International Journal of Engineering Education Vol. 28, No. 6, pp. 1294–1300, 2012
Printed in Great Britain
0949-149X/91 $3.00+0.00

Multiple Choice Answers Approach: Assessment with Penalty Function for Computer Science and Similar Disciplines*

IGOR SCHAGAEV, NEBOJSA FOLIC and NICHOLAS IOANNIDES
London Metropolitan University, Faculty of Life Sciences and Computing, 166–220 Holloway Road, London, N7 8DB, UK.
E-mails: {i.schagaev, n.folic, n.ioannides}@londonmet.ac.uk

ELISABETH BACON
University of Greenwich, School of Computing and Mathematical Sciences, Maritime Greenwich Campus, Old Royal Naval College, Park Row, Greenwich, London, SE10 9LS, UK. E-mail: e.bacon@greenwich.ac.uk

A new assessment technique, the Multiple Choice Answers Approach (MCAA) with several right and not mutually exclusive answers from a larger set of answers, is introduced. MCAA, even though is developed as an assessment technique, it also facilitates the formation of reliable coverage of subject curricula. The rigorous formation of assessment becomes an essential segment of curriculum design and development, follow-up of our previous papers. Levels of applicability are explained showing the practical effectiveness of the proposed approach. A trial application in a real module of computer science showed promising results. Schemes of student progress assessment are introduced including analysis of success and protection from guessing, using penalty functions. Schemes of further development and directions for further research are discussed.

Keywords: multiple choice answer approach; computer science; penalty function; curriculum design; curriculum development; theory of classification; formalisation

1. Introduction

Computer Science, as a teaching discipline, absorbs different theoretical and technical results from other disciplines and creates a fusion that penetrates and influences many aspects of human life. It is the theoretical base for the fastest ever growing area of technological development, that of Information Technology (IT).

Computer Science should absorb its own technical applications, IT, but previous attempts to do so have failed. It is, therefore, necessary to create a 'bridge over troubled water' and make *curriculum design*, *development* and *assessment* more connected and, where possible, rigorously designed. The need for such an effort evidently grows day by day, as one of the fathers of Computer Science, Dijkstra, in his letter [1] to the Communications of the ACM, admits: *'I would therefore like to posit that computing's central challenge, viz. 'How not to make a mess of it,' has not been met. . . . You see, while we all know that unmastered complexity is at the root of the misery, we do not know what degree of simplicity can be obtained, nor to what extent the intrinsic complexity of the whole design has to show up in the interfaces. . . . To put it bluntly, we simply do not know yet what we should be talking about. . .'*

The issue of *curriculum design* (and development) was addressed in our previous publications [2, 3]. We attempted to form a logical *core* of *curriculum design* for Computer Science and showed an approach towards the formalisation of the curriculum design process and adjustment of the teaching delivery process. This approach considers curriculum design from the *information processing* point of view and develops it further through the introduction of three main functions in discipline construction: *definitive*, *characteristic* and *predictive*.

In this work, the authors focus on the issue of *assessment* and address this as part of curriculum development. A new type of assessment, *Multiple Choice Answers Approach* (MCAA) is introduced, and is supported by a scheme for its implementation as well as suggestions for a *penalty function*. Our goal is to create a new assessment methodology, which, apart from assessing the students, will also enable us to estimate the quality of teaching delivery. The creation of an automatic (or semi-automatic) assessment procedure along with the development of an algorithm for the selection of the most important material for assessment will help in estimating the quality of teaching delivery. A result from such an assessment for a concrete module from Computer Science is also presented.

The sequence of steps towards delivering knowledge should generally be completed in scheduled time and in such a manner that the 'average' student could cope with the speed of its delivery as well as be able to continue learning. Furthermore, attention should be paid to the following sensible comments:

* Accepted 20 July 2012.

- Time spent on exams and tests is, in fact, time taken from learning and should thus not be considered part of learning time.
- The sequence of learning is not supported as it should by the use of IT as its involvement in knowledge delivery concerns primarily presentation aspects, and in assessment on the mandatory popular Multiple-Choice Questions (MCQ).
- There is no peer-to-peer evidence that knowledge is delivered successfully until the assessment was completed and passed after which feedback can be extracted, even though by then it may be too late for that.

By considering the following facts about Computer Science and other similar disciplines, one realizes that the need for this work becomes multi-fold:

- Computer science, in general, suffers from a snowball of information, useful and otherwise [1], which cannot be handled, properly processed or justified for teaching purposes
- Standard analysis of the learning outcomes mostly serves to indicate the general performance through the use of statistics and has very limited module or student meaning
- The level of competency of the students varies enormously
- Even though social demand to help students with different backgrounds get high levels of education is present, the resources to match these demands are limited.

If the strategic aim of this work is to be achieved, Computer Science as a discipline for teaching must be modified to incorporate IT, tuned for purpose, thus resulting in a more efficient and effective teaching process.

2. Assessment

Methods with which an assessment using the model of curriculum design and development proposed [2, 3] can be formed are discussed in this section. By considering the requirements for assessment, a nearly formal assessment procedure could be realized that is also free from existing drawbacks. This procedure is called *Multiple Choice Answers Approach* (MCAA) and should not be confused with multiple choice questions as it is very different. A process for the formation of a questionnaire and a scheme to maximize the efficiency of assessment, thus increasing the objectivity of the curriculum, are also introduced.

2.1 Assessment requirements

We consider the most important requirements and features of assessment to be:

- Objectivity.
- Quantitative analysis.
- Time (and other resources) efficiency.
- Concurrency with the learning process.

The popularity of assessing using MCQ has been growing very fast, especially during the last decade, driven by the help that it provided teachers with the complex and boring procedure of marking and calculation of assessment results, has been growing very fast. And even though MCQ is a relatively new invention, which enables the applicability of IT to be used as a tool for assessment, it was considered a panacea. However, even though successful in that, the MCQ approach, disappointedly, does not provide proof of efficiency and growth of knowledge. The fact is that it does not have a connection with real delivery of *knowledge* or a measure of the quality of knowledge delivery—an *assessment* of its result. The standard assessment used to be limited by definitive descriptions, i.e. without the use of the set of functions, *characteristic* {CF} and *predictive* {PF}, as described in [2, 3]. According to [4], Bertrand Russell claims that: *'to be directly acquainted with something is to be in a position to give it a name in the strict logical sense, and to know something only by description is to know only that something uniquely fits the description'.* But to give something a name and to be able to understand and use that name correctly are two different things . . .

2.2 Formation of the multiple choice answers approach (MCAA) assessment

The *Multiple Choice Answers Approach* that enables the effective use of IT for assessment and matches the requirements set in section 2.1 above, is presented here. Compared with its predecessor (MCQ), the MCAA, which is based on the process of curriculum design and development presented in [2, 3], enables the application of IT for assessment, increases the reliability of knowledge, eliminates guessing, and can help manage the assessment of thousands of students at the same time!

Assume that the discipline in question, Computer Science or other similar, has sets of definitions (Definitive Function, DF), a set of key characteristics about how definitions are connected (Characteristic Function, CF) and, therefore, a set of known predictions (Predictive Function, PF) and that all elements of these sets are known. A question q_i can then be created which covers (includes) several terms from DF_i ($DF_i \subset DF$, segment of the discipline) and $q_i \in CF_i$ ($CF_i \subset CF$, questions on the segment of the discipline DF_i) as shown in Fig. 1.

Suppose we organise a table from the elements of {DF} called *Working Table*, $WTDF_u$. Inside $WTDF_u$ we have overlapped areas in DF_1, . . .,

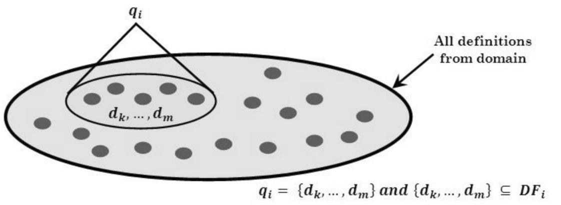


Fig. 1. Formation of a question using {DF} and {CF}.

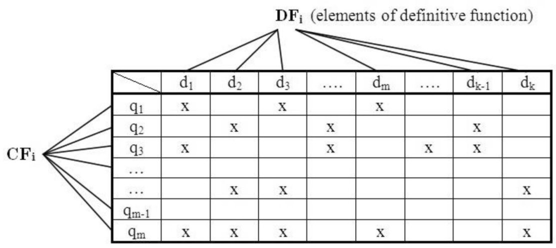
DF_u ($DF_i \cap DF_j \neq \emptyset$, $i \neq j$) and where some definitions might belong to the various questions, as shown in Equation 1, and be 'more important' than others:

$$\exists d_k \in DF_i : d_k \in DF_j, i \neq j \qquad (1)$$

By organising sub-tables $WTDF_i$ for each DF_i as an elementary segment of knowledge delivery – could be a lecture, seminar, etc. – we can create a question q_i with explanations for each term $d_k, \ldots, d_m$ from DF_i as shown in Fig. 2. All elements from DF_i should be covered by at least one question from CF_i. We can then form subsets of questions $\{q_i, \ldots, q_t\}$ to construct the full set $Q_i = \{q_1, \ldots, q_i\}$ where Q_i covers the whole area of definitions from DF_i. Very importantly, the same terms from $\{d_k, \ldots, d_m\}$ can be involved in the answers of different questions.

Once table $WTDF_i$ is developed, the number of definitions required for the correct answer of question q_i becomes available. Furthermore, the table will also allow us to create more than one question with multiple right definitions from the same domain (DF_i), if necessary. All of them (say m) could be used together with all other possible definitions' domains relating to DF_i. The correct answer will then mean 100% hit of right m from n. In fact, the full set of questions for one fragment of the course will form CF_i. Using other definitions from the same DF_i for the construction of different questions will actually enforce the student learning during assessment or the trial assessment exercise.

Formally, the procedure to create a question set is completed when the resultant questions cover all elements from DF_i. But this condition is not unique. We will, later on, show that there are other conditions with their own merits in forming a question set around DF_i.



	d_1	d_2	d_3	….	d_m	….	d_{k-1}	d_k
q_1	x		x		x			
q_2		x		x			x	
q_3	x			x		x	x	
…								
…		x	x					x
q_{m-1}								
q_m	x	x	x		x			x

Fig. 2. Connecting questions and definitions.

The level of overlapping between working tables $WTDF_i$ developed for a module assessment will indicate how directly connected with the structure of the curriculum they actually are. The formation of a question set, however, has both objective and subjective aspects: coverage of the module and the students' abilities and psychology respectively. Thus, the overlapping of WTDF itself might be an interesting subject for further research.

2.3 Formation of the questionnaire

Let's assume that several questions have been built around the same subset of terms: $q_i \in Q_i$ and $\{d_k, \ldots, d_m\} \subseteq DF_i$ and that questions from Q_i cover DF_i as shown in Fig. 3.

The procedure for the generation of set Q_i is completed when all terms of the subject (or its part) are involved in the questions and, therefore, are covered by answers of some or, at least, one question. Or formally (Equations 2 and 3):

$$(\forall d_k : d_k \in DF_i)(\exists q_i \in Q_i : d_k \in q_i) \qquad (2)$$

Alternatively:

$$\bigcup_{q_i \in Q_i} q_i = DF_i \qquad (3)$$

If the set of questions uses several terms more than once, we can say that the subset of mostly used terms forms a core of a discipline (or its part). The rule of belonging to the core of definitions in the discipline segment DF_i is fairly simple—if there exist questions created on this segment of definitions (DF_i) and the intersection of these questions in the number of terms used from DF_i are the biggest, then these questions and elements from the segment of definitions form a core of the segment, or formally (Equation 4 and 5):

$$CF_{i-core} = \{\forall q_j, q_k \in Q_j : q_k \Leftrightarrow q_k(DF_i), q_j \Leftrightarrow q_j(DF_i) \text{ and } |q_k(DF_i) \cap q_j(DF_i)| = \max\} \qquad (4)$$

$$DF_{i-core} = \bigcap_{q_i \in CF_{i-core}} q_i \qquad (5)$$

The formation of a set of questions is the formation of CF, since these questions show how terms

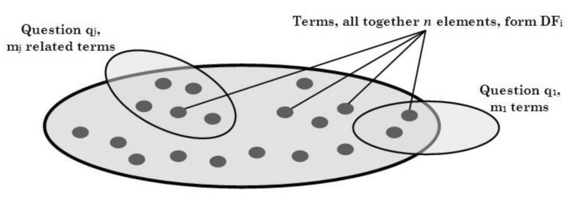


Fig. 3. Formation of DF_u.

are connected and characterised. The selection of the core terms from DF can be combined with the selection of the core questions from CF. This double core, if determined, can be really useful for the formation of the course assessment, formally (Equations 6 and 7):

$$DF_{core} = \bigcup_{i=1}^{n} DF_{i-core} \tag{6}$$

$$CF_{core} = \bigcup_{i=1}^{n} CF_{i-core} \tag{7}$$

The hierarchy of the importance of the questions can also be interesting as it helps to form an essential set of questions towards the required level of knowledge (grade) and adjust this level to different student abilities. Assume that for grade A we have one set of questions Q_a, for B we have Q_b, for C we have Q_c and that $Q_a \supset Q_b \supset Q_c$, where A, B, C are grades of student marks. But the most important aspect of the MCAA approach is that it avoids the well known problem of MCQ that of *knowing by name*. In fact, MCAA enforces students and learners to dig deeper for the meaning of definitions along with their interconnections and known predictive functions, thus pushing researchers and students to discover new knowledge—increasing the power of {PF}. R. Feynman [5], in his interview for the BBC, emphasises the importance of such an approach, in our case for curriculum and assessment design and development: *'When you finished all that naming, you know absolutely nothing what that bird is; now let us look at the bird and discover what it is doing . . .'*

3. Multiple choice answers approach (MCAA) efficiency

Suppose we have DF_i which has cardinality n, i.e. n terms in it and $n = |\{d_1, \ldots, d_n\}|$ as shown in Fig. 4. The level of their connections does not matter here.

Let's assume that a set of questions $Q_i \subseteq CF_i$ which covers and includes all elements of DF_i has been built. As an example, for memory based questions in a computer architecture course, this could be: Memory: DRAM, SRAM, VRAM, Address, Data, etc. Full understanding of the meaning of the question q_j assumes a selection of right m related to the question terms, i.e. 100% correct answer for q_j is m from n possible choices.

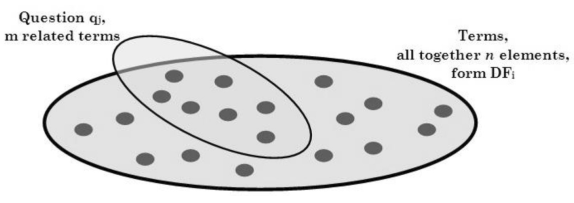


Fig. 4. Formation of questions q_i around definition domain DF_1.

To avoid guessing by the students, the best ratio of m and n can be found, using features of binomial coefficients, to be (Equation 8):

$$2 \times m = n \tag{8}$$

This relationship between the numbers of right and wrong answers in the proposed multiple choice answers approach can be proven by considering the number of different m options from n possible answers which is determined by the classic formula of binomial coefficients (Equation 9):

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} \tag{9}$$

Our task here is to find a maximum of (4) making the chances of guessing in MCAA negligible. Let's consider two extreme examples: when $m = n$-1 and when $m = 1$. For these two variants equal results can be found (Equation 10):

$$\binom{n}{1} = \binom{n}{n-1} = \frac{n!}{(n-1)!(1)!} \tag{10}$$

Also, because we have $n!$ in the upper part of the ratio in both variants it is easy to see that the minimum of $m!(n-m)!$ is achieved when $m! = (n-m)!$ and, thus, $n = 2m$.

Suppose that $n = 2a$, and $m = n/2$, then (Equation 11):

$$\binom{n}{m} = \binom{2a}{a} = \frac{(2a)!}{a!a!} = \sum_{i=0}^{a} \binom{a}{i}^2 \tag{11}$$

For $n = 8$ and $m = 4$ the probability of the occasional hit on a right answer, even if the student knows that $m = 4$, is 1/70 (1.43%) confirming the validity of this approach as a means towards excluding guessing and cheating during exams. Even semi-automatic procedures of formation of questions for assessment can decrease the workload of teachers. There is no doubt that there is much more to be done in the approbation of this approach in various universities and around different disciplines.

There will, of course, be cases where the terms in the answer would be selected incorrectly and we would thus have: Selection (q_j (DF_i)) < m. Furthermore, we may also face the case when all buttons might be pressed, i.e. n from n. For these cases a *penalty function* should be established, which will decrease the mark for each wrong answer and only count the right ones.

Various forms of penalty function do exist and selection of some *best-fit* functions can be developed even though the approach towards the selection of best-fit functions requires further research. The distribution of answers, of course, varies. The best answers on questions q_j have hit ratio 1 (m right answers from m (the same) options).

The form and severity of the penalty function is the subject of special interest and may be different whether on assessment or training purposes. A penalty function can be constructed where its effect could be 'severe' in examinations, whereas one might choose another form of 'kind testing' scheme for training purposes. During tutorials, for example, student should be able to experiment with MCAA. The penalty function in this case should be somehow forgiving or even soft to enable more wrong answers without visible penalization. On the contrary, during assessment exercises the penalty function should become severe to exclude guessing. Changing the questions for each tutorial by noting which questions a student has previously answered, provides us with the facility to automatically force student to answer all questions from the discipline, and, in fact, to receive an essential knowledge and be assessed *concurrently* with learning! We feel that this feature of MCAA makes this approach promising.

Precision of the answers depends on the n and m selections and their sizes. When a candidate does not choose all right boxes for questions correctly and does not hit any wrong answers his / her total mark can be calculated by deduction of the ratio $|m_r|$ / $|m|$ from the total mark (equal 1) (Equation 12):

$$\textit{Mark in absence of wrong answers} = (1 - |m_r|/|m|) \tag{12}$$

Where: m_r is the set of not mentioned correct answers.

Thus, a candidate who misses 2 from 4 right answers gets a total mark of:

$$1 - (2/4) = 0.5, \text{ or } 50\%$$

If the question is required to be weighted, say, cost 30 marks then the student with the sample answer above, will be awarded 15 marks. A student, in turn, can make mistakes and may respond with incorrect answers. Again, the number of wrong answers should be weighted with the total number of wrong answers (equal to the number of right answers, see above) and deducted from the total mark (Equation 13):

$$\textit{Mark when wrong answers are present} = (1 - |m_w|/|m|) \tag{13}$$

Where: m_w is the set of wrong answers identified by student.

A generalized relative mark is calculated using the following set theory equation (Equation 14):

$$T_m = 1 - (1/|m|)(|m - m_a| + |m_a \cap (n - m)|) \tag{14}$$

Where: m_a is the answer provided by the candidate; m is the right answer, $n – m$ is the relative complement. Note that $m_r \cup m_w = m_a$.

At the same time, further research regarding the transition from set theory equations into arithmetic equations to calculate the grade in the assessment is needed. Formalisation of assessment procedures by use of set theory Euler, Venn or Peirce diagrams might not be so easy. Forms of penalty function must consider two variables— the number of right answers and the number of all possible answers— as these are dependent on each other. Preliminary work on these diagrams failed to convert a decision rule from set theory into an arithmetic rule in order to get a grading rule. The first impression is that the resolution of uncertainties mentioned here might be found if models for DF, CF and PF and the scheme of assessment using MCAA could be described with more than two dimensions.

4. Trial application

A trial application of the MCAA approach to assessment in real everyday practice (module: Network Technologies) in the School of Computing at the London Metropolitan University, showed very promising results. With MCAA, students achieved much better marks than those achieved by students on the previous year when MCAA was not used either for practice or for assessment. An example of a typical MCAA question used is shown in Fig. 5.

Statistics accumulated, an essential part of module development procedures, for the specific module by the Assessments Unit of the university supports this claim. Comparative statistics of final results from this year, using MCAA, and last year are presented in Fig. 6.

The data in Fig. 6 show a real trend of improvement in the module results—less failures and more students achieving A and B grades—confirming that the discussed changes and the way of implementing these changes seem to be workable. The result, as a whole, looks very promising even though it is too early to discuss efficiency of separate elements in any changes planned and realised.

5. Further work

Further development of this work will initially focus on the formation of a framework and design of a tool for automatic curriculum design and update

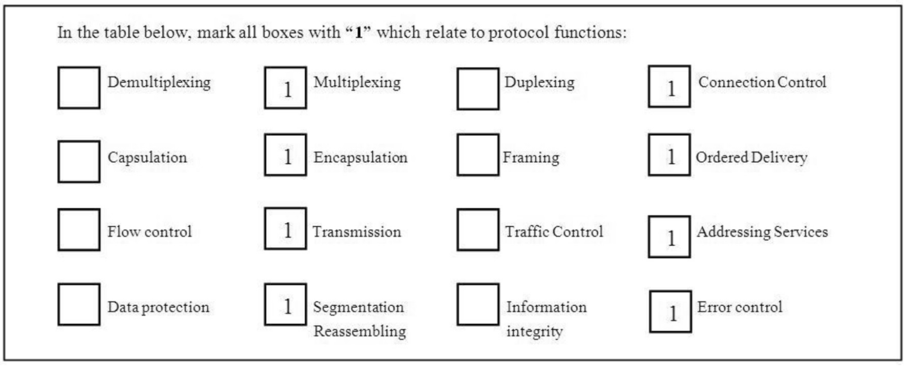


Fig. 5. An example of a typical MCAA question.

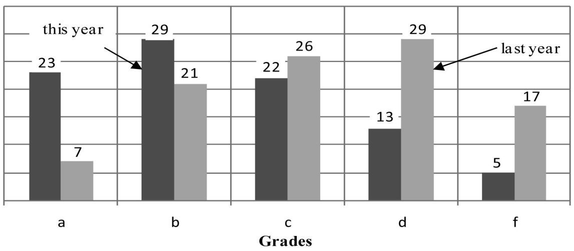


Fig. 6. Statistics of MCCA approach: 2 sequential years.

using the web and any available electronic resources. Various assessment modes including training and testing are to be introduced using various supportive/advising schemes and various penalty functions. In this way, rewarding the students for choosing the correct answers, even though they may not choose all the correct ones might make assessment a part of learning. Special forms of penalty functions for various modes of learning and testing should be introduced and analyzed. Furthermore, the possibility of a grading function will be investigated in terms of overall level of knowledge achieved and 'fine tuning' of required efforts in improving student knowledge.

6. Conclusions

A new assessment methodology, *Multiple Choice Answers Approach* (MCAA), which apart from assessing the students also enables us to estimate the quality of teaching delivery, has been introduced and analysed. This was addressed as part of curriculum design and development, where the cycle of knowledge delivery and its progress were analysed using three functions: *definitive*, *characteristic* and *predictive*.

The features of the *Multiple Choice Answer Approach* assessment were analysed and the principles of the formation of multiple choice answers around a discipline, using the *core* of that discipline, were explained. A scheme for its realisation was shown to be through the creation of an automatic (or semi-automatic) assessment procedure and through the development of an algorithm for the selection of the most important material for assessment.

A process for the formation of an MCAA questionnaire and a scheme to maximize the efficiency of assessment, thus increasing the objectivity of the curriculum were also introduced. Special consideration was paid on a penalty function which even though it will consider the correct answers given it will also consider the incorrect ones which, in turn, will determine the level of penalty imposed.

A trial application of the MCAA approach to assessment on a real module at the Faculty of Computing at London Metropolitan University confirmed the applicability of this methodology with the analysis of statistics showing the growth in efficiency of teaching in this specific field of computer science.

References

1. E. Dijkstra, The End of Computer Science? *Communications of the ACM*, **44**(3), 2001, pp. 92.
2. I. Schagaev, E. Bacon and N. Ioannides, *An Approach to Curriculum Design for Computer Science*. Proceedings of the 1st International Conference on Technology Enhanced Learning, Quality of Teaching and Reforming of Education, Athens, May, in M.D. Lytras et al. Eds. WSKS 2010, Part I, CCIS 111, 2010a, pp. 493–499.
3. I. Schagaev, E. Bacon and N. Ioannides, Curriculum Design and Development for Computer Science and Similar Disciplines. *International Journal of Knowledge Society Research (IJKSR)*, **1**(3), 2010b, pp. 17–32.
4. Analytic Philosophy 2011. In Encyclopaedia Britannica. Retrieved from: http://www.britannica.com/EBchecked/topic/22568/analytic-philosophy (accessed: 09 Jan 2011).
5. R. Feynman, BBC interview, BBC Horizon, 1981. Retrieved from: http://video.google.co.uk/videoplay?docid=8777381378502286852# (accessed: 09 Jan 2011)

Igor Schagaev is Professor of Computer Science at the London Metropolitan University. His expertise includes the design of OS for parallel systems (late 70s), 64 processor FT RT SIMD (early 80s), new processor architectures with minimum redundancy (90s), and solid state black boxes for aircrafts. He is the author of 'Theory of Redundancy for Computer Systems' implemented by a new processor architecture for parallel FT systems, 'Methods of Minimization of Recovery Overheads' using structurally and semantic oriented recovery point schemes, and 'Concept of Active System Safety' (1986) for aviation and transport recently reincarnated as Damage Prognosis Theory. He holds a patent on an active system safety method.

Elisabeth Bacon is a Professor and the Dean of the School of Computing and Mathematical Sciences at the University of Greenwich. She is Vice President and Trustee of the BCS (Chartered Institute for IT), Chair of the BCS Academy of Computing, and is a past Chair of the Council of Professors and Heads of Computing. She is also an experienced systems designer and developer, with the bulk of her research and practice activity being directly industry facing, through knowledge transfer and consultancy.

Nebojsa Folic is a Senior Lecturer in Mathematics at the London Metropolitan University. His areas of expertise include Logic (formal and informal), Set Theory, Group Theory, Philosophy of Mathematics and Philosophy.

Nicholas Ioannides is a Senior Lecturer in Communications Technology at the London Metropolitan University. He has a PhD in Optoelectronic and Optical Fibre Sensing Techniques and a Master in Business Administration (MBA). He has previously worked for Cable and Wireless Global as a Consultant in Network Economics. He is the leader of the Optoelectronics research group and the course leader of the MSc Computer Networking, MSc Network Management and Security, and MSc Computer Systems Engineering. His research interests focus in the areas of optoelectronics, computer networks, curriculum design and development, and hardware design.