

# A Cost Model for E-learning projects in the United Kingdom

**Sue Moseley.** Moseley Multimedia Ltd. United Kingdom

**Raul Valverde.** Concordia University. Canada [rvalverde@jmsb.concordia.ca](mailto:rvalverde@jmsb.concordia.ca)

**Abstract.** This research aims to address the problems of estimating e-Learning development costs particularly within the United Kingdom. Hundreds of managers with no prior experience of managing e-Learning development often find themselves needing to produce cost estimations for e-Learning development and rely on structured expert judgement to produce accurate cost estimates. The lack of prior experience in e-Learning development means that these managers will not be able to apply structured expert judgement to their cost estimations and the risk of inaccurate estimations will be high, with all the subsequent problems this will bring with it. A solution needs to be found whereby these managers can produce accurate estimates without past experience. This research seeks to create a cost model, built on data from e-Learning projects in the United Kingdom which managers can use to produce cost estimates for e-Learning development.

Keywords: Regression for Cost estimation, Cost estimation for e-learning, E-learning projects, Multimedia cost analysis, COCOMO.

## 1. INTRODUCTION

Historically cost estimation for e-Learning development has been carried out on a 'rule of thumb' basis, meaning it is a heuristic approach based on judgement borne out of experience. This has meant that those carrying out the estimation

need to have considerable knowledge of e-Learning development along with knowledge of the actual costs on previous projects in order for the estimations to have any degree of accuracy. In recent years, the use of rapid e-Learning development tools such as Lectora (lectora.com) has been gaining pace. Furthermore, the recession has forced companies to look at ways of reducing their overheads. These two factors have prompted a move away from the previous norm of outsourcing to expert e-Learning development companies, with businesses choosing instead to develop their e-Learning in-house.

This creates a specific problem in terms of cost estimation. Those in charge of estimating the costs for an e-learning project in many cases do not have the specialist knowledge needed to make reasonably accurate estimates. Often those carrying out the development work are subject matter experts rather than e-Learning developers, so there is no pool of knowledge for managers to refer to when making cost estimations for e-Learning projects. The subject matter experts that do have some e-Learning authoring experience do not have e-Learning design experience, therefore any help they can offer managers in terms of e-Learning knowledge is limited and does not reflect the ‘whole picture’.

A solution is needed to this problem, so that the managers who will be responsible for overseeing e-Learning development can create cost estimates that are as accurate as those developed by managers with expert experience.

The purpose of this research study is to develop a cost model for e-learning projects in the United Kingdom, based on data extracted from previous e-Learning projects located in the United Kingdom which can be used to predict the cost of future e-Learning projects.

## **2. LITERATURE REVIEW**

### **2.1 Industry Sources**

The foremost question asked by clients when considering e-Learning is “How much will it cost?” Kevin Kruse, an e-Learning guru and Laurie Bassi, an expert in measuring the return on investment in staff through training, both address this issue. Although they begin to answer the question with “it depends” they differ in their viewpoints from there on. Kruse’s approach is typical of most e-Learning developer’s responses in that he states that the most accurate way of arriving at a cost estimate is to specify all the design details and then apply a cost to each detail

based on the number of estimated hours each team member will spend on the project, along with their hourly rate (Kruse, 2008). Bassi's approach is to say that "Costs are irrelevant. Cost-effectiveness is what matters" (Bassi, 2000). It is true that cost-effectiveness is important; ineffective e-Learning is money wasted, but costs are relevant. Businesses don't have unlimited budgets, they do have to work within a financial framework and they will often tailor the specification of a proposed e-Learning project to fit within a budget. Therefore it is important to be able to estimate costs for e-Learning development as accurately as possible.

## **2.2 Rule of Thumb vs Structured Expert Judgement**

In his book "Winning e-Learning Proposals" Kapp proposes a similar approach to Kruse in determining estimates for e-Learning. He splits the project into phases, then the phases into tasks before assigning a time and price for each task, then totalling these costs (Kapp, 2003). Both these approaches rely on judgement for estimation. Depending on the level of experience and knowledge behind the cost estimates, this can either be a 'rule of thumb' approach, or what Steve McConnell calls "structured expert judgement" (McConnell, 2006). McConnell makes a distinction between these approaches and states that most estimates are 'rule of thumb' or intuitive and that for estimates using a structured expert judgement approach, those responsible for completing each task should be involved in estimating the costs of those tasks. He argues that the 'rule of thumb' approach is inaccurate, a view supported by Lederer and Prasad in their article on estimating the costs of information systems "Nine Management Guidelines for Better Cost Estimating" (Lederer and Prasad, 1992) and that a structured expert judgement approach can be as accurate as model based estimates; which is supported by Jorgensen in his paper "A Review of Studies on Expert Estimation of Software Development Effort" (Jorgensen, 2004).

Lederer and Prasad refer to the 'rule of thumb' or intuitive approach as a "separate function approach" (Lederer and Prasad, 1992). By this they mean that the focus is on the relationship and communication between the estimator and end user; the estimation process is 'separate' from involvement with the actual developers who will be carrying out the work. This is fairly typical in e-Learning projects, where a project manager may be the single point of contact with the client. Often a project manager may have no programming or development experience and so agrees timescales and costs with the client that may be

unrealistic and inaccurate. The managers may be driven by a motive to win a contract, or keep a client happy and so produce cost estimates that are unrealistically low. More accurate cost estimates are achieved by what they call a “combined function approach” (Lederer and Prasad, 1992). In this case the cost estimation is carried out by those who will be carrying out the development, during a feasibility study. This equates to the “structured expert judgement” (McConnell, 2006) referred to earlier. This more accurate approach was the method used to estimate the project costs used in the tender data included in this study. The rest of the project cost data comes from actual cost data from completed projects.

In his review of studies in expert estimation, Magne Jorgensen concludes that there is no substantial evidence favouring the use of estimation models over expert estimates, and that in some cases expert estimates will be more accurate than formal estimation models (Jorgensen, 2004). This may be one reason why formal estimation models have not been embraced by the e-Learning community in general.

### **2.3 Lines of Code Approach**

Formal estimation models such as COCOMO (Constructive Cost Model) typically use lines of code (LOC) as their units of measurement. This does not fit comfortably in an e-Learning context for several reasons: Firstly the unit of measurement itself is alien to most people involved in the estimation of e-Learning development. E-Learning is typically developed using authoring, WYSIWYG or rapid e-Learning development tools, so unlike normal software development, developers are not working directly with the code, but rather with a visual interface for much of the time. Secondly, much of the development involves media such as vector graphics, photographs, video and audio, which is not represented by LOC. Finally, there are other aspects, such as the analysis, design, project management, and hardware and software costs that all have a major impact on the cost of a project, but cannot be represented by LOC.

Problems such as differences in productivity between programmers and differences in coding rate between types of software prompted Steve McConnell to state that “the LOC measure is a terrible way to measure software size, except that all other ways to measure size are worse” (McConnell, 2006) and that

although counting models are flawed, they are better than rule of thumb judgements. Given that an estimation model is better than rule of thumb estimation, where structured expert judgement is not an option (as in the NHS (National Health Service) situation) the best solution would be to provide a model that could be used by anyone without the need for expert judgement; only, a model that does not rely on LOC as the unit of measurement, but rather units of measurement that have a more natural fit to e-Learning development.

Another problem with the LOC model is that the number of LOC have to be estimated before the model can be used, so an “estimate is based on an estimate” (Pengelly, 1998). Pengelly carried out a comparison of different estimation models on a completed small software project for which the costs were already known. In addition to the COCOMO model, other approaches used included function point analysis and software lifecycle management. COCOMO, the LOC model came out most accurate with a relative accuracy of 8% (ibid). This confirms Steve McConnell’s view that even with its flaws; the LOC model is more accurate than most other estimation approaches.

The US Department of Defense is involved with the use of COCOMO for estimating e-Learning costs. As with other developers of e-Learning their previous way of estimating these costs had been the ‘rule of thumb method’ (Garnsey *et al*, 2006). Since the LOC approach does not fit naturally into the e-Learning framework they have developed a derivative of COCOMO II called COSCOMO (COCOMO SCORM Cost Model). This is a cost estimating tool for Shareable Content Object Reference Model (SCORM) conformant courseware. Instead of LOC the unit of measurement is “hours of courseware” (Smith and Edwards, 2006) and the data set it has been calibrated with consists of 9 Department of Defence projects (SPARTA and General Dynamics Information technology, 2006).

It could be argued that basing the model solely on military projects makes it unlikely that the tool would be of any use to the wider e-Learning community, especially since the original data set consists of only 9 projects, some of which do not represent the entire ADDIE (Analysis, Design, Development, Implementation and Evaluation) process, but the long-term plan for COSCOMO is that the model is calibrated against more data sets, from projects across the broader Advanced Distributed Learning (ADL) Initiative (Smith and Edwards, 2006). “The ADL

Initiative develops and implements learning technologies across the U.S. Department of Defense and federal government” (ADL, 2013). It is hoped that data from forty or more projects will be used in the model eventually (SPARTA and General Dynamics Information technology, 2006). The broader scope and greater number of projects that the model could be calibrated against would mean that over time the COSCOMO model could evolve into a useful tool for the e-Learning community in general, just as has happened with COCOMO for software development. The COSCOMO tool GUI (Graphical User Interface) currently takes the form of a Microsoft Excel add-in. Whether it will eventually take the form of a stand-alone piece of software as COCOMO has done is currently unclear.

#### **2.4 cMel**

cMel is a cost model for e-Learning projects proposed by Choi *et al* (2003) of Simon Fraser University in Canada. It is designed to offer a fast estimating method, able to provide ‘just-in-time’ cost estimations for e-Learning projects (Choi *et al*, 2003). The model uses activity based costing estimations, as individual tasks in the work breakdown structure of the e-Learning development are the units of measurement for the model. The model takes into account factors such as the complexity levels of tasks, which is very much in line with how a non-model method of e-Learning cost estimation takes place. Choi *et al* (2003) have gone on to apply the same model to a Blended Learning scenario (Choi, 2004).

Two data sets were used for the evaluation of cMel for e-Learning, one low quality course and one high quality course. The predictive validity of the model when applied to each data set proved to be 100% accurate! The reason for this would appear to lie in the fact that “synthetic data” (Choi *et al*, 2003), provided by the university itself had been used for the study. It is not clear whether or not the ‘project data’ had any real basis in actual e-Learning projects or was speculative based on what a project would typically be expected to cost. Even if the data was extracted from real projects, it was insufficient to carry out a regression analysis (*ibid*). The same applies to the application of the cMeL model to Blended Learning projects.

Real world data would have to be used to validate this model for it to have any application in the e-Learning community.

### 3. RESEARCH METHODOLOGY

#### 3.1 Data Collection

A multiple variable linear regression model was used to develop a cost model that can be applied to the cost estimation of e-Learning projects. Data was collected from previous projects with known costs and from tender documents from potential projects where structured expert judgement was used to estimate the costs for different e-learning projects located through the United Kingdom.

The collected data for the cost model came from twenty e-learning projects from multiple industries and different sizes of projects.

The independent variables considered for the model are as follows:

- Number of screens: For tutorial style e-Learning this is the number of pages; for system simulations a screen is considered to be each step of the procedure.
- Interactions: These are the points where the user interacts with the e-Learning, not including the interface, such as drag and drop exercises, or multi-choice questions. No distinction has been made between simple and complex interactions as there was insufficient data available.
- Graphics: These are vector drawings, diagrams, hand drawn images that have been scanned in etc.
- Photographs: Some of the photographs were supplied by the client themselves and some required the services of a photographer. All required some editing for use in the e-Learning.
- Animations: Many of these, particularly on project A were simple animated gifs, used to illustrate processes. Others were more complex Flash animations.
- Audio clips: All the audio involved hiring voice over artists, studio time, recording and editing.
- Video: All the video consists of clips taken from videos supplied by the client, so it does not cover the costs of shooting the video originally.

Independent variables were selected by using a stepwise regression that included a set of regression models in which the choice of predictive variables was carried out by using a sequence of F-tests. Once the final data set and independent variables were chosen, a multiple variable linear regression was carried out and a cost model developed which can be applied to cost estimating projects.

The model was evaluated by applying it to some of the projects used for the data set. The same projects had also their costs estimated by a 'rule of thumb' method, COCOMO 2 and COSCOMO and a comparison made between the results of the different cost estimation methods. The objective of the evaluation was to verify that the proposed method carries value as a cost estimation model for e-learning projects in the United Kingdom.

## **4. RESULTS AND EVALUATION**

### **4.1 Data Collection**

The data for collected the cost model came from invoices of previous projects carried out by a Multimedia development company located in the UK and from documents that were used to tender for work by the same company. The projects used for the study were then chosen from those that were valued from a value of £50,000 or more. The model does not cover implementation costs to the client themselves; any future application of the model to commercial software would need to take these factors into account. The costs quoted in the tenders were all estimated by those who would have been responsible for carrying out the development and therefore fall into the category of "structured expert judgement" (McConnell, 2006) referred to earlier. The quotes were all 'fixed fee' and therefore would have been the actual costs, had the projects gone ahead.

Projects will be listed as Project A, Project B etc for the purposes of confidentiality. The following table lists the type of company and the nature of the project:

Project A	Company: International bank
	Project: Highly interactive tutorial on internet technology
Project B	Company: Distributor of heating and plumbing products
	Project: An induction course for new staff in tutorial format
Project C	Company: International organisation for the promotion of cultural relations and educational opportunities
	Project: Induction course for overseas students coming to the UK to study
Project D	Company: Government department
	Project: Training for a new financial system using system simulation
Project E	Company: Government department
	Project: Systems training
Project F	Company: Postal service
	Project: Training counter staff in cross selling
Project G	Company: Motor breakdown service
	Project: Training in multimeter usage for engineers
Project H	Company: Manufacturer of aluminium cans and aerosols
	Project: Training on waste prevention and recycling
Project I	Company: UK bank
	Project: Training for a financial system using system simulation
Project J	Company: Educational government agency
	Project: Training for farmers on diversification

Table 1. Project Descriptions

Project K	Company: UN refugee agency
	Project: Management training using system simulation
Project L	Company: Software company
	Project: Introductory e-Learning module for a new banking system
Project M	Company: Educational government agency
	Project: Training for care workers
Project N	Company: Government department
	Project: Staff training, tutorial style
Project O	Company: Government department
	Project: System simulation training for debt management
Project P	Company: Commercial company
	Project: Management training using tutorial style e-Learning
Project Q	Company: Clothes and home-ware retailer
	Project: Product knowledge training
Project R	Company: Tool hire company
	Project: Induction and equipment training
Project S	Company: Jewellers
	Project: System simulation
Project T	Company: Insurance company
	Project: Tutorial and quiz

Table 2. Project Descriptions (continued)

## 4.2 Regression Analysis

A simple linear regression model was developed for each of the independent viable as indicated in table 1. Each independent variable was evaluated by looking at a scatter plot of that independent variable with the dependent variable, to see if there was a linear relationship between the two and if so, the strength of that relationship. If there was a linear relationship that significantly explains the relationship between the independent variables and the dependent variable then that demonstrated that a multiple variable linear regression model is an appropriate model to use. It will also reveal if there are any outliers – project data that is atypical, being numerically remote from the rest of the sample. The model would be improved by removing these projects from the data set altogether as they are not representative of a typical project and therefore not useful for prediction purposes.

Number of Screens	$Y = 1.6203X + 113.68$	$R^2 = 0.2576$
Interactions	$Y = 0.9865X + 45.338$	$R^2 = 0.1111$
Graphics	$Y = 1.269X + 21.257$	$R^2 = 0.1431$
Photographs	$Y = 0.2672X + 89.536$	$R^2 = 0.0148$
Animations	$Y = 0.1104X + 5.2657$	$R^2 = 0.0321$
Audio Clips	$Y = 2.1667X - 30.498$	$R^2 = 0.3967$
Video Clips	$Y = -0.0178X + 2.7211$	$R^2 = 0.2009$

Table 2. Regression equations for each independent variable

Based on the results displayed in table 2, neither the graphics nor the photographs produced a strong relationship; the graphics indicated that about 14% of the variability of the man days could be explained by the variability of the graphics and the photographs indicated that less than 1.5% of the variability of the man days could be explained by the variability of the photographs. This is surprising since still images form a large part of any e-Learning project. It was therefore decided to merge both these variables into one new variable ‘still images’ to see if this variable would produce a stronger correlation than either the graphics or photographs variable had done. A new data set was produced and the

scatter plots re-done to see if there had been any improvement on the quality of data to be used. This new set of regression equations is shown in table 3.

Number of Screens	$Y = 1.4549X + 90$	$R^2 = 0.4658$
Interactions	$Y = 0.7826X + 22.685$	$R^2 = 0.3413$
Still Images	$Y = 1.3687X + 86.157$	$R^2 = 0.4539$
Animations	$Y = 0.1458X - 4.8986$	$R^2 = 0.3235$
Audio Clips	$Y = 2.2062X - 16.319$	$R^2 = 0.4324$
Video Clips	$Y = -0.0178X + 2.8554$	$R^2 = 0.206$

Table 3. Revised Regression equations for each independent variable

A stepwise regression was carried out starting with the independent variable with the highest  $F$  value the other independent variables will be added one at a time in the order of highest  $F$  value. The  $F$  values have been taken from ANOVA tables which can be found in Appendix A. The independent variables were added as long as they improve the model, which will be determined by an increasing  $R^2$  value. The  $F$  values for the independent variables, in order starting with the highest are as follows:

- Number of Screens  $F$  13.95
- Still Images  $F$  13.30
- Audio Clips  $F$  12.19
- Interactions  $F$  8.29
- Animations  $F$  7.65
- Video Clips  $F$  4.15

All 6 independent variables have improved the accuracy of the model, so none of them will be removed from the final equation (see table 4). Therefore the equation for the mathematical model to calculate the number of man days a project will take will be:

$$Y = 32.13 - 0.7 * \text{Number of Screens} + 0.26 * \text{Still Images} + 0.13 * \text{Audio Clips} - 0.03 * \text{Interactions} + 0.86 * \text{Animations} - 7.81 * \text{Video Clips}$$

In order to convert the total estimated man days to an estimated financial value the predicted man days figure will be multiplied by the average daily rate for a project. For the purposes of evaluation the average daily rate will be presumed to be £350.00, which was a typical daily rate in the projects used for most of the data set. Some projects were priced as low as £300.00 per day and parts of some other projects were priced as high as £400.00 per day, but £350.00 is a reasonable approximation for the purposes of evaluation.

Step 1	$Y = 15.18 + 0.32 * \text{Number of Screens}$	$R^2 = 0.47$
Step 2	$Y = 15.18 + 0.28 * \text{Number of Screens} + 0.05 * \text{Still Images}$	$R^2 = 0.47$
Step 3	$Y = 14.54 - 0.14 * \text{Number of Screens} + 0.37 * \text{Still Images} + 0.14 * \text{Audio Clips}$	$R^2 = 0.61$
Step 4	$Y = 14.64 - 0.2 * \text{Number of Screens} + 0.48 * \text{Still Images} + 0.19 * \text{Audio Clips} - 0.19 * \text{Interactions}$	$R^2 = 0.63$
Step 6	$Y = 14.43 - 0.63 * \text{Number of Screens} + 0.82 * \text{Still Images} + 0.04 * \text{Audio Clips} - 0.19 * \text{Interactions} + 2.12 * \text{Animations}$	$R^2 = 0.74$
Step 7	$Y = 32.13 - 0.7 * \text{Number of Screens} + 0.26 * \text{Still Images} + 0.13 * \text{Audio Clips} - 0.03 * \text{Interactions} + 0.86 * \text{Animations} - 7.81 * \text{Video Clips}$	$R^2 = 0.78$

Table 4. Stepwise regression

### 4.3 Evaluation

Six projects have been chosen from the original data set; three system simulations and three interactive tutorials with assessments for evaluation purposes.

Project	Cost	Man Days	Screens x <sub>1</sub>	Images x <sub>2</sub>	Audio x <sub>3</sub>	Interactions x <sub>4</sub>	Animations x <sub>5</sub>	Video x <sub>6</sub>
System Simulations								
1	£71,400.00	204	400	400	780	390	0	0
2	£63,350.00	181	300	240	420	120	60	0
3	£19,950.00	57	120	120	0	110	0	0
Interactive Tutorials								
4	£61,600.00	176	360	324	450	105	36	0
5	£38,150.00	109	250	262	250	62	8	1
6	£21,000.00	60	124	124	0	32	0	0

Table 5. Data Set of Six Projects for Evaluation

The percentage difference was calculated in the following way with the use of the proposed cost model:

1. Actual Man Days – Predicted Man Days = Difference in Man Days
2. Difference in Man Days / Actual Man Days x 100 = Percentage Difference

This was then repeated using the costs to confirm the figures:

1. Actual Costs – Predicted Costs = Difference in Costs
2. Difference in Costs / Actual Costs x 100 = Percentage Difference

The results of the comparison of the proposed cost model and actual cost is summarized in table 6.

Project	Actual Man Days	Predicted Man Days	Actual Costs	Predicted Costs	Percentage Difference
System Simulations					
1	204	197.83	£71,400.00	£69,240.50	3.02
2	181	176.13	£63,350.00	£61,645.50	2.69
3	57	51.63	£19,950.00	£18,070.50	9.42
Interactive Tutorials					
4	176	177.48	£61,600.00	£62,118.00	0.84
5	109	112.46	£38,150.00	£39,361.00	3.17
6	60	54.73	£21,000.00	£19,155.50	8.78

Table 6 Comparison of Actual and Predicted Costs for proposed Cost Model

The model was evaluated had their costs estimated by a ‘rule of thumb’ method, COCOMO 2 and COSCOMO and a comparison made between the results of the different cost estimation methods.

Kruse’s article on estimating e-Learning budgets he refers to two common methods used for ‘rule of thumb’ approaches to estimation, namely estimating by completed hour of e-Learning and estimating by the number of screens (Kruse, 2008). Since the data for completed hours of e-Learning is unavailable the latter method will be used. Kruse quotes a cost of \$60,000 per 120 to 150 screens of e-Learning (ibid). \$60,000 is £36,209.05 at the time of writing this paper; this figure will be rounded to £36,000 for the estimates and a mid-way figure of 135 will be used for the number of screens. A cost of £266.67 per screen is arrived at by dividing £36,000 by 135 screens; therefore the predicted costs will be calculated by multiplying the number of screens by £266.67 (See table 7).

Project	Number of Screens	Actual Costs	Predicted Costs	Percentage Difference
System Simulations				
1	400	£71,400.00	£106,668.00	49.39
2	300	£63,350.00	£80,001.00	26.28
3	120	£19,950.00	£32,000.40	60.40
Interactive Tutorials				
4	360	£61,600.00	£96,001.20	55.85
5	250	£38,150.00	£66,667.50	74.75
6	124	£21,000.00	£33,067.08	57.46

Table 7. Comparison of Actual and Predicted Costs for Rule of Thumb Method

The COCOMO model uses lines of code as its unit of measurement. To arrive at a figure of lines of code for each project the source code from previous projects was analysed. Three different screens of source code were looked at, one for a typical text only screen, one for an interactive screen (a multiple choice question) and one for a mixed graphic and text screen. The COCOMO manual specifies that the lines of code figure should not include any source code that has been generated with a source code generator. As e-Learning is typically created using WYSIWYG tools the only source code counted was lines of code that would have involved direct input from the developer. The number of lines for each type of screen were added together and divided by the number of screens analysed to produce an average figure for lines of code per screen. This came out at 13 lines of code per screen. This figure was then multiplied by the number of screens for each project to determine the figure for lines of code to be used for each project. The labour costs for each project were determined by multiplying the daily rate

used in the multiple regression model (£350 per day) by 22, which is the typical number of working days in an average month. The COCOMO model uses US dollars but for the purposes of this project these can be assumed to be British pounds. All other settings were left at their default.

The cost estimates produced were extremely generous so the optimistic costs will be used for comparison with actual costs (see table 8).

Project	Number of Screens	Actual Costs	Predicted Costs	Percentage Difference
System Simulations				
1	400	£71,400.00	£157,617.80	120.75
2	300	£63,350.00	£80,895.02	27.69
3	120	£19,950.00	£29,532.97	48.03
Interactive Tutorials				
4	360	£61,600.00	£98,854.72	60.47
5	250	£38,150.00	£66,198.20	73.52
6	124	£21,000.00	£30,617.33	45.80

Table 8. Comparison of Actual and Predicted Costs for COCOMO

The COSCOMO model uses person months as its unit of measurement. It does not calculate actual costs, so for the purposes of this evaluation the person months figure will be converted into man days and then into an actual cost figure by multiplying by £350 as in the mathematical model.

The COSCOMO model is e-Learning specific and takes into account a great many factors in the development of e-Learning, including factors such as how cooperative the development team are with each other; how many years

experience they have had with the development software and whether the team members are all located in the same office. A small alteration in the level of experience of the development team can make significant differences to the number of person months in the results.

COSCOMO gave its estimate as .4 person months which is the equivalent of 8.8 days whereas the actual figure is 60 days. 8.8 days at £350 per day would put the estimated cost for this project at £3,080.00. The actual cost was £21,000.00 (see table 9).

Project	Number of Screens	Actual Costs	Predicted Costs	Percentage Difference
System Simulations				
1	400	£71,400.00	£3,850.00	94.61
2	300	£63,350.00	£3,080.00	95.14
3	120	£19,950.00	£1,540.00	92.28
Interactive Tutorials				
4	360	£61,600.00	£5,390.00	91.25
5	250	£38,150.00	£4,620.00	87.89
6	124	£21,000.00	£3,080.00	85.33

Table 9. Comparison of Actual and Predicted Costs for COSCOMO

Project	Actual Costs	Regression Model Percentage Difference	Rule of Thumb Percentage Difference	COCOMO Percentage Difference	COSCOMO Percentage Difference
System Simulations					
1	£71,400.00	3.02	12.04	120.75	94.61
2	£63,350.00	2.69	5.29	27.69	95.14
3	£19,950.00	9.42	20.30	48.03	92.28
Interactive Tutorials					
4	£61,600.00	0.84	16.88	60.47	91.25
5	£38,150.00	3.17	31.06	73.52	87.89
6	£21,000.00	8.78	18.10	45.80	85.33

Table 10. Comparison of All Models

According to table 10, the multiple variable linear regression model is the most accurate cost estimator, with all predicted costs within ten percent and four of the six within four percent of the actual cost. This is followed by the rule of thumb estimations, with most of the estimates falling around or within twenty percent. This demonstrates why the use of rule of thumb estimates continues in the e-Learning environment as a way of giving quick ball park estimates; it clearly still has some usefulness for giving a rough idea of the budgets that will be needed by clients considering whether or not to pursue an e-Learning route to staff training.

The COCOMO model produced very inaccurate results which is not surprising given that its unit of measurement is lines of code and that automatically generated code is not included in the calculation. Virtually all e-Learning is developed with some sort of authoring tool that automatically generates code;

therefore COCOMO is not a practical way of estimating the costs of e-Learning development.

The COSCOMO model is an e-Learning specific derivative of the COCOMO model and looking at the very detailed input fields and the nature of the questions asked it was surprising that this was the most inaccurate model. Great care was taken to fill out each section correctly; the model comes with a user manual and help pages for each of the input fields and yet the estimates produced were wildly inaccurate, all of which were drastically underestimated. It is hard to see how, for example in project six, it could arrive at a figure of less than nine days for a developer to produce two hours worth of completed courseware, especially as one third of the finished courseware time was for highly interactive material and all of the courseware was to be built from scratch, with no reused content. Even if all the other input fields had been incorrectly filled in, the courseware hours in final product and reuse of existing courseware on the first page should have been enough information to produce a much higher person day's estimate.

The COSCOMO model was based on massive e-Learning projects within the US military and it appears that it is not a suitable model for much smaller scale projects. As the model is developed by the addition of other projects to the data set, maybe it will become more relevant over time, but it is unusable for small projects as it is.

The survey consisted of 28 questions, the total participants were 63 participants however, 34 of those successfully completed all questions in the survey, 29 participants either viewed the first pages only or filled partial information and hence these were excluded from the final analysis.

Participants were mostly from technologies departments and audit however there were some participants who were in HR and technical sales teams.

## **5. CONCLUSIONS**

The research has shown that the proposed cost model works as hoped for and that it has shown to be more successful than other tools in calculating the costs of e-learning projects. There are some limitations to this study, the cost model would benefit from being based on a much larger data set of project data and is incomplete due to the fact it does not take into account other stages of e-learning development such as the analysis, design, implementation and evaluation. Nor

does it take into account impact on staffing costs as a result of needing to take time out for using the e-Learning. A much larger data set would produce a much better mathematical model with wider application.

The project data all came from the sponsor's own information on past projects. If data could be obtained from a much wider source of information; from many companies and government departments, a more detailed model could be created that could take into account the complexity of different variables. The model as it stands, would not be sufficient for the NHS or any company to use 'as is', due to the fact it cannot be applied to the whole e-learning process. Expert judgement would still be needed to estimate analysis, design, implementation and evaluation costs. It is though a good starting point and does prove that a solution to the problem of lack of experience in e-learning managers is possible.

It could well be possible to develop this further to produce a commercial product or open source tool for e-Learning development cost estimation. Whether or not this will happen will be likely to depend on the availability of data from a wide range of projects. Companies tend to jealously guard financial information, particularly when it comes to what they charge their customers for e-Learning, compared with what it actually costs them to produce. Therefore it doesn't seem likely, unless a suitable incentive can be found that companies will be willing to part with this information. Perhaps though, an organisation such as the NHS, which would benefit greatly from having such a tool, would be willing to part with their project data, if in doing so they would gain a better ability to manage budgets for e-Learning.

## 6. REFERENCES

- ADL (2013): *ADL Initiative Introduction*. <http://www.adlnet.gov/scorm/>
- BASSI, L. (2000): *How Much Does e-Learning Cost?* <http://bit.ly/1qWwfhp>
- CHOI, A. (2004): "An integrated cost model for blended learning environments". In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Vol. 2004, No. 1: 1012-1019.
- CHOI, A.; HATOL, J.; KUMAR, V. (2003): "Cost Modeling for eLearning". In A. Rossett (Ed.), *Proceedings of World Conference on E-Learning in Corporate*,

*Government, Healthcare, and Higher Education: 1554-1557*. Chesapeake, VA: AACE.

GARNSEY, M.; EDWARDS, L.; WARD, K.; MARVIN, D. (2006): "COCOMO and SCORM: Cost Estimation Model for Web-Based Training". In *The Interservice/Industry Training, Simulation & Education Conference (IITSEC)*, Vol. 2006, No. 1. National Training Systems Association.

JØRGENSEN, M. (2004): "A review of studies on expert estimation of software development effort", *Journal of Systems and Software*, Vol. 70, No. 1: 37-60. [http://dx.doi.org/10.1016/S0164-1212\(02\)00156-5](http://dx.doi.org/10.1016/S0164-1212(02)00156-5)

KAPP, K.M. (2003): *Winning e-Learning proposals*. Florida: J. Ross Publishing Inc.

KRUSE, K. (2008): *How Much Will It Cost? Estimating e-Learning Budgets*. e-LearningGuru.com. <http://bit.ly/1mMAIH0>

LECTORA.COM missing citation from page 94

LEDERER, A. L.; PRASAD, J. (1992): "Nine management guidelines for better cost estimating", *Communications of the ACM*, Vol. 35, No. 2: 51-59. <http://dx.doi.org/10.1145/129630.129632>

MCCONNELL, S. (2006): *Software Estimation*. USA: Microsoft Press.

PENGELLY, A. (1998): "How long is a piece of string?", *The Computer Bulletin*, Vol. 40, No. 2: 14-15. <http://dx.doi.org/10.1093/combul/40.2.14>

SMITH, R.; EDWARDS, L (2006): "COCOMO-SCORM Interactive Courseware Project Cost Modeling", *International Council of Systems Engineering Conference*.

SPARTA & GENERAL DYNAMICS INFORMATION TECHNOLOGY (2006): COSCOMO Prototype Project US: SPARTA & GDIT. <http://bit.ly/1qy9G0r>

## Appendix A

### ANOVA Tables

#### A1. Number of Screens

##### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	148968.9213	148968.9213	13.95317125	0.001802491
Residual	16	170821.5787	10676.34867		
Total	17	319790.5			

#### A2. Still images

##### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	131843.701	131843.701	13.297392	0.002174466
Residual	16	158640.077	9915.00481		
Total	17	290483.778			

#### A3. Audio Clips

##### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	342541.3192	342541.32	12.19027345	0.003017503
Residual	16	449592.9586	28099.56		
Total	17	792134.2778			

## A.4 Interactions

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	43210.93	43210.93	8.28865697	0.010905785
Residual	16	83412.18	5213.261		
Total	17	126623.1			

## A.5 Animations

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	1496.351332	1496.351332	7.650482198	0.013774645
Residual	16	3129.426446	195.5891529		
Total	17	4625.777778			

## A.6. Video Clips

## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	14500.46948	14500.46948	4.152376147	0.058463777
Residual	16	55873.4333	3492.089581		
Total	17	70373.90278			