The Pyramid Model: A New Model for Evaluating E-Learning Systems

K. Hadjar¹, and F. Alawi¹
¹Department of Multimedia, Ahlia University, AL Manama, Bahrain

Abstract - In this paper we propose a new evaluation model for e-learning systems with its implementation details extracted from two standard models, the Kirkpatrick's classical model with the ISO 9126 standard for evaluating the e-learning systems. The new proposed model: "The Pyramid Model", consists of five consecutive levels, each of which is having different research tactics as inputs that will draw the output results. The e-learning system used in Bahrain is a Telecommunication Company (Batelco) was chosen to be the case study to test all the levels' implementations of the pyramid model. Based on the outcomes of each level, extra suggestions were given as customized enhancements for the two used standards. This was accomplished through reviewing the literature that is related to this context to build the theoretical foundation needed to design and conduct the research. Hence, the case study research strategy was adopted and the surveys, questionnaires, semi structured interviews, observations and secondary data research methods were used. Finally we expose concrete evaluations for our new model.

Keywords: E-Learning, Kirkpatrick e-learning evaluation model, ISO 9126 standard.

1. Introduction
E-learning system is defined as interactive software that is available for access upon request in self-pace manner [2]. E-learning offers the possibility of learning from information delivered to us electronically [13]. It is a web based personalized learning experience that provides measurable results [21]. The broadest definition refers to any distance-learning mode other than a correspondence course with printed material [17]. In [22], the author refers to E-learning as the use of internet technologies to deliver a broad array of solutions that enhance knowledge and performance. E-learning lets the students and employees learn at a distance [11], over the Internet. It also enables them to learn at any time and any place. All what they need is a computer, an Internet connection and access to the course materials that reside on the web. Depending on the topic being taught, e-learning can be instructor-led or self directed (without instructor). Some e-learning is designed to be completed at scheduled times, while other kinds of e-learning are completely unscheduled: students can take the course at their convenience without a need to attend face-to-face classes. While e-learning comes in a variety of styles and sizes, all of the e-learning varieties share the goal of getting knowledge and skills training to the learners more quickly, more effectively, and often at a lower cost but without being regarded as an alternative to the more common methods: they rather constitute a new possibility to be added to a more traditional approach [11]. There is no doubt that modern computer-based learning support must go far beyond simply delivering learning content in only one approach that fits for all students. From the teacher's point of view, educational aspects must be considered by flexible sequences of learning assets or learning activities. It should also be flexible enough to let the teacher follow the students’ progress. From the student's perspective, pre-knowledge, preferred learning styles, and learning media must be considered for personalization purposes. These requirements may contradict the available services in the e-learning system which leads to confusion in selecting the most appropriate one to be used. That's why the evaluation process for the e-learning system is needed.

2. Evaluation of E-Learning Systems
Since e-learning systems exist in different styles and aspects, the leader of the organization may find it difficult to select the most appropriate one that will help accomplishing all the organization’s educational goals. This is the main reason why it is very important to evaluate the e-learning system. The evaluation would not help to select the best system only, but furthermore it will help to measure the quality and effectiveness of various aspects of education, such as materials, instructors, facilities, and presentation techniques. It can identify areas that need improvement and ones that can serve as models of excellence. No corporation can afford training that does not have a tangible benefit to the business [25]. To measure learner satisfaction, testing and assessments are an essential part of the process. Tests based on consistent standards provide an objective way to measure the skills and knowledge of the learner. If the goal is to get employees to learn actual job skills that can translate to benefits for the employee and the organization, it is necessary to provide proof of effectiveness [4]. One benefit of evaluation is to provide feedback to subject matter experts and content designers so they can improve their decision making process in e-learning [12]. Effectiveness can be improved quality in performance. Xerox Europe Limited, for example, implemented e-learning for training its call centre staff so that they are able to demonstrate professional telephone techniques and strong communication and interpersonal skills. This will allow employees to deal with calls successfully and will enable them to direct the call to the appropriate person [8]. The ultimate goal of effective e-learning is to drive business results. Corporate performance can be enhanced through alignment of training and business strategy [11]. Managers need to demonstrate that e-learning is having a positive impact on corporate strategy and investment objectives. If the business goal of an e-learning program cannot be identified, then there should be a query on why it is there in the first place. Soft cost savings that reflect value added productivity, improved retention, and
greater satisfaction are more difficult to measure. The ultimate value of e-learning comes when e-learning is linked to achieving a company's goals [9]. There are many concerns in order to evaluate e-learning but those can be countered with a carefully designed evaluation plan [14].

3. New Model for Evaluating E-Learning Systems

The best way to go in order to evaluate an e-learning system is to use a standardized evaluation model. There are four qualitative evaluation systems on e-learning: IEEE Learning Technology Standard Committee (LTSC) reference model [15], The Sharable Content Object Model Reference Model (SCORM) [24], The ISO 9126 Quality Model developed by the International Organization for Standardization (ISO) [26], and the Kirkpatrick’s Classical Model [16]. Having in mind the previous e-learning evaluation systems, we decided to tune the Kirkpatrick’s classical model (The Four Levels) and join it with some characteristics of the ISO9126 quality model to form a new enhanced evaluation model that is more effective and practical. This decision had been taken for the following reasons:

1. Most of the training professionals trust to use “The Four Levels” for its longevity; it has become the industry evaluation standard for almost fifty years.
2. Taking a deep view at each level, will prove that each level has particular benefits and unique challenges.
3. We noticed that all the levels are evaluating a sort of relationship or interaction between the human (Trainees, Managers or Developers) and the evaluated system which means concentrating on the feedback results without an in-depth consideration of the technical details. As a result, this criterion will make this model general enough to evaluate any system from its users’ perspectives while on the other hand, specific enough to obtain the targeted results. Concisely, as it is defined: "Measuring changes in behavior that occur as a result of training programs."
4. For its simplicity and its ability to help people to understand the real meaning and purpose of evaluation, it has been the most widely used and popular standard model for the evaluation of training and learning as it has provided successful framework that has clearly met an organizational need [16].
5. For each level in the classical model, all the detailed metrics defined make it obvious its use, while in the ISO 9126 standard the characteristics and sub characteristics are defined properly to demonstrate what should be evaluated in the system without mentioning the metrics for each one nor specifying the method for measurement, rating and assessment. This makes the standard’s characteristics very broad and general and left to be used upon the evaluator perspective [7].
6. We believe that even though the metrics and evaluation tools are not defined in the ISO 9126 standard, there are some of the model’s six characteristics that are very important to be included in our enhanced model of evaluation. So, we decided to consider the following three major characteristics: functionality, usability, and efficiency.

Using these characteristics, that are more behavioral than technical, in conjunction with the levels of the classical model, will result in a homogeneous refined version of the standard evaluation model proposed by Kirkpatrick.

Our new model “Pyramid” illustrated in figure 1 consists of five consecutive layers, four of them from the classical model plus one that belongs to the ISO9126 standard. The special name of our model demonstrates the importance of the time sequence among the levels. The goal of our pyramid model is to provide a methodology that ensures a high level of quality evaluation of any given e-Learning-system. With its clearly structured process, it can serve in the future for potential quality accreditations of e-learning-systems. Our main focus in “Pyramid” is on evaluating the human-system interaction aspects without paying that much attention to the technical matters. This is because the technical problems can be overcome by the support solutions while the interaction aspects will proportionally affect the important outcomes obtained from selecting the specific e-learning system.

The Pyramid Evaluation Model

Figure 1. The Pyramid Model.

3.1 The Pyramid Evaluation Model

At each level the evaluation answers whether a fundamental requirement of the training program was met. There is no discrimination in the importance of the five levels since each predecessor level is a prerequisite for its successor. Hence, all the levels are equivalently important. In fact, the evaluation’s outcome of each level will directly affect the measurement of the next level and vice versa, i.e. if the assessment scale shows a low indicator at a given level, this is perhaps because some training skills were not gained at the previous level. As you move up the levels, the evaluation process becomes more challenging and the difficulty of tracking the exact measurement increases as illustrated in figure 2.

3.1.1 Level one: Reaction

The first step of evaluation is to know how much the participants reacted to the e-learning training program and how well they liked it. This is the minimum that should be done for all the programs. Directly after the completion of the course/program, the trainees are asked some questions in what is called “Happy Sheets” or
“Smile Sheets”. As the e-learning is a technology based system, those surveys can be served by the e-learning system itself and it can be viewed automatically and submitted online. This point can speed up the analysis of the results at the reaction level and its presence in the e-learning system can be considered as a main point in the evaluation process.

“Reaction Sheet” is the best way to measure the customer’s satisfaction. Some peers, bosses, and/or future participants will probably ask the participants, “What did you think of the training program you took?”. And anyone can guess the possible consequences of them saying such things as “It was a waste of time, “I sure didn’t learn anything that will help me do my job better,” or “Don’t waste your time taking it unless you have to.” The evaluator must not forget the main objective of making those sheets and he/she should keep in mind the trainees’ responses to raise the quality of the offered e-learning programs and make it best fit for the participants. Although a positive reaction does not guarantee learning, a negative reaction almost certainly reduces its possibility.

3.1.3 Level three: Behavior

Acquiring the knowledge is something different than having it in practice. That is the main reason why it is very important to have this level of evaluation. Here, the evaluator will measure to what extent the participants changed their behavior back in the workplace as a result of the training. If the objectives were met in (Level 2) and the trainee had all the required knowledge, skills and attitudes then the trainer surely wants to know whether this amount of knowledge will be used in the learner’s environment on a daily basis. There is no doubt that this level represents the most important and truest one that will seriously measure e-learning effectiveness. However, measuring at this level is difficult as it is often impossible to predict when the change in behavior will occur, and thus requires important decisions in terms of when to evaluate, how often to evaluate, and how to evaluate. Whatever method will be used, the trainer has to make sure to get honest answers. For example, if he or she interviews the trainees, they may be unwilling to say they have not implemented any of the learning because it will make them look bad. It is important to stress that you are measuring the effectiveness of the program and trying to determine to what extent it has been practical. The trainees should understand that their answers may help improve future programs and will remain anonymous so they feel no pressure to say they have made changes when they have not.

3.1.4 Level four: The ISO Level

The main enhancement we proposed for the classical model is actually adding this level of evaluation. From our perspective, looking from different angles can give precision. If this is to be applied to the “Pyramid Model”, merging two evaluation standards will result in a powerful model that will break all the drawbacks of both of them. We chose three main characteristics from the ISO9126 standard to be treated as one unit in a separated level: functionality, usability, and efficiency.

3.1.5 Level five: Results

As we go up the pyramid, we are moving from quite general to more specific objectives. Here is the peak level where the main objective resides with the answer to this question: “Why did the evaluator go along all the steps of the pyramid?” As the name gives strong meaning, (Results), it is at the same time so much general and relative. The trainee may –personally- achieve great results due to the e-learning program but this is worth nothing for the organization if he or she did not apply those results in his job. So, the correct name for this level should be “Organizational Results”. In this level, we are concerned with the business impact of the training program and will measure its success in terms the tangible difference as perceived by managers and executives.

4. Research Strategy: Case Study

Research strategy is a general plan of how the research questions will be answered, how research objectives will be set, how data will be collected, and what constraints a research may face [23]. The case study research method is an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used [27]. Case study can also be defined as an in-depth exploration of one situation and it involves the investigation of
this particular situation, problem, company, or group of companies, where the investigation can be done directly by interviews and observation or indirectly by studying company reports or company documentation [5]. Our case study is Batelco, Bahraini Shareholding Company. It evolved from a local telephone company with a capacity for 66,000 lines to become a regional leader specializing in a broad range of communications services including mobile, national and international telephony, business network services, internet, and satellite services. Supported e-learning was introduced by Batelco Training Centre to offer all staff the opportunity to develop both personal and professional skills in a wide range of subject areas. The Centre provides the tools and information necessary for the interactive learning experience and monitors the progress of each student offering support and encouragement to ensure the timely and successful completion of each course. Upon successful completion, participants are invited to attend a classroom-based session to complete a blended-learning solution. Applying “The Pyramid Model” on a huge and successful organization like Batelco will give a more accurate measure for each level that is needed to prove the efficiency of our model. There are different tactics that can be utilized in case studies [20, 23]. These tactics are questionnaires, interviews, observations, archival records, and the analysis of documents. For the questionnaires, we will use this method in the implementation of “Level 3: Behavior” to obtain a measurement about how and to what extent could the e-learning change the behavior of the employees/trainees. The sample size was thirty of Batelco employees. Through the questionnaire that we designed, the employees were asked some questions that can measure their change in behavior at work after they took the e-learning course(s). For the interviews we used the semi-structured interviews specifically when implementing the last level “Level 5: Results” of our model to acquire data about Batelco organizational results gained from inducting their employees in e-learning training courses. There will be at least five interviews with different line managers to get their feedback. The semi-structured interview was selected over the other types because not all the questions would be asked to all interviewers and the order of these questions may differ according to the interviewee’s characteristics. For the archival records and the analysis of documents will be used to get data at the implementation of the first two levels as follows:

- Level 1: Reaction*: All the surveys which we call “Reaction Sheets” will be queried directly from Batelco e-Learning Centre since the system is providing them with all the answers for all the surveys for each employee after the completion of each e-learning course.
- “Level 2: Learning”: All the assessments’ results for all the courses in details will be granted also automatically by Batelco e-learning system. Thus, it is easier, more reasonable, and more economic in terms of time and efforts to use the secondary data instead of conducting interviews with all the trainees in the company in order to get the same data but with less accuracy.

5. Results
This section demonstrates the real implementation of the pyramid model on the case study “Batelco”. Since our model consists of five layers, we had implemented each layer separately.

5.1 Level 1: Reaction
For the reaction level implemented on “Batelco e-learning system”, we considered the following:
1. The system provides the reaction sheet automatically for each student after completing the e-learning course.
2. The sheet consists of thirteen questions where each has a scale of six options to choose from: 5 = Strongly Agree, 4 = Agree, 3 = Neutral, 2 = Disagree, 1 = Strongly Disagree and N/A = Not applicable
3. The sheet is the same for all the courses and there is no way to alter the questions by the administrator or the evaluator.
4. The system forces the trainee to fill in the reaction form after completing the course in order to grant the completion status to him/her. Otherwise, the training status for the course will remain ‘incomplete’.

 Ninety seven random samples of answered reaction sheets for different courses were used. Twenty one samples were excluded because they were having incomplete responses. All the excluded samples were having a steady constant answer for all the thirteen questions. After excluding the bad samples, we ended up with seventy six samples and this was our sample size at this stage. All the response data were entered in an excel sheet of 13 row (representing the questions) and 76 columns (representing the sample size). Then we designed a C++ program which read the data from the input file analyze it and save the analysis in another output file which consists of the summary response table for each question, (see Table 1 below).

| Table 1. Relation between titles and mapping rules during extraction |
|-------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Frequency of | N/A - Not | (1) Strongly | (2) Disagree | (3) Neutral | (4) Agree | (5) Strongly |
| Scale | applicable | Disagree | | | | Agree |
| Q1 | 0 | 0 | 1 | 11 | 25 | 39 |
| Q2 | 0 | 0 | 3 | 11 | 40 | 22 |
| Q3 | 0 | 0 | 4 | 10 | 32 | 30 |
| Q4 | 1 | 0 | 1 | 12 | 44 | 18 |
| Q5 | 2 | 0 | 2 | 12 | 29 | 31 |
| Q6 | 0 | 0 | 2 | 9 | 43 | 22 |
| Q7 | 3 | 0 | 2 | 7 | 21 | 43 |
| Q8 | 0 | 0 | 2 | 7 | 34 | 33 |
| Q9 | 0 | 0 | 2 | 9 | 39 | 26 |
| Q10 | 1 | 0 | 2 | 10 | 38 | 25 |
| Q11 | 0 | 0 | 0 | 10 | 24 | 42 |
| Q12 | 0 | 0 | 2 | 7 | 31 | 36 |
| Q13 | 0 | 0 | 3 | 7 | 36 | 30 |
| Mean | 0.5 | 0.0 | 2.0 | 9.4 | 33.5 | 30.5 |

First, before measuring each level, we have to pay attention to the time of evaluation. Second, there should be a specific way of measurement. At this level, the most important measure is the learners’ feedback which was measured by conducting a survey. Based on our findings in implementing this level on “Batelco e-learning system”, we can say that this level of standard can be adjusted to get better reaction if we consider the following:

- We think one of the key successes for any e-learning system at this level is making this measurement automatically provided by the system. There is no doubt that doing so will save the evaluator the time and effort needed to design and substitute the sheets manually.
Rather than thinking each time what are the questions that should be asked to the trainee, alternatively, we suggest that there should be a template for the reaction sheet which is coming embedded in the system and the system should be flexible to allow making some changes in the sheet as it is required by the evaluator.

A 100% response can be guaranteed if the e-learning system was designed in a way that it is compulsory to fill the reaction sheet after completing any e-learning course. As the response rate increases, it will lead to a more realistic rate of the overall reaction of the students/trainees.

5.2 Level 2: Learning
The measurement methods of Level Two tend to require more time, effort, and care than Level One. Some methods used to evaluate Level Two are mentioned in the following list:

- Formal and informal testing
- Self assessments at the beginning (pretest) and end (posttest) of the training
- Interviews can assess a participant’s expectations of learning and confidence in learning
- Observation and feedback from participants, managers, and supervisors

The following are our findings when we implemented the learning level on Batelco’s e-learning system:

1. The system uses the “self-assessment” method in order to measure the learning extent on the trainees for each of the e-learning courses.
2. All the assessments are being displayed and graded in an automatic manner by the system itself.
3. Most of the assessments are taking place during the training time period; to be specific, immediately after each unit in each module (Post-assessments). While in some few technical courses, pre-assessments take place prior to each unit in the module.
4. In all the courses, the system obliges the trainee to get a minimum pass result of eighty percent (80%).
5. It is permitted for the trainee to repeat the assessment as much as he/she wishes and at the end the maximum score will be undertaken as the final result.

Our sample size (n) was the whole population (N) (i.e. the total number of the enrolled students in all the courses) where n = N = 1021 results. We used the same technique that we have used in level 1 and we have obtained the Assessment analysis results listed in Table 2.

Table 2. Assessment analysis results

<table>
<thead>
<tr>
<th>Grades</th>
<th>frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>80-84</td>
<td>167</td>
<td>16%</td>
</tr>
<tr>
<td>85-89</td>
<td>214</td>
<td>21%</td>
</tr>
<tr>
<td>90-94</td>
<td>291</td>
<td>29%</td>
</tr>
<tr>
<td>95-100</td>
<td>349</td>
<td>34%</td>
</tr>
</tbody>
</table>

The table shows very inspiring results; almost one third of the e-learners got the highest range (above 95) while only 16% of them had a result in the lowest range (from 80 to 84). Based on the results analysis of the case study learning implementation, we suggest the following:

- At the stage of e-learning system selection, the first thing the training manager should keep in mind is that the selected system should provide high level of learning to the trainees and this should be assured with a minimum cost of time and effort. We think the best way to assure this is to select such an e-learning system with automated learning evaluation process (i.e. the system should provide the learning assessments/tests automatically without the need to waste the time and budget to conduct the procedure manually).
- Obliging a minimum pass percent -as it is happening in Batelco e-learning system- will motivate the trainees to get higher scores than the minimum and hence will guarantee such extent of learning. But we suggest that it is very practical to make the minimum pass percent adjustable (by the system’s administrator) rather than fixed, depending on the type and the importance of the course.

Again if these suggestions were implemented as part of the evaluation model, we will be able to jump over this level directly to evaluate the change in behavior at the next level since the previous step was completely automated.

5.3 Level 3: Behavior
At this level, we used the questionnaire research tactic to evaluate the behavior of the trainees after they took the e-learning course. In [10] the authors mention that questionnaires are a good way of collecting certain types of information quickly and relatively cheaply as long as the researcher is sufficiently disciplined to abandon questions that are superfluous to the main task. The questions were carefully developed based on the research aim and objectives and were developed from the literature review. They were written and organized in an order to obtain the learners’ consideration about their change behavior. Patton [18] pointed out six kinds of questions that can be asked. Of these, the following five were adopted for the construction of the questionnaire:

- Experience/behavior questions: these are aimed at eliciting descriptions of experiences, behaviors, actions, and activities that have been encountered.
- Opinion/belief questions: these are aimed at describing cognitive and rationality issues behind values, intentions, desires, and goal setting.
- Feeling questions: these are aimed at understanding the emotional reactions, such as satisfaction or confidence, to a certain issue.
- Knowledge questions: these are aimed at eliciting factual information.
- Background/demographic questions: these are aimed at eliciting the characteristics of the respondent being questioned.

The sample size of our questionnaire was determined by the overall aim of the research. As a minimum time limit period set by the standard for the best timing of the evaluation at this level of about 3 to 6 months after the training, we had a lower sample size than what we had in the previous levels. In addition, most of the employees took more than one e-learning course, which explains the reason behind the employee scope which is tighter than the course scope. After establishing the final size needed to collect the primary data had been established, the questionnaire was sent by
The questionnaire did not change of all learning on their employees because of the lack of evaluation and follow-ups or they need more time to observe the change. From our side, we find it difficult to set ‘yes or no’ final result for this level of evaluation as we do agree with what the managers said. There is a gap because of the absence of evaluating level three (behavior) in Batelco. We think performing that level’s evaluations regularly will help the decision makers in Batelco to draw a clear picture about the company’s expected benefits because of the e-learning. This level is so hard to implement [6]; the evaluators believe that ‘Results Level’ is too difficult and costly because they misunderstand Kirkpatrick’s label of “organizational results.” Another problem is that evaluators believe they must be able to show proof beyond a reasonable doubt that business improvements were a direct result of a training intervention. This proof is often impossible to demonstrate.

5.4 Level 4: ISO 9126
Based on the standards’ characteristics of the ISO 9126 [19] we set the quality metrics requirements of “Batelco e-Learning System”. After, we checked with the system administrators if the ISO characteristics were met in the system since they are fully aware of all the systems’ capabilities and all the users’ complaints. We noticed that not many flaws were discovered with the system and all of these were not critical to user satisfaction since all of them can be resolved either by the vendors or by the support team. In this level, we observed that there is no straightforward method of evaluation. It is all dynamically being set by the organization itself depending on the required level of quality as the ISO standard did not set any. Furthermore, the concerns in the organization can also set the rating levels and the needed measures. The evaluation could be improved by having a global characteristic to summarize the overall user satisfaction. At the same time, it is difficult to determine the user satisfaction level because different users will have different priorities that will influence on which characteristics they will place more emphasis.

5.5 Level 5: Results
Level 5 evaluations measure the business results of learning. They help answer such questions as these:

- Is training contributing directly to the business objectives of the organization?
- Did the training accomplish its original business and organizational goals?
- How much had essential business processes (quality, efficiency, and productivity) improved as a result of training?
- How does the rate of return on money invested in training compare to other investments open to the organization?

We do believe that measuring some organizational factors in terms of revenues and costs is very critical and confidential plus it is very challenging to link such kinds of factors to the e-learning itself. That is why it is not practical to perform an in-depth evaluation at this level, so we decided to use the interviewing method with some managers who experienced advanced e-learning courses as trainees in order to give a clear picture about the organizational results from their perspectives and the real impact of e-learning on their employees.

According to the findings at this stage, most of the managers do not agree on the direct impact of e-learning on the organizational outcomes and even though they can feel the positive change of e-learning on their employees they find it difficult to link between e-learning and the level of performance since there may be some other reasons behind the improvement other that e-learning itself. From their perspective, it is still difficult to judge either because of the lack of evaluation and follow-ups or they need more time to observe the change.

6. Conclusion
This paper presents our new evaluation model: “The Pyramid Model” that best evaluates the e-learning systems. Our proposed model consists of five consecutive levels, each of which is having different research tactics as inputs that will draw the output results. Some limitations had affected the accuracy of the results extracted from the analysis and findings: the questionnaire did include such information such as age, gender, work experience, level of education. The main reason behind this was to minimize the amount of questions that were required to be filled by the trainees in order to motivate them and get higher response rate. We think if that information were present, that will help in finding more relative and helping factors to analyze the data. Also, even that we have followed many guidelines in designing the questionnaire for level 3 to get a maximum response rate from the trainees; we got only quite small percent of (40.7%). Future work is finding such a way to feed the e-learning system with the required evaluation model in order to embed all the levels inside it.

7. REFERENCES


