On Generalized Evaluation of PhD Competences in Computing

Dimo Dimov
Institute of Information Technologies, Bulgarian Academy of Sciences
29-A, Acad. G. Bonchev Str.
1113 Sofia, Bulgaria
+(359 2) 870 64 93
dtdim@iinf.bas.bg

Nikolay Dokev
Computer Science Dep., New Bulgarian University, and Inst. of Information Technologies, BAS
29-A, Acad. G. Bonchev Str.
1113 Sofia, Bulgaria
+(359 2) 870 04 29
niki@iinf.bas.bg

Petya Asenova
Computer Science Dep., New Bulgarian University
21, Montevideo Str.
1618 Sofia, Bulgaria
+(359 2) 811 01 63
pasenova@nbu.bg

Stanislav Ivanov
Computer Science Dep.
New Bulgarian University
21, Montevideo Str.
1618 Sofia, Bulgaria
+(359 2) 811 06 11
sivanov@nbu.bg

ABSTRACT
The paper aims at presenting an investigation concerning the evaluation of competences, which have to be achieved by PhD students in Computing. The results of an inquiry into this matter generalize the preliminary selected and approved sets of Generic and Specific competences for the four basic Computing areas – Computer Science (CS), Computer Engineering (CE), Software Engineering (SE), and Information Systems (IS). 238 researchers, university professors, employers and PhD students from 27 European countries were involved in the inquiry within the framework of the ETN-DEC project[1]. The proposed generalization of the inquiry results is based on well-known statistical and algebraic approaches. An extra classification of the competences in three categories is given taking into consideration their significance for the respective area – preferable, strongly consolidated, and "usually neglected" competences. A common visual interpretation of the four Computing areas is proposed which could be used also in an education measurement aspect.

Categories and Subject Descriptors
K.3.3 [**Computers and Education**]**:** Computers and Information Science Education – *computer science education, curriculum.*

General Terms
Documentation, Design, Experimentation, Measurement, Human factors, Algorithms.

Keywords
PhD competences, Educational objective, Computing curricula.

1. INTRODUCTION
The design and implementation of PhD educational programs is a complex and perennial process depending on a range of parameters such as the growth level of science and industry, employers' requirements, social factors, the need of personal scientific contribution, the means of communication and research tools, etc.

The ETN-DEC project promotes the idea of harmonizing PhD education in the European Computing space. The idea follows the experience of the well-known 'Tuning' project [3], which was initially oriented to the European Bachelor and Master programs' perfection. Like Tuning, the ETN-DEC adopts the well spread decomposition of Computing area into four basic sub-areas, namely – Computer Science (CS), Computer Engineering (CE), Software Engineering (SE), and Information Systems (IS).

One of the ETN-DEC objectives is to design Questionnaires to clarify what are the most important sets of generic and specific competences that successful doctors in Computing must possess. The Questionnaires' results are expected to support the present and future activities for design, modification and harmonization of the PhD curriculums in Computing.

This report presents the generalized results of an inquiry carried out among the ETN-DEC members. A brief description of the Questionnaires' forms and data gathered is given in Section 2. A simple statistical approach for generalizing the inquiry data is proposed in Section 3, on which base, some conclusions can be done separately for each of the five defined groups of competences – Generic and/or Specific ones. Using the canonical tools of linear algebra, this approach is further developed in Section 4 to a generalized interpretation of the competences in the four basic Computing specialties. We suppose that the illustrations proposed herein could be usefully applied also for educational measurements in the ETN-DEC framework. Following this line of thinking, possible ways of improving the proposed approach are commented on in conclusion. Meanwhile, one further way for possible test of the Computing decomposition consistency into four sub-areas (CS, CE, SE, and IS), as well as for looking for an eventual 5th sub-area place therein is suggested (e.g. for the Information Technologies area frequently explored recently, i.e. in Computing Curricula 2005 [4].)

2. BRIEF DESCRIPTION OF THE AVAILABLE INQUIRY DATA
The inquiry was carried out in January – March 2006, on preliminary Questionnaires designed and approved for ETN-DEC [8]. 238 persons have participated in the inquiry. These were ETN-DEC members from 76 scientific organizations – universities, academic institutes and/or companies from 27 European countries, (which represent more than 75 % of all the research teams of the project).

[1] CEC Contract SOCRATES *Erasmus Thematic Networks* No. 114046-CP-1-2004-1-BG-ERASMUS-TN, *Doctoral Education in Computing*, coordinated by the University of Rousse, Bulgaria, http://ecet.ecs.ru.acad.bg/etndec/

The Questionnaires consist totally of 7 parts, namely: (Q1) for respondent's personal data, (Q2) for evaluation of the Generic Competences, (Q3.1 - Q3.4) for evaluation of the Specific Competences in each of the four Computing areas, and (Q4) for evaluation of Computing Competences using the KSAM classification that has been additionally proposed, see [9]. The blank Questionnaire forms can be found at the ETN-DEC site. Most of the discussions about the Questionnaires' development are also documented there; see [6, 7, 8].

The present research concentrates on the data gathered only by Q2, Q3.1, Q3.2, Q3.3, and Q3.4. The Q2 form looks like the Table 1 shown herein to represent the generalized results of all 16 Generic Competences. The generalized results on Q3.1 - Q3.4 concern all the 24 Specific Competences defined for each basic Computing area (CS, CE, SE, and IS). They are shown in Table 2 below. The quantitative analysis hereinafter is based on known classifications and comments, [1 - 7].

The methods proposed for generalization and interpretation of the collected data could also be applied on the KSAM evaluation in (Q4) by analogy, and for this reason it will not be delivered here.

All significance estimations of the competences in this inquiry are made in accordance with the following integer scale:

$$1 \Leftrightarrow \text{no},\ 2 \Leftrightarrow \text{weak},\ 3 \Leftrightarrow \text{good},\ 4 \Leftrightarrow \text{excellent} \qquad (1)$$

adopted in [3].

3. STATISTICAL APPROACH TO INQUIRY DATA ANALYSIS AND INTERPRETATION BY SPECIALTIES

The following statistical methods are used:

• *Mean (average) of a competence* for a given specialty S:

$$\mu_k^{(S)} = \frac{1}{N}\sum_{i=1}^{N} x_{k,i}^{(S)}\ , \qquad (2)$$

where k is the competence number , $1 \le k \le |S|$, $x_{k,i}^{(S)}$ is the estimation given by respondent i for the competence k, $x_{k,i}^{(S)} \in \{1,2,3,4\}$, N is the number of the respondents.

• *Standard deviation* of the competence k according to the mean $\mu_k^{(S)}$ for a given specialty S:

$$\sigma_k^{(S)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i^k - \mu_k^{(S)})^2}\ . \qquad (3)$$

• *Generalized mean* for a given specialty S:

$$\mu^{(S)} = \frac{1}{|S|}\sum_{k=1}^{|S|}\mu_k^{(S)}\ , \qquad (4)$$

where $|S|$ is the cardinality of S, see (6a) below.

• *Averaged standard deviation* for a given specialty S:

$$\sigma^{(S)} = \sqrt{\frac{1}{|S|}\sum_{k=1}^{|S|}(\sigma_k^{(S)})^2}\ . \qquad (5)$$

As a rule, N varies by competences and/or specialties, i.e. $N=N(k,S)$, but the variance is negligible, so we accept $N(k,S) = N$ = const. For the given Questionnaire: $N = 238$.

Moreover, we consider the set of the Generic Competences (GC) likewise the four sets of Specific Competencies, i.e. we will work with a set of 5 sets:

$$S \in \{\text{GC, CS, CE, SE, IS}\}, \qquad (6)$$

for which cardinalities (see Table 1 and Table 2) we have:

$$|GC| = 16,\ \ |CS| = |CE| = |SE| = |IS| = 24. \qquad (6a)$$

We can compare the estimation of each competence and conclude, for example, which competences are the most significant or for which competences there is strong consolidation about their significance.

For this reason we will introduce some appropriate definitions:

• *Significance of a competence* will mean "importance, necessity" of the competence for the given specialty S. We consider that the significance of a given competence k is proportional to the value of the corresponding mean $\mu_k^{(S)}$, see (2).

• *Consolidation of a competence* will mean "degree of consensus" among respondents about the significance of the competence. We consider that the consolidation of a given competence k is reciprocal to the value of the corresponding standard deviation $\sigma_k^{(S)}$ from the mean $\mu_k^{(S)}$, see (3).

This way we can define also:

• *Preferable competences* are those having significance $\mu_k^{(S)}$, but greater than the generalized significance $\mu^{(S)}$ for the given specialty S, see (2) and (4). The preferable competences are marked by the sign '*' in the Tables 1 and 2.

• **"***Strongly consolidated" competences* are those having standard deviation $\sigma_k^{(S)}$, but lower than the averaged square deviation $\sigma^{(S)}$ for the given specialty S, see (3) and (5). The strongly consolidated competences are marked by the sign '+' in the Tables 1 and 2**.**

• *"Usually neglected" competences* are those satisfying the inequality $\mu_k^{(S)} < \mu^{(S)} - \sigma^{(S)}$, i.e. their significance is lower than the average significance of the competences for the given S, even lower than the area of the averaged standard deviation for S.

It is a curiosity to see the neglected competences in the four Computing specialties are not available, see Table 2. There are examples for neglected competences in Table 1 (Generic Competences): for k=15 (Efficient knowledge on intellectual property protection), and for k=16 (Efficient knowledge and skills in economical aspects of research).

The following *conclusions* can be made concerning the preferable and/or consolidated competences:

♦ The four *most preferable Generic Competencies* ($\mu^{(GC)}$=3.16) are: Research related skills such as creativity and intuition (μ_3=3.57), Ability to identify and refine the aims and activities necessary to solve a research problem (μ_6=3.50), Ability to acquire, analyze, evaluate and use relevant resources to the problem solution (μ_7=3.48), Ability to identify the problem and to specify user's requirements (μ_5=3.46).

♦ The four *most consolidated Generic Competencies* ($\sigma^{(GC)}$= 0.68) are: Ability to acquire, analyze, evaluate and use relevant resources to the problem solution (σ_7=0.56), Research related skills such as creativity and intuition (σ_3=0.58), Ability to discuss with other computing practitioners (σ_{13}=0.61), Ability to identify and refine the aims and activities necessary to solve a research problem (σ_6=0.61).

♦ The *most preferable Specific Competences***:**

- For Computer Science ($\mu^{(CS)}$=2.85) they are: Algorithms and data structures (μ_1=3.63), Operating systems principles and design (μ_7=3.26), Scientific computing (μ_{13}=3.13);

- For Computer Engineering ($\mu^{(CE)}$=2.79) they are: Computer architecture and organization (μ_4=3.56), Embedded systems design and implementation (μ_5=3.13), Real-time systems design and implementation (μ_9=3.11);

- For Software Engineering ($\mu^{(SE)}$=2.98) they are: Software modeling and design (μ_{10}=3.68), Software process and evolution (μ_{11}=3.55), Software validation and quality issues (μ_{12}=3.52);

- For Information Systems ($\mu^{(CS)}$=2.75) they are: Databases and information management (μ_{16}=3.52), e-Content principles, Web technologies (μ_{20}=3.32), Information security principles and policies (μ_{17}=3.23).

Table 1. Generic Competencies in Computing

No	Capacity to apply knowledge, analyze requirements and evaluate solutions in Computing	Average		Standard deviation		Count on
1	Skills for working in a team	3.33	*	0.65	+	235
2	Skills for preparation and delivery of learning content	3.29	*	0.73		235
3	Research related skills such as creativity and intuition	3.57	*	0.58	+	235
4	Interpersonal and communication skills	3.15		0.72		226
	Capacity to participate and manage a project development such as:					
5	Ability to identify the problem and to specify user's requirements	3.46	*	0.63	+	235
6	Ability to identify and refine the aims and activities necessary to solve a research problem	3.50	*	0.61	+	234
7	Ability to acquire, analyze, evaluate and use resources relevant to the problem solution	3.48	*	0.56	+	235
8	Ability to plan and organize the research and development activities	3.20	*	0.70		235
9	Ability to keep systematic records of plans, progress and achievement, and to reflect critically and constructively upon research and development	2.97		0.76		233
10	Ability to identify and assess the risks involved in the proposed solution so as to minimize them	2.97		0.73		233
11	Ability to create and/or apply concepts of quality to test and evaluate proposed solutions	3.13		0.70		232
12	Ability to effectively communicate proposed solutions to varied audiences, in both verbal and written form	3.33	*	0.63	+	232
13	Ability to discuss with other computing practitioners	3.24	*	0.61	+	233
14	Efficient knowledge on methodology of scientific and/or applied research	3.26	*	0.70		232
15	Efficient knowledge on intellectual property protection	2.48	–	0.81		235
16	Efficient knowledge and skills in the economic aspects of research	2.30	–	0.81		234
means: $\mu^{(GC)}$, $\sigma^{(GC)}$		**3,16**		**0,68**		

(*) to mark the competencies of significance grater than the Averages' mean $\mu^{(GC)}$, see (4), *S*=GC;
(+) to mark the competencies of variation (consolidation) lower than the Standard deviations' mean $\sigma^{(GC)}$, see (5), *S*=GC;
(–) to mark the competencies of significance lower than the value: $\mu^{(GC)}$ - $\sigma^{(GC)}$.

Table 2. Specific Competencies in Computing

		CS				CE				SE				IS			
No.	Capacity to apply **Specific** Competencies	Average		Standard deviation		Average		Standard deviation		Average		Standard deviation		Average		Standard deviation	
1	Algorithms and data structures	3,63	*	0,68	+	3,10	*	0,77	+	3,42	*	0,76	+	3,05	*	0,93	
2	Integrative programming	2,73		0,88		2,53		0,84	+	3,09	*	0,96		2,65		0,91	+
3	Network programming	2,80		0,90		2,81	*	0,96		3,04	*	0,91		2,84	*	0,96	
4	Computer architecture and organization	3,00	*	0,81	+	3,56	*	0,76	+	2,92		0,76	+	2,68		0,86	+
5	Embedded systems design and implementation	2,47		0,92		3,13	*	1,00		2,71		0,95		2,25		0,10	+
6	Digital devices design and implementation	2,28		0,92		3,24	*	0,96		2,25		0,86		2,01		0,90	+
7	Operating systems principles and design	3,26	*	0,79	+	3,10	*	0,81	+	3,28	*	0,80	+	2,79	*	0,96	
8	Grid computing	2,59		0,97		2,47		0,97		2,55		0,93		2,32		0,98	
9	Real-time systems design and implementation	2,53		0,89		3,11	*	1,00		2,79		0,94		2,42		0,99	
10	Software modeling and design	3,08	*	0,76	+	2,76		0,86	+	3,68	*	0,58	+	3,05	*	0,85	+
11	Software process and evolution	2,86	*	0,84	+	2,60		0,85	+	3,55	*	0,67	+	2,87	*	0,85	+
12	Software validation and quality issues	2,81		0,84	+	2,61		0,93		3,52	*	0,75	+	2,9	*	0,92	
13	Scientific computing	3,13	*	0,92		2,61		0,90		2,69		0,90		2,43		0,92	
14	Intelligent systems design and implementation	3,09	*	0,83	+	2,68		0,85	+	2,95		0,80	+	2,84	*	0,90	
15	Multidimensional data analysis and data mining	2,86	*	0,92		2,30		0,86	+	2,77		0,88		3,01	*	0,99	
16	Databases and information management	3,06	*	0,78	+	2,66		0,91		3,29	*	0,74	+	3,52	*	0,72	+
17	Information security principles and policies	2,82		0,94		2,65		0,89	+	3,02	*	0,90		3,23	*	0,91	
18	Formal languages processing	3,12	*	0,91		2,32		0,83	+	2,87		0,85		2,43		0,88	+
19	Network configuration and management	2,55		0,95		3,03	*	0,92		2,65		0,85		2,84	*	0,87	+
20	e-Content principles, and Web technologies	2,92	*	0,91		2,65		0,96		3,07	*	0,87		3,32	*	0,85	+
21	Human-computer multimodal interaction	2,52		0,91		2,56		0,93		2,89		1,00		2,87	*	0,97	
22	Platform oriented Computing	2,51		0,93		2,79	*	1,01		2,79		0,93		2,43		0,89	+
23	Multimedia technologies	2,69		0,99		2,75		0,97		2,87		0,93		2,95	*	0,92	
24	Computer based modeling and simulation	3,05	*	0,92		2,82	*	0,96		2,89		0,97		2,66		0,99	
means: ($\mu^{(S)}$, $\sigma^{(S)}$)		**2,85**		**0,88**		**2,79**		**0,90**		**2,98**		**0,85**		**2,75**		**0,92**	

(*) to mark the competencies of significance grater than the Averages' mean $\mu^{(S)}$, see (4) and (6), $S \neq GC$;
(+) to mark the competencies of variation (consolidation) lower than the Standard deviations' mean $\sigma^{(S)}$, see (5) and (6), $S \neq GC$;
(–) to mark the competencies of significance lower than the value: $\mu^{(S)} - \sigma^{(S)}$.

♦ The *most consolidated Specific Competences* are**:**

- For Computer Science ($\sigma^{(CS)}$=0.88) they are: Algorithms and data structures (σ_1=0.68), Software modeling and design (σ_{10}=0.76), Databases and information management(σ_{16}=0.78);

- For Computer Engineering ($\sigma^{(CE)}$=0.90) they are: Computer architecture and organization (σ_4=0.76), Algorithms and data structures (σ_1=0.77), Operating systems principles and design (σ_7=0.81);

- For Software Engineering ($\sigma^{(SE)}$=0.85) they are: Software modeling and design (σ_{10}=0.58), Software process and evolution (σ_{11}=0.67), Databases and information management (σ_{16}=0.74);

- For Information Systems ($\sigma^{(CS)}$=0.92) they are: Embedded systems design and implementation (σ_5=0.10), Databases and information management (σ_{16}=0.72), Software modeling and design (σ_{10}=0.85), Software process and evolution (σ_{11}=0.85), e-Content principles, and Web technologies (σ_{20}=0.85).

Generally, two of the competences only: k=1 (Algorithms and data structures) and k=7 (Operating systems principles and design) are strongly preferable for the four specialties. 5 of the competences: k=3 (Network programming), k=10 (Software modeling and design), k=11 (Software process and evolution), k=16 (Databases and information management), and k=20 (e-Content principles and Web technologies) are preferred for three of the specialties. 7 of the competences are preferred for two specialties, while the rest 10 are strongly preferred for one specialty only. It is noticed that those competences which are preferable for four or three specialties are also well consolidated.

We consider that the results received confirm the competences are appropriately formulated and their specialty specifics are well taken into consideration, in the Questionnaires.

4. INTERELATED GENERALIZATION OF THE BASIC SPECIALTIES

The definitions introduced in Section 3 are based on a comparative analysis of the competences of given specialty S and can be applied separately for each of the five specialties (6), (6a).

One might be interested in a common analysis of the specialties, for example by comparing their same name competences. Obviously, such a comparison is reasonable only for the basic specialties CS, CE, SE, and IS, see (6a).

For this purpose the *averaged significances* μ_k, could be used for each competence k, by analogy to (4):

$$\mu_k = \frac{1}{4}\sum_{S\in\{\text{CS,CE,SE,IS}\}} \mu_k^{(S)}\ , \quad k = 1 \div |S|,\ |S| = 24\ , \tag{7}$$

as well as the respective *averaged consolidations* σ_k , by analogy to (5):

$$\sigma_k = \frac{1}{2}\sqrt{\sum_{S\in\{\text{CS,CE,SE,IS}\}} (\sigma_k^{(S)})^2}\ , \quad k = 1 \div 24\ , 24 = |S|\ . \tag{8}$$

The number of the basic specialties is 4, what is too small from statistical point of view. That is why we will interpret (7) and (8) most of all in terms of an “averaged specialty”. We will adopt this averaged specialty a generalized representative of the four basic specialties, i.e. as the Computing area as a whole.

A linear vector space $\mathbf{E}^{24}$, a construction well known from Linear Algebra, can be used for common interpretation of the four basic specialties CS, CE, SE, IS, plus the *averaged Computing* as well.

The results of each response (each respondent answer on a given specialty S) can be considered as a vector (point) in this 24-dimensional space $\mathbf{E}^{24}$:

$$x_i^{(S)} \equiv (x_{1,i}^{(S)}, x_{2,i}^{(S)}, ..., x_{24,i}^{(S)})\ , \tag{9}$$

where $x_{k,i}^{(S)}$, $x_{k,i}^{(S)} \in \{1,2,3,4\}$ is the estimation given by the respondent i, $1 \le i \le N$, about the competence k, $1 \le k \le 24$, for the specialty S, see also the comments to (2).

The aim of this interpretation is to use the Euclidean norm $\|x\|$ of given vector x , $x\in \mathbf{E}^{24}$, i.e.:

$$\|x\| = \|(x_1, x_2, ..., x_{24})\| = \sum_{k=1}^{24} x_k^2\ , \tag{10}$$

having in mind that the distance $d(A,B)$ between arbitrary two vectors A and B in $\mathbf{E}^{24}$ to be given by the Pythagoras formulae:

$$d(A,B) = \|A-B\| = \sum_{k=1}^{24}(A_k - B_k)^2,\ A \equiv (A_1, A_2, ..., A_{24}) \in \mathbf{E}^{24},\ B \in \mathbf{E}^{24}. \tag{11}$$

So, the averages of the four basic specialties can be considered as points (vectors) in $\mathbf{E}^{24}$:

$$(S) \equiv (\mu_1^{(S)}, \mu_2^{(S)}, ..., \mu_{24}^{(S)})\ , \quad S \in \{\text{CS, CE, SE, IS, CoG}\}\ , \tag{12}$$

where $\mu_k^{(S)}, k = 1,...,24$ are the means (2). In this way, the respective distances among the four specialties can be calculated by (11). Obviously, they four can be considered as a tetrahedron (CS, CE, SE, IS), which Center of Gravity (CoG), i.e. the vector of the above mentioned averaged Computing, corresponds to (7).

This $\mathbf{E}^{24}$ disposition of the specialties is illustrated in Figure 1. The following comments could be worth for its interpretation:

• The region of interest is restricted to a 24-dimensional (24-D) hyper cube, which 24 basic edges coincide with the coordinate axes of $\mathbf{E}^{24}$, each axes corresponding to a Specific Competence. Each filled-in form $x_i^{(S)}$ of the Questionnaires, see (9), is considered a 24-D point inside this hyper cube.

• Each Computing specialty S is considered a class (or cloud) of points (i.e. Questionnaires’ fulfillments). The cardinality of all the classes is equal to N=238. Each class S can be considered as usual a 24-D ellipsoid with center (S), accordingly to (12) and (2), and with dimensions along its principal axes, proportional to the respective standard deviations $\sigma_k^{(S)}$, k=1,…24, accordingly to (3).

• For a simplicity herein, the ellipsoid of each class $S\in\{$CS, CE, SE, IS$\}$ is approximated by a 24-D sphere of the same center (S) and of a radius equal to the averaged standard variation $\sigma^{(S)}$, (5). Actually, two more standard deviations are also illustrated geometrically in Figure 1, namely − the minimal σ, and the maximal σ, among the competences for each specialty S. These “extra σ-s” will help on the interpretation of “gaps” among the specialties that will be introduced herein by analogy to the gaps defined in the Tuning project [3].

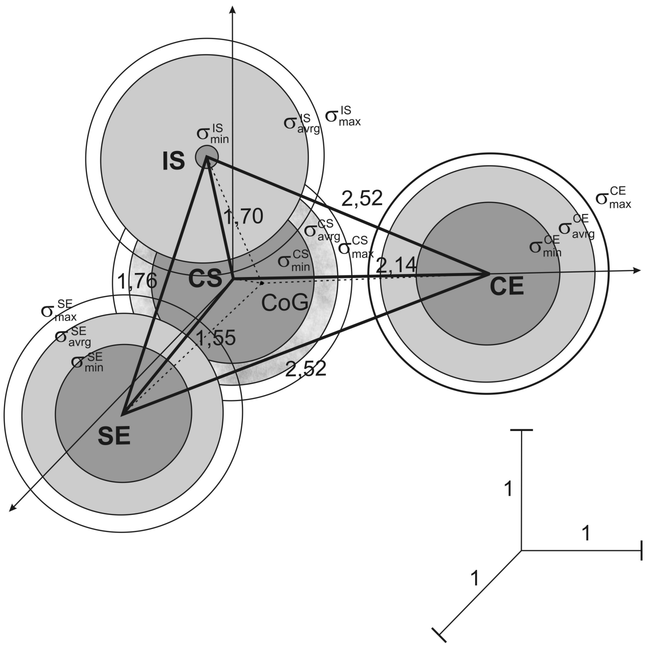

Figure 1. Graphical illustration of the Specific (CS, CE, SE & IS) competencies' disposition. The 6 distances of interest are written and the minimal, averaged and maximal σ-s are also shown to illustrate "primary gaps" among the specialties. CoG represents the averaged centre of the Computing area as a whole.

In other words, we can think about the illustration in Figure 1 as an immersion of the $\mathbf{E}^{24}$ into the $\mathbf{E}^{3}$ space in such a way that the six 24-D distances among the four vertexes (CS, CE, SE and IS) remain unchanged in $\mathbf{E}^{3}$. The resulted tetrahedron in $\mathbf{E}^{3}$ is finally projected into $\mathbf{E}^{2}$ to obtain the illustration.

A inclined parallel isometric projection (Cavalier projection, [10]), has been chosen to illustrate the tetrahedron because of its keeping unchanged the distances along the $\mathbf{E}^{3}$ coordinate axes. The $\mathbf{E}^{3}$ axes are chosen in such a way that at least 3 of the tetrahedron edges, CS-CE, CS-SE and CS-IS, to be proportional (two of them approximately) to their $\mathbf{E}^{24}$ distances, namely – the CS-CE-edge coincides the *Ox*-axis and the CS-CE-SE-triangle coincides the *Oxy*-plane.

The 3-D spheres approximating the classes, i.e. the clouds around the 4 specialty centers appear as 2D circles in the Cavalier projection. The 3 types of standard deviations, mentioned above for each specialty, are also shown in Figure 1.

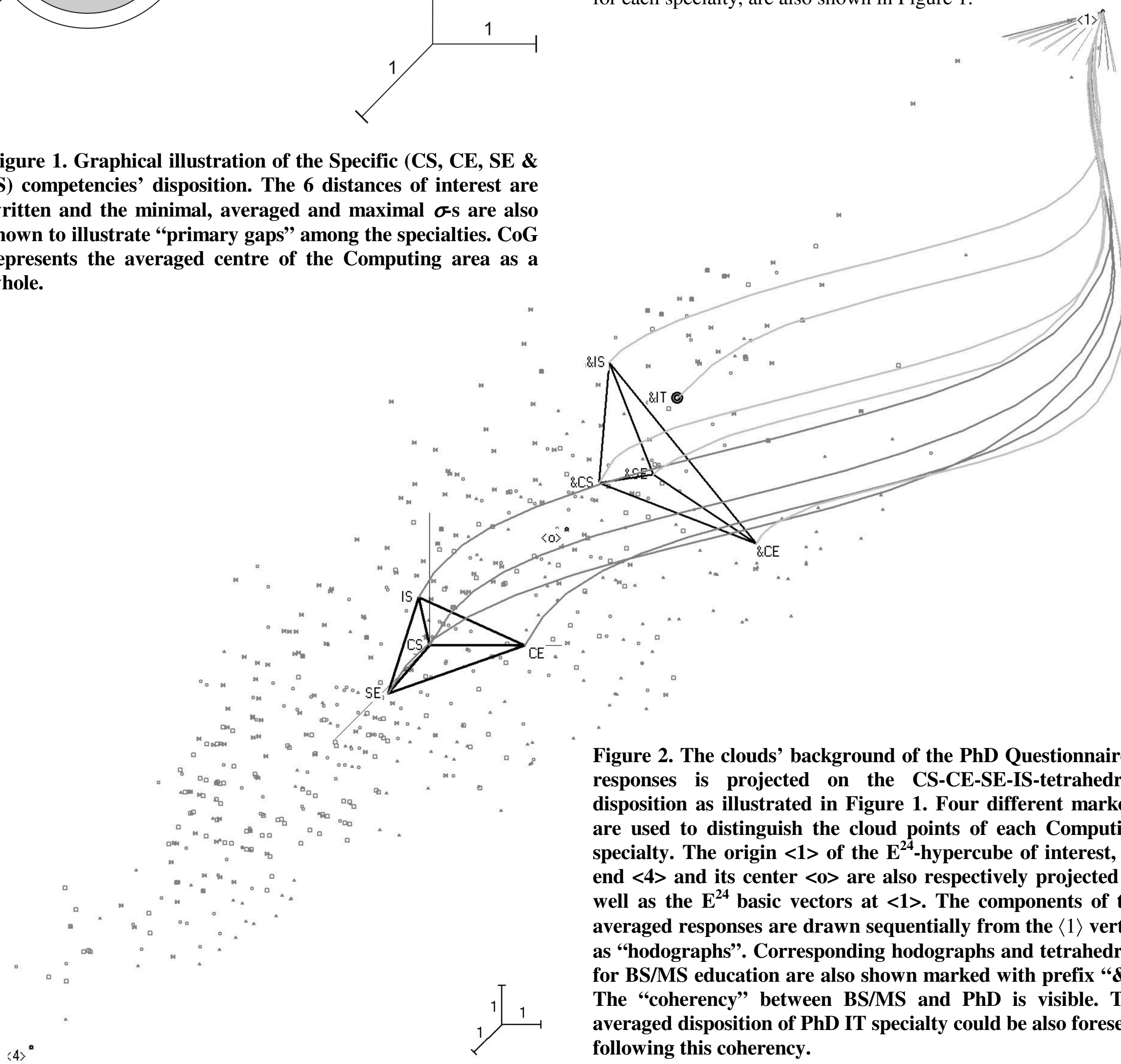

Figure 2. The clouds' background of the PhD Questionnaires' responses is projected on the CS-CE-SE-IS-tetrahedron disposition as illustrated in Figure 1. Four different markers are used to distinguish the cloud points of each Computing specialty. The origin <1> of the E^{24}-hypercube of interest, its end <4> and its center <o> are also respectively projected as well as the E^{24} basic vectors at <1>. The components of the averaged responses are drawn sequentially from the ⟨1⟩ vertex as "hodographs". Corresponding hodographs and tetrahedron for BS/MS education are also shown marked with prefix "&". The "coherency" between BS/MS and PhD is visible. The averaged disposition of PhD IT specialty could be also foreseen following this coherency.

So, the idea for "*the gaps among the specialties*" [3], is exactly visible here along CS-CE and approximated along the CS-SE and CS-IS edges. We can define the gap value $g(S_1,S_2)$ between 2 specialties S_1 and S_2 by:

$$g(S_1,S_2) = d(S_1,S_2) - (\sigma^{(S1)}+\sigma^{(S2)}) , \qquad (13)$$

where $\sigma^{(S)}$ is the averaged standard deviation (5) for S_1 and S_2 respectively, and $d(S_1,S_2)$ is the distance (11) between S_1 and S_2.

The following issues can be generally made considering Figure 1:

♦ As they were voted (i.e. evaluated/estimated) by the inquiry participants, the 4 specialties are well distinguishable in average for most of the competences.

♦ One might wish to check "if there exists enough space among the 4 specialties", within which to represent one more "virtual" Computing specialties, e.g. the IT (Information technologies) specialty that is treated in [4]. Obviously there is not enough space for that, at least in inner volume of the CS-CE-SE-IS-tetrahedron (!). Actually, only 3 averaged gaps are negative (or close to zero) that could suggest about possible area crossings, namely:

g(CS,SE)=-0.19, g(CS,IS)=-0.05, and g(SE,IS)=+0.03.

Of course, the gaps for the respective maximum deviations become strongly negative.

♦ On the other hand an idea could be raised for the averaged (generalized) Computing specialty that should be interpreted as the Computing area itself. The respective σ^{CoG}_{min} , σ^{CoG}_{avrg} and σ^{CoG}_{max} are not illustrated for simplicity, but the simple evaluation of respective gaps between CoG and each specialty shows that CoG extremely covers them. In other words, CoG is well close to the 4 specialties to interpret them as Computing components, and simultaneously CoG is enough far to decide that no existence of any other virtual specialty (of a similar range) in the common area of Computing seems reasonable, at least in the tetrahedron volume (cloud), see Figure 1.

Many extra questions could arise from looking at the Figure 1:

(?1) more exactly, where are located herein the clouds of Q-responses for a given specialty and/or how they relate each other;

(?2) what lies in the gap in the bottom right hand corner of the figure, or where lies the 5[th] Computing specialty, largely commented in [4], or where is the place of Master/Bachelor (BS/MS) education in Figure 1, etc.

To answer the first group of questions, the projection matrices for the transformations $\mathbf{E}^{24}\rightarrow\mathbf{E}^{3}\rightarrow\mathbf{E}^{2}$ discussed about Figure 1 were composed. Using these transformations the projections of all points from the four clouds were calculated with respect to the tetrahedron (CS,CE,SE,IS), and the same was done for the most interesting two points of the hypercube: the "origin" ⟨1⟩ (all competences evaluated as 1) and "the end point" ⟨4⟩ (all competences at 4). The results are shown at Figure 2. These calculations discover that:

♥ The clouds are substantially merged. At the figure all responses are marked by specialty as follows: (▣) for CS, (▲) for CE, (□) for SE, and (⋈) for IS. The clouds are visible as a unique generalized cloud approximately along the main diagonal ⟨1-4⟩ of the hypercube. Besides, the clouds for each of the four specialties have similar form.

♥ Hence, the clouds should be better regarded as 24-dimensional ellipsoids than as spheres, as was admitted discussing Figure 1. Surely, other projections differing from these of Figures 1 and 2 and with better discrimination between the clouds can be obtained, e.g. using the Fisher's linear discriminant approach, [11]. The preliminary calculus shows that Figure 2 is enough informative to this end. In this line of thinking, the evaluation of "gaps between specialties" that have been done by the proposition of spherical clouds remains actual but only for the case of observation by a single averaged competence. Thus, by observation by two competences the gaps should diminish, by three – still more, etc., and by all 24 competences the gaps go to be strongly negative. In those "most heavy" cases we can define for the gaps similarly to (13):

$$g_{\min}(S_1,S_2) = d(S_1,S_2) - (D(S_1)+D(S_2)), \qquad (14)$$

where $D(S)$ is the averaged 24-D distance from the cloud S to its center, i.e. $D(S) = \frac{1}{N}\sum_{i=1}^{N} d(S, x_i^{(S)})$, see also (9). About the 4 averaged distances of interest we have:

D(CS)= 4.29, D(CE)= 4.41, D(SE)= 4.14, D(CS)= 4.48.

To this end, the following *recommendation* could be addressed to the preparation process of PhD Computing curricula and syllabi for a given specialty S_{new} – to obey the distinguishing rule based on the "nearest mean" principle:

$$(S_{new} \to S_0) \Leftrightarrow d(S_{new}, S_0) = \min_{S\in\{\text{CS,CE,SE,IS}\}} \{d(S_{new}, S)\} . \qquad (15)$$

To answer the second group of questions, it should be written the following:

♠ The location of BS/MS education in Computing is somewhere between the tetrahedron (CS,CE,SE,IS) and the origin ⟨1⟩. To this aim a heuristic reduction has been done from the given evaluations of competences in [4] to our case of 24 competences. The comparison among the competences shows that they are composed similarly, but some competences are missing in BS/MS. Nevertheless, the obtained data are enough for to uncover the tetrahedron (&CS, &CE, &SE, &IS) for the BS/MS education, see Figure 2.

♠ In similar way, the center of IT, the 5[th] Computing specialty accordingly to [4] has been also uncovered corresponding to BS/MS education in Computing.

♠ Each specialty can be represented as a vector sum of the averaged values of its corresponding competences. Drawing these sums as polylines from the origin ⟨1⟩ on a selected projection we obtained an extended graphical representation for the specialties and their components as well. This representation (we called it "hodographs") was built both for the Questionnaires' results and for the BS/MS Computing competences as explained above. Generally, we can state that both curricula are coherent – the corresponding points are situated in similar way. A guess about the disposition of PhD IT program can be made using this similarity (see Figure 2).

5. CONCLUSION AND FUTURE WORK IDEAS

In this paper we have applied quantitative methods/approaches of Pattern Recognition to generalize the results of ETN-DEC Questionnaires. At this stage of work we have proposed a generalized evaluation (Section 3) separately for each of the 4 basic Computing specialties, and have added to them by analogy also the Generic competences evaluation. In Section 4 we propose an approach to generalized interrelated evaluation of the Specific competences in Computing. The calculations should be considered fairly labor consuming to be used only for visualization aims, so we can move naturally to the following considerations:

♣ If the above interpretations/illustrations are sufficiently informative by itself then could one expect that more precise computations in the sense of classification theory and related areas like discriminant analysis, analysis of variance, PCA, etc., will be interesting.

The answer to this question is expected from the future discussions in the ETN-DEC framework and will direct and motivate future research steps like:

° Further improvement of the proposed approach for generalized evaluation of the competences so that to support the expert groups (EG) activities on the four basic specialties for PhD Computing curricula design.

° Future adaptations to the approach for both qualitative and quantitative comparative analyses as a basis for evaluation and decision making in similar areas and/or activities.

° It would be also interesting to assemble a generalized evaluation, based on the three types of Questionnaire – Generic, Specific and KSAM, i.e. Knowledge, Skills, Attitude and Management, instead of CS, CE, SE, and IS, [8].

6. ACKNOWLEDGMENTS

Our thanks to all members of the EG on Questionnaires and to many other ETN-DEC members for the fruitful discussions around this work as well as to all those 238 respondents of the Questionnaires, which scientific, research and educational professionalism brings up essentially the results shared herein.

7. REFERENCES

[1] Bloom, B. S. *Taxonomy of Educational Objectives*. New York, Handbook I: The Cognitive Domain, David McKay Co Inc., 1956.

[2] Denning, P. J. and Durham, R. *The Core of the Third Wave Professional*, CACM, vol. 44, no.11, 2001, pp. 21-25.

[3] Tuning Educational Structures in Europe (2003), In Julia González and Robert Wagenaar (Eds.), Final Report, Pilot Project, Phase 1, 2001-2002 http://odur.let.rug.nl/TuningProject/.

[4] Computing Curricula 2005, The Overview Report, The Guide to Undergraduate Degree Programs in Computing. (April 11, 2005) ACM, AIS, IEEE-CS, http://www.acm.org/education/curricula.html#CC2005.

[5] Oldfield, S. J. Core Computing: What every student needs to understand. In *Proceedings of CompSysTech'02* (Sofia, Bulgaria, June 20-21, 2002), Avangard Print Ltd, Rousse, BG, 2002, pp. IV.8-1 - IV.8-8.

[6] Oldfield, S. J. Subject Specific Competences in the context of postgraduate studies in the academic discipline of Computing. In *ETN-DEC Forum* (2005) http://ecet.ecs.ru.acad.bg/etndec/.

[7] Vandamme, F. Suggestion concerning: Document Set EG "Questionnaires". In *ETN-DEC Forum* (2005) http://ecet.ecs.ru.acad.bg/etndec/.

[8] ETN-DEC Progress Report, 2005, Annex 1: Document Set EG "Questionnaires", ETN-DEC work meeting in Berlin, Sept. 05, 2005, Avangard Print Ltd, Rousse, BG, 2006, pp. 51-73.

[9] Ivanov, S., Maneva, N., Dimov, D., and Dokev, N. One More View on PhD Competences in Computing. In *Proceedings of 2-d e-Learning Conference (*Berlin, Germany, September 6-7, 2005), (to appear) http://www.lotuswebtec.com/e-conf/Schedule.htm.

[10] Foley, J. D., van Dam, A.: Fundamentals of interactive Computer Graphics, Vol. 2. Addison-Wesley Publ. Co., London, Amsterdam, Sydney (1982).

[11] Duda, R.O., P.E. Hart, D.G. Stork: Pattern Classification, A Wiley-Inter science Publication, John Wiley & Sons Inc., New York, Singapore, Toronto (2001).