AN EVALUATION OF COGNITIVE MODELING TOOLS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

CAN BICAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

MAY 2007

Approval of the Graduate School of Informatics Institute.

____

Assoc. Prof. Dr. Nazife Baykal
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

____

Prof. Dr. Deniz Zeyrek
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

____

Assist. Prof. Dr. Bilge Say
Supervisor

Examining Committee Members

Prof. Dr. Deniz Zeyrek	(METU, COGS)	____________
Assist. Prof. Dr. Bilge Say	(METU, COGS)	____________
Assoc. Prof. Dr. Cem Bozşahin	(METU, CENG)	____________
Assoc. Prof. Dr. Kürşat Çağıltay	(METU, CEIT)	____________
Assist. Prof. Dr. Annette Hohenberger	(METU, COGS)	____________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Can Bican

Signature :

ABSTRACT

AN EVALUATION OF COGNITIVE MODELING TOOLS

Bican, Can

MSc, Department of Cognitive Science

Supervisor: Assist. Prof. Dr. Bilge Say

May 2007, 74 pages

This thesis evaluates several aspects of the cognitive modeling tools, using a questionnaire as the survey method. We try to assess the the suitability for cognitive modeling task of the cognitive modeling tools, from the perspective of international community of cognitive modeling tool users. Part of this assessment is done with respect to general usability of software and the rest is specialized for the cognitive modeling issues. Frequency and correlation analyses reveal that there is a significant relationship between suitability as a software product and suitability as a cognitive modeling tool. Specifically, there are correlations between the features of the tool involving flexibility, presentation of input and output and the process of design, implementation and evaluation of a cognitive modeling tool, while these processes are negatively related to adversely effecting features of the tool, such as having to do extra tasks that are not related to the actual task. Our study confirms that a cognitive modeling tool can also be evaluated from the perspective of a general purpose software product, and also gives clues about directions for improvement to tool developers.

Keywords: cognitive architectures, cognitive modeling, software evaluation, cognitive modeling tools

ÖZ

BİR BİLİŞSEL MODELLEME ARAÇLARI DEĞERLENDİRMESİ

Bican, Can

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Assist. Prof. Dr. Bilge Say

Mayıs 2007, 74 sayfa

Bu tez, bilişsel modelleme araçlarını inceleme yöntemi için bir anket kullanarak, farklı açılardan değerlendirmektedir. Bilişsel modelleme araçlarının bilişsel modelleme işine uygunluğunu, uluslararası bir bilişsel modelleme araçları kullanıcıları topluluğunun görüş açısından incelemeye çalıştık. İncelemenin bir kısmı yazılımların genel kullanılabilirliğine göre yapılırken, geri kalanı da bilişsel modelleme sorunlarında özelleşmektedir. Sıklık ve bağlılaşım çözümlemeleri, bir yazılım türü olarak uygunluk ile bir bilişsel modelleme aracı olarak uygunluk arasında önemli bir ilişki olduğunu göstermektedir. Özellikle aracın esnekliği, girdi ile çıktılarının sunumu ile bilişsel modelleme aracının tasarım, gerçekleştirme ve değerlendirme süreci arasında bir bağıntı görülürken, bu süreçler aracın asıl işle ilgisi olmayan fazladan işler yapmak zorunda bırakması gibi olumsuz özellikleriyle tersine bir bağlılaşım içindedirler. Çalışmamız, bir bilişsel modelleme aracının aynı zamanda genel amaçlı bir yazılım ürünü olarak da değerlendirilebileceğini doğrulamakta ve araç geliştiricileri için ilerleme yönleriyle ilgili ipuçları vermektedir.

Anahtar Kelimeler: bilişsel mimariler, bilişsel modelleme, yazılım değerlendirmesi, bilişsel modelleme araçları

ACKNOWLEDGMENTS

It took many people to make this thesis possible, and I can only list a few here. My parents, Seyfettin Bican and Sibel Bican, always come first, for they have always convinced me that I can do better, no matter how good I have done. Bilge Say, my advisor actually made this work possible. I remember many times that I decided to shut down the word processor and walk away, but her support alone kept me going. I should also thank my colleagues in Tübitak, who were patient enough while I was busy with the thesis. So does Ayça Bulut, who always convinced me that I was doing the best thesis ever, against all counter evidence I provided. Asena Devlet helped me with my silly spelling mistakes, without her this study would look like a threatening letter from a teenager.

I also would like to thank everybody who contributed to the questionnaire of this thesis. There are so many people who generously helped with the pilot and the actual questionnaires that I cannot list all of them here. I am grateful to all who have contributed, this would have been a dream without them.

I have not addressed many people I should really have, but it has been a long way from the beginning to the end of this thesis. Thank you all, for at least not bothering me while I was thinking of how to rephrase a sentence for the fifth time.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Cognitive modeling is the modeling of cognition either as a method in cognitive science or as components of artificial intelligence applications. Cooper, Fox, Glasspool, and Yule (2002) discuss *modeling* as having the role of refining theories about cognition and putting them in a clear form for communicating ideas. A good model is distinguished by its complete description in representation of what it is trying to model. It needs to make sure that no essential features of the original have been left out and no extra components and relationships, which do not exist in the original, have been introduced. Since a model is a way of visualizing the theory, it enforces precision on the theory itself so that any vague or ambiguous aspects of the theory, or aspects that may have dual interpretations become evident and are taken out during the modeling. Moreover, the use of a common language in presentation of models supports clear communication.

Computer aided modeling not only supports but also enforces these two aforementioned features. By enforcing that no details can be left out or be vague, and by introducing a computer language for communication, a computer model has the potential to help the theory get more descriptive and explanatory.

These requirements and properties of cognitive models also serve the purpose of being predictive. This means that cognitive models are tools of the researcher to test the predictions of the theory it models. However, there are issues in assessing whether a given model has the proper power to predict. For example, the problem of "overfitting", where the model fits perfectly to the given data but fails to generalize the concepts. Similarly, a model having high complexity, even if it can model the theory, does not count for a plausible model, since a possible simpler model is favored. Therefore, evaluation of models through preferably comparative techniques that involve validating the model need to be employed.

One can view the use of computational methods in cognitive modeling in two ways. First of all, a defined language (in the form of a standard computing language or a domain-specific language) for describing the model is given in some computing system. Secondly, by viewing mind as an information processor and cognition as information processing, justification of theories is available in a computing machine[1]. Thus, new hypotheses, supporting experiments, and explanations for existing ones could be a result of such descriptive and explanatory modeling practices.

Although most modeling tasks can be accomplished by common computing languages, representing domain specific conventions in these might not always prove to be a straightforward task. It is always possible that some aspects of the model are degraded to common implementation details of this standard language, affecting the ability to supply the need of a ground for communication negatively. This issue is not specific to cognitive modeling situations. Modeling in engineering mostly makes use of customized tools or libraries, saving the researcher from reinventing the wheel with every model and establishing a common ground for improvements balanced with simplicity. But this comes with a price: Although using well-established tools is convenient, it is harder to customize a fundamental component when and if there is a need to. Likewise, computational implementations of cognitive modeling tools may limit the researcher to what they offer.

Domain-specific tools or languages claim to remedy this situation by providing their users with ready-made infrastructure of commonly used patterns of models and domain-specific languages that outline the underlying theory more clearly. For this to be convenient, the language itself and the researcher must already be in

[1] However, some forms of connectionist and dynamic models do not subscribe to this view.

sync in terms of knowledge and the modeling paradigm. The components of the infrastructure should match what the researcher requires.

We tend to view a cognitive modeling tool in two aspects, both as a software product in its general terms and as a domain-specific tool. This view also reflects to the user of the tool. The user is the one who uses the software product, having experiences shaped negatively or positively by the usability of it. She is also a researcher, modeling via the use of the tool, translating the theory to the language of the tool.

There have been attempts to evaluate and compare tools in terms of their capabilities. Langley, Laird, and Rogers (2006) provide a recent review of several cognitive architectures and frameworks and discusses capabilites that they should support. Ritter (1992) provides a questionnaire as part of his dissertation work that examines the strengths and weaknesses of the Developmental Soar Interface[2], in which questions focus on the usage frequencies of various different user interfaces of Soar, and if users think additional features should be provided. Soar and ACT-R in particular have been compared in (Ritter & Wallach, 1998) and (Jones, 1996), also in terms of capabilities. However, there is no research regarding the usability of the tools themselves.

The aim of this thesis is to provide an evaluation of the cognitive modeling tools with respect to perspectives discussed above, using a questionnaire, with target population being the active cognitive modelers using these tools. The evaluation covers questions related to the profiles of respondents, their training backgrounds on the tool they use, the support they receive for the tool, and their perceptions of the tool in terms of suitability for the task they use it for, both as a general software product and as a specific cognitive modeling tool. Questions in the parts of the questionnaire about training and suitability for the task are studied for significant correlations.

---

[2] Developmental Soar Interface (DSI) is a graphical and textual interface for Soar that displays models and lets the user modify them.

There are advantages or disadvantages of questionnaires, in terms of cost, access to the population, implementation, timing, design and administration (Bourque & Fielder, 1995). Although there are administrative disadvantages such as having no control over who responds[3], advantages of selecting the questionnaire as the method of survey makes it convenient for this study: The target population is geographically dispersed, implementation is easier via web-based methods, cost and timing is not directly relevant to this study.

The nature of our research questions include studying the effects of the cognitive modeling tools in the process of cognitive modeling practice. We also view the cognitive modeling tools as software products, and question their usability and the way they are perceived as a software product by the researcher.

The thesis is organized as follows: Chapter 2 provides a view on cognitive modeling and its issues, taking into account the paradigms in cognitive modeling and their implementations, along with the previous studies evaluating them. Chapter 3 is about the methodology used in the evaluation of the modeling tools in this thesis, explaining the design and conduct of the survey. Chapter 4 provides an analysis and discussion of the results of the questionnaire. Conclusions are discussed in Chapter 5.

[3] We have tried to compensate for this by targeting specific cognitive modeling communities and making the questionnaire invitation only.

CHAPTER 2

EVALUATION OF COGNITIVE MODELS AND ITS TOOLS

In this chapter, we will present the basic ways of how cognitive modeling is done, and introduce various architectures and frameworks used for cognitive modeling. Among these tools, we have specifically targeted ACT-R and Soar, which we also received most of the responses from their users. We then proceed with brief introductions to various cognitive architectures and other tools that the other respondents declared that they are using.

2.1 Cognitive Modeling Paradigms

A model is a re-construction of the real thing, letting the researcher perform reasoning about the relationships between the variables taken into account in the model. Cognitive models aim to represent the mental processes – mostly a part of their functionality – with the premises that mind is an *information processor* and cognition is the act of *information processing*. Computer models provide a way into explaining, predicting or strengthening postulation of possible mechanisms in the brain, to be validated by different methods in Cognitive Science. A computer model of a cognitive process also forces theoretical precision by requiring computational completeness and supporting clear communication by publicly-specified computer languages.

General perspective on cognitive modeling can be studied in three schools that often overlap in descriptions of particular models. Symbolic models suggest that information processing can be described in terms of manipulation of symbolic representations[1]. Symbolic models are systematic, in the sense that they consist of a number of parts which can be replaced with their similar counterparts. They are also compositional, that is, the function of the model is a function of its parts. The other perspective is the connectionist modeling which assumes that the properties of neural tissues are critical to understand and explain how mind works. Similar to neural configuration of the brain, models are constructed out of simple interacting units functioning in parallel. Connectionist models suggest feature-based representations which are representations of activation values in a connectionist network[2]. Dynamic view of cognition suggests that cognitive systems can be described in the same concepts as dynamic systems, described in terms of differential equations explaining how the system evolves over time[3].

Hybrid models of symbolic and connectionist models are reconciliation of two understandings of modeling. Since symbolic models are effective in high level processes of mind and connectionist models excel in describing low-level processes, a mixture of the two concepts help to generate better models. In such models, the symbolic part provides explicitness and access to abstract knowledge, while the connectionist part allows for biological plausibility, learning, fault-tolerance and generalization of input.

Cognitive architectures and tools are implementations based on one or a combination of these paradigms, and build foundations for models involving paradigms. Therefore, the underlying paradigm of a cognitive modeling tool is a definitive component on discussing the evaluation issues related to the tool itself.

---

[1] Polk and Seifert (2002) covers major approaches and architectures, including symbolic ones.

[2] A classical overview of studies on connectionist systems is (Rumelhart, McClelland, & PDP Research Group, 1986).

[3] An overview of dynamical systems in cognitive science can be found in (Beer, 2000).

2.2 Cognitive Architectures and Frameworks

Cognitive architectures are theories about large-scale structure and organization of cognitive processing. They are to be distinguished from cognitive modeling tools in the sense that the term *tool* in this thesis refers to a computer-based tool designed and implemented as an aid to do computational cognitive modeling and simulation. Implementations of cognitive architectures and frameworks, in the context of this study, are therefore a subset of tools that make it possible to develop models which are based on the corresponding theory of cognition. Evaluations of models depending on cognitive architectures and frameworks also need to address this dependence on the underlying theory.

The need and readiness of the current level of psychology for establishing such theories are outlined by (Newell, 1990). Newell defines "unified theories of cognition" as "single sets of mechanisms that cover all of cognition"(Newell, 1990, p.15) and cognitive architectures as the implementation of unified theories.

A particular cognitive architecture consists of components such as long and short term memory, language subsystem, learning mechanisms and decision components. Just as a computer architecture refers to the parts of hardware and software that is fixed for all applications, a cognitive architecture describes a group of components that allow for a unique platform for building models. Two of the most widely known examples of cognitive architectures are ACT-R and Soar. These two architectures are representative because of several reasons: Both are mature products which have been around for years, forming a community and support structures around them, having properties deemed necessary by (Newell, 1990) for unified theories of cognition and cognitive architectures.

Frameworks define a broader range of tools. Since frameworks, in this context, are the basis of software products that consist of basic facilities to model human cognition, cognitive architectures can also be considered to have frameworks which provide the underlying functions of the architecture but are constrained more strictly by underlying theories. However, the spectrum of cognitive modeling tools also contain tools which do not necessarily adhere to holistic theories as much as cognitive architectures do.

There are numerous architectures and frameworks (see Table 2.1 on page 8), which vary in degrees of comprehensiveness, cognitive constraints and commitments, and availability. We will emphasize on ACT-R and Soar and provide a general comparison for a few of the others in the following sections.

Table 2.1: Human Behavior Representation Architectures Available for Use (Gluck & Pew, 2005, p. 5)

Architecture	*For Additional Information...* (Invalid links from the original table are removed.)
ACT-R	http://act-r.psy.cmu.edu/
ART	http://web.umr.edu/~tauritzd/art/
Brahms	http://www.agentisolutions.com/brahms.htm
CHREST	http://www.psyc.nott.ac.uk/ research/credit/projects/CHREST
Clarion	http://www.cogsci.rpi.edu/ ~rsun/clarion.html
Cogent	http://cogent.psyc.bbk.ac.uk/
COGNET/iGEN	http://www.chisystems.com/
D-OMAR	http://omar.bbn.com/
EPAM	http://www.pahomeschoolers.com/epam/
EPIC	http://www.umich.edu/~bcalab/epic.html
Micro Saint, HOS, IPME	http://www.maad.com/MaadWeb/ products/prodma.htm
MIDAS	http://caffeine.arc.nasa.gov/midas
PDP++	http://psych.colorado.edu/~oreilly/ PDP++/PDP++.html
SAMPLE	http://www.cra.com
Soar	http://www.soartechnology.com

2.2.1 ACT-R

ACT-R (Adaptive Control of Thought–Rational) is a cognitive theory about how human cognition works, having a special coding language as its computational implementation. Anderson et al. (2004) describes ACT-R 5.0, along with some example applications[4]. ACT-R framework assumes that human knowledge can be divided into two parts. Declarative part stores facts and experiences. Procedural part contains knowledge about skills and procedures. In terms of ACT-R, declarative knowledge is represented as chunks that are made accessible via buffers, resembling the input/output streams in other computer languages. Buffers provide access paths to modules. Modules are largely independent systems, roughly representing brain structures. Combination of all the states of buffers cumulate to the state of the system. There are two types of modules, one for perceptual-motor activities and the other for memories. The most developed perceptual-motor modules of ACT-R are visual and manual modules. Memory modules are used for storing and retrieving either the facts (chunks) or the productions. Procedural memory, unlike the rest, has no buffers and its purpose is to access other modules' buffers. Procedural knowledge is represented as productions, representing the way how the system changes the contents of buffers. Apart from buffers and modules, the other main component of the ACT-R system is the pattern matcher. It searches for productions that match the current state of the system and executes them, modifying buffers. Figure 2.1 on page 10 is a graphical view of the ACT-R architecture.

[4] Current version is ACT-R 6.0, which can be found at
`http://act-r.psy.cmu.edu/actr6/`.

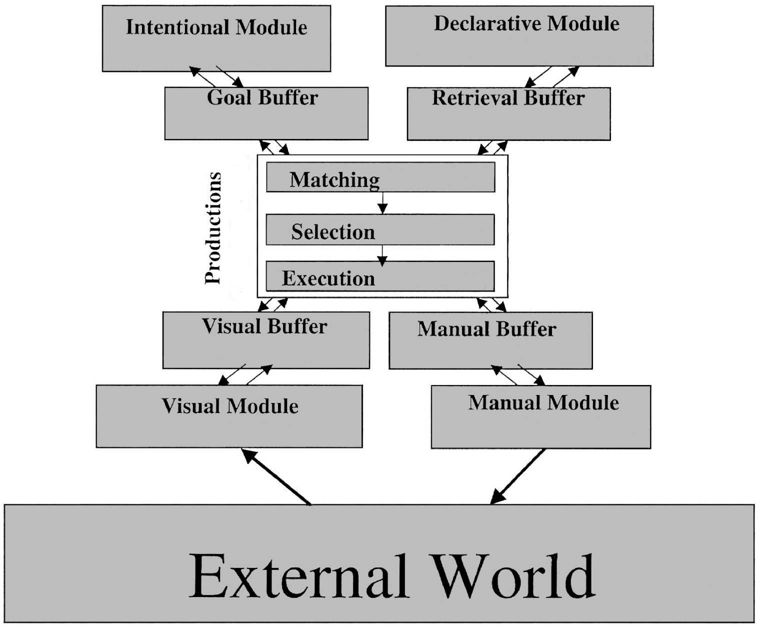

Figure 2.1: The organization of information in ACT–R 5.0.
Information in the buffers associated with modules is responded to and changed by production rules (Anderson et al., 2004, p. 1037).

ACT-R is a hybrid cognitive architecture, consisting of symbolic and subsymbolic systems. Symbolic systems in ACT-R consist of manipulation of chunks and application of production rules. On the other hand, learning in ACT-R forms a subsymbolic system.

ACT-R is used in models concerning perception and attention, learning and memory, problem solving and decision making, and language processing[5].

2.2.2 Soar

Soar (originally from State, Operator And Result) is a symbolic cognitive architecture. Soar is based on a particular theory of cognition, described in (Newell, 1990). It forces a set of principles and constraints on how the model is constructed.

Soar is also based on a production system like ACT-R, utilizing explicit rules to

[5] An index of publications related to ACT-R are listed at `http://act-r.psy.cmu.edu/publications/index.php`.

determine its execution path. Problem solving in Soar is basically a search through the problem space to find a goal state. This process is implemented by searching for transitions of states that can bring the system to one of the goals. Every transition involves a decision cycle. In the decision cycle, different pieces of knowledge related to the problem are retrieved to Soar's working memory and then the action to be taken is decided by weighing what was found from the memory to assign preferences. For a given state of the system implemented in Soar, independent productions are matched to process the state and offer operators to change the state. If more than one operator is found eligible, Soar can use different strategies to choose from alternatives, ranging from breadth and depth first search to means-ends analysis. After a new operator to change the state of the system is selected, the system gradually approaches to one of the defined goal states. Paths of execution that are found to contribute to the solution are cached, and this provides a knowledge-level learning mechanism, called chunking. A high-level description of the architecture is given in Figure 2.2, on page 12[6].

Several cognitive models in Soar[7] include studies on human memory, knowledge acquisition, learning and intelligent agents.

[6] Since the version 6 of the Soar manual was not available and the later versions of the manual did not provide the diagram, image is taken from
`http://ai.eecs.umich.edu/cogarch0/soar/arch.html`.

[7] The list is available at
`http://winter.eecs.umich.edu/soarwiki/Soar_Publications`.

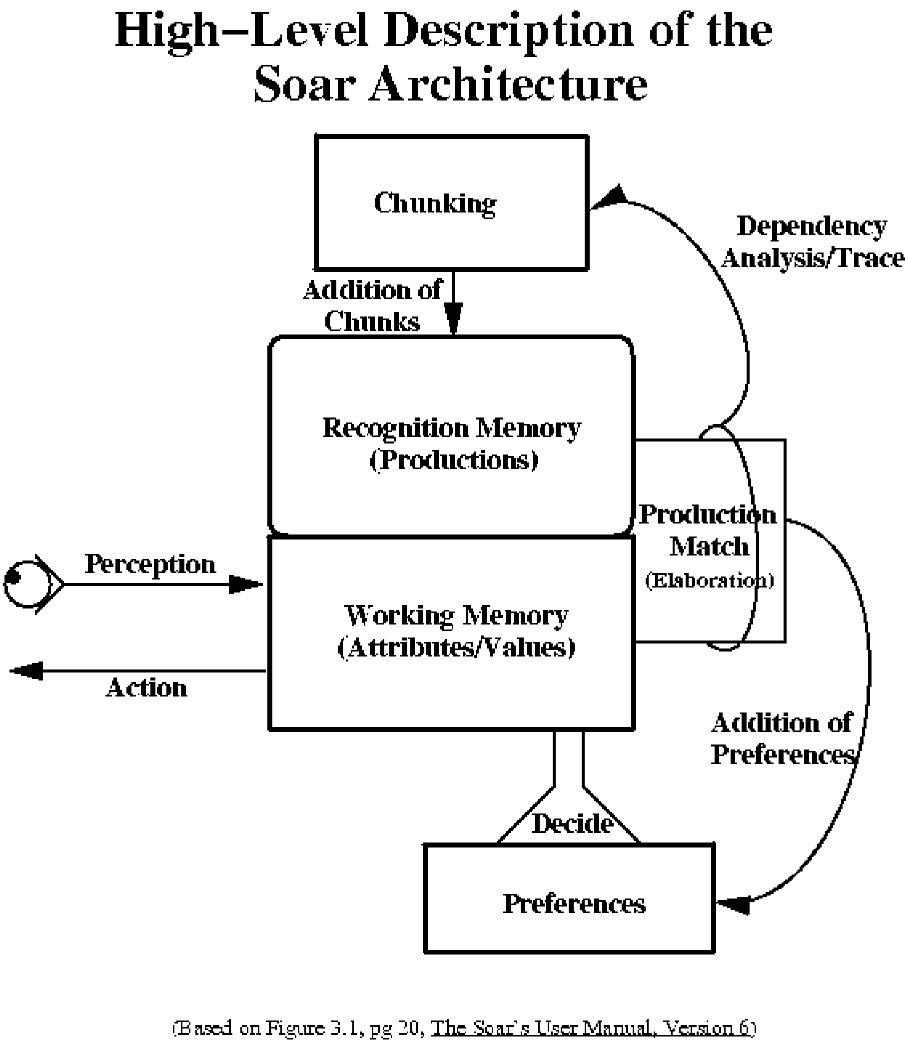

Figure 2.2: High-level description of the Soar Architecture

2.2.3 Other Cognitive Architectures and Frameworks

Other cognitive architectures and tools which were the main tools reported by the questionnaire respondents are briefly explained below.

Respondents of the questionnaire reported cognitive architectures such as AKIRA (Artificial Knowledge Interface for Reasoning Applications)[8] and DUAL[9]. AKIRA

[8] Most of the available information about AKIRA is found at `http://akira-project.org/`.

[9] DUAL homepage is at

is a development tool for building virtual worlds where agents manifesting high-level behavior reside, and DUAL is a hybrid cognitive architecture that comprises a unified description of mental representation, memory structures and processing mechanisms.

Other cognitive modeling tools differ from cognitive architectures by not necessarily imposing and constraining the user to a particular theory for implementation. These tools vary from libraries for a specific programming langauge to graphical user interfaces. Such tools provide general purpose tools for building models, even frameworks, but lack the holistic approach to cognitive modeling problems in the absence of an underlying unified theory of cognition: They do not necessarily provide a single set of mechanisms to account for all of cognition as outlined by (Newell, 1990).

The tools that respondents reported that they are using are CCMSuite (Carleton Cognitive Modelling Suite)[10], CogTool (Tools for Cognitive Performance Modeling for Interactive Devices)[11], CTAT (Cognitive Tutor Authoring Tools)[12] and Gnowsys [13]. These tools are designed for specific purposes: For example, CTAT is a specific tool for creating tutors for on-line courses and Gnowsys is an expert system used for developing semantic web content. CCMSuite is a collection of software tools for creating, analyzing and comparing computational models in cognitive science. CogTool is a software that assists in creating accurate models of skilled performance behavior, making use of the concept of design storyboard, used also for further integration to other cognitive architectures and tools.

We will also introduce COGENT, PDP++ and SOM Toolbox below. COGENT

---

http://nbu.bg/cogs/personal/kokinov/dual_i.html.

[10] CCMSuite is available at
http://www.carleton.ca/ics/ccmlab/ccmsuite.html.

[11] Project homepage of CogTool is
http://www.cs.cmu.edu/~bej/cogtool/.

[12] Homepage of CTAT is at
http://ctat.pact.cs.cmu.edu/index.php.

[13] Gnowsys homepage, also being an instance of itself, is avaliable at
http://www.gnowledge.org/gnowsys/.

is notable by its simple and intuitive interface, PDP++ is a popular example for a connectionist framework and SOM Toolbox exemplifies tools that are developed for particular language dependent needs.

COGENT (Cognitive Objects within a Graphical Environment) (Cooper, Yule, Fox, & Sutton, 1998) is a cognitive modeling tool that has been mostly used for educational purposes, without commitment to a particular architecture. COGENT has simple concepts as boxes and arrows to stand for memory and data flow processes, and it allows for an object oriented production system, mostly suitable for developing symbolic models, although it has facilities for connectionist paradigms.

PDP++ is a connectionist modeling tool, used for realizing parallel distributed processing models, featured in (O'Reilly & Munataka, 2000) along with examples. It is based on a simple object-oriented design, where the main component is the "network" consisting of layers, units and connections between them.

SOM Toolbox (Self Organizing Map Toolbox) is another modeling tool for parallel distributed processing models (Vesanto, Himberg, Alhoniemi, & Parhankangas, 1999). It grew out of the particular need to provide the functionality in Matlab.

In terms of cognitive modeling, we can say that COGENT is an example of an intuitive user interface with the added value of being generic enough to implement different paradigms, while PDP++ and SOM Toolbox are popular examples of frameworks that provide basic mechanisms for cognitive modeling.

As can be seen from this short review, there is a whole spectrum of cognitive modeling tools available, ranging from cognitive architectures with underlying theories to loosely-defined frameworks with flexible building blocks.

2.3 Evaluation Issues of Cognitive Models and Cognitive Modeling Tools

In this section, we describe the evaluation issues of cognitive modeling tools. Firstly, we discuss the issues related to evaluation and validation of cognitive modeling tools, with respect to their contribution to the modeling process. Then we discuss the evaluation of the cognitive modeling tools as software products.

2.3.1 Evaluation and Validation of Cognitive Modeling Tools

A cognitive modeling tool aids in the realization of the model. A good model needs to have two critical features:

complete so that the model does not count out important aspects the phenomenon that it is trying to model,

faithful so that the model does not introduce extra aspects that are not part of the original phenomenon.

A model must be *complete and faithful* so that the researcher can induce and deduce the properties of the real phenomenon from the model. Completeness and faithfulness can be assessed by *goodness of fit measures* which are typically measures of the difference between expected outcomes and actual results.

There are several studies regarding the evaluation of cognitive modeling tools. A recent project, called the Agent-Based Modeling and Behavior Representation (AMBR) compares, evaluates and validates several cognitive architectures in a case study of air traffic control simulation (Gluck & Pew, 2005). The project compares and evaluates different tools in two tasks[14], so that the modeling goals and the domain are the same for all tools. Metrics of performance for the models depend on data collected from actual traffic control sessions.

[14] These tasks are multitasking and category learning, both applied in air traffic control.

Gluck and Pew (2005) evaluate and compare architectures with respect to fixed tasks. The first task involves multitasking abilities and the second one is aimed at category learning. Human participants are tested against the same interface of an air traffic control, and their performances are required to be modeled by the tools. Results are compared in terms of compliance of accuracy, response time, perception of workload and penalty scores (determined initially as negative scores for failing to perform specific tests).

Evaluation of tools mainly focus on two aspects. Regarding cognitive architectures, their plausibility is addressed on the grounds of the theory itself, and capabilities are evaluated in terms of the features and facilites they provide.

Regarding Soar, Cooper and Shallice (1995) is an instance of a critical view of Soar theory and architecture which addresses the problems of depending on the results of implementation to justify the theory behind it. Ritter (1992) performs a survey addressing usage patterns of interfaces to Soar.

There are numerous comparisons in between Soar and ACT-R comparing their capabilities and suitability for the task of modeling cognitive theories ((Ritter & Wallach, 1998), (Jones, 1996)). A similar study covering other tools is (Langley et al., 2006).

Validation in models are discussed in (Gluck & Pew, 2005) and goodness-of-fit measures are not found to be the ultimate solution to validation: "Overfitting"[15], which diverts the model from making generalizations, and "complexity"[16] of the model, which makes it less plausible to accept as a valid model, are two issues of problems with models that goodness-of-fit measures alone can fail to detect. So the project introduces further validation techniques: model comparisons, ability of a model to fit complex patterns and its power to make a priori predictions. They also

[15] Overfitting is a problem with models with too many parameters. The resulting model fits both expected results and the errors of the original, while failing to make generalizations.

[16] Free parameters in a model and its functional form together make up the complexity of the model. Since a model with high complexity may fit any set of data, a less complex model which fits the same data should be favored.

apply application validation, such as the use of model in training or as a decision support system.

There are other evaluations available for specific architectures. Cooper and Shallice (1995) evaluates Soar critically on the grounds of Newell's arguments. One issue is the psychological aspects that are *lost in translation* from theory to implementation. They also assert that implementation cannot justify the psychological theory, and emphasize that the implementation in Soar's case is mingled with details which only serve to make the system more efficient. Another objection is that modular approaches can also account for convergence of separated pieces of information, so means of unification is not limited to unified theories only.

One of their objections for lack of enough criticism about Soar (and cognitive architectures) is that Soar needs specific dedication to understand, so that "...those qualified to criticize Soar are, therefore, those least likely to be critical" (Cooper & Shallice, 1995). This implies that the opinions of active users of a tool based on unified theories are an issue that begs for quantification.

Such studies in smaller scales focus on comparing the properties of tools and their cognitive validations, in terms of how much they agree with the human performance of the same task they model, and how they satisfy the philosophical notion of a model. However, to our best knowledge, there are no studies that research the conceptions of the users of tools. Evaluation of these tools perceived as a software product has the potential to reveal issues related to the user experience, and support the validation techniques described above.

2.3.2 Evaluation of Software Environments

A cognitive modeling tool can provide an architecture (as in ACT-R or Soar), or basic building blocks (as in COGENT), or libraries of functions (as in PDP++), made accessible via computer languages. Therefore, an infrastructure of a cognitive modeling tool has two aspects. First one is the architecture which outlines the capabilities and the second one is the language which determines the access to these capabilities. Cognitive modeling tools, just as every software product, provides a user interface for communicating with the user. Even in the form a library of functions, or a high level graphical user interface, the user interface is the point of communication which has either a positive or a negative impact on the success of implementation and the performance of the user.

Gediga, Hamborg, and Düntch (1999) provide a guideline for evaluation of software, which can be used as a basis for a usability review. In line with ISO 9241-10[17], as the reference document, the guideline lists the points of review for usability of a piece of software as the suitability for the task, self descriptiveness, controllability, conformity with user expectations, error tolerance, suitability for individualization and suitability for learning. As outlined in Chapter 3 of this thesis, we refer to relevant items of this list.

In line with our research objectives, we make use of the part of (Gediga et al., 1999) which addresses suitability issues. This part deals with issues during the task performance of the user while using the software. It questions the suitability of the software in terms of functional coverage of the whole process, facilities it provides for making the process of development more comfortable, familiarity of the terminology used and accessibility of functionality provided.

Suitability for the task for a given software is mostly biased, since we can expect that the person evaluating the tool is already familiar with the tool and accustomed to its particular processes. Gediga et al. (1999) also notes that when software products are in different task domains, the evaluation may also be biased. This may effect the evaluation of cognitive modeling tools, depending on the user interface they provide, and the same effect can be observed, depending on the paradigm they implement.

We focus on the usability issues introduced above. We view the cognitive mod-

[17] ISO9241-10 is the ISO standards document titled "Ergonomic requirements for office work with visual display terminals (VDTs). Dialogue principles".

eling tool as a bridge between the architecture and the user in which both ends need an evaluation. Since this process is about the implementation of the cognitive model, we need to understand the impact of the qualities of the cognitive modeling tool to the modeling process, and the features of the modeling process that effect the use of the cognitive modeling tool. We study the issues of suitability for the task, both as a software in general, and as a cognitive modeling tool, which we detail in Chapter 3.

CHAPTER 3

DESIGN AND CONDUCT OF THE QUESTIONNAIRE

This chapter aims to explain the steps taken in the choice of the type of the survey, design of the questionnaire, and its application. A copy of the questionnaire is in Appendix A for reference.

3.1 Objectives of the Survey

In line with the research questions introduced in Chapter 1, the following objectives are set for the survey:

1. Identify the distribution of usage of cognitive modeling tools among the sample that is chosen,

2. Estimate the effect of education or past experience in the perceived suitability of the cognitive modeling task,

3. Examine the contribution of user interfaces to the suitability perception of the tool,

4. Examine the effect of the design of the tool as a software product for suiting the task of cognitive modeling,

5. Examine the effect of the design of the tool as a cognitive modeling tool for suiting the tasks in cognitive modeling,

Justifications of these objectives are presented in Chapter 4; we discuss the success of these objectives in Section 4.9.2.

3.2 Design of The Survey

We have been able to reach a relatively small set of respondents, and given the nature of the target population, this was expected. The design of the survey aimed at revealing correlations between the items in the questionnaire. We published and conducted the questionnaire on the web. This provided us advantages which are relevant to the distribution of the target population which are so distributed geographically that it would be harder to implement via other means.

Since the users of cognitive modeling tools are dispersed around the world and the expected respondents are actively using the tools in their fields of interest, the questionnaire could be conducted remotely and without the need of an interviewer, provided that questions are clear and the possible answers are not too vague to be interpreted.

For the fact that an interviewer will not be present during the questionnaire, sufficient commentary is added to the beginnings of sections to make sure the questionnaire is self-contained, in terms of conveying what it expects from the contributor. We believe that this also contributed to improving the ratio of people who have returned the completed questionnaire versus the questionnaires sent out for completion.

We also had the option to supplement the questionnaire with online interviews by selected respondents. But the analysis of such interviews required techniques which we could not have completed in the given time period. However, we collected the contact information of respondents who are willing to contribute to such an interview, which we can use for a future study.

We have carried out a pilot implementation before a public announcement, which contributed to the final version of the questionnaire, by means of providing feedback from the respondents of the pilot implementation and their answers they provided for the initial questions. The changes implemented are detailed in Section 3.4.1.

3.3 Sampling

The target population of the survey is the set of researchers actively involved in cognitive modeling, without any constraint to which method they are using for modeling. Since such a community can be considered tightly-packed around universities and research centers, it is possible to reach the whole population in theory, but it is not practically possible. Therefore, samples are formed from the subscribers of the mailing lists which are primarily set up for the communication of the tools chosen as the primary focus of this thesis. We also personally invited 136 individuals who actually recently published active research using cognitive modeling[1] as a method. The template for the invitation message can be found in Appendix A.7.

We sent announcements to following mailing lists:

1. *Cogpsy*, a list of hundreds of subscribers interested in cognitive psychology.
2. *Cogsci*, a cognitive science discussion list.
3. *Soar-group*, mailing list for Soar users.
4. *Act-r-users*, mailing list for Act-r users.

Given this much of announcement, there were 35 respondents who asked for an invitation. The amount of people who asked for an invitation is small compared to the people who actually received the announcement, which is mainly due to the

[1] The list is collected majorly from the participants of ICCM 2005 and 2006, ProQuest database of dissertation and abstracts, Cognitive Systems Research.

extra step added by us by requiring the respondent to ask for the invitation. 30 of the people who asked for an invitation actually participated in the questionnaire. Therefore, the actual participation rate is 85.71% .

3.4 The Conduct of the Survey

This section explains how the questionnaire questions and answer types are prepared. We also provide implementation details of the web based questionnaire here.

3.4.1 Preparing the Questions

In order to prepare the questions for the questionnaire, we firstly prepared an outline of the questionnaire, in line with the research objectives of the study. The first part contained the questions that try to collect profile related information. We have removed questions about the reasons for choosing the tool after the pilot questionnaire because they were not sufficiently covering the subject, and changed the wording of questions to enhance understandability. The second part about cognitive modeling practice in the pilot questionnaire was changed to “training and support for the tool”. While keeping the questions related to training in this part, we moved the rest of the questions to the last part, that contains the questions about the specific usability for the task. The third part, regarding the general suitability for the task are taken from suitability section of the IsoMetrics Usability Inventory (Gediga et al., 1999). The last part in the pilot questionnaire was named “cognitive modeling”, in which we moved the question from the second part and renamed to “specific usability for the task”, in order to make the target of the section specific. In summary, we started out with preliminary questions and revised them based on the input from the pilot questionnaire for the first, second and fourth parts. The third part was taken directly from Gediga et al. (1999).

3.4.2 Questions and Answer Types

The survey is divided into 7 sections, where each section is related to one aspect of the research questions. A copy of the questionnaire can be found in Appendix A.

The first section is the introductory part, where the contributor is introduced to the survey. Sufficient information on how to complete the survey, who designed it, and how the results will be used are included here. Also the consent of the contributing person is asked here. The objective of this section is to inform the contributor about the contents and the aims of the questionnaire.

The second and third sections are about collecting general profile information. Questions here are designed to determine the individual's position in the population in terms of gender, age, professional and domain-specific experience, and tool choice. This section aims to identify the general profile of respondents.

The fourth section is about the experience of the individual and support available for the tool. We aim at determining the educational level of the respondent for the tool in this section.

The fifth section is taken from Isometrics survey, and is a generic set of questions for the suitability of a software product. The aim is to evaluate the tool as a software product, in terms of suitability for the task.

The sixth section deals with the suitability issues of the tool as a specific cognitive modeling tool. The emphasis is on the theory-model interactions and the design, and the evaluation process of the model, developed by the tool.

The last section is for general feedback, with the hope of getting additional information that will help this thesis by pointing to areas that can further be explored in the subject, such as topics that the contributor finds important but not covered in the questionnaire. Optional contact information is also asked for future feedback.

Fourth, fifth and sixth sections have two ending questions where respondents can state their opinions about the section, whether they were relevant to the topic of the section and whether the questions were effective in investigating the topic.

Answer types are multiple-choice, except the questions where the answer may not fit into our expected outcomes. For example, there is an extra question (11) in "Profile Section" that will distinguish people who use other tools, where they will input the actual tool they use, in case it's not in one of the given choices.

3.4.3 Implementation

In order to make sure that questionnaire is as compliant with standard research techniques and the layout is designed for user convenience, an online questionnaire software is used. The product of choice is Unit Command Climate Assessment and Survey System (UCCASS)[2].

UCCASS provides an interface for building questions and collecting responses. It comes ready with common answer types and ability to provide conditional questions, depending on answers from previous questions. It provides preliminary results grouped by answers to each question and bulk data for further analysis via other statistical tools.

For the purposes of the questionnaire in this thesis, most of the answer types that come ready with the software are used. In addition, answer types for comments at the end of each section (a customized multiple line text entry box), current career position (containing the choices Graduate Student, Researcher, PostDoc, Faculty Member and Other) and last degree obtained (containing Undergraduate, Masters and PhD), are added because of the absence of these answer types.

The software provides three types of ways to distinguish users:

1. Identifying the computer the questionnaire is filled in,
2. Providing each contributor with a user name and password,
3. Sending invitations to each invitee, with unique identification numbers in each invitation.

[2] UCCASS is available online at `http://www.bigredspark.com/survey.html`

For the practical requirements of the questionnaire, the third option is chosen to be able to control the qualitative properties of the sample. The first option was not reliable or secure enough both in terms of the contributor and the conductor, and the second option came with unnecessary administrative overhead. The third option enables us to estimate the performance of contributors[3] and confine the questionnaire to a focused and controlled group.

The questionnaire software used throughout the study (UCCASS) was sufficient for the purposes of the questionnaire collecting process, but a few sacrifices had to be made, in exchange of usability of the questionnaire itself. Conditionally appearing questions depending on previous responses had to be turned into normal questions which do not need to be answered, since the former one was likely to cause confusions. One of the confusions related to dependent questions was that the numbering of questions would change if the dependent question appeared. The other confusion arises from the design of the software, where conditionally appearing question is displayed on a separate page, rather than the same page with the other questions. Also, the resulting layout of forms were not modifiable to suit the needs, so there were a few misalignments on the layout. Although we received a comment confirming this was an annoyance, there are no critical problems which prevented the questionnaire from being completed. UCCASS is a mature tool, but it needs more development to support a similar questionnaire.

A copy of the resulting questionnaire is in Appendix A.

[3] The term "performance of the contributor" here refers to the ratio of invitations sent to the replies received. Higher ratio implies a higher rate of contribution to the questionnaire.

CHAPTER 4

RESULTS AND DISCUSSION

This chapter is an analysis of the results of the questionnaire. We provide demographic information about the respondents of the questionnaire and study the significant results that are obtained. This chapter is organized in line with the individual sections of the questionnaire.

We performed the conversion from UCCASS output to tabular format using Excel 2004 for Macintosh, version 11.3.3 and then we have imported the resulting table to SPSS 13 for Mac OS X for statistical analysis.

4.1 Profile Section

Responses in the profile section provide an insight to the general characteristics of the respondents, also giving clues about their tendencies. This section is an analysis of the responses collected in the profile section.

Occupations of respondents are shown in Table C.1 on page 60. Those who have responded as "other" to our question (6.7%) declared their position as "research and development scientist (private business)" and "undergraduate student". Respondents mostly (86.6%) have a master's (33.3%) or higher (53.3%) degree, as seen in Table C.2 on page 60. Age distribution is shown in Table C.5 on page 61. All but one respondents are above 26 years old (96.6%), with a majority being in between 26 and 41 years old (63.3%). The results of the group of the questions in the "profile" section that deal with the background experience of the user are shown in Tables C.6, C.7 and C.8 on page 61. Most of the respondents have been studying Cognitive Science or Artificial Intelligence for at least 5 years (86.7%), and most of them find themselves experienced in the field (79.2%). But the history of cognitive modeling experience in the sample set is rather new, as compared to experience in cognitive science: Only 50% of the respondents have been working on cognitive modeling for at least 5 years. Given distribution of data in the profile, we can say our sample mostly consists of experienced and professional individuals.

Respondents' tool of choice are listed in Table C.9 on page 62. There are 11 people who have responded as "other" to this question. Two of them responded that they implement their models from scratch. Others have listed AKIRA, CCMSuite, CogTool, Cognitive Tutor Authoring Tools, DAS (2 responses), DUAL architecture, standard Matlab programming and Gnowsys[1]. There is an evident majority of ACT-R users (46.7%) among other tools, which denotes its popularity in our sample set. On the other hand, there are users of many different tools, sufficiently ensuring that the results do not rely on the usage of one tool only.

Gender information of respondents is demonstrated in Table C.4 on page 61. One of the respondents chose not to answer this question. There is a high amount of male population, which is expected in our to represent a similar cognitive modeling community which formed our invitation list.

Distribution of nationalities of respondents are given in Table C.3 on page 61. There are respondents from different countries, with a major contribution from USA (36.7%).

Our sample set, although not exhibiting a normal distribution of profiles (as illustrated in Appendix C.4), seems to be typical of our target population, that is, individuals experienced on cognitive modeling, given that there is no profile characterization of this population known to us.

[1] Short information about these tools are presented in Chapter 2.

4.2 Training Section

Responses to questions in the training section are given in Tables C.10, C.11, C.12, C.13 and C.14 on page 62. This section also contains questions on whether respondents found the questions in the section effective and relevant, and their results are given in Tables C.15 and C.16 on page 64.

The ratio of respondents who have received formal training and those who have not received training are almost equal. An amount of 73.3% of the respondents reported that the tool has self-learning material, while 66.7% of them have studied the self-learning material. The support for the tool is studied in Tables C.13 and C.14 on page 63. A magnitude of 66.7% of the respondents receive online support, and 70% of them receive support from their colleagues. Checking these results, we can conclude that our sample set consists of trained individuals and most of the tools already contain self learning material.

Relevancy and effectiveness of questions were asked in Tables C.15 and C.16 on page 64. 83.4% of the respondents either agree or strongly agree that the questions were relevant, and 70% of them either agree or strongly agree that the questions were effective in investigating the topic.

4.3 General Suitability For The Task Section

Individual frequencies of the questions in this section are included in Table C.17 on page 65, and the responses to questions in this section regarding the relevancy and the effectiveness of questions are in Tables C.20 and C.21 on page 68. Checking individual results for each question, we can see that 80.0% of the respondents either agree or strongly agree that the functions implemented in the support them in performing their work, which is an indication that functions provided in the tools are appropriate for their work. Similarly, 73.3% of the respondents either agree or strongly agree they can easily adapt the tool for performing new cognitive modeling tasks. Other responses exhibit a distribution of responses which are not as strong as the ones we have emphasized here.

We have also calculated an aggregate score for the questions in this topic[2]. The score is calculated by first mapping the responses to points, such that "strongly disagree" is 1 and "strongly agree" is 5, and then summing them up for each respondent. Points for questions 20 and 25, which have negative statements, were reversed. By calculating such a score, we tried to come up with a score similar to the one calculated in Gediga et al. (1999). However, since the criteria for the aggregate score in Gediga et al. (1999) were not clear, we were unable to compare the results. This score helped us in Section 4.6 find the relationship between general suitability for the task as an overall concept and the responses in the specific suitability for the task section.

66.7% of the respondents found the questions in the section relevant to the section and 50% of them found the questions effective in investigating the topic. This is a lower level of confidence compared to the other sections, which is an indication that adoption of the questionnaire in (Gediga et al., 1999) is not perceived as a seamless integration to the domain of this study. This is mostly because there are questions not directly related to cognitive modeling tools, such as questions related to screen representations. However, we have not removed them as a design choice.

4.4 Specific Suitability For The Task Section

The results of the section regarding the specifics of the suitability for the task are presented in Table C.22 on page 69. One of the results that is most agreed on are the answers to the statement that the tool makes it convenient to communicate the model with colleagues, in which 76.7% of the respondents either agree or strongly agree that this is true. Likewise, 76.7% of the respondents either agree or strongly agree that using the specific tool is easier that using a general purpose programming

[2] See Appendix C.4 for an analysis of this score.

language for cognitive modeling purposes.

There are also two questions regarding relevancy and effectiveness of questions, whose results are shown in Tables C.24 and C.25 on page 70. 86.7% of the respondents found the questions relevant to the section topic, and 70% of the respondents found the questions effective in investigating the section topic.

4.5 Correlation of Training and Specific Suitability for the Task

We applied Spearman's correlation coefficient which is suitable for non-normal distributions to find the relations between the questions in the training section and the questions in the "suitability for the task - specifics section", since our results do not involve normal distributions.

Table 4.1: Correlations of question 41

Question number	r	r^2	p
14	-.389	.15	$< .05$
16	-.383	.15	$< .05$

The ability of the tool to make it convenient to communicate the resulting model with colleagues (question 41) is significantly related to whether the tool has self-learning material (question 14) ($r = -.389, r^2 = .15, p < .05$), and it is also significantly related to whether the respondent gets online support for the tool (question 16) (user groups, help lines, e-mail), ($r = -.383, r^2 = .15, p < .05$), as listed in Table 4.1. However, 73.3% of the respondents either agree or strongly agree to the statement in question 41 – that the tool makes it convenient to communicate a model with colleagues. –, while 73% of the respondents agree that the tool has self-learning material and 66.7% of the respondents receive online support for the tool. Therefore, these results may not be representative of a negative effect, since the population is biased towards the positive opinion that using the tool enhances communication.

The lack of other significant relations might be the result of a shortage of questions in the training section which could reveal more about the subject. Also, since high amount of experience diminishes the effects of training over time, and our sample set consists mostly of experienced individuals, it can be expected that training does not have much effect on the perception of the tool for the suitability for the task.

4.6 Correlations of the General and Specific Suitability for the Task Sections

We applied Spearman's correlation coefficient to observe relationships between general suitability for the task score (explained in Appendix C.4) and the questions in the specific suitability for the task section, since the related data do not exhibit a normal distribution.

Table 4.2: Correlations of aggregate score

Question number	r	r^2	p
37	.410	.16	$< .05$
38	.619	.38	$< .01$
39	.396	.16	$< .05$
40	.639	.40	$< .01$
41	.490	.24	$< .01$
42	.481	.23	$< .01$
43	.632	.40	$< .01$

We found out that the general suitability for the task score is significantly related to the opinion that the tool enforces good practices for theory-model interactions (question 37) ($r = .410, r^2 = .17, p < .05$), the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .619, r^2 = .38, p < .01$), the need to change the model to fit the constraints of the tool (question 39) ($r = -.396, r^2 = .16, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .639, r^2 = .40, p < .01$), the ability of the tool to make it convenient to communicate the model with colleagues (question 41) ($r = .490, r^2 = .24, p < .01$), the fact that the tool is easier than a general-purpose programming language (question 42) ($r = .481, r^2 = .23, p < .01$), and the possibility to express most aspects of the model in the tool (question 43) ($r = .632, r^2 = .40, p < .01$) (see Table 4.2).

The score correlates with every question in the last section. Higher correlations ($p < .01$) are in questions 38, 40, 41, 42 and 43. These are questions about the evaluation and communication of the results of the model and the relative ease of use compared to other alternatives. This fact hints a strong relation between the general suitability for the task and the perceived quality of the resulting model developed with the tool.

We also applied Spearman's correlation coefficient to each individual question in the "suitability for the task - general section".

Table 4.3: Correlations of question 20

Question number	r	r^2	p
39	.489	.24	$< .01$
40	-.484	.23	$< .01$
41	-.467	.22	$< .01$
43	-.623	.39	$< .01$

We can see a negative relationship between having to do extra tasks while modeling and the comfortability of the evaluation and communication process of the resulting model. The relations that support this observation are that there are significant correlations between the opinion that the tool forces to perform tasks that are not related to the actual modeling task (question 20) and the need to change the model to fit the constraints of the tool (question 39) ($r = .489, r^2 = .24, p < .01$), the ability of the tool to make the evaluation of the process easier (question 40) ($r = -.484, r^2 = .23, p < .01$), the ability of the tool to make it convenient to communicate the model with colleagues (question 41) ($r = -.467, r^2 = .22, p < .01$)

and the possibility to express most aspects of the model in the tool (question 43) ($r = -.623, r^2 = .39, p < .01$), as seen in Table 4.3.

Table 4.4: Correlations of question 21

Question number	r	r^2	p
38	.381	.15	$< .05$
40	.409	.17	$< .05$

An observation from the correlations is that completeness of a tool regarding tasks relates with the ease of design, implementation of evaluation processes, since the capacity of the tool to completely perform entire work routines (question 21) is significantly related to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .381, r^2 = .15, p < .05$) and the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .409, r^2 = .17, p < .05$), as listed in Table 4.4.

Table 4.5: Correlations of question 22

Question number	r	r^2	p
38	.555	.31	$< .01$
40	.512	.26	$< .01$
42	.611	.37	$< .01$
43	.503	.25	$< .01$

Another aspect of the facilities provided by the tool itself is about supportive functions for performing the work. We found that this is related to the ease of design, implementation and evaluation processes, preference over a generic programming language and the ability to express the aspects of the model. The support facilites of the functions implemented in the tool for performing the work (question 22) is significantly related to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .555, r^2 = .31, p < .01$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .512, r^2 = .26, p < .01$), the fact that the tool is easier than a general-purpose programming language (question 42) ($r = .611, r^2 = .37, p < .01$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .503, r^2 = .25, p < .01$), as listed in Table 4.5.

Table 4.6: Correlations of question 32

Question number	r	r^2	p
32	.380	.14	$< .05$

In terms of supporting the development process, having functions for the disposal of the user easily accessible correlates with enforcing good practices for theory-model interactions. This may also be a relationship between having the functions that eases theory-model interactions accessible and actually using them. To be able to easily find the important commands (question 32) is significantly related to the fact that the tool enforces good practices for theory-model interactions (question 37) ($r = .380, r^2 = .14, p < .05$), also listed in Table 4.6.

Table 4.7: Correlations of question 27

Question number	r	r^2	p
38	.425	.18	$< .05$
40	.427	.18	$< .05$
42	.548	.30	$< .01$
43	.475	.22	$< .01$

Having a tool compatible with work correlates with the design, implementation and evaluation processes, preference over a general-purpose programming language and the ability to express the aspects of the model, since the suitability of the tool to the requirements of the work (question 27) is significantly correlated to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .425, r^2 = .18, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .427, r^2 = .18, p < .05$), the fact that the tool is easier than a general-purpose programming language (question 42) ($r = .548, r^2 = .30, p < .01$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .475, r^2 = .22, p < .01$), as listed in Table 4.7.

Table 4.8: Correlations of question 31

Question number	r	r^2	p
38	.444	.20	$< .05$
40	.389	.15	$< .05$
43	.701	.49	$< .01$

A tool being flexible enough to adapt new tasks is related to the design, implementation and evaluation processes, and the ability to express the aspects of the model, as the flexibility of the tool for adapting new cognitive modeling tasks (question 31) is significantly correlated to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .444, r^2 = .20, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .389, r^2 = .15, p < .05$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .701, r^2 = .49, p < .01$), as listed in Table 4.8.

Table 4.9: Correlations of question 25

Question number	r	r^2	p
38	.363	.13	$< .05$
40	-.485	.25	$< .01$
43	-.449	.20	$< .05$

We can also see that there is a negative relationship between the steps needed to perform a task and the design, implementation and evaluation processes, and the ability to express to aspects of the model. The tool's requirement of too many different steps to perform a task (question 25) is significantly related to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = -.363, r^2 = .13, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = -.485, r^2 = .24, p < .01$) and the possibility to express most aspects of the model in the tool (question 43) ($r = -.449, r^2 = .20, p < .05$), as listed in Table 4.9.

Table 4.10: Correlations of question 23

Question number	r	r^2	p
37	.383	.15	$< .05$
38	.547	.30	$< .01$
40	.516	.27	$< .01$
41	.375	.14	$< .05$
42	.430	.18	$< .05$
43	.569	.32	$< .01$

Regarding the input and output facilities, we first see that the the way the data is entered correlates with the ease of design, implementation, evaluation and communication processes, preference over a general purpose programming language and the ability to express most aspects of the model. The results supporting this relation is that the suitability for the task of the way the data is entered (question 23) is significantly related to the opinion that the tool enforces good practices for theory-model interactions (question 37) ($r = .383, r^2 = .15, p < .05$), the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .547, r^2 = .30, p < .01$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .516, r^2 = .27, p < .01$), the ability of the tool to make it convenient to communicate the model with colleagues (question 41) ($r = .375, r^2 = .14, p < .05$), the fact that the tool is easier than a general-purpose programming language (question 42) ($r = .430, r^2 = .18, p < .05$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .569, r^2 = .32, p < .01$), as listed in Table 4.10.

Table 4.11: Correlations of question 26

Question number	r	r^2	p
38	.436	.19	$< .05$
40	.437	.19	$< .05$
41	.505	.26	$< .01$

Similarly, the suitability of the output of the data (question 26) is significantly related to the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .436, r^2 = .19, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .437, r^2 = .19, p < .05$) and the ability of the tool to make it convenient to communicate the model with colleagues (question 41) ($r = .505, r^2 = .26, p < .01$). We again see a relationship between the format of the output and the design, implementation and evaluation processes. There is also a relationship with the convenience to communicate the model, as seen in Table 4.11.

Table 4.12: Correlations of question 24

Question number	r	r^2	p
40	.469	.22	$< .01$
43	.391	.15	$< .05$

There are also findings related to the representation of data on-screen, which is correlated with the evaluation process and the ability to express the aspects of the model in the tool. The perception of the on-screen representation as sensible for the work (question 24) is significantly related to the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .469, r^2 = .22, p < .01$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .391, r^2 = .15, p < .05$), as seen in Table 4.12.

Table 4.13: Correlations of question 34

Question number	r	r^2	p
39	.364	.13	$< .05$
40	.409	.17	$< .05$
43	.431	.19	$< .05$

There is also a relationship between the presentation of information by the tool with the evaluation process and the ability to express most aspects of the tool. It also correlates with the increased need to change the model to fit the constraints of the tool. This can also be a relationship between the comfortability of the presentation and the trade-off of the user to fit the model to the particular presentation. Related correlations are that supportive facilities of the presentation of information on the screen (question 34) is significantly related to the need to change the model to fit the constraints of the tool (question 39) ($r = .364, r^2 = .13, p < .05$), the ability of the tool to make the evaluation of process of the model easier (question 40) ($r = .409, r^2 = .17, p < .05$) and the possibility to express most aspects of the model in the tool (question 43) ($r = .431, r^2 = .19, p < .05$), as listed in Table 4.13.

Table 4.14: Correlations of question 33

Question number	r	r^2	p
37	.380	.15	$< .05$
38	.554	.31	$< .01$

Another result related to the presentation of results is about the flexibility of the output, which is related to the design and implementation processes, but not the implementation process. To be able to adjust the presentation of results to various work requirements (question 33) is significantly correlated with the fact that the tool enforces good practices for theory-model interactions (question 37) ($r = .380, r^2 = .14, p < .05$) and the fact that the tool eases the design and implementation process of a cognitive model (question 38) ($r = .554, r^2 = .31, p < .01$). This finding hints an iterative development process while developing a model, as listed in Table 4.14.

We could find no correlation for “in a given screen, I find all of the information I need in that situation” (question 28), “the terminology used in the tool reflects that of my work environment” (question 29) and “the tool provides me with a repeat function for work steps that must be performed several times in succession” (question 30). These questions make sense in a consistent fully graphical environment, therefore not finding any relations is a reasonable outcome.

4.7 Effect of Choice of Tool or Gender

There is a massive amount of respondents who use ACT-R (46.7%), which could make a difference in the results. We applied Mann-Whitney test to determine these differences, where the grouping variable is whether the respondent uses ACT-R or not. The results are given in Table C.26, where there are no significant differences.

There is also an imbalance for gender in the sample set – 80% of the respondents were male. We also applied Mann-Whitney test to determine if there is a difference in responses depending on gender. The results are in Table C.27. We did not find any significant results as a result of this test.

4.8 Agreements of Respondents Regarding Effectiveness and Relevancy of the Questionnaire

We have also measured the level of agreement in questions where the effectiveness and relevance of the questions of the section are asked. The responses were grouped in two types, first one being the responses that are "strongly agree" or "agree" and the second one being the rest of the response types. We have applied Cochran's Q test for answers to questions 18, 36 and 44 for relevancy. For this set, Cochran Q = 4.133, df = 2 and p = .127, therefore a significant result could not be obtained. Similarly, we applied the same test to questions 19, 37 and 45 for effectiveness. Cochran Q = 4.800, df = 2 and p = .091, therefore there is no significance for this set, either. These results show that the respondents are not in agreement regarding the effectiveness and relevancy of the different sections in the questionnaire. These figures may be the result of differences in perception of different sections in terms of effectiveness and relevancy. As we have listed the percentages of the opinions of respondents regarding effectiveness and relevancy in Sections 4.2, 4.3 and 4.4, degrees of agreement are different for each section. This may arise from the fact that the training section requires more coverage with questions, and not all questions in the general suitability for the task section are related to a specific tool. Therefore, the respondents' perceived coverage of a section not being uniform over the sections may be the reason for this absence of significant agreements in responses to given questions.

4.9 Discussion

The questionnaire also contains a question where respondents can add their comments about the questionnaire. We will discuss the points obtained from comments from respondents in the first part of this section. In the second part, we will discuss about the results obtained in this chapter, commenting on how the results matched our objectives and providing a general interpretation of the analysis.

4.9.1 Evaluation of Individual User Comments

Respondents are given the opportunity to comment on the questionnaire in the final section. There are 13 comments by respondents, which accounts for 43.3% of the sample.

There are respondents that have developed their own tools, who have reported the concern that they may not be fit for the questionnaire since they may be biased. Three respondents left similar comments. One of them expressed concern about her neutrality, because she was the principle investigator of the tool she was using. But we expect such biases, since we require experience on cognitive modeling tools from respondents. Therefore we decided to include these results, since the opinions of the developers themselves also make sense in the context of this questionnaire. However, we agree that the questionnaire could be enhanced with questions studying the distinction between the developer and the user perspective, to tell apart such respondents.

Another respondent agrees that this questionnaire helps to point out places for improvements in the software he is developing. This is a supportive comment that the questionnaire can further be exploited as a guideline for the design of cognitive modeling or a feedback mechanism to improve the tool building process.

Another group of responses address concerns about what we mean by tools and whether their choices are to be included in our definition. There are four similar comments that fall into this category. One of the respondents asserted that things like ACT-R and Soar are actually architectures that form a basis for tools that build and evaluate models and collect data. We believe that our definition of a tool as "a computer-based tool designed and implemented as an aid to do computational cognitive modeling and simulation."[3] covers implementations of cognitive architectures as tools but fails to distinguish between the tool as modeling environments versus the tool as implementations of the theoretical cognitive architecture. We decided that although the respondents are confused about the definition of a tool, their assumptions agree with our definitions, so they were eligible for the questionnaire. As a working definition, conflating this distinction did not create a hazard for the conduct of the questionnaire since it covered architectures as tools. However further work should elaborate and work on this distinction.

There were also comments regarding the layout of the questionnaire. Two respondents commented about the layout. One of them commented that we should have repeated the option headings every 10 lines, so one can still see them when scrolling down. Another respondent suggested distributing the radio buttons evenly, which do not align properly on the on-screen presentation of the questionnaire, which is handed to respondents. We agree that these were issues effecting the comfort of the respondent during the questionnaire, in which we were limited by the options offered by the questionnaire software (UCCASS).

There was a comment that because the tool lacks a graphical user interface, usability section was not relevant. This fact also effected the results: Questions related to the screen layout show little or no relation to the suitability for the task responses.

[3] This is also emphasized in the questionnaire (in Appendix A.3).

Another response was that the tool needed more open ended questions. For the time constraints of this study, we preferred not to include such questions, but a similar future study needs to provide more space for open-ended questions.

4.9.2 Results of the Analysis

For the purpose of correlational analysis of results of the questionnaire, we have applied Spearman's rank correlation to the results we found. Results obtained via this method enabled us to find out relationships between responses to questions in sections about general suitability for the task and training. This type of analysis enabled us to study the objectives of estimating the effect of education or past experience in the users selection of a specific tool, examining the advantages or disadvantages of particular user interfaces in the modeling process, examining the effect of the design of the tool as a software product for the suitability of the task of cognitive modeling, and examining the effect of the design of the tool as a cognitive modeling tool for suitability of the tasks in cognitive modeling, as outlined in Chapter 3, which correspond to objectives 2, 3, 4 and 5 on page 20. By analyzing the frequencies of responses, we were able to identify the distribution of usage of cognitive modeling tools among the sample that is chosen, which was our objective 1.

We observed that the distribution of usage of cognitive modeling tools shows more variability than we had anticipated. Such a variability in terms of tools did not cause a problem for this study since we tried to design the questionnaire as neutral as possible with regard to specific tools. However, there seems to exist groups of mainstream, specialized and homegrown tools. An extended study may need to address different types of tools while studying their suitability factors.

We have not been able to find decisive correlations between training or support for the tool and suitability for the task. Our correlation analysis did not reveal relations, but this does not mean there are no relations, but suggests directions for improvement for a similar study: A more comprehensive set of questions, possibly supported by interviews and a sample with a wider range of educational background can provide more meaningful results.

Our objective of examining the advantages or disadvantages of particular user interfaces in the modeling process is blended into the other objectives. We have been able to find correlations related to input, output and presentation facilities – they are related to a enhanced process of design, implementation and evaluation. However, since the idea of a graphical user interface and screen presentations are not relevant for most of the tools, questions related to these facilities did not reveal correlations as useful as the other items, such as input and output facilities.

Effects of considering the tool as a software product and as a specific cognitive modeling tool constituted most of the analysis. Specifically, we were able to find significant relations about every question in the specific suitability for the task section. In order to get a broader view, we applied an aggregate score for the general suitability for the task section for to relate with the questions in specific suitability for the task section. We observed that a higher score is significantly related to the perceived quality of the resulting model developed with the tool. After this, we studied the relations of individual questions in the general suitability for the task section. To summarize the significant results, we can group the results into correlations that are related to supportive facilities provided by tool, facilities provided for development purposes and properties of input, output and presentation. We found out that supportive facilities, such as easily accessible functionality and having the tool compatible with the requirements of the work is significantly correlated with an improved design, implementation and evaluation process of the model. Similarly, facilities provided for development purposes are significantly related to preferring the tool to a general purpose programming language and ability to express most aspects of the model, as well as enhancing the design and implementation processes. Finally, we found that input, output and presentation facilities significantly correlates with the ease to communicate the model, enhanced development process and ability to express most of the aspects of the model.

A finer analysis of answers could be performed, by making use of exploratory factor analysis. However, we need a sample set at least as much as ten times of what we had in this study to perform such an analysis, otherwise the results would not be meaningful. We applied principal component analysis to extract components, which did not produce results that are sufficiently meaningful – with a very high rate of collinearity. Therefore, this analysis is excluded from the study. However, it will be interesting to explore components of interest with a larger sample, which may be one of the methods of analysis of a future survey.

Overall, we can conclude that our results support our initial objectives set for the questionnaire. The results additionally showed us places that are open for improvement, which we will also discuss in the next chapter.

CHAPTER 5

CONCLUSION

In this thesis, an initial attempt for an international survey was made for an evaluation of cognitive tools in terms of suitability for the task. For this purpose, a questionnaire is prepared, targeting individuals who are experienced in cognitive modeling. Selected individuals who have contributed to related conferences and publications were personally invited and similar invitations have been sent to related mailing lists. In total, responses from 30 selected participants have been collected and the results are based on these responses.

An analysis of the questionnaire, covering correlations between the answers to the questions is provided in Chapter 4. We have failed to find significant relations between training for a particular tool and the suitability for the task. One reason for this may be that there have not been sufficient questions to uncover effects related to the usage of the tool and training – our set of questions were rather brief and addressed general concepts related to training. Also, the effects of training may diminish with experience, therefore our sample set may not be fit for revealing these effects. Future research may address these two issues and reinvestigate the issue with more complete questions and a sample set consisting of individuals from a wider spectrum of educational and experience backgrounds on the training for the tool and experience.

The effects of general suitability for the task on suitability for the cognitive modeling task were also analyzed, checking for correlations between individual questions and an aggregate score, computed as explained in Appendix C.4. The analysis reveals sufficient results to conclude that there are convincing relationships between good practices in terms of keeping usability in mind (providing supportive functions, convenient input and output facilities and customizable results) during the development of a tool and suitability of the tool for the task of developing cognitive models. Since most of the software that was subject to the questionnaire either had no graphical user interface or did not have a constant user interface available to every user, fewer correlations compared to other questions were observed related to the on-screen representations of user controls and input/output data. Another topic of interest for a future study will possibly be the impact of visual user interfaces and their effect on the modeling practice.

Due to the time constraints for this study, I chose not to present open ended questions in the questionnaire, as the classification of the responses to such questions require a different path of analysis. An alternative questionnaire complementing the results found in this questionnaire should inevitably make use of such questions to be able to receive more qualified responses. Likewise, making use of interviews where respondents can provide subjective insight to the modeling process will improve a subsequent survey on the topic.

The questionnaire and the results altogether show the value of a questionnaire based method to evaluate cognitive modeling tools. This study has revealed information regarding the connection between the view of a cognitive tool as a general purpose software product and a specific software product for modeling cognition, that an alternative view of a cognitive modeling tool as a software product will improve the quality of the software and perception of the tool by the user.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Beer, R. D. (2000). Dynamical Approaches to Cognitive Science. *Trends in Cognitive Sciences*, *4*(3), 91–99.

Bourque, B. L., & Fielder, E. P. (1995). *How to Conduct Self-Administered and Mail Surveys*. Thousand Oaks, California: SAGE Publications, Inc.

Cooper, R. P., Fox, J., Glasspool, D. W., & Yule, P. (2002). Modeling High Level Cognitive Processes. In (chap. Modelling Cognition). New Jersey, USA: Lawrence Erlbaum Associates.

Cooper, R. P., & Shallice, T. (1995). Soar and the case for unified theories of cognition. *Cognition*, *55*, 115–149.

Cooper, R. P., Yule, P., Fox, J., & Sutton, D. (1998). A Cognitive Science Approach to Reasoning, Learning and Discovery. In (pp. 55–82). Lengerich, Germany: Pabst Science Publishers.

Gediga, G., Hamborg, K. C., & Düntch, I. (1999). The IsoMetrics Usability Inventory: An Operationalisation of ISO 9241-10. In *Behaviour and information technology* (Vol. 18, pp. 151–164). Oxford, UK: Taylor and Francis.

Gluck, K. A., & Pew, R. W. (2005). *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation*. New Jersey, USA: Lawrence Erlbaum Associates.

Jones, G. (1996). The architectures of Soar and ACT-R, and how they model human behaviour. *Artificial Intelligence and Simulation of Behaviour Quarterly, Winter 1996*(96), 41–44.

Langley, P., Laird, J. E., & Rogers, S. (2006). *Cognitive architectures: Research issues and challenges* (Tech. Rep.). Stanford University, CA.: Computational Learning Laboratory, CSLI.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA, USA: Harvard University Press.

O'Reilly, R. C., & Munataka, Y. (2000). *Computational Explorations in Cognitive Neuroscience*. Cambridge, MA, USA: MIT Press.

Polk, T. A., & Seifert, C. M. (2002). *Cognitive Modeling*. Cambridge, MA, USA: MIT Press.

Ritter, F. E. (1992). *Tbpa: A methodology and software environment for testing process models' sequential predictions with protocols*. Doctoral dissertation, Psychology, Carnegie-Mellon University.

Ritter, F. E., & Wallach, D. (1998). Models of two-person games in ACT-R and Soar. *Proceedings of the Second European Conference on Cognitive Modelling*, 202–203.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group the. (1986). *Parallel Distributed Processing*. Cambridge, Massachusetts: The MIT Press.

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (1999, November). Self-organizing Map in Matlab: the SOM toolbox. In *Proceedings of the Matlab DSP Conference* (pp. 35–40). Espoo, Finland.

APPENDICES

APPENDIX A. QUESTIONNAIRE FORM

A.1 Introduction Section

The aim of this study is to analyze the attitudes of users of cognitive modeling tools, with a particular perspective for suitability of the tools for the cognitive modeling task from the perspectives of general usability and cognitive modeling specifics.

This questionnaire is part of the MSc Thesis work of Can Bican, a student of Cognitive Science Program at Middle East Technical University, supervised by Dr. Bilge Say.

The questionnaire consists of 7 sections and 45 questions. Answering questions that are marked with [*] are required. The rest are optional or dependent on a previous answer. The whole questionnaire takes 15-20 minutes to complete.

You can contact them with any further questions at can@bican.net and bsay@-ii.metu.edu.tr respectively.

Your contribution to the study by filling up the questionnaire is entirely voluntary. You can decide to stop answering the questions at any time. In this case, simply closing your browser window without finishing the questionnaire will be sufficient.

All the answers and personal details will be kept strictly confidential and the aggregate results only will be published in a master's thesis or possibly, in a relevant academic conference proceedings.

If you consent with these terms, please select "yes" below and continue with the questionnaire.

We appreciate your contribution to this questionnaire by devoting your valuable time. As an acknowledgment to your contribution, we will present one of the participants of this questionnaire with a gift certificate from amazon.com.

1. [*] I consent with these terms.

(Check Box)

(Yes/No)

A.2 Profile Section, Part 1

In this page, we will ask a few specific questions to determine your profile.

2. [*] What is your current position?

(Graduate Student / Researcher / Postdoc / Faculty Member / Other)

3. If you answered the previous question as 'other', please state your current position:

(Text Field)

4. [*] What is the last degree you obtained?

(Undergraduate / Masters / PhD)

5. [*] What is the country you currently reside in?

(Text Field)

6. What is your gender?

(Male / Female)

7. [*] What is your age?

(18 years to less than 26 years / 26 years to less than 41 years / 41 years or more)

A.3 Profile Section, Part 2

In this page, will ask a few more questions to determine your profile.

We take the term tool as a computer-based tool designed and implemented as an aid to do computational cognitive modeling and simulation.

We understand from computational cognitive modeling as modeling of cognitive processes on computational platforms either as a method in cognitive science or as components of artificial intelligence applications.

8. [*] How long have you been studying and researching in Cognitive Science or Artificial Intelligence?

(Less than 1 year / 1 year to less than 5 years / 5 years to less than 10 years / 10 years or more)

9. [*] How experienced are you in general computer programming, apart from computational cognitive modeling?

(Highly experienced / Moderately experienced / Not so experienced)

10. [*] How long have you been working on cognitive modeling?

(Less than 1 year / 1 year to less than 5 years / 5 years to less than 10 years / 10 years or more)

11. [*] What is the cognitive modeling tool you use most frequently?

(ACT-R / COGENT / Soar / A Neural Network Toolkit / Other)

12. If you answered the previous question as 'A Neural Network Toolkit' or 'Other', please state the cognitive modeling tool you use most frequently:

(Text Field)

A.4 Training and Support for the Tool

This section refers to the details of the training and the support associated with the cognitive modeling tool that you specified as your most frequent choice in the previous section. From this section onwards, the tool will always refer to same cognitive modeling tool.

13. [*] Did you have formal training on the tool?

(Yes / No / Not Applicable)

14. [*] Does the tool have self learning material?

(Yes / No / Not Applicable)

15. If you have answered the previous question as 'yes', have you studied the self learning material of the tool?

(Yes / No / Not Applicable)

16. [*] Do you get online support for the tool (user groups, help lines, e-mail)?

(Yes / No / Not Applicable)

17. [*] Do you get support from your coworkers for the tool?

(Yes / No / Not Applicable)

18. [*] The questions in this section were relevant to the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

19. [*] I found the questions in this section effective in investigating the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

A.5 Suitability for the Task - General

In this section there will be some questions about the suitability of the tool for the cognitive modeling task from the general usability perspective.

20. [*] The tool forces me to perform tasks that are not related to my actual cognitive modeling task.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

21. [*] The tool lets me completely perform entire work routines.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

22. [*] The functions implemented in the tool support me in performing my work.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

23. [*] The way in which data is entered is suited to the cognitive modeling tasks I want to perform with the tool.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

24. [*] I perceive the arrangement of the fields on-screen as sensible for the work I do with the tool.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

25. [*] Too many different steps need to be performed to deal with a given task.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

26. [*] The way in which data is output is suited to the tasks I want to perform with the tool.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

27. [*] The tool is well suited to the requirements of my work.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

28. [*] In a given screen, I find all of the information I need in that situation.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

29. [*] The terminology used in the tool reflects that of my work environment.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

30. [*] The tool provides me with a repeat function for work steps that must be

performed several times in succession.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

31. [*] I can easily adapt the tool for performing new cognitive modeling tasks.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

32. [*] The important commands required to perform my work are easy to find.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

33. [*] I am able to adjust the presentation of results (on the screen, to printer, plotter, etc.) to my various work requirements.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

34. [*] The presentation of the information on the screen supports me in performing my work.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

35. [*] I found the questions in this section relevant in investigating the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

36. [*] I found the questions in this section effective in investigating the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

A.6 Suitability for the Cognitive Modeling Task - Specifics

In this section there will be some questions about the specifics of cognitive modeling task and the suitability of the tool.

37. [*] The tool enforces good practices for theory-model interactions.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

38. [*] The tool eases the design and implementation process of a cognitive model.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

39. [*] I had to change my model to fit the constraints of the tool.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

40. [*] Using the tool makes the evaluation process of my model easier.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

41. [*] The tool makes it convenient to communicate my model with my colleagues.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

42. [*] Using the tool is easier than using a general-purpose programming language for cognitive modeling purposes.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

43. [*] It is possible to express most aspects of my model in the tool.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

44. [*] I found the questions in this section relevant in investigating the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

45. [*] I found the questions in this section effective in investigating the section topic.

(Strongly disagree / Disagree / Neither disagree nor agree / Agree / Strongly agree / No opinion)

A.7 Final Section

46. Thank you for completing our questionnaire. Please feel free to add some comments about the questionnaire and its topic.

(Text Field)

47. We want to further our analysis in this questionnaire by conducting interviews (around 45 minutes) with researchers actively involved in cognitive modeling by means of some kind of teleconferencing. If you would be interested in participating in such an interview, please tick the box below and leave your contact details in item 48.

(Check Box)

48. Please leave your contact information (email or postal address) below, if you would like to be notified about the results of this survey. We will also add your name to our prize draw.

(Text Field)

”Suitability for the Task - General” section of this questionnaire was adopted from the IsoMetricsS usability questionnaire with permission.

APPENDIX B. INVITATION MESSAGE FOR THE QUESTIONNAIRE

Invitation messages are sent to selected individuals and mailing lists. Below is the template of this message.

B.1 Template for the Invitation Message

Dear (name)[1],

We got your name from from (source)[2].

We are conducting a study addressing the attitudes of users of cognitive modeling tools. Our particular emphasis in the study is on the suitability of the tools for the cognitive modeling task from the perspectives of general usability and cognitive modeling specifics.

As a part of this work, we have prepared a questionnaire. We kindly ask if you can help our study by participating in the questionnaire, which should take no more than 15 minutes to complete. Our only requirement is experience with a cognitive modeling tool.

Access to the questionnaire is by e-mail invitation. If you agree to fill in our questionnaire, please send an e-mail to can@bican.net with subject 'questionnaire'. You will receive an reply consisting of instructions on how to access the questionnaire.

All the answers and personal details will be kept strictly confidential and the aggregate results will only be published in a master's thesis or possibly, in a relevant academic conference proceedings.

We appreciate your contribution to this questionnaire by devoting your valu-

[1] Name/surname for individuals and ”Colleagues” for mailing lists.
[2] This paragraph is skipped for mailing lists.

able time. As an acknowledgment to your contribution, we will present one of the participants of this questionnaire with a gift certificate from amazon.com.

This questionnaire is part of the MSc Thesis work of Can Bican, a student of Cognitive Science Program at Middle East Technical University, supervised by Dr. Bilge Say. You can contact them with any further questions at can@bican.net and bsay@ii.metu.edu.tr respectively.

Best Regards,

–

Can Bican

APPENDIX C. STATISTICS FOR THE QUESTIONNAIRE

This part of the appendix contains tables of statistics for the questionnaire questions.

C.1 Profile Section

Table C.1: Current position (Question 2)

	Frequency	Percent
Other	2	6.7
PostDoc	2	6.7
Researcher	7	23.3
Graduate Student	8	26.7
Faculty Member	11	36.7
Total	30	100.0

Table C.2: Last degree obtained (Question 4)

	Frequency	Percent
Undergraduate	4	13.3
Masters	10	33.3
PhD	16	53.2
Total	30	100.0

Table C.3: Country (Question 5)

	Frequency	Percent
Canada	1	3.3
India	1	3.3
Japan	1	3.3
Bulgaria	1	3.3
Israel	1	3.3
Turkey	1	3.3
Macedonia	1	3.3
United Kingdom	2	6.7
The Netherlands	3	10.0
Germany	6	20.0
USA	11	36.7
Total	30	100.0

Table C.4: Distribution of gender (Question 6)

	Frequency	Percent
Unanswered	1	3.3
Female	5	16.7
Male	24	80.0
Total	30	100.0

Table C.5: Age distribution (Question 7)

	Frequency	Percent
18 to less than 26	1	3.3
41 or more	10	33.3
26 to less than 41	19	63.3
Total	30	100.0

Table C.6: Distribution of experience in cognitive science or artificial intelligence (Question 8)

	Frequency	Percent
1 year to less than 5 years	4	13.3
5 years to less than 10 years	12	40.0
10 years or more	14	46.7
Total	30	100.0

Table C.7: Distribution of experience in general computer programming (Question 9)

	Frequency	Percent
Not so experienced	5	20.8
Moderately experienced	7	29.2
Highly experienced	12	50.0
Total	30	100.0

Table C.8: Distribution of experience on cognitive modeling (Question 10)

	Frequency	Percent
Less than 1 year	2	8.3
5 years to less than 10 years	5	20.8
10 years or more	7	29.2
1 year to less than 5 years	10	41.7
Total	24	100.0

Table C.9: Distribution of cognitive modeling tool usage (Question 11)

	Frequency	Percent
COGENT	1	3.3
Soar	4	13.3
Other	11	36.7
ACT-R	14	46.7
Total	30	100.0

C.2 Training Section

Table C.10: Distribution of respondents who had formal training (Question 13)

	Frequency	Percent
N/A	3	10.0
No	13	43.3
Yes	14	46.7
Total	30	100.0

Table C.11: Distribution of tools that have self learning material (Question 14)

	Frequency	Percent
N/A	2	6.7
No	6	20.0
Yes	22	73.3
Total	30	100.0

Table C.12: Distribution of respondents who studied the self learning material (Question 15)

	Frequency	Percent
No	2	6.7
N/A	3	10.0
Yes	20	66.7
Missing	5	16.7
Total	30	100.0

Table C.13: Distribution of respondents who receive online support (Question 16)

	Frequency	Percent
N/A	4	13.3
No	6	20.0
Yes	20	66.7
Total	30	100.0

Table C.14: Distribution of respondents who receive help from coworkers (Question 17)

	Frequency	Percent
N/A	4	13.3
No	5	16.7
Yes	21	70.0
Total	30	100.0

Table C.15: Distribution of respondents who found the questions in the training section relevant (Question 18)

	Frequency	Percent
No Opinion	1	3.3
Neither Disagree nor Agree	4	13.3
Strongly Agree	8	26.7
Agree	17	56.7
Total	30	100.0

Table C.16: Distribution of respondents who found the questions in the training section effective (Question 19)

	Frequency	Percent
No Opinion	2	6.7
Disagree	2	6.7
Strongly Agree	4	13.3
Neither Disagree nor Agree	5	16.7
Agree	17	56.7
Total	30	100.0

C.3 General Suitability for the Task Section

In the following table, abbreviations mean: SD: Strongly disagree, D: Disagree, N: Neither disagree nor agree, A: Agree, SA: Strongly Agree.

Table C.17: Frequencies for the questions in general suitability for the task section

		No Opinion	SD	D	N	A	SA
The tool forces me to perform tasks that are not related to my actual cognitive modeling task (Question 20)	Count	0	2	10	4	11	3
	%	.0%	6.7%	33.3%	13.3%	36.7%	10.0%
The tool lets me completely perform entire work routines (Question 21)	Count	3	1	6	4	12	4
	%	10.0%	3.3%	20.0%	13.3%	40.0%	13.3%
The functions implemented in the tool support me in performing my work (Question 22)	Count	0	0	3	3	15	9
	%	.0%	.0%	10.0%	10.0%	50.0%	30.0%
The way in which data is entered is suited to the cognitive modeling tasks I want to perform with the tool (Question 23)	Count	1	1	6	6	14	2
	%	3.3%	3.3%	20.0%	20.0%	46.7%	6.7%
I perceive the arrangement of the fields on-screen as sensible for the work I do with the tool (Question 24)	Count	6	1	4	10	8	1
	%	20.0%	3.3%	13.3%	33.3%	26.7%	3.3%

Table C.18: Frequencies for the questions in general suitability for the task section – Continued

		No Opinion	SD	D	N	A	SA
Too many different steps need to be performed to deal with a given task (Question 25)	Count	1	1	12	4	8	4
	%	3.3%	3.3%	40.0%	13.3%	26.7%	13.3%
The way in which data is output is suited to the tasks I want to perform with the tool (Question 26)	Count	1	1	5	6	15	2
	%	3.3%	3.3%	16.7%	20.0%	50.0%	6.7%
The tool is well suited to the requirements of my work (Question 27)	Count	0	0	4	6	15	5
	%	.0%	.0%	13.3%	20.0%	50.0%	16.7%
In a given screen, I find all of the information I need in that situation (Question 28)	Count	5	1	13	5	6	0
	%	16.7%	3.3%	43.3%	16.7%	20.0%	.0%
The terminology used in the tool reflects that of my work environment (Question 29)	Count	0	0	3	4	18	5
	%	.0%	.0%	10.0%	13.3%	60.0%	16.7%

Table C.19: Frequencies for the questions in general suitability for the task section – Continued

		No Opinion	SD	D	N	A	SA
The tool provides me with a repeat function for work steps that must be performed several times in succession (Question 30)	Count	5	2	5	8	6	4
	%	16.7%	6.7%	16.7%	26.7%	20.0%	13.3%
I can easily adapt the tool for performing new cognitive modeling tasks (Question 31)	Count	0	2	5	1	13	9
	%	.0%	6.7%	16.7%	3.3%	43.3%	30.0%
The important commands required to perform my work are easy to find (Question 32)	Count	0	2	10	9	8	1
	%	.0%	6.7%	33.3%	30.0%	26.7%	3.3%
I am able to adjust the presentation of results (on the screen, to printer, plotter, etc.) to my various work requirements. (Question 33)	Count	1	2	10	4	12	1
	%	3.3%	6.7%	33.3%	13.3%	40.0%	3.3%
The presentation of the information on the screen supports me in performing my work (Question 34)	Count	2	0	6	6	15	1
	%	6.7%	.0%	20.0%	20.0%	50.0%	3.3%

Table C.20: Distribution of respondents who found the questions in the general suitability for the task section relevant (Question 35)

	Frequency	Percent
No Opinion	1	3.3
Neither Disagree nor Agree	4	13.3
Disagree	5	16.7
Strongly Agree	6	20.0
Agree	14	46.7
Total	30	100.0

Table C.21: Distribution of respondents who found the questions in the general suitability for the task section effective (Question 36)

	Frequency	Percent
No Opinion	2	6.7
Strongly Agree	5	16.7
Neither Disagree nor Agree	6	20.0
Disagree	7	23.3
Agree	10	33.3
Total	30	100.0

C.4 Specific Suitability for the Task Section

In the following table, abbreviations mean: SD: Strongly disagree, D: Disagree, N: Neither disagree nor agree, A: Agree, SA: Strongly Agree.

Table C.22: Frequencies for the questions in specific suitability for the task section

		No Opinion	SD	D	N	A	SA
The tool enforces good practices for theory-model interactions. (Question 37)	Count	1		7	5	11	6
	%	3.3%		23.3%	16.7%	36.7%	20.0%
The tool eases the design and implementation process of a cognitive model. (Question 38)	Count	1	1	5	2	14	7
	%	3.3%	3.3%	16.7%	6.7%	46.7%	23.3%
I had to change my model to fit the constraints of the tool. (Question 39)	Count	1	4	9	4	7	5
	%	3.3%	13.3%	30.0%	13.3%	23.3%	16.7%
Using the tool makes the evaluation process of my model easier. (Question 40)	Count			5	3	14	8
	%			16.7%	10.0%	46.7%	26.7%
The tool makes it convenient to communicate my model with my colleagues. (Question 41)	Count		1	3	4	16	6
	%		3.3%	10.0%	13.3%	53.3%	20.0%

Table C.23: Frequencies for the questions in specific suitability for the task section – Continued

		No Opinion	SD	D	N	A	SA
Using the tool is easier than using a general purpose programming language for cognitive modeling purposes. (Question 42)	Count			5	2	9	14
	%			16.7%	6.7%	30.0%	46.7%
It is possible to express most aspects of my model in the tool. (Question 43)	Count		1	6	2	14	7
	%		3.3%	20.0%	6.7%	46.7%	23.3%

Table C.24: Distribution of respondents who found the questions in the specific suitability for the task section relevant (Question 44)

	Frequency	Percent
No Opinion	1	3.3
Neither Disagree nor Agree	3	10.0
Strongly Agree	5	16.7
Agree	21	70.0
Total	30	100.0

Table C.25: Distribution of respondents who found the questions in the specific suitability for the task section effective (Question 45)

	Frequency	Percent
No Opinion	2	6.7
Disagree	2	6.7
Neither Disagree nor Agree	5	16.7
Strongly Agree	5	16.7
Agree	16	53.3
Total	30	100.0

Table C.26: Mann-Whitney test using ACT-R as grouping variable

Question	Mann Whitney U	Z	Asymp. Sig. (2-tailed)	Exact Sig. [2*(1-tailed Sig.)]
20	93.500	-.806	.420	.448
21	89.500	-.973	.330	.355
22	94.000	-.813	.416	.473
23	90.500	-.951	.342	.377
24	100.500	-.495	.621	.637
25	76.500	-1.544	.123	.142
26	90.500	-.962	.336	.377
27	99.000	-.582	.560	.608
28	89.500	-.985	.325	.355
29	89.500	-1.061	.289	.355
30	90.500	-.911	.362	.377
31	104.500	-.331	.741	.759
32	103.000	-.391	.696	.728
33	112.000	.000	1.000	1.000
34	91.500	-.919	.358	.400
37	23.501	-1.840	.066	.078
38	111.000	-.044	.965	.984
39	99.500	-.533	.594	.608
40	72.500	-1.756	.079	.101
41	68.000	-1.998	.046	.070
42	101.500	-.469	.639	.667
43	91.500	-.909	.363	.400

Table C.27: Mann-Whitney test using gender as grouping variable

Question	Mann Whitney U	Z	Asymp. Sig. (2-tailed)	Exact Sig. [2*(1-tailed Sig.)]
20	40.000	-1.254	.210	.270
21	29.500	-1.840	.066	.078
22	47.000	-.811	.417	.482
23	42.000	-.772	.440	.482
24	52.000	-1.073	.283	.323
25	55.500	-.484	.628	.674
26	55.500	-.282	.778	.801
27	55.500	-.282	.778	.801
28	55.500	-.272	.785	.801
29	48.000	-.798	.425	.518
30	32.500	-1.620	.105	.114
31	29.000	-1.892	.058	.078
32	37.500	-1.356	.175	.201
33	56.500	-.213	.832	.845
34	40.000	-1.254	.210	.270
37	58.000	-.120	.904	.933
38	58.500	-.093	.926	.933
39	59.500	-.030	.976	.978
40	49.500	-.645	.519	.556
41	47.000	-.814	.416	.482
42	42.000	-1.111	.267	.323
43	56.500	-.214	.830	.845

APPENDIX D. DISTRIBUTION OF THE GENERAL SUITABILITY FOR THE TASK SCORE

We have included the graphs for the general suitability for the task score in this appendix. This score is calculated by first mapping the responses to points, such that “strongly disagree” is 1 and “strongly agree” is 5, and then summing them up for each respondent. Since questions 20 (“the tool forces me to perform tasks that are not related to my actual cognitive modeling task”) and 25 (“too many different steps need to be performed to deal with a given task”) imply negative statements about the tool, we have reversed their scores by setting “strongly agree” to 1 and “strongly disagree” to 5. The Q-Q plot of the score is shown in figure D.1, which hints a normal distribution. But the histogram of scores which is presented in figure D.2 shows a fairly normal distribution, with a secondary peak for the interval 20.00 – 25.00. This is a deviation from the normal distribution.

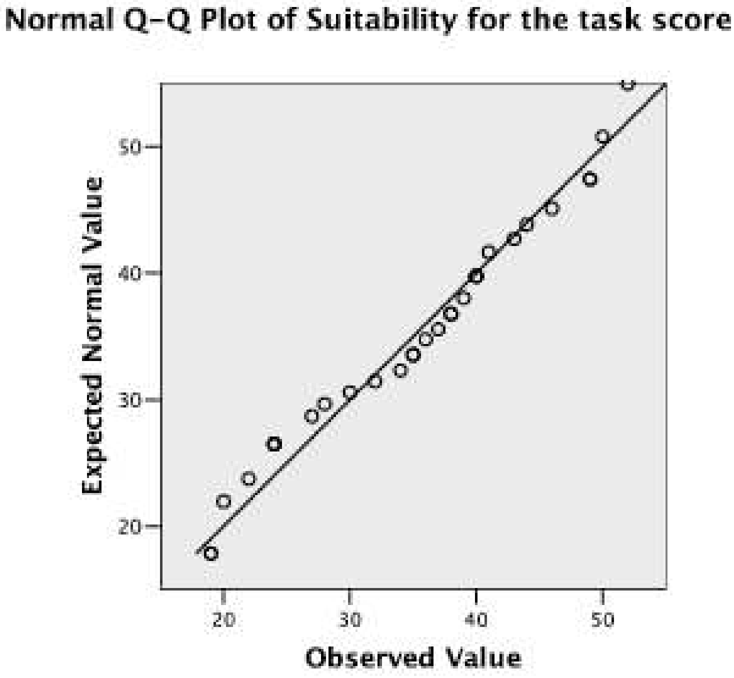

Figure D.1: Q-Q plot of general suitability for the task score.

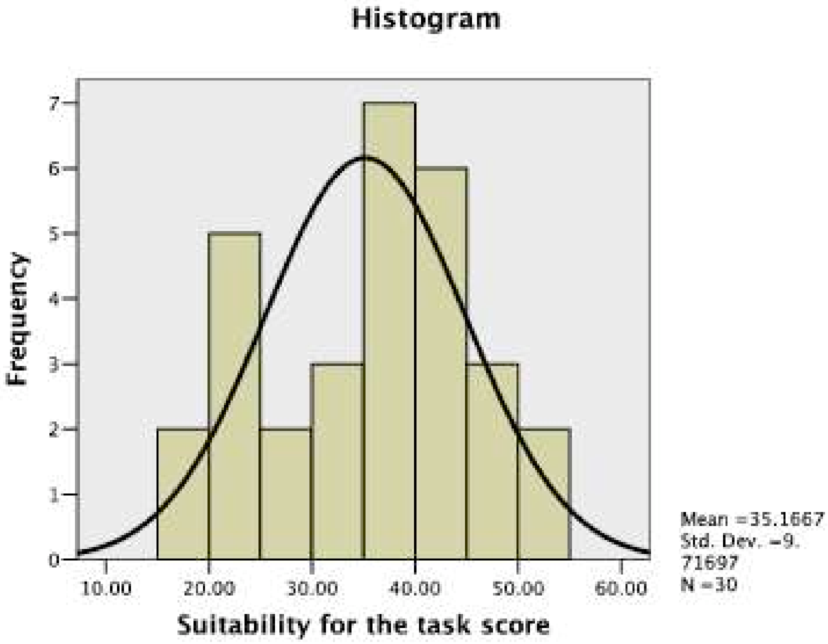

Figure D.2: Histogram of the general suitability for the task score.