

Cognitive Dimensions Questionnaire Applied to Visual Modelling Language Evaluation - a Case Study¹

Anna E. Bobkowska
Technical University of Gdańsk, Poland
annab@eti.pg.gda.pl

Abstract

In order to achieve more maturity in modelling languages engineering, effective methods of visual modelling language evaluation are required. There is growing understanding of human factors' impact on the quality of results and thus cognitive perspective seems to be a good frame of reference for evaluation. In the paper, a case study of applying Cognitive Dimensions Questionnaire for visual modelling language evaluation is presented. Discussion of cognitive dimensions in context of its proposed use, analysis of the filled in questionnaires and responders' opinions about the method as well as cross-comparison between set of cognitive dimensions and a set of criteria for visual modelling languages are described. Conclusions suggest need of customisation of the questionnaire with special focus on precision.

1. Introduction

Software becomes more and more complicated and role of visual modelling languages (VML) increases. There is a need for advanced methods of VML engineering and as a part of them, methods of VML evaluation that could verify whether VMLs really support software developers' work. The purpose of the case study was to find out empirically about usability of the Cognitive Dimensions (CD) Questionnaire [2] for evaluation and improvement of visual modelling languages. CD framework [1] and questionnaire were originally designed for evaluation of notational systems, which include evaluation of notations. Idea of cognitive dimensions seems to give a new perspective in system evaluation and is likely to be universal enough to allow for application in the area VML evaluation.

Cognitive dimensions questionnaire was filled in by two responders for Unified Modelling Language (UML) [3] supported by a Computer Aided Software Engineering (CASE) tool. The responders were asked also to express their opinions about the method and comment on quality of the results. The interesting questions in that case study can be expressed as follows: Is the CD questionnaire easy to understand and use? What is quality of results? Does it cover

¹ This research was performed during author's visit at the University of Kent at Canterbury, UK supported by the SegraVis project.

well cognitive perspective? Are the questions adequate, complete and not overlapping according to user's needs? Is gathered information precise enough? Do the users find the results useful? What is efficiency of using CD questionnaire as a method of evaluation? Does it really show things that could be improved? What is really new in using the method?

Collected data were analysed to check out whether the CD questionnaire delivers useful information for improvement of visual modelling languages. The following results were expected in the questionnaire analysis:

- Analysis of the CD questionnaire and its background research should give an idea of scope of notions covered by the questionnaire;
- Categorisation of CD questions according to their type and subject as well as analysis of responders' answers in that context should indicate which categories of questions are most useful;
- Comparison of responders' answers should show whether the questions are precise enough to allow for common understanding by all responders.

During analysis of the responders' opinions related to the method and its results the following aspects has been examined:

- Problems that responders met should indicate weak points in the method;
- Amount of time spent on filling the questionnaire together with level of details in gathered answers should give an idea about precision of the results and efficiency of the method;
- Information about level of satisfaction and proposed improvements should indicate general attitude to using CD questionnaire as a method of VML evaluation as well as provide some insights into its customisation;
- Finally, question about really new aspects in the method and elicitation of unknown before results should be helpful in evaluating the uniqueness of the method.

The most difficult to verify is scope of CD questions. The problem is in discovering the right perspective. A set of VML quality criteria (VMLQC) based on quality criteria in the software modelling literature was proposed and cross-comparison between VMLQC and CDs was made.

Section 2 includes brief description of the CD questionnaire and discussion of its application to VML evaluation, section 3 provides characteristics of responders and evaluated systems as well as analysis of filled in questionnaires, section 4 presents responders' opinions after filling in the questionnaire, section 5 describes results of cross-comparison between VMLQC and CDs, and section 6 draws final conclusions for customisation of the CD questionnaire for VML evaluation and suggestions for improvement of the CD framework.

2. Description of the CD questionnaire

In the introduction to the cognitive dimensions questionnaire authors write it aims to collect views about usability of 'notational systems' in context of user needs. It is a product of CD framework, which covers many research results concerned with CDs and their background and application. The intention of the authors of cognitive dimensions was to provide a discussion tool that will allow for evaluation and improvement of notational systems in an effective and cheap way. Later CDs were adopted as a research tool. The CD questionnaire consists of 5 sections: section 1 collects background information about evaluated system and experience of responder; section 2 provides set of basic definitions, which are presented briefly together with description of notation system in the Table 1; section 3 collects information about the notational system: products, main notations, sub-devices, activities it is used for, and percentage of time spent on performing five groups of activities; section 4 contains questions about the main notation; and section 5 – questions about sub-devices.

The most important part of the questionnaire is the section 4, which consists of about 48 questions associated with 14 cognitive dimensions and system improvements. (Question introducing a problem and a question of type 'Which ones?' were counted as a single question.) The list of cognitive dimensions in the questionnaire with short descriptions is shown in the Table 2.

Notational systems	Notational systems include many different ways of storing and using information – books, different ways of using pencil and paper, libraries or filing systems, software programs, computers, and smaller electronic devices.
Product	Event or object being the ultimate reason why notational system is used. This can be thing that happens as an end result, or thing that is produced as a result of using the notational system.
Notation	The notation is how users communicate with the system - special format they describe end result and structure of information they can read. Notations' structure should correspond in some way to the structure of the product. Notations can include text, pictures, diagrams, tables, special symbols or various combinations of these. Some systems include multiple notations.
Sub-devices	Complex systems can include several specialized notations, which normally are not considered to be part of the system. There are two types of sub-devices: helper devices and redefinition devices. Both of them need their own notations that are separate from the main notation of the system.

Table1. Basic definitions in cognitive dimensions questionnaire

Is there any chance that CD questionnaire can be successfully applied in the area of VML evaluation? There is growing understanding of the impact of human factors on quality of the results. Activity of software modelling is much more complex than just using a piece of software. In the phase of analysis, it is a creative task of domain modelling in order to capture such fragment of reality that is not overcomplicated, yet mirrors user's domain knowledge. It is strongly influenced by information gathered from the stakeholders and should support communication with them. In the phase of design, software solution is elaborated and this task is affected by available technology. In the cognitive dimension framework's classification of activities for usability profiles this activities can be very often assigned to the most advanced 'exploratory design' category. The exploration of possibly new information structures depends on the language we use to express them, not only on developer's decisions or tool support. In similar way like a theory influence experiment, the language has a great impact on

developer's observations and design decisions. Aspect of cognitive fitness of VML for the developers and their tasks is crucially important.

Viscosity: resistance to change.
Visibility: ability to view components easily.
Premature commitment: constraints on the order of doing things.
Hidden dependencies: important links between entities are not visible.
Role-expressiveness: the purpose of an entity is readily inferred.
Error-proneness: the notation invites mistakes and the system gives little protection.
Abstraction: types and availability of abstraction mechanisms.
Secondary notation: extra information in means other than formal syntax.
Closeness of mapping: closeness of representation to domain.
Consistency: similar semantics are expressed in similar syntactic forms.
Diffuseness: verbosity of language.
Hard mental operations: high demand on cognitive resources.
Provisionality: degree of commitment to actions or marks.
Progressive evaluation: work-to-date can be checked at any time

Table 2. List of cognitive dimensions and their short descriptions.

However, is the CD framework a good tool to provide solutions? Having the name of 'cognitive dimensions' one could expect they should cover essential part of cognitive psychology with characteristics known from our experience with other dimensions, especially 3-dimensional space and algebra, and that they should eliminate unnecessary details. While investigating in more details CD bibliography one can discover it is still under development and that metaphor of gardener trying to provide their plants with all substances needed for their healthy growth is more adequate nowadays. So, the next question is arising: Is the method mature enough to provide useful results in the area of VML evaluation? It is said, that theory answers the questions it was designed to answer. So other questions could be asked: Is CD framework universal enough to be helpful for the purpose of evaluation of other artefacts then notational systems? Is its scope limited to discussion tool providing some kind of terminology or this terminology has really strong cognitive background and is really a powerful tool? And answers for those questions are very difficult to find on other way than by empirical investigations.

3. Analysis of the questionnaire and filled in data

Cognitive dimensions questionnaire was filled in by two responders. Notational systems being evaluated were CASE tools supporting modelling in the UML with focus on subset of UML in a given context of use. Notational system evaluated by the first responder was Select Enterprise [5]. Context of use was system analysis during laboratory of Information System Analysis, and subset of UML included use case diagram, class diagram and sequence diagrams. The first responder filled in the questionnaire in 70 minutes. The second responder evaluated Poseidon for UML [4]. Application of this system was meta-modelling with class diagrams and Object Constraint Language. The second responder filled in the questionnaire in 20 minutes. Both responders considered themselves proficient with using the tool and having experience with other CASE tools. There are a few additional responders' characteristics that could influence results. The first responder had previous experience with developing questionnaires. The second responder was highly motivated to achieve useful feedback and generally is a very busy person. Both of them at the time of filling in the questionnaire had a very brief idea about cognitive dimensions.

In the table 3 metrics related to types of questions in the section 4 of the questionnaire and the responders' answers are presented. 45 out of 48 questions are CD questions and last 3 questions ask about unusual way of using the system and improvements. All the questions were categorised in four groups depending whether they ask about simplicity (i.e. they start with 'how easy', 'how close'), reasons (i.e. they start with 'why?'), specifications (i.e. they include expressions 'which', 'what kind', 'what ways', 'what sort'), and possibility (i.e. they contain words 'can', 'possible', 'let').

Numbers of questions in each category should indicate what kind of information the questionnaire really collects. Comparison of numbers of questions of each category indicates that a half of them are open questions asking responder to find things with specific characteristics. This means that they force user to reflect on the system from suggested in the question perspective. Other questions have similar character asking whether something is easy or possible and what are the reasons of that.

Assuming that useful questions are answered and reasons of missing answer is lack of understanding, importance or relevance, percentage of answered questions in each category

should give a sketch idea of usefulness of those questions. Numbers show that about 55% of specification and possibility questions were answered, almost 80% questions about simplicity and just about 28% of questions about reasons. This indicates that questions in category of simplicity were most useful, then specifications and possibility questions, and the most problematic were the questions about reasons. Another problem is level of details of those answers and their value for language and system improvement, but this will be discussed in further part of the paper.

Type of question	Nr of questions	Responder 1	Responder 2
Simplicity	5	4	4
Reasons	9	3	2
Specifications	24	11	15
Possibility	10	6	5

Table 3. Metrics of types of questions and answers

The next focus in the questionnaire analysis was subject of questions. The goal of applying such perspective was to examine what are subjects of the questions and thus what questionnaire really helps to evaluate. Notational systems are complex entities that include notations, products, media, and sometimes also sophisticated functions of software tool. Sometimes influence of them can be contradictory (notation is difficult but tool very good). There can also appear a kind of misfit between product and notation or product and a system in context of system’s application. And finally sophisticated tools can eliminate deficiencies of notations. So, evaluation of the entire system without considering its parts is not very likely to give precise results.

The purpose of questions’ subject analysis was to find out which parts of the system are evaluated. Metrics concerned with use of words indicating subject in questions and the subjects of responders’ answers are shown in the table 4. Words that have similar meaning (e.g. product and result; or notation, language, and possibility of expression) were also included. 17 out of 45 CD questions do not have explicit indication what are they about. Default subject of evaluation is ‘notational system’ which includes all of them. In the questionnaire there is no clear distinction for different parts of the system with further combination of the results. A problem in subject’s analysis is that existence of the word is in a question that does not automatically mean that the question is about that notion. For example,

use of ‘product’ had just auxiliary purpose. The first responder answered about half questions about the system and a half about notation, the second responder answered about 77% of questions about notation. Numbers in parentheses indicate answers to the questions in which exist word concerned with the subject and they are the same or subject is not indicated. This confirms that existence of a words indicating subject does not necessarily indicate what answers were about. Co-relation of subjects in both responders’ answers was also checked. It indicated that there were 14 questions answered by both of them and they were concerned with the same subject. This indicates that questions had common interpretation. However 40% of answers was related to different questions, so based on that case study it is not possible to conclude which of them are useful or not.

Another problem with application of CD questionnaire to VML evaluation is that concept of ‘notation element’ is slightly different from that of ‘language element’. Notation as a set of symbols used in communication is just a part of the language. Considering both ‘language element’ and ‘notation element’ as a single concept is not correct just because in complex modelling languages, like UML, with one ‘language element’ can be associated many ‘notation elements’. They have different representations depending on context of use and type of view they appear on. E.g. in the class diagram containment can be represented as association with end of diamond, or classes within a container class.

Words concerned with ‘subjects’	Nr of ‘words’	Responder 1	Responder 2
‘System’	4	10 (7)	5 (3)
‘Notation’	12	11 (10)	17 (16)
‘Product’	5	0	0

Table 4. Metrics of ‘subjects’ in questions and answers

A closer look at the quality of answers has shown that generally they were not very detailed. In many cases some examples were put in rather than all possibilities carefully examined. The answers not always were clear, for example first responder’s answer for the question of ‘how easy is it to make a change?’ was ‘It depends on the kind of change...’ or for the question of ‘Are dependencies visible?’ – ‘Yes, generally they are’, which was followed by one example of hidden dependency that ‘could be also visible’ in the next answer. The first responder five times indicated something more could be added there (used a three-dot sign: ‘...’) and three times had not enough space to write full answer. Some interesting examples from second

responder's answers include answer for the question of 'how closely related is the notation to the result that you are describing?', which was 'Hard to answer this. UML class have become so ingrained' or for the question of 'how easy is it to make a change?' answered with 'Sometimes doesn't work' or the question of 'what features of notation help [to sketch]?' commented with 'or Tools?' and also three times didn't have enough space to write the answer.

To summarise, the questionnaire collects information about possibilities, specifications, reasons and simplicity concerned with cognitive dimensions. The questions often don't have a clear indication what is their subject however the responders' answers are concerned with the same subject. About a half of questions were answered and the answers were not very detailed. Additional questions arise while analysing correctness of application of this questionnaire to VML. Some idea whether this is what responders really needed, reasons why results are like that, and how responders themselves consider the method, is described in the next section.

4. Responders' opinions about the method

Responders' opinions after filling in the CD questionnaire allowed to comment on application of cognitive dimensions questionnaire for evaluation of visual modelling languages as well as achieved results. The responders were asked about time needed to fill in the questionnaire (which was given in section 2), level of details and evaluation of usability, main problems with the questionnaire as well as general description of level of satisfaction in terms of thoughts and emotions after using the questionnaire. Background assumption in collection that kind of information was observation that 'emotions influence motivations, and motivations influence quality of results'. Since cognitive perspective should bring unique results, additional question was asked about how many answers have brought really something new? (Not remainder of already known problems.)

The first responder evaluated level of details and usability for quite poor (3 in 10 point scale) and noticed that:

- The questionnaire is too general and it doesn't support well notational system improvement;
- There are some 'hard mental operations' that need to be performed while filling it in;

- It doesn't capture contradictory or complementary influence of several aspects of a notational system and doesn't allow for presentation how different stakeholders influence different aspects at different moments in time. There is a need to distinguish what questions are about.
- Sometimes working with the questionnaire was annoying because a lot of time was needed to find out what was the question really about and how to fit it for visual modelling language – about ten times was that kind of problem.
- Not a good questionnaire – too many open questions. Experience of questionnaires for software engineering tasks shows that people work in hurry, and need support of questionnaires that combine pieces of information from different conceptual models otherwise it is very likely that useful information will be missed. (This observation is confirmed by time of filling in questionnaire by the second responder.)

The results were considered of little practical use, just surface ones. It would be preferred possibly more longer questionnaire that could deliver more detailed results. The vocabulary of the questionnaire should fit to the mental representations of the developers. None of the results was really new, and the first responder was a little bit disappointed with what CD questionnaire can offer.

Second responder noticed the following problems:

- There is not a clear distinction between tool and notation, i.e. sometimes questionnaire refers to 'system', sometimes to 'notation';
- Questionnaire was not very helpful in section of sub-devices because of too many cross-references;
- He 'couldn't answer' percentage of time spent on doing certain types of activities – there is need for observing oneself which is not done naturally during software modelling;
- Questions about diffuseness, closeness of mapping, and role-expressiveness were considered vague.

For the question about the level of details, answered that it caused to reflect a bit and think about some aspects more clearly, however there is a need for a tool, which could easily eliminate the problem of 'not enough space to provide details'. And usability evaluation was expressed as 'useful as a starting point, but that's all; need many more hints on how to 'measure' some of the factors and preparations of subject.' He found nothing really new, but

considered the questionnaire useful in explaining the given fact as a part of general problem and found useful the cognitive dimensions terminology.

In the opinions' section the following important notions were captured: precision, usability scope and need of customisation. Software engineers require much more precision that it used to be acceptable in social sciences. The subjects of evaluation should be clearly distinguished, and some measures of success are suggested. The questionnaire is helpful as terminology provider (even the first responder who didn't like much the questionnaire used 'hard mental operations' in description of the opinions). The intention of the authors that cognitive dimensions are good 'discussion tool' and 'artificial creator of empathy' can be confirmed. However the case study didn't prove any unique results for improvements. One can argue that those unique results are not possible at all to achieve without participation of cognitive experts in evaluation. So, why application of original CD questionnaire for VML evaluation was not a full success? Is that a problem of cognitive dimensions themselves or just customisation of the questionnaire? Discussion of scope of cognitive dimensions is provided in the next section, however the questionnaire itself strongly influenced achieved results, so more investigations with customised one should be made.

5. Comparison of CD and a set of VML quality criteria

Key issue in evaluation of scope of cognitive dimensions is discovering a perspective that could be a good frame of reference. Perfect cognitive frame of reference should be superset of cognitive dimensions that covers universal cognitive aspects of human perception and information processing. However because of large number of psychological theories that answer only selected questions nowadays as well as holistic nature of human's mind elicitation of the complete set of right aspects is almost impossible. Another proposal for a good frame of reference can be perspective that comes out from totally independent source and is concerned with context of use. In this section, a set of VML quality criteria used in software engineering community is described and cross-comparison between its scope and CDs scope is made.

One of important parts of software engineering is quality assurance. As the description of quality understanding in the project is a basis of all quality activities, there are several works and standards that supply the community with quality terminology. In a few papers there are

also trials of description what is high quality notation. A selection of the criteria with short descriptions grouped into categories of model creation and their further usage can be found in the table 5.

Criteria related to easy and efficient model creation:

- Expression power – language provides mechanisms for modelling all the cases developer wants to model;
- Natural correspondence – it is easy to see correspondence between models and modelled fragment of reality;
- Simplicity – constructs are expressed in the possibly simple way;
- Changeability – changes are easy to introduce;
- Abstraction mechanisms – there are mechanisms that allow operation at several levels of abstraction;
- Precision – all users have the same interpretation of the model.

Criteria related to models usage:

- Include all the criteria of creating models with more emphasis to reading and understanding those models;
- Universal usage – language is easy to learn and use to satisfy needs of all stakeholders;
- Inspect-ability – models are easy to verify by human reviewers;
- Automation-ability – models are precise enough to allow for automatic processing.

Table 5. A set of quality criteria for VML.

The results of comparison are shown in the figure 1. One should remember, VMLQC are set of specific criteria for VML without considering cognitive perspective, and CD were originally developed for evaluation of notational systems. However it is expected that CDs are universal enough to indicate what VML designers find valuable.

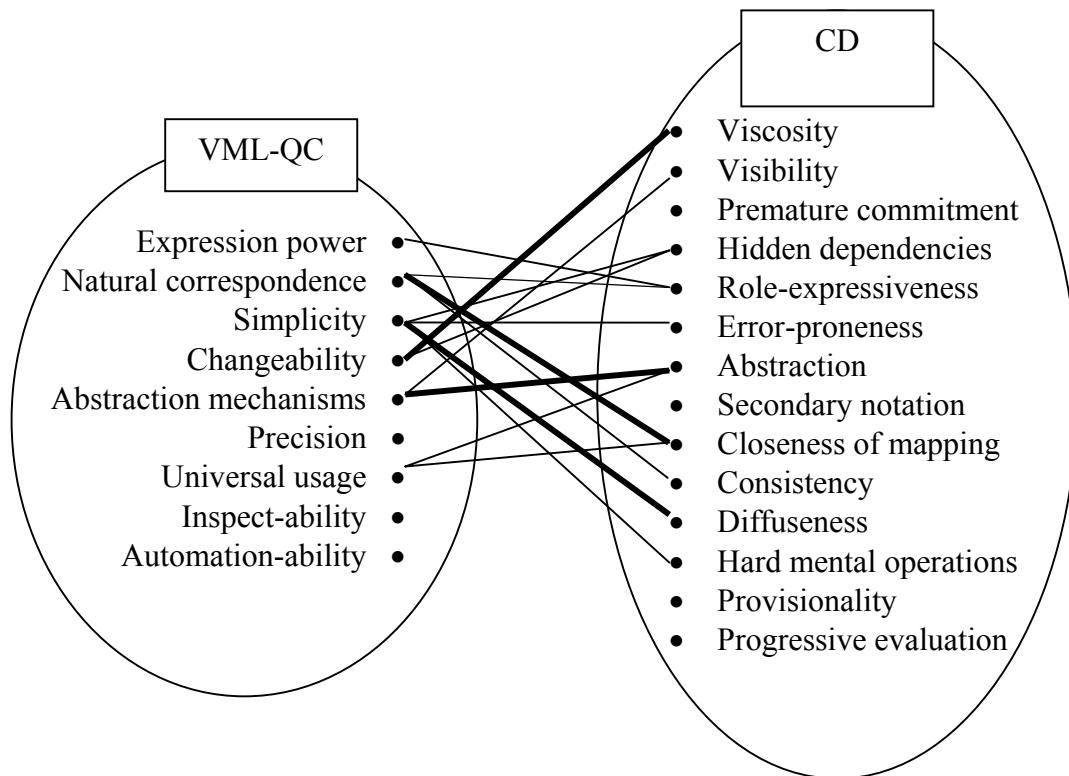


Figure 1. Connections between VML quality criteria and CD. Thickness of lines indicates strength of connection.

Comparison has shown that there are some concepts expressed by both VMLQC and CD, but some concepts are different. Concept of change was covered in both sets in the same scope - names expressing them are VMLQC–changeability and CD-viscosity. They differ only in the sign of positive/negative expression. Very close to each other are also notions of VMLQC-abstraction mechanisms and CD-abstraction. They are based on the common assumption that since humans have mental representations of things at different level of abstraction as well as capability of navigation between them the same features should be also covered by modelling languages or notational systems. (Partly it includes also CD-visibility). Another clear concept is association between reality and the model. The representations of this are VMLQC–natural correspondence and CD–closeness of mapping, although the first concept is also slightly related to CD–role expressiveness and CD–consistency. Yet another strong connection exists between VMLQC–simplicity and CD–diffuseness. Both of them are trying to answer the question of length of description needed to express a single concept. But VMLQC-simplicity is broader and includes also part of CD-hard mental operations and a part of CD-error proneness. There is some relation between VMLQC-expression power and CD-role expressiveness, however the first is wider and includes not only notion of clear understanding what is expressed by a given symbol, but also possibility to express all needed constructs.

CD-hidden dependencies influence in some way VMLQC-simplicity and VMLQC-changeability, and there is some relation between VMLQC-universal usage and CD-abstraction and CD-closeness of mapping, but it additionally indicates for need to take into account different types of users.

There are five VMLQC that do not have a strong connection with CDs. One can argue that two last: VMLQC-Inspect-ability and VMLQC-Automation-ability are very specific to modelling and no one could expect them to be covered in any general approach to evaluation, but notion of VMLQC-precision and VMLQC-universal usage probably are important also in other notational systems. Also VMLQC-expression power that indicates a need of fitness of the language (or notation) to the user's context of use could be important criterion of usability evaluation.

There are 10 CDs that have no strong connections with VMLQC. In the case of three of them: CD-premature commitment, CD-hard mental operations, and CD-progressive evaluation, the reason is they are related to the activities, not models or notations themselves. An example of CD, which is important for VML, but was not covered by these VMLQC is CD-secondary notation, which in UML has name of extension mechanisms. Others CDs, including CD-consistency, CD-error-proneness, CD-role-expressiveness, CD-hidden dependencies and CD-visibility can be adopted by the set of criteria and provide insights how those criteria should be extended, understood and measured.

The notions of precision and terminology require more comments. Since software developers have strong background in logic and maths, precision is very important for them. Some requirements for useful cognitive dimensions framework can be derived from that. It should be complete and precise, concepts should be orthogonal and possibly methods of measuring/evaluating them should be added; and as they are going to be 'discussion tool' description of them should be precise and easy to understand. To eliminate minor deficiencies of CD questionnaire use tool support is suggested. It is very difficult to discuss terminology since it is earlier education dependent. Terminology that seems more natural depends on which framework was learned first and on time spent while learning and using it. So, the longer people work with a given terminology and more connections to other knowledge they have, the framework seems to be more intuitive for them.

6. Conclusions

Cognitive dimensions seem to be a good discussion tool, providing terminology in perception of problems with notational systems. They are universal enough to allow seeing also some problems of visual modelling languages. However case study of their application revealed also some imperfections. Some of them are caused by generality of the questionnaire and probably can be eliminated by customising the questionnaire, but some of them require more work on cognitive dimensions themselves. As cognitive dimensions are the basis for the evaluation and their quality influence quality of results of the VML evaluation there are requirements of more precision and better description of them. CD framework should deliver orthogonal, complete and easy to understand set of concepts.

Not all questions were possible to answer in this case study. Customisation of the cognitive dimensions questionnaire for VML evaluation is planned as a next step of investigation. The most important factors are increase in precision, usage of language definition terminology in the questions, and clear specification of subjects of evaluation. Some more work is also suggested to increase efficiency of the questionnaire's usage.

7. References

[1] A Bibliography of Cognitive dimensions

<http://www.cl.cam.ac.uk/~afb21/CognitiveDimensions/CDbibliography.html>

[2] Blackwell A.F., Green T., A Cognitive dimensions questionnaire (v. 5.1, 2000),

<http://www.cl.cam.ac.uk/~afb21/CognitiveDimensions/>

[3] Object Management Group, Unified Modeling Language v. 1.4, <http://www.omg.org/uml/>

[4] Poseidon for UML v.1.4.1, <http://www.gentleware.com/>

[5] Princeton Softech, Select Enterprise v. 6.1, <http://www.princetonsoftech.com>.