method of application: Operate same task number of times. First, the data will be received from server and stored in device’s memory. Next, the data will be retrieved locally. The time to get data as output is measured, starting from second iteration.

Metric type: External

Attributes measured:

InRsT_i - Input Response Time of message i

TrPNum - Number of Transactions performed during test

i – Serial number of message, from 1 to TrPNum

Computation:

$$X' = \frac{\sum_{i=1}^{TrPNum} InRsT_i}{TrPNum}$$

where $TrPNum \geq 1$ and $InRsT_i > 0$ for each i

Transformation:

$$X = \frac{\max - X'}{\max} \text{ or } X = \frac{1}{X'}$$

Interpretation of measured value: The higher the better and the response time lower.

Definition of “Device memory cleanup” Metric

Purpose of the metric: Measures the degree of cleaned-up memory after completing a task. The small resources of memory in the devices have to be carefully handled. Any “garbage” left after completing a task may decrease any subsequent work performance because less free memory is available.

Method of application: Check free memory size before and after performing task. Check number of times to see if size decreases.

Metric type: External

Attributes measured:

- MmET_i – free memory size at end of task i
- ApS – Application Size
- MmS – Memory Size
- TNum - Number of Tasks in System
- i – serial number of task, from 1 to TNum

Computation:

$$X = \frac{\sum_{i=1}^{TNum} MmET_i}{TNum (MmS - ApS)}$$

where $MmS - ApS > 0$

$$TNum \geq 1$$

Interpretation of measured value: The higher the better

Definition of “Display load” Metric

Purpose of the metric: Measures the burden degree of the displays. Because of the screen’s size, the understandability is lower when the display is overloaded.

Method of application: Calculate the ratio between display size (information displayed) and screen size for each display in system. The ratios are categorized and given a score, in inverse relationship to the display size, thus smaller displays categories receive greater scores. Then a weighted average is calculated.

Metric type: Internal

Attributes measured:

- ScrS – Device screen size
- PgS – Display size
- DisNum – Number of displays in the system
- i –serial number of display, from 1 to DisNum

Computation:

Definition of categories and scores:

$$X_i = \begin{cases} 1 & \text{if } 0 < \frac{PgS_i}{ScrS} \leq \frac{1}{4} \\ \frac{2}{3} & \text{if } \frac{1}{4} < \frac{PgS_i}{ScrS} \leq \frac{1}{2} \\ \frac{1}{3} & \text{if } \frac{1}{2} < \frac{PgS_i}{ScrS} \leq \frac{3}{4} \\ 0 & \text{if } \frac{3}{4} < \frac{PgS_i}{ScrS} \leq 1 \end{cases}$$

$$X = \frac{\sum_{i=1}^{DisNum} X_i}{DisNum}$$

where $DisNum \geq 1$, $ScrS \geq 1$ and $PgS_i \leq DisS$ for all i .

If the size of the display is greater than the size of device screen, the display will be partitioned according to the number of pages to be displayed, and the calculation will be performed for each one of these pages.

Interpretation of measured value: The higher the better

Definition of “Messages conciseness” Metric

Purpose of the metric: Measures the conciseness of operational and error text messages. This save place in memory and the messages are easy to read from screen.

Method of application: Calculate ratio between message size and maximum lines in screen. Distribute ratios to inversed weighted categories, and calculate weighted average.

Metric type: Internal

Attributes measured:

ActMNum - Number of Active Messages.

ErrMNum - Number of Error Messages.

OutL - Output Length (in lines) of message i .

ScrL - Number of Lines in Screen

i – serial number of message, from 1 to ActMNum+ ErrMNum

Computation:

Weight each message:

$$X_i = \begin{cases} 1 & \text{if } 0 < \frac{OutL_i}{ScrL} \leq \frac{1}{5} \\ \frac{4}{5} & \text{if } \frac{1}{5} < \frac{OutL_i}{ScrL} \leq \frac{2}{5} \\ \frac{3}{5} & \text{if } \frac{2}{5} < \frac{OutL_i}{ScrL} \leq \frac{3}{5} \\ \frac{2}{5} & \text{if } \frac{3}{5} < \frac{OutL_i}{ScrL} \leq \frac{4}{5} \\ \frac{1}{5} & \text{if } \frac{4}{5} < \frac{OutL_i}{ScrL} \leq \frac{5}{5} \end{cases}$$

where $ActMNum + ErrMNum \geq 1$, $ScrL \geq 1$ and $OutL_i > 0$ for each i

$$X = \frac{\sum_{i=1}^{ActMNum+ErrMNum} X_i}{ActMNum + ErrMNum}$$

Interpretation of measured value: The higher the better

Definition of “Ease of input entering” Metric

Purpose of the metric: Measures the ratio of easy to fill input fields, like usage of default values, list of values or self-completing fields.

Method of application: Count number of closed input fields, default value fields, and self-completing fields. Calculate ratio between counted fields and total of input fields.

Metric type: Internal

Attributes measured:

FdClsNum - Number of Close Fields

FdDefNum - Number of Fields with Default Values

FdFillNum - Number of Fields with automatic or self-completing Filling

FdNum – Total Number of input Fields

Computation:

$$X = \frac{FdClsNum + FdDefNum + FdFillNum}{FdNum}$$

where $FdNum \geq 1$

Interpretation of measured value: The higher the better. The input of data is simpler, and the risk of error in input is minimal.

Definition of “Ease of output use” Metric

Purpose of the metric: Measures the suitability between length of outputs and device screen size.

Method of application: Calculate average displays for message. Count number of outputs needing leafing by checking ratio between number of lines in output message and number of lines in screen.

Metric type: External

Attributes measured:

OutMNum - Number of Output Messages

OutNumL - Number of Lines in Output Message i

i – serial number of output message, from 1 to OutMNum

ScrL - Number of Lines in Screen

Computation:

Average for all messages:

$$X' = \frac{\sum_{i=1}^{OutMNum} \left[\frac{OutNumL_i}{ScrL} \right]}{OutMNum}$$

where $ScrL \geq 1$, $OutNumL_i > 0$ for each i and $OutMNum \geq 1$

Transformation:

$$X = \frac{\max - X'}{\max} \quad \text{or} \quad X = \frac{1}{X'}$$

Interpretation of measured value: The higher the better, because it needs fewer leafing through the screens and the operation is simpler.

Definition of “Ease of use - displays per task” Metric

Purpose of the metric: Measures the number of screens involved in one task. The effectiveness and usability is greater when the system is easier to operate.

Method of application: Calculate number of iterations between user and system till end of task.

Metric type: Internal

Attributes measured:

DisNum - Number of Displays

TNum - Number of Tasks in System

Computation:

$$X = \frac{DisNum}{TNum}$$

where $TNum \geq 1$, $DisNum \geq 1$ and $DisNum \geq TNum$

Transformation:

$$X = \frac{\max - X'}{\max} \quad \text{or} \quad X = \frac{1}{X'}$$

Interpretation of measured value: The higher the better, and the system is more attractive because it demands lower effort from user.

Definition of “Secure messages and information on device” Metric

Purpose of the metric: Measures the usage degree of security mechanisms, like identification, authorization, and confidentiality.

Method of application: Check if the system uses mechanisms to secure information. Categorize the degree of usage.

Metric type: Internal

Attributes measured:

Id - Identification

Authe - Authentication

Cryp - Crypto

Confid - Confidentiality

Autho - Authorization

Computation:

For each mechanism:

$$X = \begin{cases} 1 - full \\ 0.6 - largely \\ 0.3 - little \\ 0 - not\ at\ all \\ N/A - not\ applicable \end{cases}$$

Interpretation of measured value: The higher the better.

Definition of “Use of user profile” Metric

Purpose of the metric: Measures the extent in which the system relies on user profile to fit the inputs and outputs to the user and device.

Method of application: Check if the system uses mechanisms to fit to user's profile. Categorize the degree of usage.

Metric type: Internal

Attributes measured:

Uprof - User Profile

Computation:

$$X = \begin{cases} 1 - full \\ 0.6 - largely \\ 0.3 - little \\ 0 - not\ at\ all \\ N/A - not\ applicable \end{cases}$$

Interpretation of measured value: The higher the better, because the functions, the contents, and formats are adapted, automatically, according to user's profile.

Definition of “Installations success” Metric

Purpose of the metric: Measures the degree of successful installation in the different planned devices.

Method of application: Calculate number of successful installations according to devices types and total installations.

Metric type: External

Attributes measured:

DevTyNum - Number of Devices Types

DevNum – Number of Devices of each type

i – serial number of device, from 1 to DevNum

j – serial number of device type, from 1 to DevTyNum

Computation:

For each device, Z_{ji} indicates whether the installation succeeded or not:

$$Z_{ji} = \begin{cases} 1 - \text{success} \\ 0 - \text{failure} \end{cases}$$

Calculation for a specific device:

$$Y_j = \frac{\sum_{i=1}^{DevNum} Z_{ji}}{DevNum}$$

where $DevNum \geq 1$

Calculation for all devices:

$$X = \frac{\sum_{j=1}^{DevTyNum} Y_j}{DevTyNum}$$

where $DevTyNum \geq 1$

Interpretation of measured value: When Y is higher the system can be installed in more types of devices, so Installability is of higher quality.

When X is higher the degree of success of installations is higher, so Installability is superior.

Definition of “Ease of installation” Metric

Purpose of the metric: Measures the duration time and ease of installation.

Method of application: Calculate time of system installation on devices.

Metric type: External

Attributes measured:

DevTyNum - Number of Devices Types

DevNum – Number of Devices of each type

i – serial number of device type, from 1 to DevTyNum

j – serial number of device, from 1 to DevNum

InstTij - Installation Time in Device j of Type i

Computation:

Average installation time for devices of type i:

$$Y_i = \frac{\sum_{j=1}^{DevNum_i} InstT_{ij}}{DevNum_i}$$

Total weighted average:

$$X' = \frac{\sum_{i=1}^{DevTyNum} Y_i}{DevTyNum}$$

where $DevTyNum \geq 1$, $DevNum_i \geq 1$ and $InstT_{ij} > 0$

Transformation:

$$X = \frac{\max - X'}{\max} \quad \text{or} \quad X = \frac{1}{X'}$$

Interpretation of measured value: The higher the better. When X is greater the mean time to install the system is smaller, so the installation is easier and the system is of better quality.

The “PDA” Experiment

Each metric defined by the research was validated theoretically, mathematically, and empirically. The aim of the validation was to prove that the metrics behave in a consistent and logical mode (Zelkowitz & Wallace, 1997). Thus, it is possible to rely on them to quantify the quality of mobile-wireless information systems, for example, by showing that the value of the metric grows when quality increases and vice versa.

This research was part of a broader research, to define new quality metrics, in which different kind of mobile-wireless applications were inspected. Some of the metrics were validated empirically using other experiments or compared between different experiments. This paper focuses on the PDA application and the validation in the “PDA” experiment alone.

The "PDA" experiment is based on a system developed by Dooblo, a commercial company. The experimental system contains a set of questionnaires, which allow conducting surveys, using a

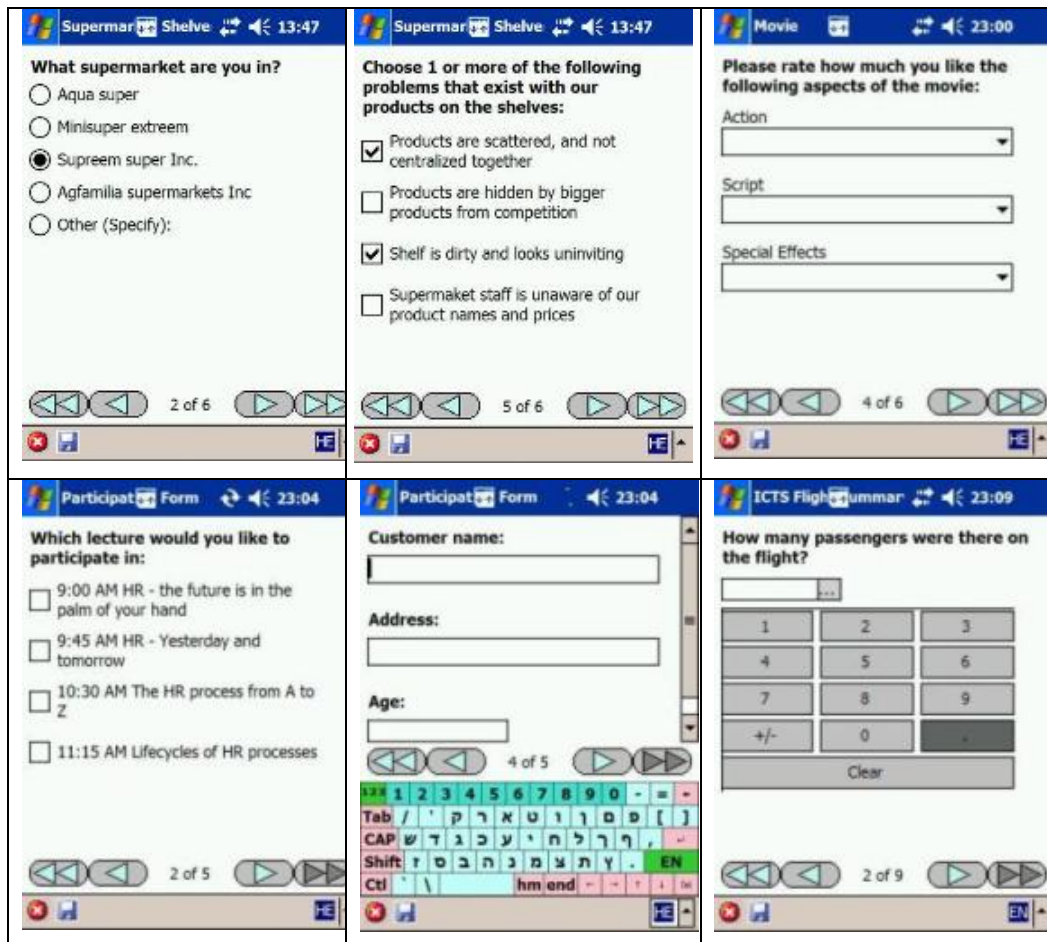


Figure 2 – Samples of “PDA” experiment displays

PDA device based on Windows Mobile CE operation system. The system contains five different surveys, "Supermarket", "Purchase", "Participation", "Movie" and "Flight". Each survey includes 6-10 displays (samples in Figure 2). These questionnaires include different kinds of questions, such as open questions or selecting an answer from a number of possibilities, similar to performing quizzes in a PDA-based m-learning system. The questionnaires were loaded to the PDA device's memory in advance. The surveys were performed, accumulating the data on the device, and further downloaded to a server for elaboration.

This is a realistic experiment, using real tasks and environment (Sjoberg et al., 2002), and the observation was done as a "black box", with no possibility to see or interfere with the code of the system. Each survey was performed thirty times, and the data needed for the different metrics was collected.

The theoretical proof and the empirical experiments, summarized in the Appendix, successfully validated the new metrics defined in the research.

Practical Implications

Some practical implications can be deduced from this research, including technical, implementation, and content issues, to assist the developers of PDA-based m-learning systems.

The technical developers must be conscious of the different kinds of devices which will be used, including differences in memory size, screen size, and attributes, to design an application suitable to all. This can be accomplished by using user profiles mechanisms. Applications must be aware of the small memory and low processing resources restrictions and optimize their utilization.

The content developers must be aware of the restrictions in size and usability, limiting input to minimal required data, determining pre-defined options and automatic filled-in fields, using location aware information when possible, instead of user's input. They need to consider output length, by defining concise messages and adapting the content of the application to screen size, taking into consideration that the operation of leafing through multiple pages in a mobile device decreases the usability.

Restrictions and Future Enhancements

The purpose of the "PDA" experiment was to validate the definition of the metrics, and this is the reason the small sample was enough. The experiment can be enhanced with an extended sample. In this experiment, only one text-based PDA application was examined, in one kind of device. The experiment can be broadened by enlarging the number of kinds of devices and the number and types of applications. Moreover, it can be performed again, with a real m-learning information system, instead of the survey system.

Summary

This article describes a methodology to enable quality measuring of m-learning systems that allow interactive learning "any place, any time" using PDA devices. PDA-based m-learning systems are a sub-field of mobile-wireless information systems, which face unique problems and challenges that affect quality attributes. These attributes were analyzed and accurate metrics were defined. These metrics enable objective quality evaluation and comparison of mobile-wireless information systems.

These metrics are useful when the quality of a mobile-wireless information system must be analyzed and quantified, for example, when comparing two proposed systems or when a system has

to be developed or bought. When the metrics are used to compare systems, the higher the metric value, the higher the system's quality. However, when only one system has to be measured, the metrics need an external value to compare to. These values can be defined in advance according to the requirements of the system. Moreover, practical implications for the development of PDA-based m-learning systems were defined based on insights gained from the metrics definition and methodology.

References

- Alexander, B. (2004). Going nomadic: Mobile learning in higher education. *EDUCAUSE Review*, 39(5), 28–35.
- Asunmaa, P., Inkinen, S., Nykänen, P., Päivärinta, S., Sormunen, T., & Suoknuuti, M. (2002). Introduction to mobile internet technical architecture. *Wireless Personal Communications*, 22, 253–259.
- Basili, V. R., & Rombach, H. D. (1988). The TAME project: Towards improvement-oriented software environments. *IEEE Transactions on Software Engineering*, 14(6), 758–773.
- Boehm, B. W.; Brown, J. R.; Kaspar J. R.; Lipow M., & McCleod, G. (1978). *Characteristics of software quality*. Amsterdam: North Holland.
- Brady, A., Conlan, O., & Wade, V. (2004). Dynamic composition and personalization of PDA-based eLearning – Personalized mLearning. In G. Richards (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2004, Chesapeake, VA: AACE*, 234-242.
- Calero, C., Ruiz, J., & Piattini, M. (2004). A web metrics survey using WQM. *Proceedings ICWE 2004, LNCS 3140, Springer-Verlag Heidelberg*, 147–160.
- Covella, G., & Olsina, L. (2006). Assessing quality in use in a consistent way. *Proceedings of the 6th International Conference on Web Engineering, ICWE'06, Palo Alto, California, USA*, 1-8.
- Di Pietro, R., & Mancini, L. V. (2003). Security and privacy issues of handheld and wearable wireless devices. *Communications of the ACM*, 46(9), 74-79.
- Green, R. (2003). Wap 2.0: What is it? Wap 2.0 is coming, sooner than you might think, and it's going to change everything. *Wireless Business and Technology*, 3(6), 34-38.
- Gafni, R. (2008). Framework for quality metrics in mobile-wireless information systems. *Interdisciplinary Journal of Information, Knowledge, and Management*, 3, 23-38. Retrieved from <http://ijikm.org/Volume3/IJKMv3p023-038Gafni358.pdf>
- Goh, T., & Kinshuk. (2006). Getting ready for mobile learning - Adaptation perspective. *Journal of Educational Multimedia and Hypermedia*, 15(2), 175-198.
- Herzberg, A. (2003). Payments and banking with mobile personal devices. *Communications of the ACM*, 46(5), 53-58.
- Hordijk, W., & Wieringa, R. (2005). Surveying the factors that influence maintainability. *Proceedings of the 10th European Software Engineering Conference held jointly with 13th ACM SIGSOFT International Symposium on Foundations of Software Engineering ESECFSE'05, Lisbon, Portugal*, 385-388.
- Huber, J. F. (2004). Mobile next-generation networks. *IEEE Multimedia*, 11(1), 72-83.
- ISO/IEC 15939, (2002). *Software engineering – Software measurement process*. Geneva, Switzerland: International Organization for Standardization.
- ISO/IEC 9126, (2001). *Software engineering - Product quality - Part 1: Quality model*. Geneva, Switzerland: International Organization for Standardization.
- ISO/IEC 9126, (2004). *Software engineering – Product quality – Part 4: Quality in use metrics*. Geneva, Switzerland: International Organization for Standardization.

- Lau, J. (2006). *The State of European enterprise mobility in 2006*. Forrester Research.
- Lee, P. (2003). Mobile data comes of age: Wireless is big, we know that, but how big? *Wireless Business and Technology*, 3(9), 16-18.
- Kan, S. (2002). *Metrics and models in software quality engineering* (2nd ed.). Addison-Wesley.
- Kim, G., & Ong, S. M. (2005). An exploratory study of factors influencing m-learning success. *The Journal of Computer Information Systems*, 46(1), 92-97.
- Kitchenham, B., Pfleeger, S. L., & Fenton, N. (1995). Towards a framework for software measurement validation. *IEEE Transactions on Software Engineering*, 21(12), 929-944.
- Malladi, R., & Agrawal, D. (2002). Current and future applications of mobile and wireless networks. *Communications of the ACM*, 45(10), 144-146.
- McCall, J. A., Richards, P. K., & Walters, G. F. (1977). *Factors in software quality*. RADC TR-77-369, Vols I, II, III, US Rome Air Development Center Reports NTIS AD/A-049-014, 015, 055.
- Pandian, C. R. (2004). *Software metrics: A guide to planning, analysis, and application*. Auerbach Publications.
- Parsons, D., & Ryu, H. (2006). A framework for assessing the quality of mobile learning. *Proceedings of the 11th International Conference for Process Improvement, Research and Education (INSPIRE)*, Southampton Solent University, UK.
- Sjøberg, D. I. K., Anda, B., Arisholm, E., Dybå, T., Jørgensen, M., Karahasanovic, A., Koren, E. F., & Vokác, M. (2002). Conducting realistic experiments in software engineering. *Proceedings of the 2002 International Symposium on Empirical Software Engineering (ISESE '02)*, 17-26.
- Spriestersbach, A., & Springer, T. (2004). Quality attributes in mobile web application development. *LNCS 3009, Proceedings of PROFES, Berlin*: Springer-Verlag, 120-130.
- Tarasewich, P. (2003). Designing mobile commerce applications. *Communications of the ACM*, 46(12), 57-60.
- Terho, M. (2002). Mobile web services and software quality. *LNCS 2349, Proceedings of ESCQ, Berlin*: Springer-Verlag, 2-6.
- Terrenghi, L., Kronen, M., & Valle, C. (2005). Usability requirements for mobile service scenarios. *Proceeding of HCI International Conference, Las Vegas, USA*, 1-10.
- Varshney, U., & Vetter, R. (2002). Mobile commerce: Framework, applications and networking support. *Mobile Networks and Applications*, 7, 185-198.
- Vaughan-Nichols, S. J. (2004). Wireless middleware: Glue for the mobile infrastructure. *IEEE Computer*, 37(5), 18-20.
- Venkatesh, V., & Ramesh, V. (2002). Usability of web and wireless sites: Extending the applicability of the Microsoft usability guidelines instrument. *Information Systems Technical Reports and Working Paper, TR134-1*. Retrieved Nov. 20, 2006, from <http://www.kelley.iu.edu/ardennis/wp/tr134-1.doc>
- Wood, K. (2003, March). Introduction to mobile learning (m-learning). Retrieved on October 20, 2005 from <http://ferl.becta.org.uk/display.cfm?page=65&catid=192&resid=5194&printable=1>
- Zelkowitz, M. V., & Wallace, D. (1997). Experimental validation in software engineering. *Information and Software Technology*, 39, 735-743.

Appendix – Metrics Validations

Table 2 summarizes the validation of the new defined metrics.

Table 2 – Results of Metric Validations in PDA Experiment

Metric Name	Result of validation
Response time to get information from cache	For displays with longer response time a lower value was received for the metric.
Device memory cleanup after transaction	In the cases where the memory was better cleaned after completing the task, the metric value was higher, as shown in Figure 3.
Display load	For less loaded displays the metric value received was higher, i.e., the displays of the “Flight” survey which are the less loaded, got a metric value of 1, the highest possible. (Figure 4).
Messages conciseness	Surveys, in which the messages were shorter in proportion with display, got a higher metric value. (Figure 5)
Ease of input entering	For surveys where the ratio of “closed” questions was higher, the value received was higher, as shown in Figure 6.
Ease of output use	When the number of leafing needed to see all the output is greater, the value of the metric received was smaller.
Ease of use - displays per task	When the number of iterations needed to finish a task is greater, the value of the metric received was smaller.(Figure 7)
Secure messages and information on device	This metric was validated through other experiments.
Use of user profile	This metric was validated through other experiments.
Installations success	This metric was validated through other experiments.
Ease of installation	This metric was validated through other experiments.

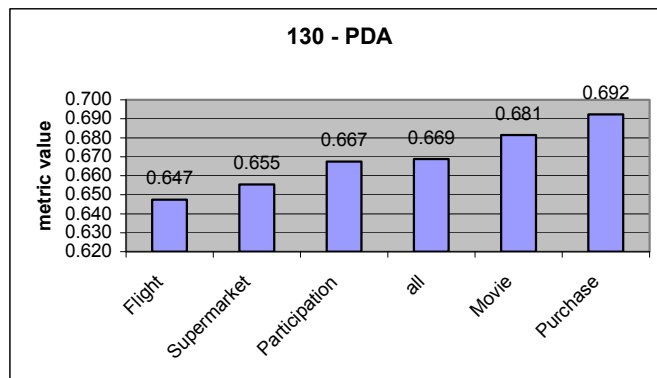


Figure 3 – Validation of “Memory Cleanup” metric in “PDA” experiment

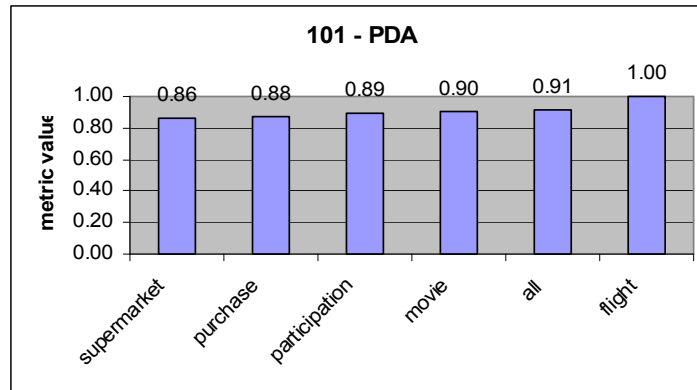


Figure 4– Validation of “Display load” metric in “PDA” experiment

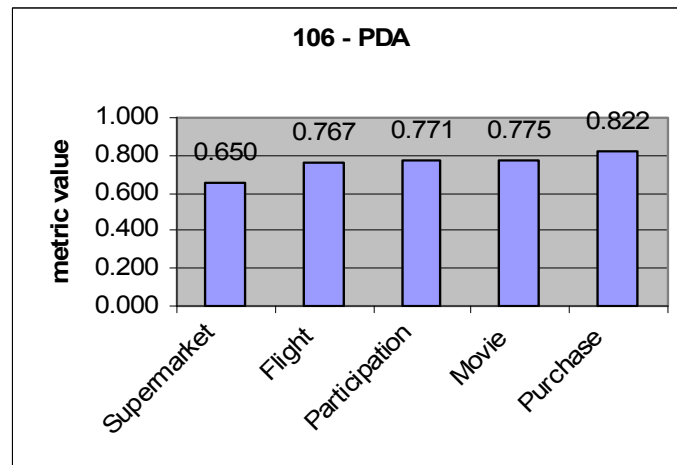


Figure 5 – Validation of “Messages conciseness” metric in “PDA” experiment

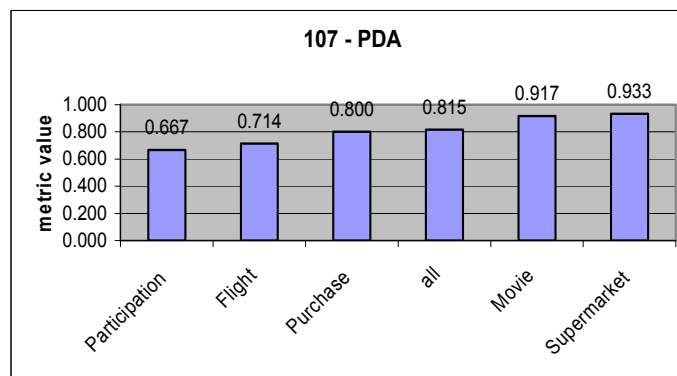


Figure 6 – Validation “Ease of input entering” metric in “PDA” experiment

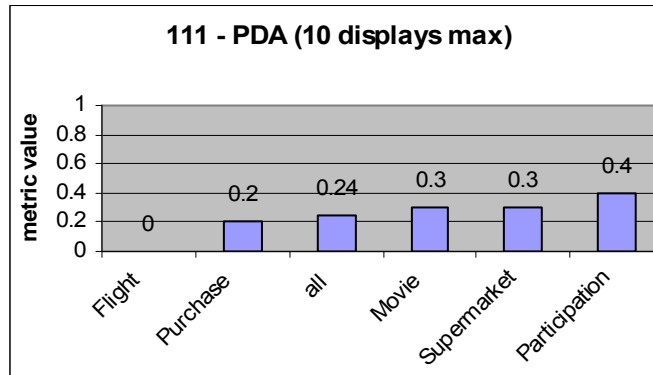
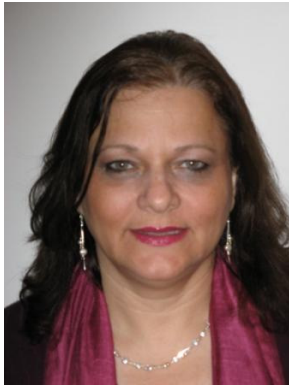


Figure 7 – Validation “Ease of use - displays per task” metric in “PDA” experiment

Biography



Dr. Ruti Gafni holds a PhD from Bar-Ilan University, Israel in the Business Administration School, focusing in Information Systems. She holds an M.Sc from Tel Aviv University and a BA (Cum Laude) in Economics and Computer Science from Bar-Ilan. She has more than 30 years of practical experience as project manager and analyst of information systems. She teaches in the Management and Economics MBA program, both at the Open University of Israel and at Netanya Academic College.