The Effect of Formal Representation Formats on the Quality of Legal Decision-Making

Tom van Engers[1] Liesbeth van Driel Margherita Boekenoogen[2]
Belastingdienst
Postbus 18280, 3501 CG
Utrecht, The Netherlands
[1]*t.m.van.engers@acm.org,* [2]*m.boekenoogen@planet.nl*

Abstract. Accessibility of legal sources is crucial to both jurists and citizens. As the development of E-government is speeded up the number of legal information systems including knowledge-based services is likely to increase accordingly. In the Program for an Ontology-based Working Environment for Rules and legislation (POWER) the Dutch Tax and Customs Administration (Belastingdienst) developed a formal modelling approach for modelling legal sources. The results of applying this modelling process are formal models expressed in UML/OCL which can be used as the basis for amongst others verification, simulation and application generation.

Since the legal knowledge to be applied in the operational units and on the Web is represented in these formal models, their quality needs serious attention. To be able to determine their quality, inspection by (legal) experts should be possible (validation). In most cases however these experts are not able to read these UML/OCL-models. Therefore a representation is required that can easily be derived from these models and that is easy to understand.

Based on previous research we compared two different knowledge representations and tested their learnability and usability. Furthermore we examined the effect of representation on legal decision-making. The two representation-forms used are production rules and scenarios. The results of our experiment show that the performance of scenarios is significantly better then the performance of production rules.

Keywords. Legal knowledge-based systems, Legal knowledge representation, Development of legal systems, Validation, Usability testing.

1 Introduction

Drafting and implementing new legislation is a rather time, energy and money consuming process. The many inter-connected processes and the large number of people involved make it very vulnerable to errors. Varying interests have to be aligned and communication difficulties due to differences in technical jargon have to be overcome in both drafting and implementing changes to legislation or even completely new legislation. The knowledge and experience needed to create new laws, specify, design and implement procedures and systems in legislative domains are very scarce. Getting the right knowledge at the right place at the right time is a critical success factor for the ability to effectuate the legislative power to regulate and control.

The Dutch Tax and Customs Administration (DTCA) conducts a research program called POWER (Program for an Ontology-based Working Environment for Rules and regulations), aimed at developing a method and supporting tools for the whole chain of processes from legislation drafting to executing the law by government employees and supporting citizens (see e.g. [3, 4, 5, 6]).

Tom van Engers, Liesbeth van Driel and Margherita Boekenoogen, ‘The Effect of Formal Representation Formats on the Quality of Legal Decision-Making’ in T.J.M. Bench-Capon, A. Daskalopulu and R.G.F. Winkels (eds.), *Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference*. Amsterdam: IOS Press, 2002, pp. 63-71.

To support the governments' employees and citizens the DTCA builds knowledge-based applications that run on the governments' infrastructure or on the Web. The formal models that contain the legal knowledge on which these knowledge-based systems are based upon are hard to understand for non-ICT people. To be able to determine their quality inspection by (legal) experts should however be possible (validation). Therefore a representation that besides containing the same information can easily be derived from these models and that is easy to understand is required.

1.1 A Preliminary Study on Learnability and Usability of Representational Formats

In an internal study ([8]) we examined five different commonly used knowledge representations:

1. Horn clauses;
2. Decision trees;
3. Production rules;
4. Grammars and
5. Decision tables.

We examined the learnability and usability of these representation forms by presenting ten questions to 15 different jurists (6 legislation drafters and 9 other jurists). Half of the questions consisted of a concrete piece of legal text and a multiple-choice (4 choices) presentation of the content of that text in one of the above-mentioned representational formats (one representing the text the other three containing errors). The other half of the questions consisted of a formal representation in one of the above-mentioned representational formats and a multiple-choice (4 choices) presentation legal text (one concordant with the presented representation). From that experiment we found that legislation drafters performed better (97% for finding a legal text that corresponds with a representation and 100% for finding a representation with a legal text) then other jurists (80% for finding a legal text that corresponds with a representation and 89% for finding a representation with a legal text).

We also measured what was the most preferred knowledge representation. The subjects indicated that decision tables were preferred by 60%, production rules by 27% and decision trees by 13% (no preferences for Horn clauses and grammars). Since decision tables proved to have lower overall performance (82%) compared to production rules and decision trees (both 94%) we decided to include production rules and decision trees in a consequent experiment (although we used scenarios in stead of the original decision trees as explained in the next section).

1.2 POWER Scenarios

Within the POWER research program we developed scenarios for mapping the reasoning strategies of legal experts in order to obtain a better understanding of how experts apply the law or regulations. In the projects run at the DTCA we experienced that at the start of the knowledge modelling process knowledge analysts, and in case of modelling new or complex legislation, legal experts want to have a quick and global overview of the legal domain at hand.

This is best obtained by looking at how some (prototypical) cases are solved that correspond to a certain target group. The knowledge analyst asks the experts to explain how he or she applies the legislation (within the domain of interest) to a certain prototypical case. The reasoning strategy is then mapped onto a decision tree. The decision tree is subsequently elaborated until all cases within the range of interest can be ”handled” by the decision tree. The join of all scenarios corresponding to solving a case using the legal source(s) forms the final scenario: a map of the legal domain expressed in the form of a kind of decision tree. The nodes of the map correspond to questions or decisions that follow from applying the legislation: a node typically contains a reference to a part of the legal source it is based on. A decision needs to be taken by a (certain) yes or no. Traversing the scenario a result or conclusion is reached. Figure 1.2 presents an example of a part of a scenario that was made of a bill concerning subsidies for children’s day nursery.

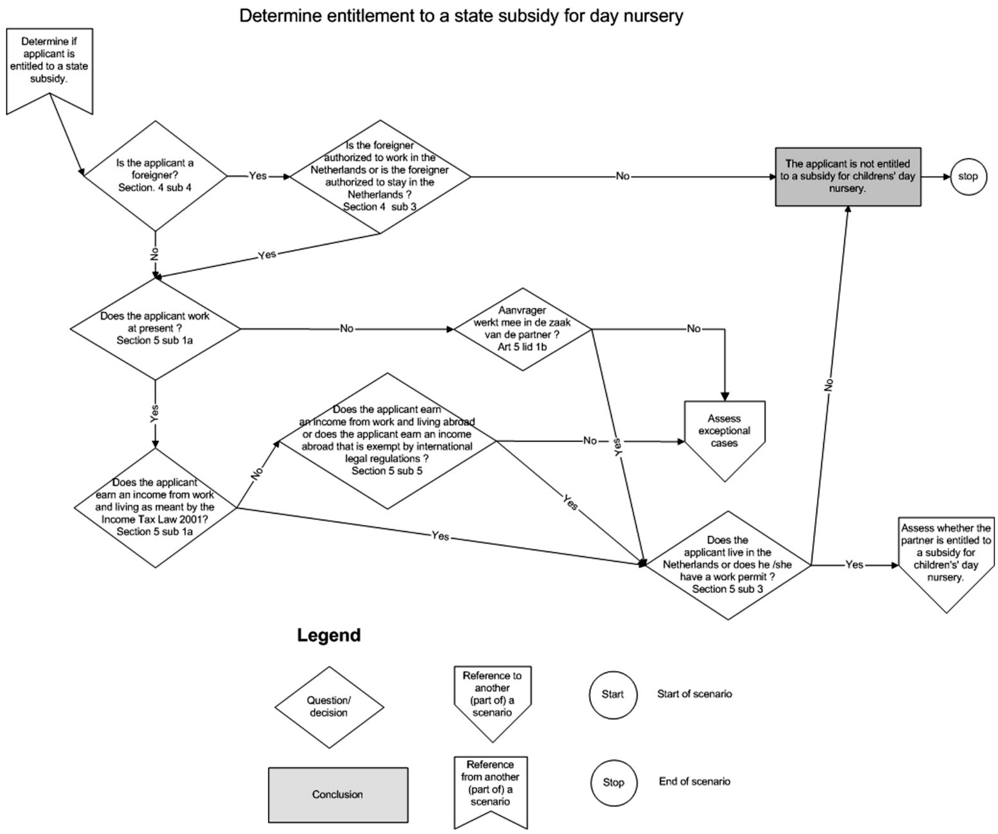

Figure 1: Example of a (part of) a scenario called ”determine entitlement to a state subsidy for children’s day nursery”.

For computational issues the expressiveness of a UML type model using OCL constraints is better suited ([6]) than scenarios because scenarios lack the benefits of a strict formal model. But nevertheless scenarios seem to provide us with a good insight for the legal domain represented, especially when the legislation involved is to be used in a categorisation or assessment task.

The questions or decisions in scenarios are kept as global as possible because we aim to provide no more than a quick overview of how a certain legal domain functions. If everyone involved in the modelling process shares a global understanding of the domain we might for example ask whether a certain section applies instead of posing several separate questions, each spelling out the exact conditions of the specific section. As with calculations, we leave the details to POWER UML models of legal sources ([6]).

POWER scenarios provide a pure functional view on how legislation is applied by experts to solve certain cases: it does not provide a ”system view” on how a decision support system would function. This is where a POWER scenario differs from a UML scenario, which is a story about how a system will be used. A UML scenario describes a prototypical sequence of interactions in a business-collaboration or the system context ([2]). The main difference between POWER scenarios and UML scenarios is that UML scenarios are used to define the boundary of a system, whereas POWER scenarios can be considered as a (global) specification of the knowledge intensive process (which could be supported by a system). However different, UML scenarios and POWER scenarios match when it comes to the goal of capturing the task flow. UML scenarios are used to capture the ideal task flow as perceived by end users. POWER scenarios provide us with the means for discovering the implicit tasks and task flow within legal domains. Legislation typically is declarative in nature. Task and task flow are revealed when asking the expert to apply the legislation to solve cases. In POWER scenarios, tasks are represented by grouping questions concerning one issue on one diagram (or more if necessary) and naming the diagram accordingly to the issue at hand. Note that such a diagram not necessarily contains an end point of the reasoning path: the decision tree may be connected to another issue (”task”). In the next sections scenarios refer to POWER scenarios, not UML scenarios (unless explicitly stated otherwise).

1.3 Comparing Scenarios with Production Rules.

In the empirical study presented in this paper we compare two different knowledge representations; production rules (see Figure 1.4) and scenarios.

In the POWER research program scenarios are commonly used for establishing a global view on the experts reasoning strategy. Experts have said that these scenarios were very understandable and easy to validate. The rationale behind this study is that we want evidence that scenarios are adequate for our purposes (easy to understand by non-ICT people, i.e. experts, communicable between domain experts etc.). In the near future we hope to generate these scenarios from our formal POWER-models (representing legislation) so that these scenarios can be easily used for validation. The final goal is to improve the quality of law-enforcement and optimally support legal experts when drafting legislation.

We want to quantify the performance of the two selected knowledge representations, in other words: we want to measure the effect of the representational format on the expert’s performance in a problem-solving task. Performance is expressed in both duration of the task (time needed to come to a solution) and correctness. As our 0-hypothesis we expect no significant difference in performance between these representational formats.

1.4 Research Approach

We use a representative part of the scenario about deciding into which fiscal category a capital-insurance contract would fit (which is a topic in the Dutch Income Tax Law). We model the knowledge about this domain in the form of a scenario and in production rules.

Both knowledge representations are presented to 20 subjects all having a juridical background. Every subject is given four different cases that have to be categorised. To do so all cases need more than seven steps (respectively 11, 9, 12 and 14 steps) which makes the cases complex enough to not fit in working memory. Consequently it is very unlikely that a subject can decide what the solutions are at a glance.

The subjects have to categorise two cases using the production rule representation and two cases using scenarios. The subjects are asked to think aloud. Before starting the experiment the subjects are given an instruction task.

```
A.  ALS   NIET de verzekeraar een professionele is
          EN
          verzekering uitsluitend bij ziekte, ongeval of invaliditeit
          EN
          lijfrenteclausule
          EN
          NIET reeks van uitkeringen
          EN
          vast aantal termijnen
    DAN   geen kapitaalverzekering

B.  ALS   NIET de verzekeraar een professionele is
          EN
          verzekering uitsluitend bij ziekte, ongeval of invaliditeit
          EN
          lijfrenteclausule
         ((EN
          NIET reeks van uitkeringen
          EN
          NIET vast aantal termijnen)
          (OF
           reeks van uitkeringen)
          EN
          datum van totstandkoming is tussen 01-01-1992 en 15-09-1999
          EN
          jaarlijks vanaf aanvang premiebetaling
          EN
         ((tenminste 15 jaar jaarlijks premiebetaling
           (OF
            NIET tenminste 15 jaar jaarlijks premiebetaling
            EN
            binnen bandbreedte 1:10))
          (OF
          NIET jaarlijks vanaf aanvang premiebetaling
          EN
          binnen bandbreedte 1:10))
    DAN   verplicht box 1, niet KEW (1992 tot 15-09-1999)

C.  ALS   NIET de verzekeraar een professionele is
          EN
          verzekering uitsluitend bij ziekte, ongeval of invaliditeit
          EN
          lijfrenteclausule
         ((EN
          NIET reeks van uitkeringen
          EN
          NIET vast aantal termijnen)
           (OF
            reeks van uitkeringen))
          EN
          datum van totstandkoming is tussen 01-01-1992 en 15-09-1999
          EN
          jaarlijks vanaf aanvang premiebetaling
          EN
          NIET tenminste 15 jaar jaarlijks premiebetaling
          EN
          NIET binnen bandbreedte 1:10
           (OF
            NIET jaarlijks vanaf aanvang premiebetaling
            EN
            NIET binnen bandbreedte 1:10)
          EN
          NIET bezit eigen woning
    DAN   box 1 of  3 (1992 tot 15-09-1999)

D.  ALS   NIET de verzekeraar een professionele is
          EN
          verzekering uitsluitend bij ziekte, ongeval of invaliditeit
          EN
          lijfrenteclausule
         ((EN
          NIET reeks van uitkeringen
          EN
          NIET vast aantal termijnen)
           (OF
            reeks van uitkeringen))
          EN
          datum van totstandkoming is tussen 01-01-1992 en 15-09-1999
          EN
          jaarlijks vanaf aanvang premiebetaling
          EN
          NIET tenminste 15 jaar jaarlijks premiebetaling
          EN
          NIET binnen bandbreedte 1:10
           (OF
            NIET jaarlijks vanaf aanvang premiebetaling
            EN
            NIET binnen bandbreedte 1:10)
          EN
          bezit eigen woning
          EN
          NIET verzekerd kapitaal op einddatum in geld
          EN
```

Figure 2: Example of a (part of) the production rules used to determine the fiscal category (box) of a capital insurance contract.

We use two conditions determined by the order of presenting the different knowledge representations. In one condition the *instruction* case is solved using a scenario representation followed by two *test* cases that have to be solved using production rules then followed by two *test* cases that have to be solved by using a scenario.

The second condition consisted of an *instruction* case solved by using production rules followed by two *test* cases that have to be solved by using a scenario then followed by two *test* cases that have to be solved using production rules. Ten subjects were tested according condition one and ten subjects according condition two.

To prevent from contamination with possible pre-existing knowledge of the domain the representations are modified so they don't represent the original law, which consequently helps to prevent the subjects in our experiment from using their internal mental model instead of our knowledge representations. Eventual pre-existing knowledge thus becomes useless in the problem-solving task.

The think-aloud protocol is recorded on a tape recorder. The reasoning processes executed while solving the task will be analysed from a written-out version of the protocol. The solution time is measured using a stopwatch while mistakes are counted.

Subjects are instructed that the representations are not representing the actual law and that using pre-existing knowledge will be of no use. Also the subjects are told that we want the subjects to think aloud, that everything will be recorded and that we will measure the time. Subjects are furthermore instructed that they should use all the time that they need to come to a good solution.

2 Results

2.1 General Remarks

First some general remarks about the conduct of the experiment. One subject refused to execute the task using the production rules. This subject told us (before even making any attempt to solve a case) that he was not capable to extract any information from this representational format at all! This subject was excluded from the analysis.

The production rules caused some confusion because the use of brackets that separated parts of the rules from other parts. One subject insisted that we didn't put the brackets at the right place. We had to explain how to read production rules (especially the meaning of the brackets) and we also had to make the subjects aware of the 'or'-clauses. We gave all subjects the advice to first check all the rules for a certain conclusion before excluding such conclusion. Subjects used to stop looking for further conditions once they discovered a rule that didn't match the case-data.

In general when production rules are used, their meaning is affected by the conflict strategy used. Also production rules can have several forms: e.g. or can be eliminated by using multiple production rules with just and. In this case we took care that the rule set didn't contain conflicting rules.

2.2 Statistics

Performance was measured using total solution time and the number of errors made. We also measured the effect of the experimental condition (presentation order) and the learnability of the two representational formats (i.e. the difference between the solution time of the first and the second case using one form of representational format).

Two Wilcoxon tests were used to calculate the effect of the representational format on performance. One was used to calculate the effect of representation on solution time, the other to calculate the effect of presentation on the number of errors. We used Wilcoxon (in stead of T-test) because we had dependent groups, non-equal distribution, outliers and the groups contained less than 30 observations.

This test showed that with a significant level of .05 the solution time using production rules is significantly longer than using scenarios ($z =$ -3.920; $p =$ .000).

The number of errors using production rules is significantly higher than using scenarios ($z =$ -2.530; $p =$ .011).

For the same reasons as mentioned before we applied Wilcoxon to calculate the learnability of the two representational formats. The test, using a significance level .025 and just looking at solution time, showed that the second presentation (same format) results in significant shorter solution time (production rule: $z =$ -3.136; $p =$ .002; scenario: $z =$ -3.249; $p =$ .001).

We used a Mann-Whitney test to calculate the effect of the test condition (i.e. the order of presentation) on performance (both solution time and number of errors). The reasons for using this test are concordant with the reasons already mentioned except for in this case the groups were independent.

We found no effect of order on solution time (significance level 0.025, z= -.681; $p =$ .496 for production rules; $z =$ -2.080; $p =$ .038 for scenarios). We didn't find an effect on the number of errors either (significance level 0.025, $z =$ -1.129; $p =$ .259 for production rules; $z =$ -1.453; $p =$ .146 for scenarios).

We checked if we could find an interaction effect between the effect of representation (solution time and number of errors) and condition. We also checked if we could find an interaction effect between the effect of learnability (production rules and scenarios separately) and condition.

The four Mann-Whitney tests applied show no interaction between representational format and condition. We however found interaction effects between learnability and condition for both representational formats. Because of this interaction effect we have to interpret learnability per condition, implying that we have to recalculate using Wilcoxon tests (since the groups showed to be dependent with respect to this variable). This calculation shows a learning effect only in condition one for production rules (z = -2.803; p = .005) and a learning effect only in condition two for scenarios (z = -2.805; p = .005).

Table 1: Results of the statistical analysis

			z-value	*p-value*
effect of repr. on solution time			-3.920	.000[1]
effect of repr. on number of errors			-2.530	.011[1]
effect of condition on solution time		*prod. rules*	.681	.496[2]
		scenario	-2.080	.035[2]
effect of condition on number of errors		*prod. rules*	-1.129	.259[2]
		scenario	-1.453	.146[2]
learnability		*prod. rules*	-3.136	.002[2]
		scenario	-3.249	.001[2]
interaction repr. (solution time) and condition			.000	1.000[1]
interaction repr. (number of errors) and condition			-.587	.557[1]
interaction between learnability and condition		*prod. rules*	-2.268	.023[1]
		scenario	-3.480	.001[1]
learnability after check for interaction effects	*condition 1*	*prod. rules*	-2.803	.005[3]
		scenario	-1.173	.241[3]
	condition 2	*prod. rules*	-1.070	.285[3]
		scenario	-2.805	.005[3]

[1]: *significant if* $p < .05$
[2]: *significant if* $p < .025$
[3]: *significant if* $p < .0125$

2.3 Verbal Protocol Analysis

We executed verbal protocol analysis to get an impression of how subjects think when using different representational formats in a problem-solving task. To be able to do so we wrote out the complete recordings of the first ten subjects (half of them condition one the other half condition two). For the analysis we used the 'search-inference framework' described by Baron ([1]):

- The search for options
- The search for goals
- The search for prove
- Options found
- Prove found
- Conclusion made

These aspects don't occur in a specific order. Overlap exists and one may switch form one to another continuously. Baron aims with this framework to identify basic elements of thinking.

We used the first three protocols to make a coding schema. During this step it became clear that some interesting aspects weren't included in Baron's original framework:

- Evaluation of the task and one's own performance
- Moments of confusion or complete silence

- Irrelevant remarks

Analysis of the protocols showed differences between production rules and scenarios. The protocol-analysis shows that more steps are needed to come to a conclusion using production rules compared to scenarios. Furthermore we found a greater diversity of aspects in the analysis of tasks in which production rules were used. In tasks using scenarios we observed that just searching for options and comparing the cases to them were used. Moments of silence and confusion were more often detected in tasks were production rules were used.

3 Conclusion

We found a number of effects of knowledge representation on the performance of a legal problem-solving task. First of all the solution time of scenarios is significantly shorter than production rules while the number of errors made is significantly less. In other words scenarios prove to be a significantly better knowledge representation.

From the fact that only learning effects were found in case the representation wasn't already used in the instruction task we conclude that both representations can be learned quickly and are in other words easy to understand which is concordant with the study of Roose ([8]).

From the analysis of the verbal protocols we tend to conclude that production rules demand greater cognitive effort compared to scenarios (because the number of steps needed is much higher when using production rules and so is the diversity of aspects of the coding schema applicable to the protocol). This should be expected since much of the work of the rule interpreter is "compiled into" the scenarios. But in our research we were at this point focussing on differences in performance rather than establishing the effect of implicit or explicit interpretation.

The conclusions from the verbal protocol analysis have to be handled prudently for we didn't evaluate the coding schema that we used (i.e. we didn't calculate e.g. the reliability of the schema). Normally one should check the sensibility of the schema for different analysts (by comparing their scores for the same cases). This 'inter-coder' reliability indicates the ambiguity of the coding schema ([9]).

4 Epilogue

The differences found between different knowledge representations are significant. The mental processes that take place in human's heads when solving a legal-reasoning task still intrigue us. These mental processes seem to be influenced dramatically by the knowledge representations that are provides. Scenarios for example seem to reduce the time needed to come to a solution because reasoning steps are limited. We however should further examine the reasons behind that phenomenon. As stated before the interpretation of the rules is as it were 'incorporated' in scenarios which raises questions as 'what is the mental load of this interpretation process' and 'what's the effect of scenarios in freezing alternative reasoning paths that would be more effective'.

The goal of the experiment described in this study was to check if scenarios could be adequate representations of the legal knowledge described in formal POWER-models. Furthermore these scenarios should be accessible and learnable by non-ICT people especially domain experts in order to validate the legal knowledge represented by these formal models.

In the near future we hope to be able to automatically generate scenarios from the POWER-models. Currently scenarios are used as a top-down analysis that facilitates our knowledge analysts to get a global picture of the domain of discourse. These scenarios should be accessible by the domain experts for scenarios are meant to be a communication vehicle, but also to allow them to validate the knowledge represented in such scenarios.

Both jurists and ICT-people have developed different ways of expressing themselves. This has thus far lead to a lot of misunderstanding and consequent problems in the execution of the law. The scenario representation we developed may bridge the 'language-gap' between jurists and ICT-personnel and consequently improve law enforcement.

References

[1] Baron, J. (1994). *Thinking and Deciding. Second Edition.* Cambridge University Press: Cambridge.

[2] D'Souza, Desmond Francis & Alan Cameron Wills. (1999). *Objects, Components and Frameworks with UML – The Catalysis Approach.* Reading Mass. Addison-Wesley Longman, Inc, 1999

[3] Van Engers, T.M., Kordelaar, P.J.M., 1999, POWER: Programme for an Ontology based Working Environment for modelling and use of Regulations and legislation, in ISMICK'99 proceedings, ISBN 2-913-923-02-X.

[4] Van Engers, T.M., Kordelaar, P.J.M., Den Hartog, J., Glassée,E., 2000, POWER: Programme for an Ontology based Working Environment for modeling and use of Regulations and legislation. In Proceeding of the 11th workshop on Databases and Expert Systems Applications (IEEE) eds. Tjoa, Wagner and Al-Zobaidie, Greenwich London, ISBN 0-7695-0680-1, pp. 327-334.

[5] Van Engers, T.M., Glassée, E., 2001, Facilitating the Legislation Process using a shared Conceptual Model. In: IEEE Intelligent Systems January 2001, p. 50- 57.

[6] Van Engers, T.M., Gerrits, G., Boekenoogen, M.R., Glassée E., 2001, Kordelaar, P. (2001). *POWER: Using UML/OCL for Modeling Legislation – an application report*. Proceedings of the ICAIL 2001, St. Louis.

[7] Pagano, R.R. (1994). Understanding Statistics in the Behavioral Sciences. Fourth Edition. West Publishing Company: St. Paul.

[8] Roose, H.M. (2000). Stageverslag B/CPP, domein vakontwikkeling klantbehandeling. Internal publication of Project group POWER. Belastingdienst/ Centrum voor Proces- en Produktontwikkeling: Utrecht.

[9] Van Someren, M.W., Barnard, Y.F. & Sandberg, J.A.C. (1994). *The think aloud method. A practical guide to modelling cognitive processes*. Academic Press Limited: London.