AVALIAÇÃO DO MÉTODO DE INSPEÇÃO SEMIÓTICA

SORAIA DE SOUZA REIS

AVALIAÇÃO DO MÉTODO DE INSPEÇÃO SEMIÓTICA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais — Departamento de Ciência da Computação. como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Raquel Oliveira Prates

Belo Horizonte

Março de 2012

SORAIA DE SOUZA REIS

ASSESSMENT OF THE SEMIOTIC INSPECTION METHOD

Dissertation presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Raquel Oliveira Prates

Belo Horizonte

March 2012

Reis, Soraia de Souza

R375a Avaliação do Método de Inspeção Semiótica / Soraia de Souza Reis. — Belo Horizonte, 2012

xxi, 224 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais — Departamento de Ciência da Computação.

Orientadora: Raquel Oliveira Prates

1. Computação – Teses. 2. Engenharia Semiótica – Teses. I. Orientadora. II. Título.

519.6*75(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Avaliação do método de inspeção semiótica

SORAIA DE SOUZA REIS

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROFA. RAQUEL OLIVEIRA PRATES - Orientadora
Departamento de Ciência da Computação - UFMG

DRA. CARLA FARIA LEITÃO
SERG/DI - PUC-RJ

PROFA. LUCILA ISHITANI
Instituto de Ciências Exatas e Informática- PUC-MG

Belo Horizonte, 23 de março de 2012.

Acknowledgments

Agradeço aos meus pais Raimundo e Lourdes pelos ensinamentos, apoio e amor concedidos não só ao longo do mestrado, mas também ao longo da minha vida. Às minhas irmãs Andréa (também Dindinha), Vanessa e Graciele por me ampararem nos momentos difíceis e me darem forças para superar as dificuldades. À minha família deixo o meu agradecimento especial.

Ao meu namorado Flávio pelo companheirismo, incentivo e amor. Também o agradeço pela paciência em me ajudar nas dúvidas, ler e reler meus textos dando suas sugestões e pela sua compreensão quando foi necessário passar finais de semana e noites estudando.

Ao meu afilhado e sobrinho Victor e a minha sobrinha Giovana que sempre me motivam com muito carinho.

Aos meus tios(as) e primos(as) por terem me incentivado e torcido por mim.

À Deus por ter me dado a oportunidade, proporcionado condições e iluminado meu caminho durante esta jornada.

A minha amiga Glívia pela sua atenção e disposição em me escutar, ajudar e aconselhar. Agradeço principalmente pela sua disponibilidade em ler e reler meus artigos, apresentações e dissertação dando valiosas sugestões.

À minha Professora e Orientadora Raquel Oliveira Prates por ter acreditado em mim e me acolhido. Também a agradeço pela paciência, pelos ensinamentos, dicas e, principalmente, pelas horas de leituras gastas revisando os artigos, apresentações e dissertação (inclusive nas madrugadas). O seu apoio e orientação tornou possível a realização desta dissertação. Portanto, muito obrigada por tudo!

Ao Professor Rodolfo Resende (inicialmente foi meu Orientador) que, apesar de eu não ter realizado meu trabalho na sua área de pesquisa, me passou ensinamentos muito importantes para conduzir não só o atual trabalho, mas para todas as pesquisas acadêmicas que eu realizar futuramente.

À Daniela Cascini pela oportunidade de realizarmos pesquisas juntas, estas que me proporcionaram ideias para a realização desta dissertação. Agradeço também pelas sugestões e apoio ao longo do mestrado.

À todos os amigos e colegas do DCC/UFMG e do laboratório do Pensi, Ismael, Natália, Luiz Damilton e Flávio Roberto, pelo incentivo e apoio constantes. Agradeço ao Ismael também pelas leituras e sugestões dadas aos meus textos.

À todos os meus amigos, em especial Wellerson, Isabela e Rhyana, e à aqueles que de alguma forma contribuíram torcendo por mim durante esta jornada.

A todos os professores do DCC/UFMG pelos conhecimentos transmitidos ao longo do curso e que foram importantes para o desenvolvimento deste trabalho.

À todas as pessoas que participaram e contribuíram voluntariamente desta pesquisa, seja respondendo questionário ou participando das avaliações e entrevistas, pela disponibilidade.

Às Professoras Carla Leitão e Lucila Ishitani que se dispuseram a estar presentes na banca final e contribuir com o presente trabalho.

Aos Professores Clodoveu Davis, Ricardo Poley e Rosilane Mota que me apoiaram na realização deste mestrado. Um agradecimento especial à Professora Rosilane que desde a graduação sempre me incentivou, ajudou e orientou.

Às secretárias da Pós-graduação do DCC/UFMG, Renata, Linda e Sheila, e demais funcionários do DCC/UFMG pelo convívio e disponibilidade para ajudar nos momentos necessários.

Ao DCC/UFMG e a CAPES pelo apoio financeiro a realização desta dissertação.

Resumo

Avaliar a qualidade da interface de um sistema interativo é uma tarefa importante durante o processo de desenvolvimento de software. No entanto, como diferentes métodos de avaliação têm sido propostos, torna-se cada vez mais necessário avaliá-los a fim de identificar qual método é o mais apropriado para certas situações. Os métodos da Engenharia Semiótica podem ser considerados métodos recentes e ainda estão sendo avaliados. O Método de Inspeção Semiótica (MIS), que foi proposto em 2006 e é o nosso objeto de estudo, é um desses métodos. Embora o MIS tenha sido utilizado em diferentes domínios desde a sua formalização, um estudo das suas capacidades e limitações ainda não foi realizado. Portanto, este trabalho teve como objetivo realizar uma avaliação do método a fim de compreender suas capacidades e limitações. Esta avaliação consistiu em apreciá-lo sob três diferentes perspectivas: (1) obter informações sobre sua aplicabilidade; (2) identificar vantagens e desvantagens sob a perspectiva daqueles que usam o método (i.e., os avaliadores); e (3) identificar custos, diferenças e semelhanças em relação a outro método, neste caso, o Método de Avaliação de Comunicabilidade (um método de avaliação da Engenharia Semiótica baseado no usuário). Como resultado, encontramos grandes vantagens do método, especialmente em relação à sua aplicabilidade a uma grande variedade de tecnologias e domínios. Alguns custos do método também foram identificados. Entretanto, os custos associados ao MIS são compensados pelos resultados obtidos com a aplicação do método. Portanto, o MIS pode ser considerado um método com uma boa relação custo-benefício. Os resultados apresentados neste estudo contribuem para a Teoria da Engenharia Semiótica, uma vez que geram dados a respeito de como um método baseado em teoria está sendo utilizado e percebido pelos avaliadores. A pesquisa é relevante não apenas para a Engenharia Semiótica, mas também para a área de IHC como um todo, uma vez que já foi identificada a necessidade de novos métodos de pesquisa, bem como as teorias em IHC e métodos baseados nas mesmas. Para a área de IHC as contribuições são: (1) fornece uma visão geral das principais vantagens e desvantagens de um método baseado em teoria; e (2) as principais vantagens e desvantagens encontradas são importantes para apoiar pesquisadores e profissionais na decisão sobre se devem utilizar o MIS ou quando fazê-lo.

Palavras-chave: Método de Inspeção Semiótica, Avaliação, Engenharia Semiótica, Método de Avaliação, Comunicabilidade.

Abstract

Assess the interface quality of an interactive system is an important task during the software development process. However, since different evaluation methods have been proposed, it becomes increasingly necessary to evaluate them to see which method is the most appropriate for certain situations. Semiotic Engineering methods can be considered recent methods and are still being evaluated. The Semiotic Inspection Method (SIM), which was proposed in 2006 and is our study object, is one of these methods. Although SIM has been used in different domains since its proposal, a study of its capabilities and limitations has not yet been performed. Therefore, this work aimed at conducting an assessment of the method in order to understand its capabilities and limitations. This assessment consisted of evaluating the method under three different perspectives: (1) gather information about its applicability; (2) identify advantages and disadvantages from the perspective of those who use the method (i.e., evaluators); and (3) identify costs, differences and similarities compared to another Semiotic Engineering method, in this case the Communicability Evaluation Method (a user-based method). As a result we encountered great benefits of the method, especially in relation to its applicability to a wide variety of technology and domains. Some costs of the method were also identified. However, the costs associated with the SIM are offset by the results obtained with the application. Therefore, SIM could be considered a cost-effective method. The results presented in this study contribute to the Semiotic Engineering Theory research, since it generates data regarding how a theory-based method is being used and perceived by evaluators. This study is relevant not only for Semiotic Engineering research, but also to the HCI field as a whole, since the need to research new methods, as well as HCI theories and methods based on them, have already been identified. The contributions to the HCI field are: (1) providing an overview of the main advantages/disadvantages of a theory-based method; and (2) informing researchers and professionals about the main advantages/disadvantages found which can support them in assessing whether or when to use SIM.

Palavras-chave: Semiotic Inspection Method, Assessment, Semiotic Engineering, Evaluation Method, Communicability.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Since the use of computer systems became popular, there is a great concern about the quality of use of interactive systems, which can be defined as the properties of a system that allow users to perform the intended tasks with efficiency and satisfaction [Prates and Barbosa, 2003; Preece et al., 2007; Barbosa and Silva, 2010]. Before declaring that the software is ready to use, it is important to know whether it properly supports users in their tasks and in the environment in which it will be used [Prates and Barbosa, 2003]. Among the existing quality of use properties we can mention usability, which is the most well-known, and that qualifies how easy to use is an interface [Nielsen and Landauer, 1993]. Recently, another property that has been highlighted is communicability, which is the property regarding how efficiently and effectively an interface conveys to its users its design intent and interactive principles [Prates et al., 2000].

Systems' evaluations are typically performed using evaluation methods which aim at helping evaluators assess the quality of use of an interface and identify problems that could affect the user-system interaction [Preece et al., 2007; Barbosa and Silva, 2010]. We can classify the evaluation methods as empirical or theoretical. The empirical methods are those based on the best practices defined by experts over many years of study. The Heuristic Evaluation and Usability Testing are the most well-known usability methods that fall into this category. The disadvantage associated with empirical methods is that the interaction with computer systems is constantly changing, becoming increasingly difficult to use an applied science for the development of interfaces in Human-Computer Interaction (HCI), because the problem to apply it to is intrinsically unstable [de Souza, 2005].

Due to this reason, theoretical approaches in HCI have been gaining power and the needed to advance research and contribute to improve the quality of information technology artifacts has been emphasized [Shneiderman et al., 2002; Carroll, 2003;

de Souza, 2005]. The advantage associated with the use of methods based on theory is that they allow computer interface designers to explain the effects that certain design choices can cause during user-system interaction, enriching the analysis of an interface design. It also supports the proposal of solutions to the problems identified during the interface evaluation [de Souza, 2005].

Among the methods based on a theory we have the Cognitive Walkthrough [Polson et al., 1992], Semiotic Inspection Method (SIM) [de Souza et al., 2006; de Souza and Leitão, 2009; de Souza et al., 2010], Communicability Evaluation Method (CEM) [Prates et al., 2000; de Souza, 2005; de Souza and Leitão, 2009] and Intermediated Semiotic Inspection Method (ISIM) [Oliveira et al., 2008]. The Cognitive Walkthrough is based on Cognitive Psychology and in line with Cognitive Engineering [Norman and Draper, 1986] – a theory that aims at identifying the quality of the system based on its usability, tightly focused on factors associated with users' cognition during the interaction [Salgado et al., 2006].

The SIM, CEM and ISIM are based on the Semiotic Engineering Theory, which is an explanatory theory of HCI that evaluates the interface based on its communicability. The theory understands the interaction as a communication process in which the system's interface is perceived as a message being conveyed from designer to users [de Souza et al., 2010]. The users understand the designer's message as they interact with the interface itself, and, therefore, the designer-to-user communication is in fact a meta-communication [de Souza, 2005].

SIM is an inspection-based method that aims at identifying potential breakdowns in the user-system communication by inspecting the message being sent by the designer [de Souza and Leitão, 2009]. CEM, unlike SIM, is a method that involves users in the evaluation [Prates and Barbosa, 2007] and its goal is to identify potential user-system communication breakdowns based on how the message is received by users [de Souza and Leitão, 2009]. ISIM, in turn, was proposed recently in order to allow the evaluation of user-systems communication considering the perspective of other stakeholders of the system (e.g., a teacher's perspective on a learning support system) [Oliveira et al., 2008].

Given that there are a large number of evaluation methods that can be used to assess the quality of interfaces, and that they, in turn, differ in many aspects, it is necessary to identify which method is more suitable to be used in certain situations. In order to choose which one is more appropriate for a specific domain professionals and researchers must seek to understand the advantages and disadvantages of each method with respect to time, costs and human factors [Karat, 1990; Prates and Barbosa, 2003; Hartson et al., 2003]. To provide useful guidance, studies have been conducted outlining the capabilities and limitations of some methods.

In the usability context, as methods were being proposed, studies aiming at assessing those methods emerged. Taking as an example the Heuristic Evaluation (HE) and Usability Testing methods (the most consolidated usability methods) assessments emerged with different purposes. HE was initially assessed in relation to the severity and number of problems encountered, in terms of the how many evaluators were needed in order to find a significant number of problems at a reasonable cost, and the influence of evaluators' experience [Nielsen and Molich, 1990; Nielsen, 1992]. Then, comparative studies were performed comparing the different existing evaluation methods, such as the work done by Jeffries et al. [1991] comparing the methods HE, guidelines, Cognitive Walkthrough and Usability Testing; Desurvire et al. [1991, 1992] comparing the impact of different evaluator profiles in the application of HE and Cognitive Walkthrough; and Karat et al. [1992] comparing Usability Testing with individual and team walkthrough methods.

Recently, evaluations have been conducted in order to present a comparison of methods in specific domains. For instance, Steves et al. [2001] present a comparison of a user-based method with HE using a set of groupware heuristics to evaluate groupware systems; Yen and Bakken [2009] present a web tool evaluation for nurse scheduling using HE and Think-Aloud protocol; and Khajouei et al. [2011] presented an effectiveness evaluation of Cognitive Walkthrough and Think-Aloud in identifying different types of usability problems in the domain of healthcare information system.

Despite the efforts to provide information about interface evaluation methods, researchers criticized the way in which some of the usability methods assessments were conducted Gray and Salzman [1998]. In this context, motivated by the lack of criteria to assess methods, some researchers Sears [1997]; Hartson et al. [2003] proposed measures to compare usability evaluation methods such as, for example, validity, thoroughness and effectiveness. However, Lindgaard [2006] argues that, although these measures have been mandatory for researchers seeking to establish the efficacy of a given procedure, especially the notions of thoroughness and effectiveness are irrelevant to HCI practice. The procedures and measures used more recently are still being criticized Hornbæk [2010], and, therefore, this indicates that the HCI field still needs further studies related to methodologies to support researchers to conduct assessments of the methods that have been emerged recently, as well as those that take into account properties other than usability.

Semiotic Engineering methods have been proposed recently and, besides the critics, criteria applied to comparing usability evaluation methods may also not be the most appropriate for communicability evaluation methods [Salgado et al., 2006]. Since the methods are recent, only a few studies have been conducted to evaluate them. A study discussing the cost-benefits of CEM in comparison to well-known usability evaluation methods discusses relevant criteria to compare these methods and presents the results [Salgado et al., 2006]. The Intermediated Semiotic Inspection Method (ISIM) was proposed and evaluated in regard to its contributions to educational systems evaluations [Oliveira et al., 2008].

SIM was proposed in 2006 [de Souza et al., 2006] and, although it has been used, since its proposal, a study to delineate its costs and benefits has not yet been performed. SIM is a recent theory-based method and in order to better understand the benefits of theory-based methods and this specific method it is important to evaluate it. Some studies have investigated SIM's applicability to specific domains, namely educational domain [Oliveira, 2010], collaborative systems [Mattos et al., 2009] and Human-Robot Interaction [Bento et al., 2009]. All these works have shown that the method could be successfully applied in the specific domain being investigated. Recently, a research involving SIM was conducted by Villela et al. [2011] in which an evaluation of the interaction and information quality of a system was made through the application of SIM and an approach for assessing the quality of information (AQI), respectively. The goal of this study was to make a comparative analysis between the results of each of these perspectives (i.e., SIM x AQI). This study showed that the methods identify different aspects and, therefore, are complementary.

These efforts in assessing SIM are important, provide useful information and support practitioners in deciding whether use it or not in the situations investigated so far. However, since these efforts are specific assessments of the method they are not sufficient to support practitioners in a wider range of situations. Therefore, an assessment that demonstrates the capabilities and limitations of SIM with a broader view is necessary. For this reason, this work aimed at conducting an assessment of SIM in order to understand its capabilities and limitations. This assessment consisted of evaluating the method under three different perspectives, they are: (1) gather information about its applicability; (2) identify advantages and disadvantages from the point of view of those who use the method (i.e., evaluators); and (3) identify the costs, differences and similarities over another method, in this case CEM.

To demonstrate the applicability of SIM a Systematic Literature Review (SLR) was conducted to identify publications that have reported the use of the method. We have identified several publications that report the use of SIM and through their analyses we yielded positive indicators regarding the applicability of the method, especially that SIM is technology and domain independent. We continued SIM's assessment through an investigation from the evaluators' perspective, aiming at identifying their perception about SIM's costs, benefits, advantages and disadvantages. In order to do so, we applied a questionnaire to novices' evaluators and made interviews with the authors of SIM (which represents the experts' evaluators) comparing, contrasting and consolidating their views. The analysis of the responses provided interesting insights about SIM's advantages and disadvantages.

In the assessment carried out from evaluators' perspective evaluators had contradictory views of some of the advantages and disadvantages of the method (i.e., easy/difficult to learn, high/low effort needed, high/low cost-effectiveness and theory advantage/disadvantage). Then, to better understand SIM's characteristics related to some of the evaluators' contradictory views we performed an assessment in an empirical way, comparing SIM with CEM. The purpose of this assessment was to collect information, mainly, related to the application effort and cost-effectiveness of SIM compared to CEM.

Since SIM and CEM focuses on different perspectives (i.e., SIM is an inspection-based method and CEM a user-based method) the aim of this assessment was not to say which one is the best method (or the most cost-effective), but rather to show their differences and similarities. Therefore, the goal was to raise information regarding types of problems found by SIM and CEM, contrast the severity of the problems encountered by each method, analyze the influence of the evaluator experience in the application of SIM, and compare the time and effort needed to apply it. It is noteworthy that although the empirical assessment occurred with SIM and CEM, all the analysis done focused on SIM since the research aims at evaluating it. As a result we perceived characteristics of SIM in terms of effort needed to apply it, its cost-benefits and also realize that SIM and CEM, as expected, are complementary methods, since they allow finding different types of problems.

Finally, after carrying out all phases of the proposed assessment, we made a final discussion and consolidation of the results obtained. In this step of the research we aimed at discussing the results convergences and divergences obtained in each assessment carried out.

This work is organized as follows. In next Chapter, Theoretical Framework, we present the Semiotic Engineering Theory and the Semiotic Inspection Method, which is the object of our study. Then, in Related Work Chapter, we present studies done to assess usability and communicability evaluation methods. Next, we present the methodology used to assess SIM. In the following Chapters we present the assessments carried out. First we present the applicability assessment and its results. Next we present an assessment of SIM from evaluators' point of view. Then, we present an empirical evaluation of the method. These Chapters are followed by the Final Discussion

Chapter where we present a final discussion and consolidation of results. Finally, in the Conclusion Chapter we present the final considerations of this work and future works.

Chapter 2

Theoretical Framework

In this Chapter we present the Theoretical Framework of this research. In Section 2.1 we present the Semiotic Engineering theory [de Souza, 2005], the theory in which the Semiotic Inspection Method (SIM), object of this study, is based on. Then, we present SIM in Section 2.2.

2.1 Semiotic Engineering

Semiotic Engineering is an explanatory theory of HCI that explains the phenomena involved in the design, use and evaluation of systems, as well as aspects related to these phenomena. The theory perceives the user-system interaction as a designer's message to users. This message is *unidirectional* and *indirect*. *Unidirectional* because it conveys a complete and immutable content encoded and made available through the system interface [de Souza, 2005]. *Indirect* because the user must understand it while interacting with the system's interface [Prates and Barbosa, 2007].

The designer's communication with users is only fully achieved if users generate meanings that are compatible with the meanings encoded in the designer's message. When the user is not able to understand the designer's intended communication, a *communication breakdown* takes place [Prates and Barbosa, 2007]. In Semiotic Engineering the interface itself is responsible for communicating the designer's vision, then assuming the role of its representative or the *designer's deputy* [de Souza, 2005; Prates and Barbosa, 2007]. As the users understand the designer's message while they interact with the interface itself, the designer-to-user communication is in fact a meta-communication that can be *paraphrased* as follows:

"Here is my understanding of who you are, what I've learned you want or

need to do, in which preferred ways, and why. This is the system that I have therefore designed for you, and this is the way you can or should use it in order to fulfill a range of purposes that fall within this vision." [de Souza, 2005, p.25]

The meta-communication dynamics of the Semiotic Engineering theory is demonstrated in Figure 2.1.

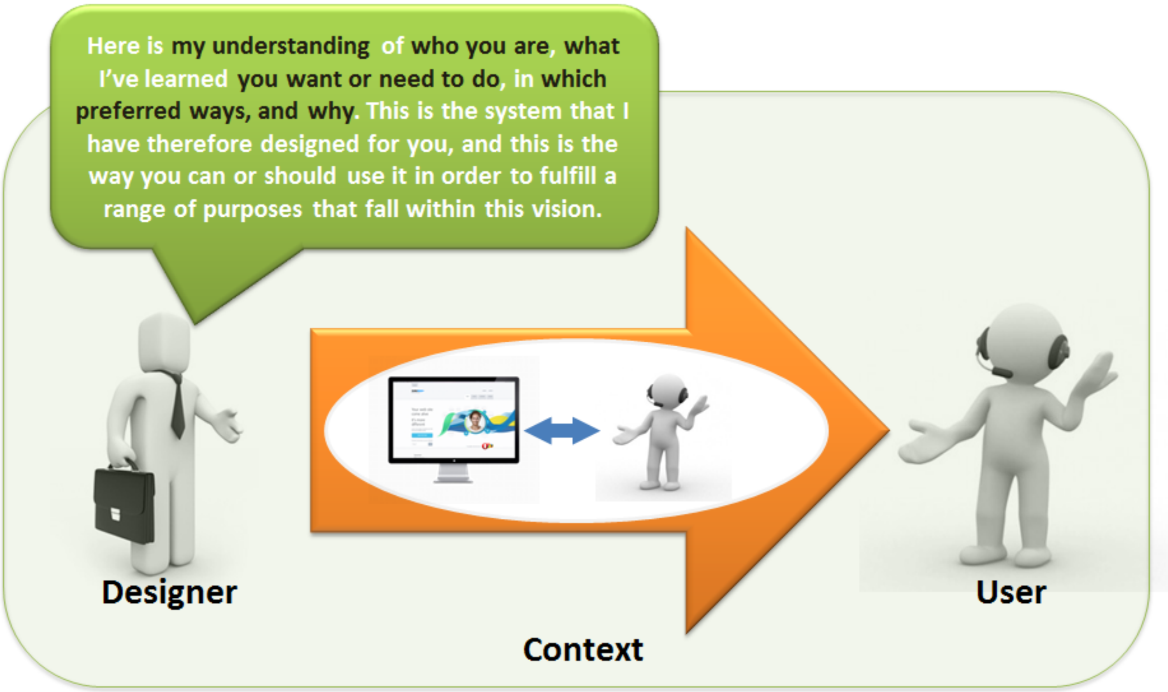

Figure 2.1. Overview of the Semiotic Engineering Theory.

The message is composed of signs, which can be defined as anything that represents something for someone [Peirce, 1992]. In Semiotic Engineering the sign can be classified into three classes [de Souza et al., 2006; Prates and Barbosa, 2007; de Souza and Leitão, 2009; de Souza et al., 2010]: *metalinguistic*, *static* and *dynamic*.

- *Metalinguistic signs* signs are explanations from designer to users about the system or other interface signs. In other words, are signs that designers use to explicitly communicate to users the meanings encoded in the system and how they can be used (e.g., instructions, tips, help and system documentation). For example, in Figure 2.2 the metalinguistic sign is explaining to the user that no tasks have been created, what he can use tasks for, and that he can create one by clicking on the "Add task" link.

- *Static signs* are signs whose meaning is interpreted independently of temporal and causal relations. In other words, the interpretation is limited to the elements present at the interface at a given time, thus, it expresses the system's state (e.g., button's state, interactive element used, selected options). For example, in Figure 2.2 the static sign ⚙ is presented, and the evaluator, when inspecting the system, must assign possible meanings to this sign only by looking at it, in other words, without actually interacting with the sign.

- *Dynamic signs* represent the system's behavior. They are bounded to temporal and causal aspects of the interface and communicate the processing that leads to transitions between system states. In other words, they can only be perceived through the interaction (e.g., action triggered by a button, the impact of selecting a particular value in a field). In the example presented in Figure 2.2, the dynamic sign is represented by the action of clicking in the sign ⚙ and opening a menu to sort (by date or priority) the tasks list.

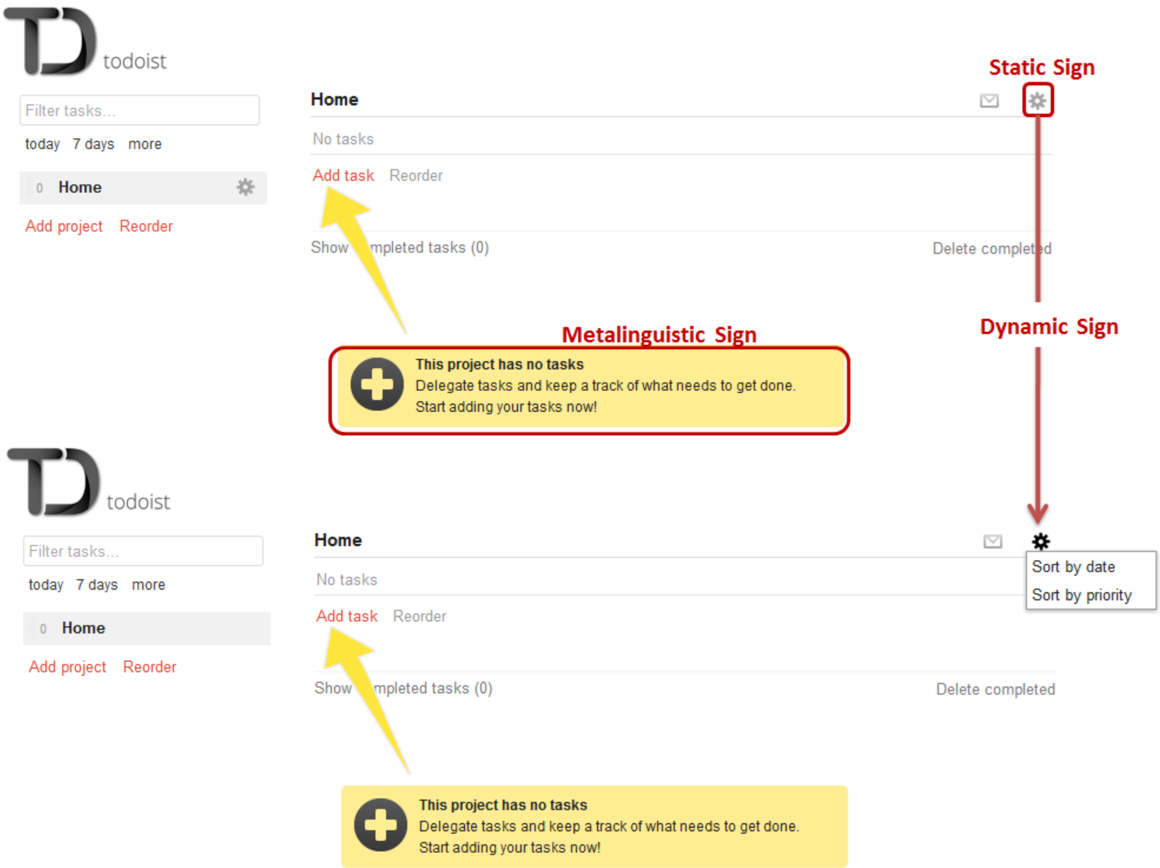

Figure 2.2. Example of Sign Classes.

Within the Semiotic Engineering theoretical framework, a relevant interface property is its *communicability*, which was first defined by Prates et al. [2000] as a distinctive quality of interactive systems that communicate efficiently and effectively to users its underlying design intent and interactive principles. Then, more recently, de Souza and Leitão [2009] complemented with a clearer definition of "efficiently" and "effectively" communication in which the former refers to a communication done in an organized and resourceful way, while the latter refers to a communication that achieves the desired result.

A more technical rereading of this definition was done by de Souza [2005]:

> *"Communicability can (...) be more technically defined as the designer's deputy capacity to achieve full meta-communication, conveying to users the gist of the original designer's message. (...) Communicability applies to both interpretative and expressive codes that designer's deputy handles for generating and interpreting messages during situated interaction with users."* [de Souza, 2005, p.114]

It is noteworthy that the Semiotic Engineering is not a predictive theory, but an explanatory theory of HCI. In other words, it does not intend to predict how users will get the designer's message, but rather it aims at explaining the observable phenomena of HCI. According to de Souza [2005], *"the theory basic ontology* [1] *contains the elements necessary to structure an explanation for any use situation of interest"* [de Souza, 2005, p.105]. It should provide the means necessary to the formulation of problems and design issues of HCI and elaborate their solutions and answers [de Souza, 2005].

The methods used in the Semiotic Engineering research should always increase the theory's explanatory power and help in the elaboration and evaluation of the designer to user meta-communication [de Souza, 2005]. Based on Semiotic Engineering, methods to evaluate the communicability of interfaces have been proposed: Semiotic Inspection Method (SIM) [de Souza et al., 2006; de Souza and Leitão, 2009; de Souza et al., 2010], Communicability Evaluation Method (CEM) [Prates et al., 2000; de Souza, 2005; de Souza and Leitão, 2009] and, recently, the Intermediated Semiotic Inspection Method (ISIM) [Oliveira et al., 2008].

SIM is an inspection-based method which evaluates the intended message being sent by designers and identifies potential breakdowns. CEM involves observing users interacting with the system in a controlled environment (e.g., a user testing lab) and

[1] *"The categories of things that exist, from which follows a series of the relations among them."* [de Souza, 2005, p.95]

identifying the communicative breakdowns that take place and their impact in the designer to user communication. Finally, in the ISIM the evaluator guides a stakeholder (who is not a user) in inspecting the system and through a semi-structured interview identifies potential breakdowns based on stakeholder's point of view. Next we detail SIM since it is our research object.

2.2 Semiotic Inspection Method

SIM is an inspection-based method in which evaluators can analyze the communicability of interactive computer-based artifacts [de Souza et al., 2006]. The method must be performed by a specialist in HCI that should analyze the interface's communicability based on his/her HCI and Semiotic Engineering knowledge [Bim, 2009]. SIM was designed to explore the designer's deputy with the emphasis in the message being sent. It aims at reconstructing the meta-communication using the Semiotic Engineering paraphrase (previously described) as a guide [de Souza and Leitão, 2009] identifying inconsistencies and ambiguities.

Like other inspection-based methods, SIM requires a preparation step before starting the procedures [de Souza et al., 2010]. The preparation is carried out in four steps: (1) the evaluator defines the purpose of inspection, taking into account the specific domain in which the method is being used; (2) the evaluator makes an informal inspection of the system in order to define the intended focus of the evaluation; (3) the evaluator navigates in the system to verify which are the main users of the system and to understand what are the main objectives and activities that the system supports; and (4) the evaluator prepares the inspection scenarios that provide the contextual structure necessary for the communication analysis.

SIM has basically five steps as shown in Figure 2.3 [de Souza et al., 2006; Prates and Barbosa, 2007; de Souza and Leitão, 2009; de Souza et al., 2010]: (1) inspection of metalinguistic signs through system's documentation and help; (2) inspection of static signs; (3) inspection of dynamics signs; (4) compare and contrast the meta-message identified in steps (1), (2) and (3); and finally (5) carry out the appreciation of the meta-communication quality.

Steps (1), (2) and (3) are done iteratively. In these steps the evaluator makes a segmented analysis of the system, one for each of the three classes of signs: metalinguistic, static and dynamic. This segmented analysis actually reconstructs the meta-message being conveyed by the designer through each type of sign, allowing the evaluator to inspect in detail what and how the designer communicates with each of

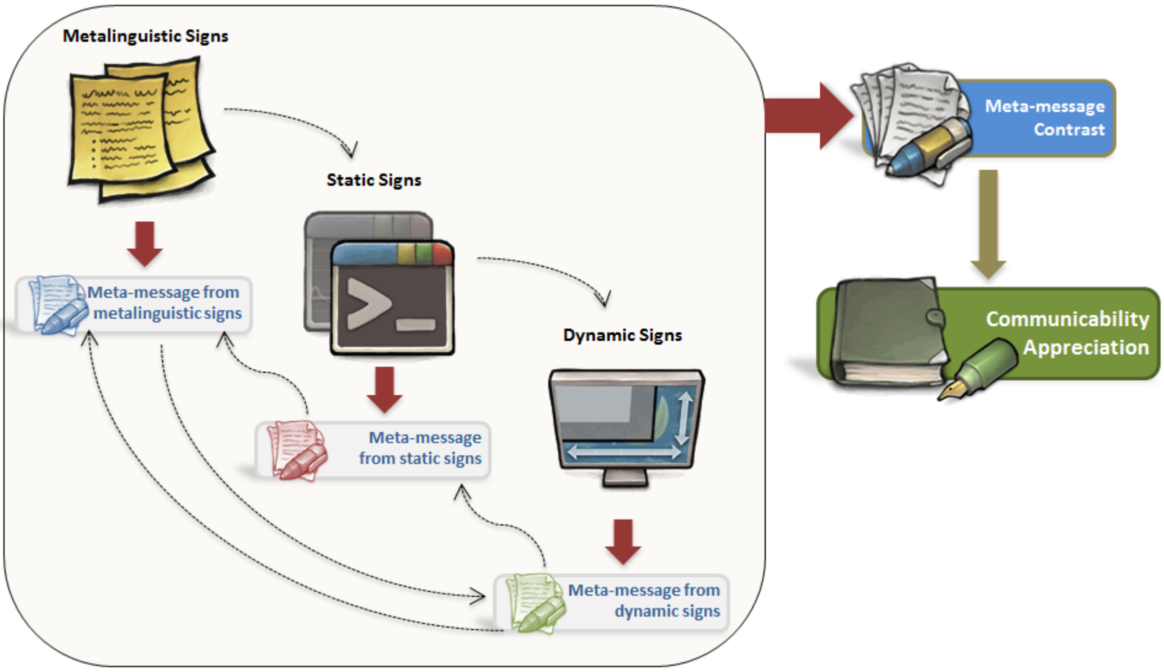

Figure 2.3. Semiotic Inspection Method Steps – Technical Contexts [Adapted from Bento [2010]].

these types of signs.

In step (4) the evaluator should contrast and compare the meta-message generated in the first three steps identifying potential breakdowns that may occur. The evaluator should explore the possibility of the user assigning different meanings to the signs or even identify cases where the meta-message is incomplete due to the lack of signs that clarify the designer's intent.

Finally, in step (5) the evaluator must reconstruct a complete/unified meta-communication message by comparing, integrating and interpreting the data collected in previous steps of the method. The evaluator is also expected to articulate his/her findings about the communication quality (i.e., communicability) of the system by judging the communicative strategies identified in previous steps.

SIM can be applied in two different contexts: *technical*, as already shown, and *scientific* [de Souza and Leitão, 2009; de Souza et al., 2010]. In *technical contexts*, it can be used to improve the quality of designer-user communication and the main aim of evaluation is dictated by business needs, as well as industrial or commercial interests. In *scientific contexts*, though, the main aim of evaluation is dictated by the research questions. The objective of using the method in the scientific context is primarily to advance knowledge.

In a scientific context, some activities must be conducted besides the activities planned in the application of the method in a technical context, as shown in Figure 2.4. In the preparation step, researchers must work with a research question, choose an application instance that serves the purpose of research and clearly define the research objectives, as well as a careful analysis of the results that SIM can help you achieve [de Souza and Leitão, 2009; de Souza et al., 2010].

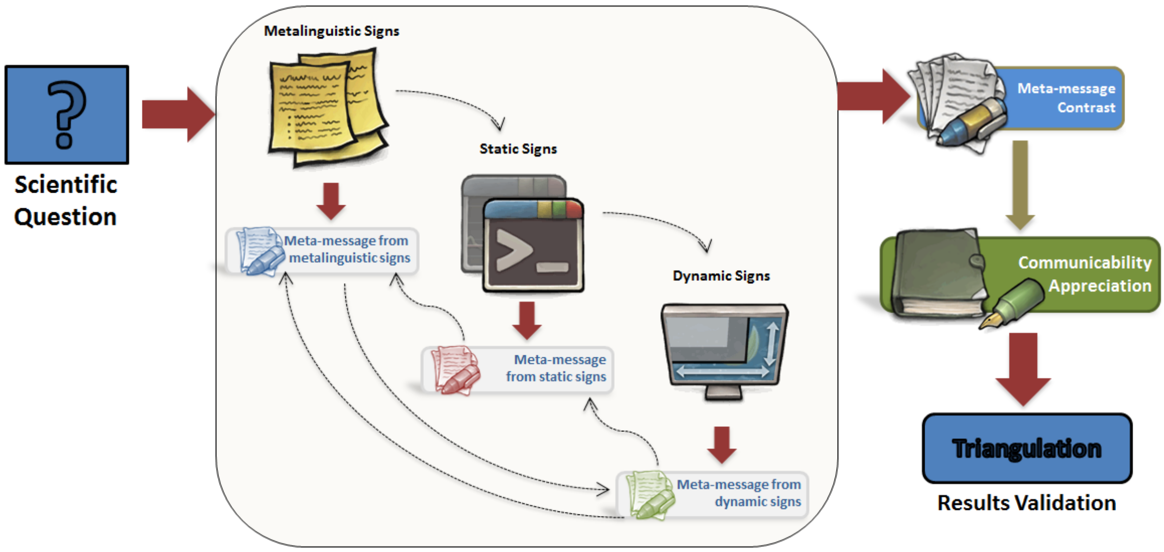

Figure 2.4. Semiotic Inspection Method Steps – Scientific Contexts [Adapted from Bento [2010]].

Moreover, in scientific contexts, SIM requires a final validation step to qualify the results, which is optional in technical contexts: a *triangulation* of results [de Souza and Leitão, 2009] *Triangulation* is necessary to ensure the scientific validity of results. In this final step, *endogenous* and/or *exogenous* sources can be used. *Endogenous* sources refer to the same type of artifact or artifacts that share the same domain model. *Exogenous* sources refer to conception of artifacts that do not share the same domain model, however, share certain relevant characteristics of the project [de Souza et al., 2010].

Chapter 3

Related Work

The introduction of new methods with different approaches and the lack of understanding of their capabilities and limitations have intensified the need to determine which methods are the most effective, in which ways and for what purposes [Hartson et al., 2003]. In order to provide useful guidance to allow the selection of the most appropriate method to be used in a given situation, studies should be conducted outlining the advantages and disadvantages of each method [Karat, 1990].

In this context, several studies have been carried out in an effort to demonstrate the capabilities and limitations of existing and new methods. In this chapter we present the efforts that have taken place in this regard. At first we present some of the works done so far in assessing usability evaluation methods. Since usability is the most consolidated quality of use property in the HCI field we present the related works of this context in order to demonstrate the evolution of the research purposes (related to assessments of usability methods) over the years, the procedures used and analysis performed. Then, we present the studies done in context of communicability evaluation methods, which is our research focus, in order to outline the state of the art.

3.1 Procedures Used and Analyzes Performed to Assess Usability Evaluation Methods

In the usability context, as methods were proposed studies aiming at assessing them emerged. Taking as an example the Heuristic Evaluation (HE) and Usability Testing methods (the most consolidated usability methods) assessments emerged with different purposes. Figure 3.1 shows a timeline that illustrates the researches focuses regarding the assessment of these usability methods in certain periods of time. As we can observe

HE was initially assessed in relation to the severity and number of problems encountered, in terms of the how many evaluators were needed in order to find a significant number of problems at a reasonable cost, and the influence of evaluators' experience [Nielsen and Molich, 1990; Nielsen, 1992].

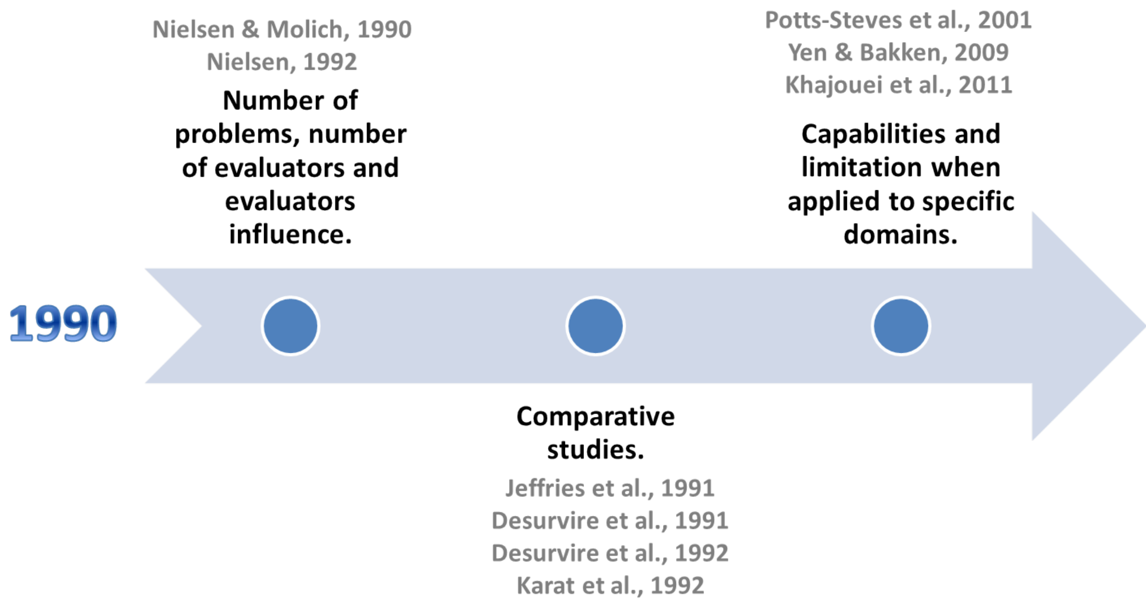

Figure 3.1. A timeline of studies carried out to assess usability evaluation methods.

To evaluate the applicability of HE Nielsen and Molich [1990] did four experiments in different systems where people with no experience in usability analyzed the interfaces heuristically. The four systems were empirically evaluated by 37, 77, 34 and 34 evaluators respectively, and they were asked to generate a report detailing all the problems identified. In possession of the report generated by the evaluators, the problems were consolidated and then compared with an initial list of problems generated by the researchers. According to Nielsen and Molich [1990], the initial list of problems, used for comparison, had to be updated because the evaluators were able to find problems that had gone unnoticed by the researchers.

An analysis of the validity (i.e., will the issues identified in fact present problems to real users?) was not performed and, according to Nielsen and Molich [1990], there were two reasons for not performing this analysis. The first argument is simply that most problems identified are clearly problems. In other words, they are in accordance to an established knowledge in the usability field and, therefore, can be considered real problems. The second argument is that since the experiment was conducted by a considerable number of evaluators it forms a kind of empirical support for the usability issues identified.

Instead of a validity analysis, Nielsen and Molich [1990] examined the following aspects of the HE: (1) percentage of problems that the evaluators were able to identify in each system evaluated; (2) distribution of the number of usability problems found by evaluators; (3) the individual differences in the evaluators' ability to find usability problems; and (4) proportion of problems found by evaluators aggregates (i.e., aggregates of evaluators are formed by having several evaluators conducting an evaluation and then collecting the problems found by each of them to form a larger set).

Continuing the assessment of HE Nielsen [1992] did another experiment where a Banking System was assessed by three groups of evaluators with different levels of experience in usability: (1) novice; (2) "regular" specialists; and (3) "double" specialists. The novice group consisted of 31 students of Computer Science who had completed their first programming course but had no formal knowledge of user interface design principles. The "regular" specialist group consisted of 19 usability experts who did not have expertise on the system context. The "double" specialist group consisted of 14 experts in interface evaluation and also within the context of the system. In this study, Nielsen [1992] looked more closely at some factors that influence the probability of finding usability problems. The factors considered were: (1) the expertise of the evaluators; (2) the severity of the usability problems; (3) the individual heuristics; and (4) the activities needed to identify the problems.

Following the independent assessments [Nielsen and Molich, 1990; Nielsen, 1992] comparative studies started to be performed and presented as, for instance, the work done by Jeffries et al. [1991]. In this study the authors compared four interface evaluation methods: Heuristic Evaluation, guidelines, Cognitive Walkthrough and Usability Testing. The main goal in this study was to determine what kinds of interface problems these methods enabled evaluators to find, whether developers who are not interface experts could use them or not, and make a comparison of methods in relation to their cost-benefit.

Four distinct groups applied each one of the methods in the same system, and reported all the problems encountered in a common form so they could be compared. The list of problems generated during the experiment was analyzed by three evaluators who worked independently and then consolidated the information. The main goal of the experiment was to analyze: (1) the total number of problems encountered by problem type; (2) how the problems have been found (i.e., using the method, side effect or evaluator's prior experience); (3) the severity of the problems on a scale from 1 (trivial) to 9 (critical); and (4) the methods' cost-benefits.

Desurvire et al. [1991, 1992] conducted a research comparing the impact of different evaluator profiles. In an initial study they [Desurvire et al., 1991] compared results obtained with Heuristic Evaluation with those obtained in a user-controlled environment. This work was extended in 1992 [Desurvire et al., 1992], in order to include the Cognitive Walkthrough method. In this experiment the authors aimed at comparing the effectiveness of each method (i.e., Heuristic Evaluation and Cognitive Walkthrough) performed by three different evaluators' profiles (i.e., interface experts, not experts and interface developers) using as a standard of comparison lab testing results. The main goal was to evaluate the percentage of problems found by each group of evaluators.

Karat et al. [1992] criticized the validity of these previous works mentioning that they leave open questions about the number and types of problems that were not identified, and information on how the problems were interpreted and analyzed. The authors then conducted a study in which they compared User Testing with two variations of walkthrough methods (i.e., individual and team).

In this work, six users participated in Usability Testing. The inspection-based methods used were proposed by the authors and were denominated Walkthrough methods. These methods involve a free inspection by the evaluator (i.e., exploratory), scenario-based inspection, and the use of 12 usability guidelines. The methods were conducted by six evaluators individually, and six pairs of evaluators conducted the evaluation as a team. The usability problems identified through the use of the three methods were categorized using common metrics. Thus, the data could be compared among the methods using the defined dimensions (e.g., number and severity of usability problems identified in the interface).

The aim of this study was: (1) understand the relationship between user-based evaluations and results of inspection-based methods; (2) determine whether the results regarding the efficacy of user-based methods and inspection-based methods were reliable (i.e., the results apply to various systems or are dependent on the system); and (3) understand how well the inspection-based methods work in evaluating interfaces and how to improve their effectiveness.

As we can observe the evaluations in the 90's aimed at presenting initial results on the evaluation methods created. The assessments, in most cases, aimed at making a specific evaluation of the method, as was the case of Nielsen and Molich [1990]; Nielsen [1992], and compare methods to generate evidence regarding the characteristics of each one in relation to the others, as was the case of Jeffries et al. [1991]; Desurvire et al. [1991, 1992]; Karat et al. [1992]. Although comparative assessments have still been conducted [Koutsabasis et al., 2007], recently, most of the assessments of those methods have been conducted in order to present their capabilities and limitations when applied to specific domains.

For instance, Steves et al. [2001] present a comparison of a user-based method with an inspection-based method using a set of groupware heuristics to evaluate groupware systems. In the user-based evaluation collaborators used a groupware system for real work over several months. The authors collected log data, had users self-report diaries, and conducted survey questionnaires and interviews. In the evaluation using an inspection-based method, independent evaluators assessed the groupware system by trying several different scenarios, and examined how well the system comply several inspection criteria. The evaluators consolidated the problems identified into a single report. In this study the authors compared the differences between the usability problems identified with the inspection method and the users' evaluation. The main goal was to identify the differences and commonalities of the methods to help practitioners in deciding when and where to use them.

Yen and Bakken [2009] present an evaluation using Heuristic Evaluation and Think-Aloud protocol of a web tool for nurse scheduling. In this study five HCI experts with no training on the system assessed the interface using the Heuristic Evaluation. The Think-Aloud protocol was conducted with three users of the system. In the comparison the authors assessed the differences, commonalities and severity of the problems identified by each method.

Finally, Khajouei et al. [2011] presents an effectiveness evaluation of Cognitive Walkthrough and Think-Aloud protocol in identifying different types of usability problems in the domain of healthcare information systems. In the study the Cognitive Walkthrough was performed by two evaluators and the Think-Aloud protocol was conducted with ten users. Each problem identified was categorized (according to a framework) and its severity was determined (according to Nielsen's severity ratings). The thoroughness (i.e., the extent to which a usability evaluation method can identify real usability problems), validity (i.e., the extent to which a usability evaluation method accurately identifies usability problems) and effectiveness (i.e., the ability of the usability method to identify usability problems) of the methods were then compared.

3.2 Assessments of Communicability Evaluation Methods

In the context of Semiotic Engineering methods, since they are recent methods, only a few studies have been conducted to evaluate them. Figure 3.2 shows a timeline that illustrates the researches focuses regarding communicability methods assessments.

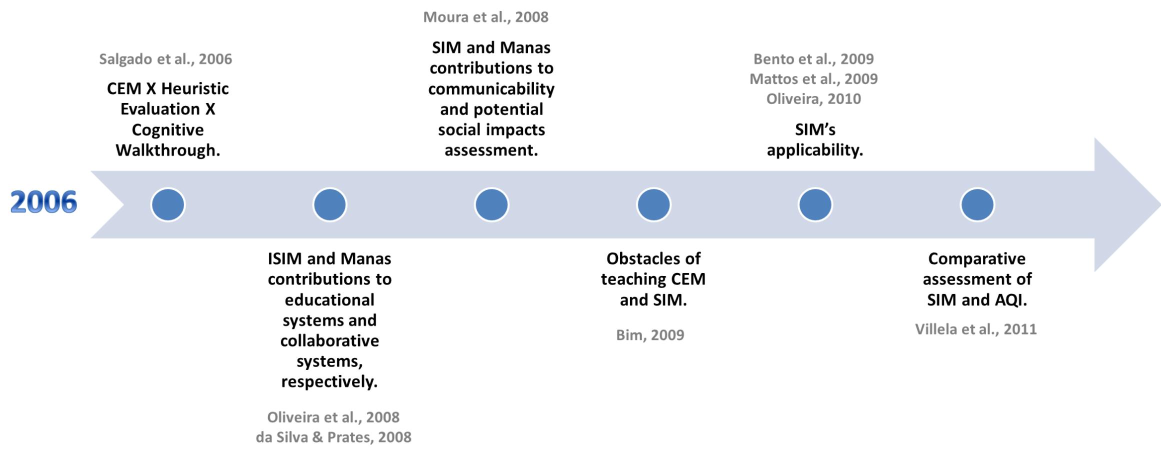

Figure 3.2. A timeline of studies carried out to assess communicability evaluation methods.

In 2006 Salgado et al. [2006] conducted a study discussing the cost-benefits of CEM in comparison to Heuristic Evaluation and Cognitive Walkthrough. The aim of this study was to compare the cost-benefits of the three methods in order to raise information about which one to choose when the time and feedback to (re)design are the most critical factors. The three methods were applied to the Real Player system. The evaluation using the Cognitive Walkthrough was performed by one evaluator, while Heuristic Evaluation and CEM were applied by two evaluators. The parameters of comparison among the methods were essentially the time spent (quantitative) and the information to (re)design (qualitative). The results obtained in this evaluation suggested that regarding the three methods, CEM produces more informative results, while Heuristic Evaluation is the most cost-effective.

Some case studies were also performed in order to present the contributions of other communicability methods for specific contexts. For instance, the Intermediated Semiotic Inspection Method (ISIM) was proposed and evaluated in regard to its contributions to educational systems evaluations [Oliveira et al., 2008], and the use of Manas model as an evaluation tool for collaborative systems has also been performed [da Silva and Prates, 2008].

Moura et al. [2008] did a study with the Semiotic Inspection Method (SIM) and Manas model where the methods were applied to assess the communicability and potential social impacts on Orkut, and identify the benefits of using each of these methods for this type of evaluation. Moura et al. [2008] discussed the types of problems encountered by each method and difficulties during the evaluation. As a result it was possible to realize that the methods allowed finding different problems and, therefore, the combination of the methods to make this kind of analysis is interesting.

In 2009 a study focusing on the obstacles to the teaching of the Semiotic Engineering evaluation methods was conducted by Bim [2009]. The purpose of this research was to identify the difficulties of teaching and learning two Semiotic Engineering methods: SIM and CEM. The research involved interviews and analysis of HCI disciplines where basically three sets of results were found: (1) difficulties related to the limited time to work a large volume of HCI topics; (2) difficulty in developing skills necessary to learn the methods (i.e., systematic interpretation, abstraction and broad view); and (3) teaching initiatives to minimize the difficulties.

Beside the previous studies mentioned, some studies have investigated SIM's applicability to specific domains, namely educational domain [Oliveira, 2010], collaborative systems [Mattos et al., 2009] and Human-Robot Interaction [Bento et al., 2009]. The authors of these studies applied the method in the domain investigated and evaluated how applicable SIM was. All these works have shown that the method could be successfully applied in the specific domain being investigated. Although only three domains were investigated so far, these studies together with the hypothesis raised by authors that the method is technology and domain independent [de Souza et al., 2006] have pointed to a possible great advantage of SIM: that it is technology and domain independent method.

The latest research involving SIM was conducted by Villela et al. [2011] where an evaluation of the interaction and information quality was made through the application of SIM and an approach for assessing the quality of information (AQI), respectively. The goal of this study was to make a comparative analysis between the results of each of these perspectives. By identifying the different aspects considered by each of these perspectives and the contrast between the results obtained, the authors concluded that their results are complementary, and that, ideally, both should be applied.

These first results about SIM make it even more relevant to continue evaluating it. Since no other study has been carried out to evaluate it, the goal of this research was to continue an assessment of SIM identifying its capabilities and limitations. We focused on evaluating the method under three different perspectives, they are: (1) gather information about its applicability; (2) identify advantages and disadvantages from the point of view of those who use the method (i.e., evaluators); and (3) identify the costs, differences and similarities over another method, in this case CEM. This assessment generated indicators about SIM that may help evaluators in selecting the method in a given situation. In the next Chapter we present how this research was conducted.

Chapter 4

Methodology to Assess the Semiotic Inspection Method

As shown in the previous Chapter, we found several efforts to evaluate usability evaluation methods that could support us in the preparation of a methodology to assess SIM. However, despite these efforts, some researchers criticized the way in which some of the usability methods assessments were conducted [Gray and Salzman, 1998]. Gray and Salzman [1998] reviewed the design, procedures and data interpretation of some works and concluded that all of them suffered from absence of validity in some way.

Motivated by the lack of criteria to assess methods, some researchers [Sears, 1997; Hartson et al., 2003] proposed measures to assess usability evaluation methods. The proposals present metrics for assessing the performance of the methods in a comparative way such as validity (i.e., to which extent the method accurately identified usability problems), thoroughness (i.e., to which extent the method identifies real usability problems) and effectiveness (i.e., is the ability of a method to identify usability problems).

Lindgaard [2006] argues that while the notions of validity, thoroughness and effectiveness are mandatory for researchers seeking to establish the efficacy of a given procedure, especially the measures thoroughness and effectiveness are irrelevant in HCI practice. The author argues that these measures requires the knowledge of all existing usability problems of the system under test, and it is impossible to know whether all the usability problems were actually identified unless the evaluation reaches an asymptote (i.e., a point where no new problem can be encountered). Asymptotic evaluations are unfeasible and, therefore, the notions of thoroughness and effectiveness are meaningless and impossible to calculate. Thus, irrelevant in HCI practice [Lindgaard, 2006].

The works carried out so far indicate that the HCI field still needs further studies related to methodologies to support researchers to conduct assessments of the methods. In this direction, Hornbæk [2010] investigated some studies carried out after Gray and Salzman's [Gray and Salzman, 1998] critique and pointed out that recent works have still been making some errors regarding the procedures and measures used. In his study, Hornbæk [2010] extends Gray and Salzman [1998] critiques, presents some dogmas in the assessments of usability evaluation methods carried out so far and proposes approaches that may help move beyond the dogmas (shown in Table 4.1).

Table 4.1: Dogmas in the assessment of usability evaluation methods.

Dogma	Description	Summary
#1	Problem counting as the main approach to assess usability evaluation methods.	This dogma is related to the assessments of usability evaluation methods that only count the number of problems found by each method (the type of assessment most commonly performed). According to Hornbæk [2010], this approach has some limitations (e.g., does not differentiate potential problems from real problems, and does not consider the types of problems identified) that limits the quality assessment of the method, except as a rough measure of how easily problems can be discovered with a particular method. To extend this approach a content analysis of the problems can be performed, for example, classification of problems according to their severity. Hornbæk [2010] also suggests complementing the analysis with an assessment of the evaluators satisfaction with the method used (through questionnaires, for instance). However, he points out that such data (i.e., satisfaction responses) can hardly stand on their own, but may capture orthogonal dimensions to the effectiveness of a method.

Continued on Next Page…

Dogma	Description	Summary
#2	Matching problem descriptions is straightforward.	In most of the methods assessment the problems encountered are matched to identify similar problems. However, often, little or no information is given about how the matching procedures are conducted. Thus, it seems that the matching of problems is straightforward. Hornbæk [2010] mentions that there are reasons why this is not so. For instance, the lack of rigor in describing the problems, and the brief and context-free descriptions allows different interpretations of what constitutes a match. Some studies have been conducted in order to provide structure to the reporting of problems, which may facilitate matching procedures, but Hornbæk [2010] mentions that some other difficulties regarding structured reporting remains. Therefore, research on how to report and match problems continues to be necessary.
#3	Usability evaluation proceeds as prescribed and directly identifies problems.	The usability evaluation methods suggest a way of proceeding and some activities to be undertaken. However, studies that assess methods assume that the methods have well defined activities and that the evaluators' decisions play a small role. In other words, it is assumed that using the method will directly leads to the identification of problems. According to Hornbæk [2010], most of the assessments present a coarse detail about the role of expertise in evaluation and suggests that it is important to understand better how expert evaluators perform their evaluations to be able to develop tools and methods to support novices.
#4	The individual usability problem as the unit of analysis.	In most of the methods assessments each problem identified by the evaluator is considered as a unit of analysis. Rarely sets of problems are prioritized or synthesized, for example, by listing the most critical problems. Hornbæk [2010] mentions that addressing individual problems as the unit of analysis has several advantages (i.e., it is easier to count and more observation can be analyzed statistically), but he mentions a few reasons to complement this type of analysis with an analysis of sets of problems. For instance, considering a larger unit of analysis is possible to ensure that individual problems are not contradictory.

Continued on Next Page…

Dogma	Description	Summary
#5	Look at evaluation in isolation from design.	The great value of evaluation methods is the ability to propose improvements in the interface. However, this is ignored in the assessments of methods as most of the studies look at evaluation in isolation from design. Hornbæk [2010] mentions that a possible solution would be to conduct studies that go closer to the context in which the results are being applied (e.g., ask developer about the utility of the problems).
#6	A single best usability evaluation method exists.	Several studies focus on finding a single best evaluation method and their conclusions are usually based on the number of problems identified by each method and not on an in-depth analysis of the kinds of problems found or utility of the method in evaluating a particular kind of system. Most of the studies assess the methods in a particular context, and, therefore, may not help evaluators to design an evaluation using the method in a different context. Hornbæk [2010] considers the search for a single best method is unfortunate, and suggests looking not only at individual techniques but also at the combination of techniques, and also characterizing the differences in the kind of problems found by each method.
#7	Usability problems are real.	Most studies seem to be based on the view that usability problems are somehow definite, unambiguous and unchanging, that is, the usability problems are real problems. This dogma is related to the belief that any interface contains an exact number of problems or that the problems detected by user-based methods are more real than the problems encountered by inspection-based methods. Hornbæk [2010] mentions that this assumption is doubtful; because we cannot be sure that we have not missed any problems, and that the evaluator's interpretation did not influence the problems encountered with user-based methods.

The work done by Hornbæk [2010] presents some examples of studies done so far that minimize the dogmas mentioned, but none of them presents a consolidated methodology that could avoid all these dogmas. Given that there is no consolidated methodology to assess evaluation methods, we opt for a methodology that would allow us to conduct an assessment of SIM from different perspectives, which will be presented next.

It is noteworthy that the dogmas presented by Hornbæk [2010] are related to usability evaluation methods, thus, they might not be valid for this research context which is to assess a communicability evaluation method. Therefore, we did not treat the dogmas presented as completely true for our assessment context, but we used some of the approaches suggested by Hornbæk [2010] (e.g., assess the evaluators satisfaction, classify the problems encountered by severity and consider the differences of the methods in the empirical assessment) in an attempt to provide a more detailed and complete assessment.

4.1 Methodology

The methodology of this research is divided into three steps, as shown in Figure 4.1: (1) assessment of SIM in relation to its applicability; (2) assessment of SIM from the evaluators' perspective; and (3) assessment of SIM in an empirical way.

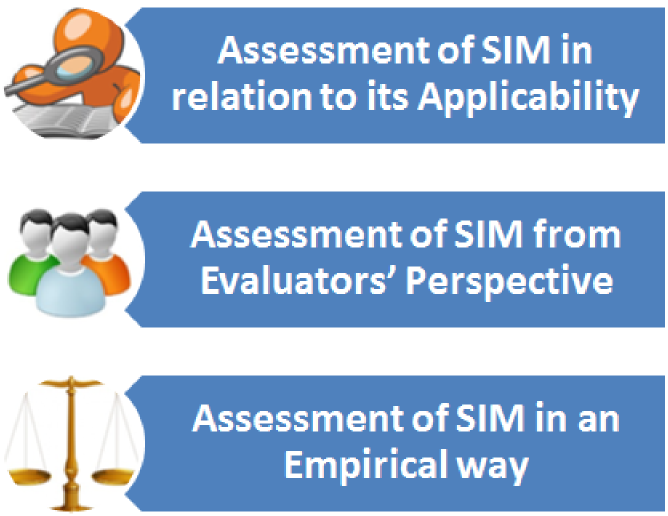

Figure 4.1. Research methodology.

Step (1) **Assessment of SIM in relation to its Applicability** aimed at evaluating SIM's applicability through the analysis of works that present applications of the method. To identify the works that present an application of SIM we carried out a Systematic Literature Review, which is a method to identify, evaluate and interpret available researches relevant to a specific research question or subject area, or phenomenon of interest [Kitchenham, 2004]. Based on the selection and analysis of the works it was possible to demonstrate the applicability of SIM to different technologies and domains, and that the method did not need to be adapted to be applied in these domains. We also discovered that SIM allows identifying issues specific to the domain to which the method was applied. In addition, SIM allowed identifying issues related to other properties, besides communicability, such as, sociability, accessibility and privacy.

We continued to evaluate the method through step (2) **Assessment of SIM from the Evaluators' Perspective** that aimed at analyzing SIM from the evaluators' perspective. At this stage we applied a questionnaire (to novices) and conducted interviews (with experts) to collect data on the experience of evaluators who used SIM aiming at identifying their views on the advantages, disadvantages, costs and benefits of the method. At this stage it was possible to identify the perceptions of the evaluators about the method and raise interesting features regarding SIM.

Although being able to justify most of the advantages and disadvantages presented in the evaluation done in step (2), some contradictions occurred and could not be confirmed only with the evaluators' responses analysis. For this reason, the step (3) **Assessment of SIM in an Empirical way** aimed at evaluating SIM empirically in an attempt to answer some contradictions and raise other features of the method. At this stage SIM was evaluated and compared to the Communicability Evaluation Method (CEM) (which will be described in the Chapter 7 – Section 7.1) in order to gather information over time/effort needed to apply, evaluators' influence in the results, and differences and similarities of the problems found by each method. Since SIM and CEM have different focus (i.e., inspection-based X user-based), at this stage our goal was not to treat them as interchangeable. Rather, our goal was to identify the differences and similarities between the methods (focusing more on the analysis of SIM). The analysis of the results obtained in this step pointed out that, although SIM allowed finding problems beyond CEM scope, as expected, they are complementary methods since some problems could not be found only using SIM (e.g., CEM allowed finding problems regarding user's experience). In addition, expert evaluators performed better than novices using SIM. Moreover, SIM, despite having demanded more time than CEM (considering total effort per person), had a better cost-effectiveness than CEM when considering the time spent to encounter each problem and the severity of the problems. SIM allowed finding a higher number of problems and the evaluators were also able to identify the same major problems identified by CEM.

Chapter 5

Applicability of the Semiotic Inspection Method

In 2006, when SIM was formally presented [de Souza et al., 2006], the authors raised the hypothesis that the method could be technology independent due to its focus on communicative aspects. Authors also mentioned that the method should be systematically and repeatedly applied in different classes of systems to delineate its strengths and weaknesses. Furthermore, a systematic application would allow one to investigate whether the method should be adapted or not to specific domains [de Souza et al., 2006]. Since after SIM's formalization in 2006 several studies have applied the method and reported their results, making it possible to investigate the authors claim.

As presented in the Chapter 3, some studies were conducted to assess the applicability of SIM. However, these studies assessed its applicability in specific domains, namely, educational domain [Oliveira, 2010], collaborative systems [Mattos et al., 2009] and Human-Robot Interaction [Bento et al., 2009]. In this study we perform a broader assessment of SIM's applicability aiming at collecting evidence regarding the hypothesis raised about the method being technology and domain independent [de Souza et al., 2006]. In order to do so, a Systematic Literature Review (SLR) was conducted to identify studies that have reported using SIM in interface evaluations. Based on this review we analyzed the evidences regarding the domains SIM has been applied to so far, and what adaptations (if any) have been necessary.

In our view, Systematic Literature Review (SLR) was an appropriate way to investigate the applicability of SIM, since it allows us to perform a broad search of any studies that may have been conducted using the method and identify the domains and technologies in which it was applied. Next we explain SLR, the process followed, the results obtained and the analysis [Reis and Prates, 2011].

5.1 Systematic Literature Review Process

Kitchenham et al. [2004] have argued that software engineering researchers should adopt "Evidence-based Software Engineering". The term *evidence* is defined as a synthesis of best quality scientific studies on a specific subject. The main method of synthesis is a Systematic Literature Review (SLR), which defines methodologically the scope and selection procedures to be followed in the review of research results (as opposed to a literature review using an ad-hoc methodology) [Kitchenham et al., 2004].

The main advantage of SLR is that it can provide information about the effects of some phenomenon across a wide range of settings. SLR is a way to identify, select and synthesize available studies *relevant to a specific research question or subject area, or phenomenon of interest* [Kitchenham, 2004].

According to Kitchenham [2004] SLR has three main phases: (1) planning the review; (2) conducting the review; and (3) reporting the review. In phase (1) we identify the goal of the review and develop a review protocol. A review protocol specifies, for example, the SLR goal, the strategy to select the primary studies[1] and how it will be conducted. Then, in phase (2), we conduct the review by selecting the primary studies, assessing their quality, doing a data extraction and synthesizing results. Finally, in the phase (3), we report our findings, which, in our case, are presented next.

5.1.1 Research Questions

This study aims at answering the following research question (RQ):

RQ1 What is the applicability of the Semiotic Inspection Method (SIM)?

To better answer this research question [RQ1], it was broken down into specific questions (SQ):

SQ1 To which domains SIM has been applied?

SQ2 In which domains was SIM adapted to meet the goal?

SQ3 What were the challenges in applying SIM in these domains?

SQ4 Does SIM allow evaluators to identify issues specific to the domain in which it has been applied or just general issues?

[1] Individual studies contributing to a systematic review are called primary studies; a systematic review is a form a secondary study [Kitchenham, 2004].

In order to answer these questions, a review of the publications that present interface evaluations carried out using SIM was conducted. During the reading of the selected studies, we identified and collected information that we thought could be interesting to be analyzed. Then, based on the additional information collected some other questions were defined. Although they are not directly related to the applicability of SIM in different domains, they are relevant to better understand aspects regarding its application. The questions considered interesting are (IQ):

IQ1 What other properties of interactive systems can SIM generate indicators of in addition to communicability?

IQ2 Which SIM approaches (i.e., technical or scientific) have been applied to these domains?

Once the questions of interest had been identified the next step was identify the relevant publications.

5.1.2 Research Process

The publications used in this review were obtained automatically in four databases: (1) IEEE Electronic Library; (2) ACM Digital Library; (3) Science Direct; and (4) SpringerLink.

Table 5.1 presents the search parameters defined in order to narrow the search to bring only studies that are related to our research question. We defined parameters to limit the search to the Computer Science area in the databases IEEE Electronic Library, Science Direct and SpringerLink.

Table 5.1. Search parameters for automatic search.

Databases	Search Parameters
ACM Digital Library	-
IEEE Electronic Library	**Subject:** Computing & Processing (Hardware/Software) **Search:** Full Text & Metadata Table of Content were not considered.
SpringerLink	**Collection:** Computer Science
Science Direct	**Subject:** Computer Science

We performed an automatic search in these databases using the following research string. The research string is "semiotic inspection" and its translation to Portuguese (i.e., inspeção semiótica), as the authors of the method are Brazilians, and Spanish (i.e., inspección semiótica), as the expert who monitored this research mentioned that there was a probability of use of the method in Hispanic countries.

"semiotic inspection" OR "inspeção semiótica" OR "inspección semiótica"

We have also conducted manual searches. These manual searches occurred as follows:

- **SERG (Semiotic Engineering Research Group):** all the studies cited on the group website (`http://www2.serg.inf.puc-rio.br/`) in the *Published Work* Section published from 2001 to 2011 were considered.
- **Digital Library of Informatics Department of PUC-Rio[2]:** all technical reports, dissertations and theses listed on the website (`http://www.inf.puc-rio.br/?page_id=112`) which had at least one of the words of the research string in its title or abstract.
- **Computer Science Department of UFMG:** all dissertations done in HCI (`http://www.dcc.ufmg.br/pos/cursos/mestrado.php`).
- **Brazilian Symposium on Human Factors in Computing Systems (IHC):** all studies published in this conference from 2001 to 2004 and 2011. The other years were available at the ACM Digital Library[3].
- **Others:** works nominated by an expert that monitored the research (known publications that had not been found in the databases searched).

The selection of the repositories to search for publications was based on the knowledge of an HCI and Semiotic Engineering expert that monitored the review, henceforth referred to as expert, who is also co-author of SIM. We are aware that there may be works that used SIM in other databases. However, somehow we had to limit our scope. Thus, we selected the main databases (or those considered most expressive) of the HCI field (from Computer Science perspective); databases that are known to have published the greatest amount of works done with SIM and that would allow us to undertake a study with high quality and relevance.

We searched for studies published between 2001 and 2011, in other words, in the last 11 years. Although the paper that formalized the method was published in 2006

[2]It is noteworthy that in the day we did the search in this database, information was presented that the last update occurred on October 18th 2011.

[3]From 2006 on the works of this conference have been published in the ACM Digital Library. However, when search occurred the proceedings of 2011 were not available yet.

[de Souza et al., 2006], there were evidences, based on the expert knowledge, that an early version of the method had been previously used. In order to make the selection as complete as possible, we then defined the period of the last 11 years.

Table 5.2: Inclusion/exclusion criteria

Inclusion Criteria	Studies that used SIM to evaluated interfaces.
	Consider full papers, short papers, technical reports, dissertations and theses.
Exclusion Criteria	Eliminate tutorials, editorials, posters, panels, lectures, roundtables, workshops and demonstrations.
	Works with title, abstract and text in languages other than English, Portuguese and Spanish should be eliminated.[4]
	Duplicated work talking about the same study will be considered equivalent, and the most recent work will be used in the analysis.
	Works that did not reach the minimum quality score established.

The searches in the databases were first done in June 2011, and the last update occurred in January 13^{th} 2012. The search resulted in a large number of works: **434**. To narrow the search down to contain only relevant works we used inclusion/exclusion criteria, as shown in Table 5.2. The purpose of defining inclusion/exclusion criteria is to identify all works that provide direct evidence about the research question [Kitchenham, 2004]. The following steps were performed by applying the inclusion/exclusion criteria established:

1. Reading the title to eliminate irrelevant documents (i.e., not related to HCI field);
2. Reading the abstract and keywords to eliminate works that were not related to the research question;
3. Diagonal reading (i.e., skim reading) in order to confirm whether the work was really related to the research question; and

[4] It is noteworthy that we did not exclude any work using this criterion.

4. Complete reading of the selected works in the previous step, except for dissertations and theses. In these latter cases (i.e., theses and dissertations), due to the length, only the chapters that described the application of SIM were considered in the analysis.

Figure 5.1 presents a quantitative summary of selected works in each step.The *Initial Search* corresponds to the initial total amount of selected works that were obtained using the research string or, the manual search, as mentioned above. Then, after carrying out the steps 1, 2, 3 and 4 we selected **27** studies, which are the studies considered in our analysis.

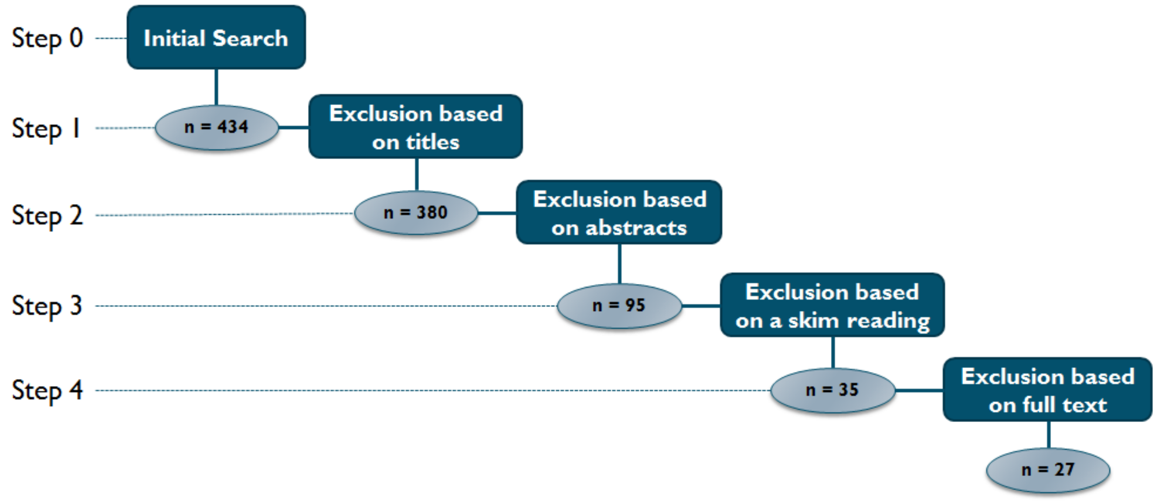

Figure 5.1. Quantitative summary of works.

Table 5.3 presents the quantitative summary from the perspective of the selected studies in each database. The *Initial* column corresponds to the initial total amount of selected works. The *Candidate* column corresponds to the number of selected works after the completion of steps 1, 2 and 3. And the *Final* column represents the total studies selected in step 4, which are the studies considered in our analysis.

In step 4, data was collected from the works selected for analysis purposes. A data extraction form containing fields regarding pieces of information that had to be collected was created. The purpose of generating such a form is that it represents an accurate way to record information reducing the chances of bias in the results [Kitchenham, 2004]. The fields that composed the data extraction form were based on the research questions and are presented in Table 5.4.

Besides the data collection, the studies were evaluated regarding their quality. Ensuring a quality level of the studies is important for several reasons: to provide

Table 5.3. Quantitative summary of works.

Databases	Initial	Candidate	Final
ACM Digital Library	31	8	4
IEEE Electronic Library	7	3	1
SpringerLink	6	2	2
Science Direct	3	2	2
PUC-Rio	53	7	7
SERG	139	3	1
IHC	183	1	1
UFMG	7	4	4
Others	5	5	5
TOTAL	**434**	**35**	**27**

Table 5.4. Data Extraction Form.

General Data	Work Title
	Source (i.e., journal, conference).
	Year of Publication
Research Data	Objective
	System evaluated / Domain
	Aspect evaluated (i.e., communicability, sociability).
	Number and experience of the evaluators
	Duration of the evaluation
	Approach (i.e., technical or scientific).
	Changes made in the method
	Challenges encountered in applying the method
	Results achieved using the method
	Are the generated results domain-specific, generic or both?
	Example of a domain-specific and generic problem encountered
	Additional notes

more details to the inclusion/exclusion criteria, to investigate whether a difference in the quality differentiates the results, provide a way to weigh the importance of each work evaluated, to guide the interpretation of the results and to guide future work recommendations [Kitchenham, 2004].

Table 5.5 presents the quality criteria used to evaluate the selected studies. Each *Criterion* was rated as "yes", "partial" or "no" and was scored with the values 1, 0.5 and 0, respectively. Each criterion has also a *Weight*, which corresponds to the criterion importance degree (1 - low, 2 - medium, or 3 - high) to answer the research questions.

Table 5.5. Quality Criteria.

ID	Criterion	Weight
C1	Does the study clearly define its purpose, aim or a research question?	1
C2	Does the study answer the research questions defined or presents the results in a clear way?	1
C3	Is the study relevant to answer the research questions of this SLR?	3
C4	Does the study mention the domain in which the evaluation was conducted?	2
C5	Does the study report the steps that were followed using SIM?	2
C6	Does the study present challenges in applying the method?	1
C7	Does the study report communicative breakdowns encountered using the method?	2
TOTAL		**12**

The criteria **C1** and **C2** aims at evaluating the general quality verifying if the study reports its purpose and results in a clear way, respectively. This is important because it interferes in the interpretation of the study. These criteria are not related to the research questions of this paper, and, for this reason, they were defined with a low importance degree. The criterion **C3** aims at evaluating if the study is, in general, relevant to answer the research question RQ1. As the question RQ1 is the main question of this study the criterion **C3** was defined with a high importance degree. The criteria **C4**, **C5**, **C6** and **C7** aim at evaluating if the study is relevant to answer the specific questions SQ1, SQ2, SQ3 and SQ4, respectively. We defined these criteria, except for **C6**, as a medium importance degree because although they do not answer RQ1 directly, they contribute to aspects involved in answering it. In the case of **C6** criterion, we defined it with a low importance degree because when a study reports a use of the method it is not required that the authors discuss the challenges encountered while using the method.

The maximum score was 12 points and we established a minimum score of 7.5

points (approximately 60%). That is, to ensure a minimum quality of the data presented in this paper, all the works with scores lower than the minimum established would be eliminated. It is noteworthy that none of the studies analyzed were eliminated by this criterion. All papers presented scored higher than the minimum established.

5.2 Results

Table 5.6 presents the **27** studies analyzed and their reference identifier assigned to be referenced in this study. It also summarizes the data collected during the research. This information will be used in the analysis which is presented next.

In addition, we have categorized the studies according to its focus:

- **Interface Quality:** studies that seek to use SIM to evaluate an interface focusing on the quality of its interface;
- **Application of SIM:** studies that intend to *demonstrate* the use of the method; and
- **Specific Domain:** studies that seek to evaluate the applicability of SIM in specific domains, as mentioned in *Related Works* (Chapter 3) of this study.

Table 5.6: References and data collection summary of the studies analyzed.

Study ID	Reference	Quality	Type	Change in the Method?	Challenges?	Generic, Specific or Both?	Technical or Scientific?	Property Evaluated[5]
S1	de Souza et al. [2004] da Silva et al. [2003]	9 (75%)	Interface Quality	No[6]	No	Both	Technical	Sociability
S2	Seixas [2004]	11 (92%)	Interface Quality	No	No	Generic	Technical	-
S3	de Souza et al. [2006]	11 (92%)	Application of SIM	No	No	Generic	Technical	-
S4	Leitão et al. [2007]	11 (92%)	Interface Quality	No	No	Specific	Scientific	Sociability
S5	de Souza and Cypher [2008]	9 (75%)	Interface Quality	No	No	Generic	Technical	-
S6	Guimarães and de Souza [2008]	11 (92%)	Interface Quality	No[6]	No	Both	Technical	-
S7	Moura et al. [2008]	11,5 (96%)	Interface Quality	No	Yes	Generic	Technical	Sociability
S8	Mattos et al. [2009]	12 (100%)	Specific Domain	No[6]	Yes	Both	Technical	-
S9	Salgado et al. [2009]	11 (92%)	Interface Quality	No	No	Generic	Technical	-
S10	de Souza and Leitão [2009]	11 (92%)	Application of SIM	No	No	Generic	Scientific	-
S11	Castro and Fuks [2009]	11 (92%)	Interface Quality	No[6]	No	Generic	Technical	-
S12	Bim [2009]	10,5 (88%)	Application of SIM	No	No	Generic	Technical	-
S13	de Souza et al. [2010]	11 (92%)	Application of SIM	No	No	Both	Scientific	-

Continued on Next Page...

[5] All the studies analyzed in this paper evaluated communicability since it is the main property of SIM. So in this column we present only the properties evaluated besides communicability. We collected this information from studies that explicitly discussed these issues.

[6] There was no change in the method, but there were changes in the Semiotic Engineering paraphrase.

Study ID	Reference	Quality	Type	Change in the Method?	Challenges?	Generic, Specific or Both?	Technical or Scientific?	Property Evaluated[5]
S14	Peixoto et al. [2010]	11 (92%)	Interface Quality	No	No	Both	Scientific	-
S15	Bento [2010] Bento et al. [2009]	11 (92%)	Specific Domain	No	Yes	Both	Scientific	-
S16	Souza [2010]	11 (92%)	Interface Quality	No	No	Generic	Technical	-
S17	Carvalho et al. [2010]	11 (92%)	Interface Quality	No	No	Generic	Technical	-
S18	Oliveira [2010]	11 (92%)	Specific Domain	No	No	Both	Technical	-
S19	Abreu [2010]	11 (92%)	Interface Quality	No	No	Generic	Scientific	Accessibility
S20	dos Santos Pereira [2011] dos Santos and Prates [2010]	11 (92%)	Interface Quality	No	No	Both	Scientific	-
S21	Brandão et al. [2010]	11 (92%)	Interface Quality	No	No	Both	Technical	Accessibility
S22	Coopamootoo and Ashenden [2011]	11 (92%)	Interface Quality	No [7]	No	Generic	Scientific	Privacy
S23	Ferreira and de Souza [2011]	9 (75%)	Interface Quality	No	No	Generic	Technical	-
S24	Barbosa et al. [2011a] Barbosa et al. [2011b]	11 (92%)	Interface Quality	No	No	Generic	Technical	Sociability
S25	Coutinho et al. [2011]	11 (92%)	Interface Quality	No [8]	No	Specific	Scientific	Accessibility
S26	Capelão et al. [2011]	10 (83%)	Interface Quality	No	No	Generic	Technical	Accessibility
S27	Villela et al. [2011]	11 (92%)	Interface Quality	No [6]	No	Generic	Technical	-

[7] There was no change in the method, but the authors adapted the meta-message to consider privacy issues.
[8] The author adapted the classes of signs to investigate the audio communicability of a computer game.

It is important to note that two of the studies selected, namely **S1** [de Souza et al., 2004; da Silva et al., 2003] and **S2** [Seixas, 2004], carried out an evaluation using steps of SIM, but before its formalization. In these studies the inspection performed SIM's steps, but in different order: static signs, dynamic signs and metalinguistic signs. However, since the analysis of each step was possible we have considered the studies in our analysis.

By collecting the information related to the systems evaluated we found **38** systems, which are listed in Table 5.7. Each system was categorized as follows:

- **Collaborative Systems (CoSys):** Systems that support the execution of activities done in groups.
- **Document Editors:** Systems with document editing features.
- **Educational Systems (EduSys):** All types of educational systems were considered. Some evaluations occurred in educational systems that were also collaborative, and, therefore, we categorized as both EduSys and CoSys.
- **Games:** All playing systems undertaken for enjoyment. We also considered educational games, which were then categorized as both EduSys and Games.
- **Geographic Information Systems (GIS):** Systems that enable and facilitate the analysis, management and representation of space and phenomena that occur in it. In this category we found studies that were conducted in interfaces based on maps.
- **Human-Robot Interface (HRI):** Any system that aims at supporting human-robot interaction.
- **Miscellaneous:** In this category were placed all the other systems that are not part of the categories listed above and also was the only one in a category.

5.3 Analysis

This Section aims at answering the specific questions (SQ) and the issues that we found interesting (IQ) to be answered in this study, as presented in Section 5.1.1. The research question (RQ) will be answered by a general discussion to be held in the Section 5.4. We next present each SQ and IQ question and our conclusions based on the research conducted.

Table 5.7. Systems evaluated.

#	Systems Evaluated	Domain	Study ID
1	AgentSheets	EduSys	**S23**
2	Amazon	Miscellaneous	**S22**
3	Audacity	Miscellaneous	**S10**
4	Bipide	EduSys	**S18**
5	ColabWeb	CoSys and EduSys	**S11**
6	Conecta	CoSys	**S6**
7	CoScripter	Miscellaneous	**S5**
8	ePic	HRI	**S15**
9	e-Puck Monitor	HRI	**S15**
10	Google Docs	CoSys and Document Editors	**S8**
11	Google Groups	CoSys	**S13**
12	Google Maps	GIS	**S17**
13	HalfLife 2	Games	**S25**
14	InForum	CoSys	**S4**
15	Internacional Children Digital Library (ICDL)	Miscellaneous	**S9**
16	Internet Explorer	Miscellaneous	**S22**
17	JECRIPE	EduSys and Games	**S21**
18	Librasnet	EduSys and Games	**S19**
19	Microsoft Dreamspark Portal	Miscellaneous	**S22**
20	Microsoft Streets & Trips	GIS	**S2**
21	Microsoft Word	Document Editors	**S3**
22	Moodle	CoSys and EduSys	**S26**
23	MSN Groups	CoSys	**S1 and S4**
24	Multi-trilhas	EduSys and Games	**S19**
25	NossoGrupo	CoSys	**S1**
26	Notepad	Document Editors	**S12**
27	Orkut	CoSys	**S7 and S24**
28	Preço dos Combustíveis	GIS	**S17**
29	Primeiras Frases em Libras	EduSys and Games	**S19**
30	ProfesSort	EduSys	**S18**
31	ResearchGate	CoSys	**S27**
32	Simple CSS	Miscellaneous	**S13**
33	SimSE	EduSys and Games	**S14**
34	Smart Groups	CoSys	**S1**
35	SME	Miscellaneous	**S16**
36	VCalc	EduSys	**S18**
37	Wikipedia	CoSys	**S20**
38	Yahoo Groups	CoSys	**S1 and S4**

5.3.1 [SQ1] To which domains SIM has been applied?

We can observe in Figure 5.2 that most of the evaluations carried out with SIM occurred in the collaborative systems domain, with **28.26%**. Along with educational systems these domains represents more than half of the evaluations performed using the method (**52.17%**). The other domains in which SIM has been applied range from CSS editors to interfaces for human-robot interaction.

Figure 5.2. Domains evaluated.

The method has been applied not only to applications in different domains, but also with different focuses and technologies. Among the systems evaluated in the collaborative systems domain we can observe that these range from systems focusing on collaborative document editing (i.e., Google Docs) to collaborative encyclopedia (i.e., Wikipedia). In the case of educational systems, analyses were conducted on learning support systems (e.g., VCalc, Bipide) and on an educational simulation game (i.e., SimSE).

Besides analyzing the domain and types of systems evaluated, we looked at the systems technology. We found that the systems evaluated can be categorized in three different technologies: web-based (e.g., ICDL, ColabWeb, Orkut, Google Docs), desktop (e.g., Audacity, Microsoft Streets & Trips) and robot (e.g., e-Puck Monitor).

The analysis shows that systems in a wide range of domains were evaluated using

SIM. The technology they were based on did not include a large number of different types, but certainly a wide range. These findings support the hypothesis that SIM is a domain and technology independent evaluation method.

5.3.2 [SQ2] In which domains was SIM adapted to meet the goal?

We can observe in Table 5.6, in the column *Change in the Method?*, that the works in general do not report changes in the method. However, some studies such as **S1** [de Souza et al., 2004; da Silva et al., 2003], **S6** [Guimarães and de Souza, 2008], **S8** [Mattos et al., 2009], **S11** [Castro and Fuks, 2009] and **S27** [Villela et al., 2011] mention that they used the Semiotic Engineering paraphrase adapted to the collaborative systems domain. The extended paraphrase was presented in the discussion of the application of the Semiotic Engineering Theory to Collaborative Systems, since in this domain there are specific concerns regarding interaction among users through the system that are addressed in the designer's meta-message [de Souza, 2005]. In the work done by **S22** [Coopamootoo and Ashenden, 2011] the authors also reported changes in the meta-message, however in order to identify issues related to privacy.

Using or adapting the paraphrase does not characterize an adaptation in the method. SIM requires the meta-communication message to be reconstructed, regardless of specific aspects being conveyed by the designer in the message being considered. Hence, the evaluator performs the same steps in any domain when examining the designer to user communicative act. What changes is the content of the meta-message being conveyed.

In the study **S25** [Coutinho et al., 2011] the authors did a game evaluation focusing on the communicability of audio sounds. Since no previous studied had done an evaluation focusing on this type of signs, Coutinho et al. [2011] framed SIM concepts to this context. Based on the original definitions of static and dynamic signs, the authors redefined the static signs as the game music and environment sounds (since they can be heard independently of players action), and dynamic signs as the sound effects (since they require players interaction or may be triggered by time-based events). One may argue that this redefinition is a change in the method. However, as it was possible to redefine SIM concepts to this context (i.e., the concepts were applicable), we argue that it cannot be taken as an adaptation of SIM.

5.3.3 [SQ3] What were the challenges in applying SIM in these domains?

As shown in Table 5.6, in the column *Challenges?*, most studies analyzed did not report any challenges in applying the method. This does not necessarily mean that they did not face any challenges, since as shown in Table 5.6, most studies focused on the software being evaluated (all those categorized as *Interface Quality*), and not on the method itself.

The three studies categorized as *Specific Domain* (**S8** [Mattos et al., 2009], **S7** [Moura et al., 2008] and **S15** [Bento, 2010]) reported some of the challenges experienced in applying the method. In collaborative systems, one of the difficulties identified was the need to simulate the interaction among users through the system during the inspection [Mattos et al., 2009]. In order to evaluate this interaction and examine the signs that represent the system's behavior it may be necessary for the evaluator to create multiple logins or profiles in a system. In synchronous systems it may even be necessary to involve another participant in the evaluation with whom the evaluator could interact. It is important to highlight that although the authors discussed these difficulties regarding the application of SIM, they represent challenges to any inspection-based method (e.g., [Baker et al., 2001]) applied to the collaborative systems' domain.

No other study mentions challenges related to the domain in which the SIM was applied. However, a challenge related to SIM's first step has been mentioned in cases in which the system does not have a help system or metalinguistic signs that explain the system as a whole [Moura et al., 2008; Bento, 2010]. In this case, the reconstruction of the meta-communication message based only on metalinguistic signs available as part of the interaction may be very restricted and not very informative. In these cases, SIM can still be applied and the poor meta-message being conveyed through metalinguistic signs will probably be characterized as a communicative breakdown.

5.3.4 [SQ4] Does SIM allow evaluators to identify issues specific to the domain in which it has been applied or just general issues?

We understand that since SIM is a qualitative method and its evaluation scope is always referred to a given context, the problems encountered by the method will always be related to the evaluation *context*. However, the goal of this question was to collect data on weather SIM could be applied to different *domains* and identify problems specific to that *domain*, or if it could only identify generic communicative breakdowns. In other words, if applied to a collaborative system, could it identify problems regarding users' interaction through the system, or only general problems, such as static elements that were not consistent with the behavior associated to them?

As we can observe in Table 5.6, in the column *Generic (G), Specific (S), or Both (B)?*, SIM allows to identify not only generic issues but also specific issues related to the domain to which it is applied. Most of the studies found generic issues and **11** out of **27** studies, which represent **41%**, also reported finding specific issues. This indicates that although the method does not need to be adapted to specific domains, it still allows evaluators to identify problems that are specific to them. We next illustrate some of the specific issues mentioned[9].

In collaborative systems' domain, in the studies done by **S1** [de Souza et al., 2004; da Silva et al., 2003] the authors found issues related to *awareness* which is the system's ability to convey information about who, what, and where users are working in the system. While inspecting YahooGroups chat they found that a user could tell the system what he is doing, but the other members do not promptly receive this information. In the same domain **S8** [Mattos et al., 2009] not only found issues related to *awareness*, but also to *feedthrough* (i.e., provide information to the users about the actions taken by other users). While inspecting Google Docs the authors found that when a group of users is editing the same document the system informs who is present in the edition, however the system does not inform in which part of the document each user is working on. Thus two users could be editing the same part of the document without knowing it.

We can also find specific issues related to other domains such as in the educational systems' domain, more specifically simulation games. In the study done by **S14** [Peixoto et al., 2010] the authors found issues in a Software Engineering simulation game (i.e., SimSE) related to *informative feedback* (information about the effects of the player's action) and *performance feedback* (information regarding how close players are to achieving their goal). The game does not properly provide those types of feedback while the user is playing.

Another example was found in the Human-Robot Interface domain. In the study done by **S15** [Bento, 2010] authors reported issues regarding *robot movement signs*. In one of the interfaces examined authors identified a breakdown regarding how linear and angular velocities of the robot movement were communicated to users.

Finally, in the games domain, **S25** [Coutinho et al., 2011] reported issues related,

[9]It is noteworthy that the issues illustrated may not (currently) exist, because the systems and their interfaces may have changed since the studies were done. However, what we want to illustrate here is that SIM allowed identifying them.

for instance, to how *sound volume* (gives clues about how far the emission source is from the players character) and *sound distribution* (how the sound is split among stereo channels to simulate the direction of its source relative to players character) in a computer game affect deaf users performance.

These examples illustrate how using SIM to analyze the designer to user communication can result in the identification of breakdowns specific to a domain.

5.3.5 [IQ1] What other properties of interactive systems can SIM generate indicators of in addition to communicability?

The goal of SIM is to inspect the designer to user meta-communication, in other words appreciate a system's communicability. During the analysis of the publications in this research, an interesting outcome was to notice that some of these papers explicitly discussed aspects regarding other system properties (besides communicability). As shown in Table 5.6, in the column *Property Evaluated*, papers reported results regarding:

- **Sociability:** studies **S1** [de Souza et al., 2004; da Silva et al., 2003], **S4** [Leitão et al., 2007], **S7** [Moura et al., 2008] and **S24** [Barbosa et al., 2011b,a];
- **Accessibility:** studies **S19** [Abreu, 2010][10], **S21** [Brandão et al., 2010], **S25** [Coutinho et al., 2011] and **S26** [Capelão et al., 2011];
- **Privacy:** study **S22** [Coopamootoo and Ashenden, 2011].

In other words, by analyzing the communicability evaluators were also able to identify aspects conveyed by designers regarding aspects related to other properties, namely sociability, accessibility and privacy.

It is noteworthy that being able to identify problems related to other properties sure is an interesting feature of the method; however, this information should be used with caution. The goal of SIM is to identify communicability breakdowns and it was developed for this purpose. The fact that SIM allows other properties to be encountered does not mean that it is the best method to do so. For instance, using SIM to find accessibility problems may end up frustrating the evaluator, because SIM may be weaker to identify accessibility problems than other methods specific to accessibility. Therefore, the evaluator when selecting SIM must take this issue into account.

[10] The accessibility analysis carried out in this work occurred in a very specific domain of educational games for deaf children.

5.3.6 [IQ2] Which SIM approaches (i.e., technical or scientific) have been applied in these domains?

Table 5.6, in the column *Technical or Scientific?*, presents the approach that has been used in the evaluation of each study and Table 5.8 presents a summary. We can notice that in total there have been **21** scientific applications of the method and **28** technical applications.

This indicates that although the scientific approach has been more recently presented (2009) [de Souza and Leitão, 2009] it already counts with a reasonable amount of applications.

Table 5.8. Approaches used by domain.

Domain	Scientific	Technical
Collaborative Systems (CoSys)	5	11
Document Editors	0	3
Educational Systems (EduSys)	4	6
Games	5	1
Human-Robot Interface (HRI)	2	0
Geographic Information Systems (GIS)	0	3
Miscellaneous	5	4
TOTAL	**21**	**28**

Another important observation is that every application of the scientific approach also involves all the steps that comprise the technical approach. However, since the focus of the inspection changes (interface improvement vs. knowledge advancement) the difference may be in what aspects of the system are registered. At any rate, every scientific application also potentially entails in a technical application. Thus, we can say that independently of which approach has been applied to a specific domain we can conclude that the method can be applied to that domain.

5.4 Discussion

In this subchapter we discuss the results obtained in this study according to the main research question (RQ) being investigated:

[RQ1] What is the applicability of Semiotic Inspection Method (SIM)?

In order to investigate this question we have broken it into different aspects that have been analyzed individually. We have examined the domains in which SIM has been applied, as well as the needs for adaptation, challenges faced in the application, type of breakdowns identified or issues discussed and which approach of the method was applied.

We have found that SIM has been frequently applied to collaborative systems and educational domains. It has also been applied to GIS, HRI and in a number of other different domains, as different as audio edition and digital libraries. Thus, it has clearly been applied to a broad set of domains that deal with very different specific issues. The detailed analysis of the application of SIM has shown that the method does not require adaptation in each domain, and is able to identify breakdowns or issues which are specific to the domain at hand.

Some of the papers have discussed specific aspects experienced in the application of SIM. One aspect that has been identified regarding its application in different domains is the fact that evaluators have often included in the reconstruction of the designer to user communication items that are specific to the domain. However, we argue that this does not represent an adaptation in the method (which requires the meta-message to be reconstructed), but rather on the content of the message being conveyed.

In some collaborative systems' evaluation it might be necessary to have different viewpoints of the system to be able to explore all static or dynamic signs. For instance, when an action of a group member generates changes in another's members interface. In this case, evaluators might have to perform an analysis for each of the different role members being considered. This is not different from considering distinct user profiles for single user systems. However, the evaluator may probably have to take into account how each member's interaction may impact the others. This means that the analysis of the system behavior (i.e., dynamic signs) will probably be more complex when compared to single users. Once again since SIM requires that an analysis of the dynamic signs be performed, but does not define how to proceed to perform it, these are considerations to be taken into account, but not changes in the method.

Another aspect is related to the redefinition of class of signs representation to evaluate the communicability of audio sounds. We have also argued that this does not characterize an adaptation since the evaluator must frame the concepts and materials to the context in which the evaluation will occur. Therefore, the interpretation of metalinguistic, static and dynamic signs is part of the preparation step of the method.

Based on these findings we can conclude that the evidences collected strongly support the claim that SIM can be considered a domain independent method. Regarding its independency of technology, the method has been applied to web based systems, desktops and HRI. Although the method has been applied only to three different technologies, the fact that they have varied as widely as from web systems to HRI is an evidence of how independent it is.

Our findings point to a great benefit of SIM, especially when considering how quickly technology evolves and how modern societies have widened the domains and situations in which technology is adopted. Furthermore, this is not a feature that most methods present. For instance, Heuristic Evaluation (HE), which is probably one of the most popular evaluation methods, requires new heuristics to be proposed for each new technology or domain, such as mobile computing [Bertini et al., 2006] or collaborative systems [Baker et al., 2001].

We have also found that SIM can be used to extend the interface evaluation to analyze other properties of interactive systems, besides communicability, such as sociability, accessibility and privacy. The reason for this is its theoretical basis and focus on communication. By analyzing the designer to users message the method allows evaluators to examine not only the quality of this communication, but also the content regarding specific issues such as how social relations are encouraged, what is being conveyed to users with special needs and how privacy issues area being communicated to users.

5.5 Limitations of this Assessment

The use of a systematic methodology itself helps to reduce problems in the selection and analysis of the works if compared to ad-hoc methodologies. However, even though this SLR has been supported by a pre-defined study protocol, it has some limitations. We will discuss these limitations in terms of *Construct Validity*, *Reliability*, *Internal Validity* and *External Validity*. *Construct Validity* reflects to what extent the study phenomenon represents what the researchers have in mind and what is investigated. *Reliability* focuses on whether the data collected and the analysis are conducted in a way that it can be repeated by other researchers with similar results [Engström et al., 2010]. *Internal Validity* concerns to the extent to which the design and conduct of the study are likely to prevent systematic error. *External Validity* concerns to the extent to which the effects observed in the study are applicable outside of the study [Kitchenham, 2004].

5.5.1 Construct Validity

5.5.1.1 Terminology

Since the search for primary studies is based on the search string, each SLR is likely to miss relevant studies if this string is not properly chosen. Therefore, in our research, there is a risk of missing relevant papers since we defined a fixed expression as a search string (i.e., "semiotic inspection" or "inspeção semiótica" or "inspección semiótica") and studies that has the words in different order or separated would not be selected. Nevertheless, we think that the risk of construct validity threat were minimum because we aimed at finding studies that used the Semiotic Inspection Method and the fixed expression used as our research string is part of the name of the method.

5.5.1.2 Completeness of the Selected Studies

By basing the review on a fixed set of repositories, as mentioned previously, we excluded certain types of publications or work published through other channels different from the defined set of repositories. We can therefore not claim to have included all relevant publications. However, the selection of the repositories to search for publications was based on the knowledge of an HCI and Semiotic Engineering expert that monitored the review and guided the definition of the repositories. The repositories selected are the main databases (or at least the ones that include the main journals and conference proceedings) of the HCI field (from the Computer Science perspective); databases that are known to have published the greatest amount of works done with SIM and that would allow us to undertake a study with high quality and relevance.

5.5.2 Reliability

Since we followed Kitchenham [2004] procedures to conduct a SLR, we believe that the reliability threats were minimized. These procedures request that we define a research question, the selection process (the steps to select the studies), inclusion/exclusion criteria and quality criteria. Although these definitions allowed conducting a research in a systematic way reducing reliability threats, the adoption of systematic procedures itself does not guarantee reliability. Therefore, we will discuss Reliability in terms of classification of studies.

5.5.2.1 Classification of Studies

In the steps 1 and 2 we classified the studies based on their titles and abstracts. This is sometimes hard to do since the titles and abstracts do not always give enough information to make a decision of including or excluding the paper from the review. In order to minimize the probability of errors and bias in the results some precautions were taken. The main precaution was: if at some step there were doubts about the exclusion of a work, the work was included to be checked at a subsequent step. For example, if during the first step, reading the title, we had doubts whether the work should really be excluded it was kept to be analyzed in the second step – reading the abstract.

5.5.3 Internal Validity

To address this issue we will discuss the risk of author bias and manual search.

5.5.3.1 Author Bias

Since our research was done manually there is a potential problem of author bias. Moreover, only one reviewer participated in the SLR review process. These problems can increase the risk of internal validity threat. Our countermeasures taken to reduce this were related to an expert in HCI and Semiotic Engineering that supported the research by reviewing the selection of the works and analysis of results (as suggested by Kitchenham [2004]). These countermeasures decreased the risk of threat because the expert that supported the study is one of the co-authors of the method that is being studied and knows most of the studies done so far using the method (and could help reviewing the analysis done).

5.5.3.2 Manual Search in the Repositories

We searched four repositories manually, which can increase the risk of threat to internal validity: (1) SERG (Semiotic Engineering Research Group), (2) Digital Library of Informatics Department of PUC-Rio, (3) Computer Science Department of UFMG and (4) Brazilian Symposium on Human Factors in Computing Systems (IHC).

In the repository (2) we searched for the words “semiotic” or “inspection” (in English and Portuguese) in the title and abstract of the studies. This brings some limitations to our research because some studies may not have these words in the title or abstract and still be relevant to our review. However, we decided for this approach since the cost-benefit of reviewing the full-text of all technical reports, dissertations and thesis of this repository was low. In total we would have to review more than 800 studies in this repository which have studies from all the different areas in Computer Science.

In the repository (3) we had a similar situation of the repository previously mentioned. Since the Computer Science Department of UFMG has researchers working in different areas, we limited our scope to review only the dissertations done in the HCI field, searched by the name of the professor of the institution that works with HCI and Semiotic Engineering. There are no limitations in the search of the repositories (1) and (4) since we reviewed all the published studies in these repositories.

5.5.4 External Validity

The conclusion presented in this paper can only be generalized for the sample (set of studies analyzed) we evaluated, in other words, our conclusion regarding the method can be generalized only for the domains mentioned. The scope of our SLR cannot be generalized for other domains since we cannot guarantee that we selected all the studies in the area and we cannot state that the domains we found in this study in which SIM was applied are enough to generalize the results.

Chapter 6

Assessment from Evaluators' Perspective

Although SIM has been used in different domains since its proposal, a study of its capabilities and limitations has not yet been performed. Thus, the second assessment of the Semiotic Inspection Method (SIM) aimed at understanding its costs, benefits, advantages and disadvantages from the evaluators' perspective. The reason for having performed this step by obtaining information from the evaluators' perspective is due to the fact that the evaluators are responsible for deciding which method to use for an interactive system evaluation. Therefore, gathering information about the costs and benefits from the evaluators' perspective may be important in defining (or revising) a strategy to present and/or teach the method, and also conduct deeper assessments of the method.

In order to assess the evaluators' perspective, we applied a questionnaire to novice evaluators and made interviews with the authors of the method (representing the expert evaluators view) aiming at comparing, contrasting and consolidating their (i.e., novices and experts) perception about the costs, benefits, advantages and disadvantages of SIM. An analysis of the responses shows interesting insights and characteristics of the method.

Before presenting how this assessment was conducted and the results obtained we present the Grounded Theory [Strauss and Corbin, 1998] – a well-known interpretative method used in qualitative researches – since some of its techniques were used in the analysis.

6.1 Grounded Theory

As a formal methodology Grounded Theory was first presented by Glaser and Strauss in their 1967 book *The Discovery of Grounded Theory*. According to [Strauss and Corbin, 1998, p.24] Grounded Theory is a qualitative research methodology that uses a *"systematic set of procedures to develop an inductively derived grounded theory about a phenomenon"*. In other words, this theory was conceived as a methodology for developing theory based on data that are systematically collected and analyzed [Goulding, 2001]. It is necessary to clarify what Glaser and Strauss [1967] understood by theory. According to the authors theories can be: formal or substantive. The former is composed of what the authors call the "big" theories, conceptual and broad, while the latter refers to explanations for everyday situations and are, therefore, simpler and more accessible.

Over the years the authors of the method, Glaser and Strauss, did not come into agreement on the objectives, principles and procedures associated with the implementation of the method. The branching of the method was marked by the publication of the book *Basics of Qualitative Research: Grounded Theory Procedures and Techniques* by Strauss and Corbin [1998]. The differences are not only in style and terminology. The Strauss version of the method was redesigned to incorporate a rigorous and complex process of systematic coding, but, although there are different perceptions about the theory, the fundamental principles remained the same [Goulding, 2001; Egan, 2002] and will be presented next.

The first step of the process is to *identify the area of interest* [Fernandes, 2002]. Usually researchers adopt the grounded theory when the topic of interest has been relatively ignored or have only been given superficial attention in the literature. Therefore, the goal of the researchers is to build the theory from the ground [Goulding, 2001].

Data collection can be performed from different sources; this includes, but is not limited to: interviews, observations and life experiences. As data is collected it should be *analyzed simultaneously looking at all possible interpretations*. This particularly involves coding procedures that usually begin with Open Coding. Open Coding is the process where data is broken into discrete units of meanings [Glaser and Strauss, 1967] and then analyzed, compared, conceptualized and categorized [Strauss and Corbin, 1998]. This process begins with a complete transcript of an interview, or other source of data, where each line of text is analyzed to identify keywords or phrases that connect the information to the one being investigated. Besides the process of open coding, it is important to incorporate the use of memos (notes that are performed immediately after data collection as a way to document the researcher's impression and describe the situation) [Goulding, 2001].

Another feature of the method is related to a *sampling of informants.* Initially, the researcher goes to the most obvious places and looks for informants who are most likely to provide early information. However, as the theory evolves, other individuals, situations and places may be needed to strengthen the results [Goulding, 2001]. This process is called Theoretical Sampling, which, according to [Glaser and Strauss, 1967, p.45], *"is the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyses his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges".*

Beyond the Theoretical Sampling, a key feature of grounded theory is the *constant comparative method*, which involves the comparison of equals, in order to find patterns and themes [Goulding, 2001]. This process facilitates the identification of concepts, and to perform it, it is necessary the use of a technique commonly called *Axial Coding*, which takes place at a later stage than the Open Coding and involves the process of rearranging the data already categorized based on the establishment of links between categories [Fernandes, 2002]. Axial Coding is the *appreciation of concepts in terms of their interrelations.* In turn, once a concept has been identified, its attributes can be explored in greater depth, and their characteristics can be dimensioned in terms of its intensity or weakness [Goulding, 2001].

Finally the data is gathered in a central category, and the researcher has to justify as the basis for the emerging theory. The central category brings together all the strands in order to offer an explanation of the behavior under study. It has theoretical significance and development should be traceable through the data [Goulding, 2001]. However, one theory is usually considered valid only if the researcher has reached the saturation point. This involves staying in the field until there are no new data from subsequent data.

This research aimed at outlining the advantages and disadvantages of SIM from the evaluators' perspective. That is, our goal was not to generate a theory. Although the purpose of Grounded Theory is theory building, its use does not need to be restricted to researchers who have this goal. According to Strauss and Corbin [1998], *"the researcher can use some, but not all, of the procedures to satisfy his or her research purposes."* [Strauss and Corbin, 1998, p.288]. Therefore, some of the techniques proposed by Strauss and Corbin [1998] were used for the analysis of results. In our analysis we were concerned with reading all the responses, and identifying, naming, categorizing and describing natural groupings of phenomena found in the text, thus we used especially Open Coding and Axial Coding techniques in our research.

6.2 The Assessment

The assessment from the evaluator's perspective was conducted in two steps. In the first step we applied a questionnaire to students and researchers who had used SIM at least once with the goal of obtaining initial information on the characteristics of the method from novice evaluators' perspective. In the second step we conducted interviews with the authors of the method. The reason for having done the interviews with the authors was due to the fact that they are the most experienced people with the method and, therefore, represent the expert evaluators' perspective. The overall goal of this assessment was to compare, contrast and consolidate their (i.e., novices and experts) perception about the costs, benefits, advantages and disadvantages of SIM.

6.2.1 The Survey

The survey was conducted through a questionnaire containing **31** questions divided between multiple-choice and open-ended questions. Multiple-choice questions were mandatory and open-ended questions were optional. The questions were divided according to their focus as follows:

- **Participants' profile:** 10 questions regarding general information (e.g., gender or age), training and professional experience;
- **Learning:** 2 questions to identify the courses in which participants had learned SIM;
- **Experience with SIM:** 15 questions about participant's experience in applying SIM, and the challenges experienced in learning and applying the method; and
- **Experience with HCI:** 4 questions about participants' experience using other HCI evaluation methods.

The reason for using a questionnaire was that it would be possible to reach a larger number of people, including evaluators who authors of this study did not know of. The questions were designed to allow us to collect information about SIM, as well as to have an overview of the general evaluation experience of the participants. Regarding SIM the idea was to see which steps were considered difficult (if any) in the application and analysis of the method, and also collect the participants' opinions regarding advantages and disadvantages and any other aspects they considered relevant about the method (through open-ended questions).

The questionnaire was applied from June 29^{th}, 2011 to July 29^{th}, 2011 and distributed by e-mail to: (1) the Brazil national HCI e-mail list, since to the best of our knowledge, at the time of the research, all publications regarding SIM and its use involved at least one Brazilian researcher[1]; and (2) HCI researchers and professors who were known to have worked with the Semiotic Engineering theory asking them to distribute the questionnaire to their students and other people who they knew had applied SIM. It is noteworthy that the authors of the method were asked not to answer the questionnaire. The questionnaire was answered by 25 participants. We considered a good number of responses since SIM is a relatively recent method, and it is not widely used or taught (yet). Besides, although the number of responses is small, we believe that this initial research is necessary to raise the main issues regarding the method to be further investigated using deeper approaches (e.g., empirical assessments).

6.2.1.1 Participants Profile

Among participants, **12** were female and **13** male. Their age varied from 19 to 40, but most (**16**) were between 19 and 25 years old. Professionally[2], as shown in Figure 6.1, more than half of the respondents were students (**16**), **11** were professionals in the Information Technology (IT) field; **4** work as researchers; and **1** is a professor at a university. Only **8** participants have a professional experience (i.e., working in industry as professionals or interns) in the HCI field. Regarding their highest educational level **15** are undergraduate students or have a university degree; **8** have a MSc degree or is a MSc candidate; and **2** are PhD candidates.

Participants mostly – **16** – have their current education level in Computer Science, **8** in Information Systems and only **1** person said that was studying both Information Systems and Computer Science. Participants are from different places, where most of them – **15** – studied or is attending a course at the *Universidade Federal de Minas Gerais*. The other participants are from Brazilian universities in other states, such as *Paraná* and *Rio Grande do Norte*, or abroad.

Graduate students (MSc and PhD) were asked to inform their research area[3]: **8** participants do research in HCI field and **2** in Software Engineering. Although not

[1]At the time this research was conducted all the studies with the method had the participation of at least one Brazilian. However, as shown in the previous section, when updating the systematic literature review we discovered a publication by a group not from Brazil in which the method was applied [Coopamootoo and Ashenden, 2011], which can be taken as an evidence of a broader adoption of the method.

[2]Participants could choose more than one option.

[3]Participants mentioned more than one research area.

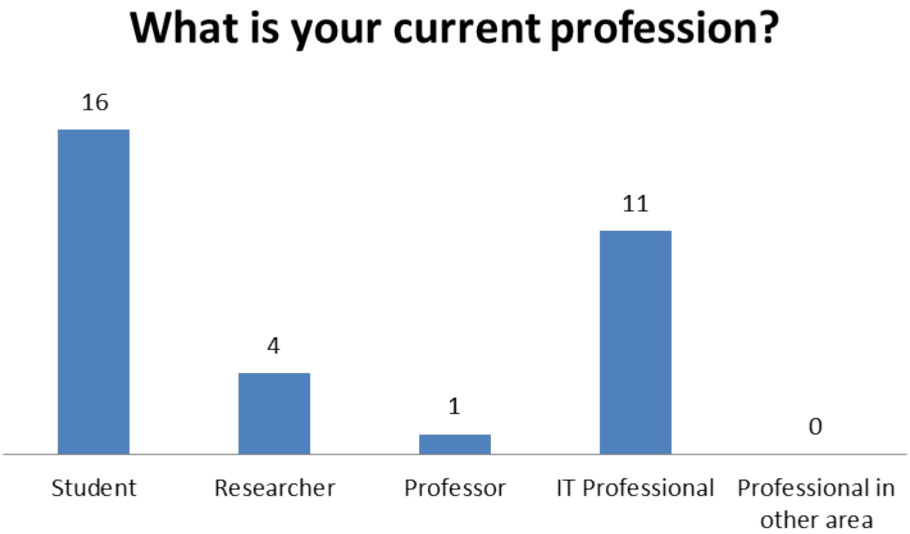

Figure 6.1. Participant's current profession.

requested, some participants also mentioned their research sub-area, which are Learning Support Systems, Learning Objects and Games.

Among participants, **5** used SIM once; **16** used SIM twice; **2** used SIM three times; and **2** used SIM four or more times (Figure 6.2). All participants had also applied other evaluation methods (Figure 6.3): **23** applied Communicability Evaluation Method (another Semiotic Engineering evaluation method); **22** applied Heuristic Evaluation; **18** applied Usability Test; **5** applied Cognitive Walkthrough; and **4** applied Think Aloud Protocol. In this question participants could not mention other methods. It is noteworthy that most participants (**24**) had experience with these methods using them in a course context; **12** used for research purposes; and **7** applied them professionally.

6.2.1.2 Results and Analysis

The analysis of the questionnaire (as well as the analysis of the interview) aimed at identifying the evaluators' perception about SIM's costs, benefits, advantages and disadvantages. The analysis performed is of a qualitative nature. The reason for conducting a qualitative analysis is that the goal of the research was not to get to a conclusion regarding the cost-effectiveness of the method, but rather understand, compare, contrast and consolidate the main difficulties and issues being experienced by evaluators applying the method (i.e., raise their perception on the main costs and benefits of the method).

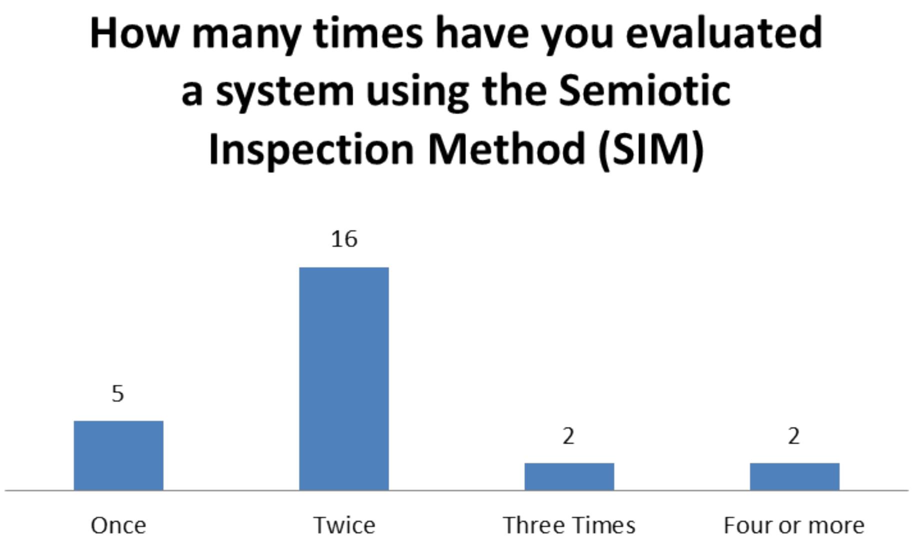

Figure 6.2. Number of application of SIM done by participants.

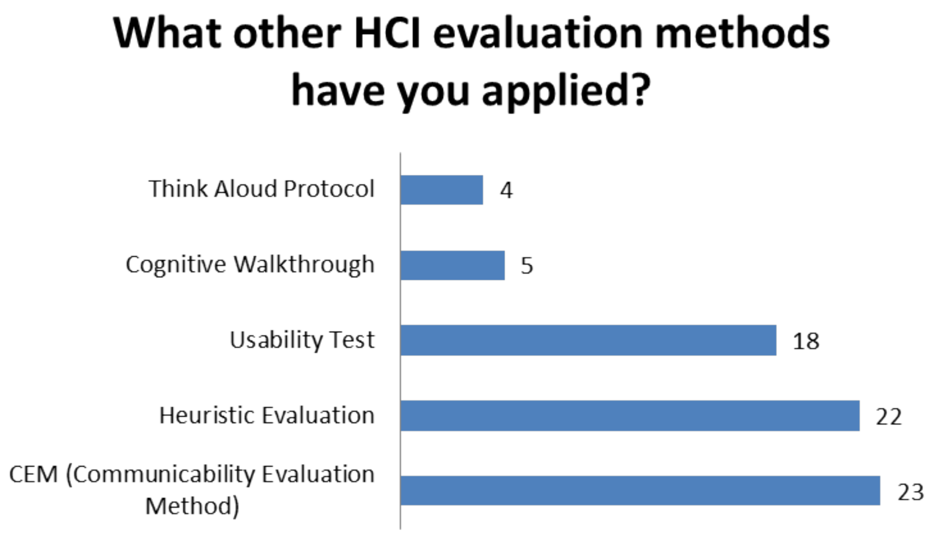

Figure 6.3. Methods applied by participants at least once.

Costs and Effectiveness Analysis The questionnaire participants assessed the difficulty in applying SIM on a scale from 1 to 5 (with 1 being very easy and 5 very difficult). Most participants evaluated the method as Medium or Difficult to apply (see Figure 6.4). Participants were also asked about which step they believed were the most difficult (if any). Among the steps of the method, the fourth step was the one that participants found the hardest (**48%**) . The fourth step is the one in which the evaluator compares the meta-messages obtained by the analysis of each sign class (steps 1 to 3), consolidating them in a single meta-message and identify inconsistencies among them. If evaluators are not able to perform well the fourth step, they may not be able to identify relevant problems that could outcome from contrasting the meta-message generated from the analysis of each sign class.

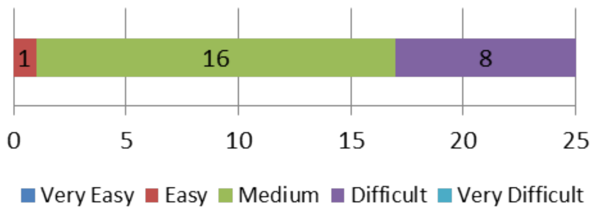

Figure 6.4. Difficulty to apply SIM.

They were also asked about what was their greatest difficulty in applying the method (if any): difficulty to learn the method; difficulty to understand the theory that underlies the method; or difficulty related to time and concentration required to conduct a good evaluation. They also had the option to answer that they had no difficulty or mention another difficulty they might have experienced. All available choices were selected by at least 1 participant. However, two of them stood out significantly: (1) time and concentration required; and (2) learning the theory. Almost half the participants indicated that their major difficulty was related to the time and concentration required. This is an indication that for these evaluators the time and concentration needed in the evaluation is a cost to be considered. However, by looking at the profile of the evaluators who chose this option, we notice that the more experienced SIM evaluators (i.e., applied SIM more than 4 times), which were two participants, did not point this option as a problem. They mentioned to have no difficulty with SIM. Thus, this could be an evidence that as evaluators become more experienced with SIM, they either take less time to apply it or make a better use of the time spent.

About one third of the participants indicated that the underlying theory is their biggest challenge in applying the method. Theory based methods usually require an understanding of the underlying theory, so this could be expected. The good news is that once evaluators learn the theory they would not face this challenge any longer. Furthermore, learning the theory could be motivated by reasons other than just applying SIM. As mentioned, participants who applied SIM more than 4 times mentioned not having difficulties with the method, and this reinforces our argument that the difficulties decreases as the evaluators become more experienced.

Due to the requirement that understanding the theory is necessary to apply the method (see Figure 6.5), participants were asked how important they felt the knowledge of Semiotic Engineering theory was in the ability to apply SIM (on a scale from 1 to 3, with 1 being low and 3 high). All participants believed the theory was necessary or useful in the application of the method. The great majority (over **70%**) believed that the theory was essential in achieving good results, whereas the others thought it was not essential, but could help. In analyzing this result combined with the previous question, we notice that all of them understand the importance of knowing Semiotic Engineering has on the method, but do not consider learning the theory the major cost of applying SIM.

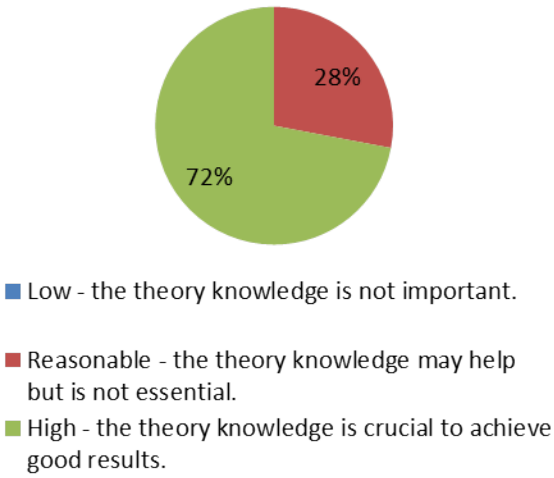

Figure 6.5. Participants' perception regarding the importance of the Semiotic Engineering theory to apply SIM.

Regarding the effectiveness of using the method, participants were asked how they evaluated the relevance of the problems identified using the method. They had three options and could select one or more of them: SIM helped in identifying problems that had no impact in the system's use; SIM helped to formalize problems that had been identified before; and SIM helped in identifying relevant problems. Almost all participants (**92%**) answered that SIM helped finding relevant problems. In addition, **72%** believed that SIM also allowed the formalization of known problems. None of them reported that the method could only find problems that did not impact the system's use.

In conclusion, we can mention that in the evaluators' perspective the costs of SIM are mainly related to (1) high time/effort demanded in applying the method; and (2) knowledge and skills needed to do so, since knowledge of Semiotic Engineering theory is required. Regarding the effectiveness of SIM, we can conclude that it is an effective method since it allows identifying relevant problems of the system being evaluated and formalizing perceived problems.

Advantages and Disadvantages Analysis We now present the advantages and disadvantages of SIM from the novices' evaluators' perspective. This information was identified based on five optional open-ended questions. The participants were asked to:

- comment whether they would or not apply SIM in other interface evaluations (22 responses);
- list the main advantages of SIM (23 responses);
- list the main disadvantages of SIM (21 responses);
- list SIM's costs and benefits related to other methods (20 responses); and
- make any other comments (5 responses).

We used Grounded Theory [Strauss and Corbin, 1998] techniques to analyze the responses. We present in Table 6.1 and Table 6.2, respectively, a summary of the advantages and disadvantages encountered. We also present in the column "#" the number of participants that mentioned it. As mentioned earlier, the analysis is of a qualitative nature and the reason for presenting the number of participants who reported an advantage/disadvantage is only to illustrate which issues had more or less agreement upon.

Analyzing Table 6.1 we can see that according to participants, the major advantage of SIM is that it allows identifying problems related to communicability **[A1]**. Although communicability evaluation is SIM's main goal, it shows that participants understand it as a relevant property of an interactive system. Also in **[A2]** participants list as an advantage the fact that the method can be used to analyze the impact of each sign class independently. This advantage may be directly related to **[A9]** in which a participant states that SIM allows a thorough analysis of the system, even though participants themselves did not indicate this relationship.

Another advantage of SIM is associated with the fact that it allows formalizing, arguing and explaining the problems encountered **[A3]**. We understand that any evaluation method allows evaluators to explain the problems encountered. However, one participant (P20) has explicitly stated that:

Table 6.1. SIM advantages.

	Advantages	**Participants**	#
A1	Identifies problems related to communicability.	P5; P15; P16; P22; P24; P25	6
A2	Allows the analysis of the impact of each sign class (i.e., metalinguistic, static and dynamic) independently.	P2; P9; P12; P18; P20	5
A3	Allows formalizing, arguing and explaining the problems encountered while interacting with the system.	P1; P4; P20	3
A4	It has a good cost-benefit relation (i.e., good results at a relatively low cost).	P4; P6; P17	3
A5	Only one evaluator is needed to perform an evaluation.	P17; P22; P24	3
A6	It is a theory-based method.	P5; P18; P25	3
A7	It is a simple method to learn/understand and apply.	P6; P11; P17	3
A8	It is a method that produces good results and allows finding important problems.	P3; P7	2
A9	Allows a thorough analysis of the system and identification of its problems.	P1	1
A10	Allows generating new knowledge using a scientific application of SIM.	P7	1
A11	It does not require an expert evaluator to understand the problems reported.	P1	1
A12	It is a method that, unlike other interface evaluation methods, performs an assessment of the help and system's documentation.	P2	1
A13	Allows identifying the impact of problems related to other properties of the system (i.e., accessibility, sociability).	P5	1
A14	Can be used to evaluate system's interface in the beginning of the development process.	P10	1
A15	Allows evaluating the system from evaluators' point of view (i.e., meta-message emission).	P25	1
A16	It can be applied to different domains and technologies without adaptations.	P5	1
A17	It is a low cost method (i.e., does not demand too much time and effort).	P22	1
A18	It is a good guide to find problems (i.e., provide procedures that guide the evaluator during the evaluation).	P17	1

> *"To formalize analysis (getting more concrete arguments against the problems under consideration); as well as helping to define more precisely where the interaction and communicability problem of the system is (it is possible to define if a problem occurred because of a particular static sign and not by others, for example)."*[4]

This statement shows the advantages related to how SIM leads evaluators to describe the problems due to its steps and to the underlying theory. Thus, **[A3]** could be closely related to the advantage **[A6]** in which participants stated that an advantage is being a theory-based method, even though they did not explain why they believed this was an advantage. The fact that participants have perceived through the application of the method that it has benefits due to its theoretical basis indicates that participants' views strengthen the researchers' claims that theoretical approaches in HCI are needed [Carroll, 2003; de Souza, 2005]. In addition, the advantage **[A6]** may also be related to **[A10]** where participants think that the possibility of generating new knowledge using a scientific application of SIM is an advantage.

The participants mentioned that SIM does not require an expert evaluator to understand the problems reported **[A11]**. This advantage could be argued to apply to any method, however, since SIM requires the learning of Semiotic Engineering theory in order to apply the method, the fact that it is possible to generate a report that does not necessarily require this knowledge to be understood by the reader could be perceived as an advantage.

At last, the participants mentioned the applicability of SIM to different domains as an advantage of the method **[A16]**. Their view is aligned with the investigation presented in the previous chapter that showed that SIM can be applied to different domains and technologies without adaptations [Reis and Prates, 2011].

The biggest drawback raised by participants (shown in Table 6.2) is that SIM is a method that demands a lot of time and effort to be applied **[D1]**. This disadvantage is related to **[D6]** where participants stated that SIM requires the generation of an extensive and detailed documentation which increases the evaluation time, and also to **[D8]** where participants stated that the method is laborious, repetitive and tiring at times which may cause the evaluator to overlook some problems. These three disadvantages may be some of the factors that explain one of the main costs identified in the closed questions: time and concentration required by the method. Also, they could

[4]Original statement in Portuguese by P20: *"Formalizar análises (obtendo argumentos mais concretos contra os problemas analisados); além de ajudar a definir com mais exatidão onde é o problema de interação e comunicabilidade do sistema (é possível definir que um problema ocorre por causa de um determinado signo estático e não por outros, por exemplo)".*

justify the need to offer an evaluation supporting tool. The inexistence of such tool has also been identified as a disadvantage **[D5]**.

Table 6.2. SIM disadvantages.

	Disadvantages	**Participants**	#
D1	Demands a lot of time and effort to apply it.	P5; P7; P10; P12; P20; P23	6
D2	It is based on a complex theory.	P2; P5; P9; P16; P17; P18	6
D3	It is a difficult method to apply with a high learning curve.	P2; P4; P14; P22; P25	5
D4	The experience of the evaluator is very important to get good results. Novices may have difficulties and may not generate good results.	P3; P17; P23	3
D5	There is no tool available to support the evaluation.	P1; P5; P18	3
D6	It is a method which requires the generation of an extensive and detailed documentation, which increases the evaluation time.	P1; P7	2
D7	Lack of material to support learning, such as case-study examples on how to apply the method.	P2; P18	2
D8	It is a laborious and repetitive method, and is tiring at times. This may cause the evaluator to overlook some problems.	P11; P20	2
D9	It has a low cost-benefit relation.	P7	1
D10	Lack of experts, which makes difficult to arrange a team to apply the method.	P12	1

In addition, participants said that SIM is a method based on a complex theory **[D2]**. Although participants did not make associations between disadvantages, we can notice that this could probably explain the disadvantage that states that it is a difficult method with a high learning curve **[D3]**. A participant's (P9) statement *"Difficult, too theoretical."*[5] explicitly creates the link between the two. Again these disadvantages reinforce the cost regarding knowledge and skills necessary to apply the method that was mentioned previously.

Regarding learning the method, most of the participants said to have learned it through classes and educational materials. However, in **[D7]** participants mentioned the lack of material to support learning the method as a disadvantage. This probably indicates that the existing materials are not sufficiently thorough or complete. The

[5] Original statement in Portuguese by P9: *"Difícil, muito teórico."*

deficiency of the existing material could also increase the cost of learning the method, pointed by participants.

Another disadvantage mentioned by the participants is related to the influence of evaluators' expertise in the results **[D4]**. Although no study of the kind has yet been performed for SIM, we would argue that expertise could influence positively the results, especially taking into account that the knowledge of the theory has already been identified as necessary in applying the method and getting good results. Furthermore, a participant cited that there is a lack of experts in the method, which makes it difficult to arrange a team to apply the method **[D10]**. The participant did not offer any comments regarding the context in which he experienced these difficulties. At any rate, SIM is a recent method and it requires time before any method could have a large number of experts. This disadvantage also justifies the need of studies to assess the method such as the ones being presented in this work which allows people to better understand its costs and benefits before investing in learning.

By contrasting the results found, we notice that participants do not always agree on the advantages and disadvantages of the method. This is the case, of the advantage stated by some participants that the method is easy to learn **[A7]** and the disadvantage that it has a high learning curve **[D3]**, as well as the cost of the required knowledge to apply the method. By looking at the answers of those participants who mentioned **[A7]**, they were consistent along the questionnaire in answering that for them SIM is not difficult. This could be a positive result, since it may indicate that for some people the method is easy. However, a deeper investigation is needed to identify if the causes of these differences are personal learning-style, or due to their experience with the method, for instance evaluating a simple versus complex system.

The second contradiction was regarding the cost of the method (among **[A17]**, **[D1]** and the analysis of the closed questions presented previously). Only one participant (P22) mentioned that SIM does not demand too much time and effort:

> *"It is a very good method to analyze the communicability of a system and it requires less time and resources compared to others."*[6]

The answers given by the participant were consistent with this view along the questionnaire. In the question in which participants were asked what was their greatest difficulty in applying the method he answered that he had difficulty to understand the theory that underlies the method and did not check the option about the difficulty

[6]Original statement in Portuguese by P22: *"É um método muito bom na análise da comunicabilidade de um sistema e que demanda pouco tempo e recursos comparados a outros mais."*

related to time and concentration required. Notice that in his/her answer he/she mentioned that SIM requires less time compared to other methods. However, he did not mention which methods he was comparing SIM to, but indicated in this questionnaire that he had experience with HE and CEM. This contradiction shows that further investigation of the cost associated with the time/effort is needed, perhaps involving the measurement of the time spent by different users' profiles.

Finally, the third contradiction occurred between **[A4]** and **[D9]** regarding SIM's cost-benefits. Only one participant (P7) mentioned that SIM has a low cost-benefit ratio **[D9]** and he stated the following:

> *"I believe that the communicability evaluation of a system could be made somewhat more superficial, placing greater emphasis on other qualities of use such as usability. After all, a system may have high communicability and still have low usability. Therefore, by being a method that evaluates only the communication, in particular, only the communication emission, I believe that the effort to apply it is not justified."*[7]

We can argue that communicability is a property that was proposed to add to the HCI field and never intended to substitute usability [de Souza, 2005]. In other words, they are two distinct properties that ideally should be combined to achieve a better quality of use in interactive systems. Therefore, the fact it does not evaluate properties that it was not meant to evaluate should probably not be considered a cost of the method. We understand that in real contexts evaluators may not be able to conduct more than one evaluation due to available resources, and will have to take into account whether communicability or usability is more relevant to the system being evaluated. However, deciding the goal of the evaluation and the specific questions to be answered are steps to be taken before deciding the method to apply, as has been described in the DECIDE framework [Preece et al., 2007]. Furthermore, although we could infer that participant P7 may not think that the communicability property is as relevant as usability; other participants have mentioned being able to evaluate communicability issues as an advantage **[A1]** and that it has a good cost-benefit relation **[A4]**. In addition, participants have also mentioned that SIM produces good results and allows the identification of relevant problems **[A7]**.

[7]Original statement in Portuguese by P7: *"Acredito que a avaliação da Comunicabilidade de um sistema poderia ser realizada de forma um pouco mais superficial, dando maior ênfase a outras qualidades de uso, como a Usabilidade. Afinal, um sistema pode ter alta Comunicabilidade e ainda assim ter baixa Usabilidade. Portanto, o MIS sendo um método que avalia apenas a comunicação, em especial, apenas a emissão da comunicação, considero que o esforço dispendido para aplicá-lo não é justificado."*

It is noteworthy that some issues mentioned by the participants were not specific to SIM, but rather inherent to any inspection-based method. For instance, issues related to the fact that SIM does not involve users in the evaluation. Therefore, we did not include these issues in the Tables 6.1 and 6.2.

Comparative Analysis Although the questionnaire mentioned advantages and disadvantages of SIM specifically, one of the optional open questions asked them how they compared SIM to other methods they had applied. Table 6.3 summarizes the comparisons made by the participants. The column *"Advantage or Disadvantage?"* shows if the item corresponds to an advantage or disadvantage regarding SIM. We also present in the column "#" the number of participants that mentioned it.

Regarding the advantages and disadvantages presented in Table 6.3, done in a comparative way, we can see that some of them (**[C2]**, **[C4]** and **[C9]**) are once again general advantages of inspection-based methods when compared to methods involving user observation. All of them compare SIM, which is an inspection-based method, with CEM, which involves observing users in a controlled environment. As mentioned by Gray and Salzman [1998] a comparison of inspection-based methods with user-based methods must be made with care because they have different purposes. Besides, it is expected that user-based methods require more time, effort and resources than inspection-based methods since they need an infrastructure and users to carry out the evaluations.

The comparison made more frequently by participants is that SIM has a higher cost than other methods **[C1]**. Some participants did not mention which methods they were comparing SIM to, but one mentioned that SIM has a higher cost than Heuristic Evaluation and Usability Testing. This particular participant (P2) had some experience with the three methods (i.e., applied Heuristic Evaluation three times and applied SIM and Usability Testing twice). A comparison between SIM and Heuristic Evaluation, as discussed before, should take into account that they evaluate different properties. However, since there is an overlap of problems that can be identified with both properties [de Souza, 2005], it could be an interesting direction for future work. We would expect SIM to have a higher learning and cost application, but it would be interesting to investigate the nature of problems identified and how they inform a possible redesign of the method [Salgado et al., 2006]. This could be useful in giving a better understanding of cost and benefit of SIM, since most HCI researchers and practitioners have a good knowledge and understanding of HE.

The biggest advantage of SIM compared to other evaluation methods is related to its results **[C3]**. The participants think that SIM generates better results than CEM

Table 6.3. SIM advantages and disadvantages in comparison to other evaluation methods.

	Comparison	**Advantage or Disadvantage**	**Participants**	#
C1	SIM has a higher cost than other methods (e.g., Heuristic Evaluation and Usability Test).	Disadvantage	P2; P11; P15; P20; P24; P25	6
C2	SIM has a lower cost compared to CEM.	Advantage	P4; P11; P17	3
C3	SIM generates better results than CEM and Heuristic Evaluation.	Advantage	P4; P10; P25	3
C4	SIM is easier and simpler to apply than CEM.	Advantage	P3; P19	2
C5	SIM generates better results when applied to knowledge generation.	Advantage	P7; P9	2
C6	SIM generates worse results (i.e., less detailed and effective) than CEM.	Disadvantage	P19; P22	2
C7	SIM is more difficult compared to other evaluation methods (e.g., Heuristic Evaluation).	Disadvantage	P3; P9	2
C8	SIM is not a good alternative for technical evaluations compared to Heuristic Evaluation.	Disadvantage	P7	1
C9	SIM is less appropriate than CEM when we want information from user's perspective.	Disadvantage	P10	1
C10	SIM has a better cost-benefit than other methods.	Advantage	P24	1
C11	SIM requires fewer resources than other evaluation methods.	Advantage	P22	1
C12	SIM requires more theory knowledge than other methods.	Disadvantage	P16	1

and Heuristic Evaluation, for instance:

> *"I found SIM more complete than the heuristic evaluation because it makes a deeper analysis of the relationship between system (as deputy) and user."*[8](P10)

SIM has also been recognized to generate better results when applied to knowledge generation in the HCI field **[C5]**. Again, there were contradictions among evaluators perception of how SIM compared to other methods. Directly contradicting **[C3]** some participants mentioned as a disadvantage that SIM generates worse results than CEM **[C6]**.

In addition, SIM was considered a difficult method compared to other methods **[C7]** and not as appropriate for technical evaluations as Heuristic Evaluation **[C8]**. Moreover, SIM was considered to have a better cost-benefit **[C10]**, to require fewer resources than other evaluation methods **[C11]** and require more theory knowledge **[C12]**. The participants who mentioned these three advantages/disadvantages did not mentioned which methods they were comparing SIM to.

It is important to note that in most of the comparisons made by the participants they did not mentioned the method they were comparing SIM to. Therefore, it was very hard to generate conclusions about SIM in relation to other evaluation methods. Besides, there is no way of knowing whether participants had any data to support their statements, or whether these statements resulted from feelings they had based on their experiences. Therefore, these issues raised in comparing the methods are not taken as final results about the method, but rather as aspects that could be interesting to further investigate.

We actually knew that the comparisons would be based on an informal perspective and experience of each evaluator. However, if there had been a consensus it could be an important indicator. The inconsistencies generated may be related to both the experience of the evaluators and the contexts in which the methods were applied, and points to a direction that it would be important to make formal assessments to compare the methods according to relevant criteria.

6.2.2 The Interview

The interview was conducted with four authors of SIM. They will be referenced as I1, I2, I3 and I4. All the authors interviewed have used several HCI methods. They all have

[8]Original statement in Portuguese by P10: *"Achei o MIS mais completo do que a avaliação heurística por fazer uma análise mais profunda da relação sistema (como preposto) e usuário."*

experience with Semiotic Engineering methods, as expected, and at least one usability evaluation method. All of them have experience with Heuristic Evaluation, but some have also experience with Usability Testing and Cognitive Walkthrough. Their overall experience in the HCI field ranges from 9 to 20 years.

Appendix B presents the guide used in the semi-structured interview conducted. The interviews were conducted through an instant message system (i.e., Skype, MSN Messenger or GTalk) in December 2011 and lasted, on average, 1 hour and 30 minutes.

6.2.2.1 Results and Analysis

The analysis of the interview aimed at investigating the advantages, disadvantages, costs and benefits of SIM from the authors' perspective, which represents the expert evaluators' view. In addition, some issues identified during the survey with novice evaluators were also placed on the agenda during the interview.

SIM's Applicability The applicability of SIM to a wide range of domains was already expected by the authors of the method who raised this hypothesis when they formalized the method [de Souza et al., 2006]. All the authors also mentioned in the interview to have applied SIM to different domains, for instance, collaborative systems, educational systems and text editors. This feature was also reinforced by I2 which said that *"the cool thing about SIM is that it does not dependent on the domain"*[9].

In addition, the authors emphasized that to apply SIM in the domains mentioned there was no need to adapt the method. According to I1, *"adjustments generally fall into a common characteristic of interpretive methods, which is how the evaluator 'fits' the concepts and materials which he/she will work with"*[10]. One may argue that this could be considered an adaptation of the method, however, the same author complements mentioning that the step of framing the concepts and materials is always required when applying SIM: *"... before any implementation of SIM the evaluator must look at the case and wonder how he/she will interpret the static, dynamic and metalinguistic signs"*[11]. It is important to mention that this expectation and experience of the authors related to SIM's applicability is consistent with the previous study described in

[9] Original statement in Portuguese by I2: *"O barato do MIS é que ele não é dependente de domínio."*

[10] Original statement in Portuguese by I1: *"As adaptações em geral recaem na parte característica dos métodos interpretativos, que é como o avaliador 'enquadra' os conceitos e materiais com os quais ele vai trabalhar."*

[11] Original statement in Portuguese by I1: *"Antes de qualquer aplicação do MIS o avaliador tem de olhar para o caso em questão e se perguntar como ele vai interpretar o que são signos estáticos, dinâmicos e metalinguísticos."*

Chapter 5 that showed that SIM can be applied to different domains and technologies without adaptations [Reis and Prates, 2011].

Another point mentioned by the authors of the method is the possibility to investigate, besides communicability, other properties. In other words, SIM allows evaluators to expand their analysis beyond the scope of communicability, allowing the identification of breakdowns related to other properties. According to I1, it is always possible to identify issues related to other qualities of use *"... because communication is the support process of other processes (cognitive, productive, etc.). Thus, when faced with communication problems (or even with certain characteristics of communication, which is not itself a PROBLEM) we always end up anticipating or just taking a glimpse of issues related to usability, productivity, etc."*[12]. Note that capital letters were included by participant.

While most of the authors agreed with this feature, author I2 believes that this characteristic should not be highlighted. He/she argues that the method should be good at doing what it is proposed to do. For instance, using SIM to find usability problems may end up frustrating the evaluator, because SIM is sure to be weaker to identify usability problems than other methods developed with focus on evaluating usability. Although I2 recommends this caution, I1 makes it clear (in the above excerpt) that the fact that SIM evaluates the communication process makes it possible to identify problems related to other properties because the communication also support other processes. Therefore, allowing other properties to be investigated is certainly a feature and can be a benefit in certain application contexts. However, the evaluator, when selecting SIM as a method, should take into account that the focus of SIM is evaluating communicability and consider if it would be an appropriate method if other properties are of interest.

Regarding the possibility of identifying domain-specific issues (i.e., if applied to a collaborative system, could it identify specific issues regarding users' interaction through the system, or only general problems, such as static elements that were not consistent with the behavior associated to them?) most of the authors think it is possible to identify domain-specific issues and mentioned to have identified issues specific to the domain in their evaluations with SIM. However, I1 mentioned that this characteristic was never the goal in the application of SIM and he/she prefers to be cautious in confirming this feature of the method. The author believes that a responsible answer

[12]Original statement in Portuguese by I1: *"... pois a comunicação é o processo de sustentação de outros processos (cognitivos, produtivos, etc.). Então, ao nos depararmos com problemas de comunicação (ou mesmo com certas características da comunicação, que não são propriamente um PROBLEMA) sempre acabamos antevendo ou entrevendo problemas de usabilidade, de produtividade, etc."*

would depend on having more knowledge than in fact he/she has. At the same time he/she mentioned to be working on a scientific application of SIM and thinks that he/she is coming close to stating a potentially new concept in the domain in which SIM is being applied. But still he/she thinks that it is too early to say.

Evaluators' Experience With regard to the influence of the evaluators experience the authors agree that the evaluator experience has an influence on the results. Author I4 mentions that it is the evaluator's HCI experience that will allow him/her to make associations between communicability problems and other HCI problems. If evaluators do not have HCI experience they might not be able to analyze what the problems identified in the signs mean regarding interaction.

Author I3 went further mentioning that in addition to HCI experience the evaluator's skills and abilities that come from life experience also influence the results. *"The ability to criticize, analyze and be attuned to the context of what is happening, I think all this ends up influencing the results of SIM application"*[13].

Although the authors agree that the evaluator's experience influence on SIM's application, some of them argue that this does not differ from other existing methods as mentioned by I1: *"... in any evaluation method – and PARTICULARLY in inspection-based methods – the evaluators experience is everything"*[14]. For instance, I4 mentioned that he/she works with Heuristic Evaluation and *"students also find it difficult to assign a heuristic to a problem identified"*[15].

But at the same time author I3 believes that *"SIM, in a way, by being a deeply interpretive method, ends up instigating this more than others"*[16]. According to I2 the Heuristic Evaluation, for instance, gives an illusion that it is easy, fast and simple, but the HCI experience is an implicit prerequisite of the method. SIM is different because while the other methods leave the necessity of expertise in HCI implicit, SIM, because it is based on theory and has chained steps, makes it clear and explicit that the experience is essential. In other words, the fact that SIM is a deeply interpretive method makes it depend more on the evaluators' experience than other methods.

[13]Original statement in Portuguese by I3: *"A capacidade crítica, analítica e estar antenado ao contexto do que está acontecendo, acho que isso tudo acaba influenciando nos resultados da aplicação do MIS."*

[14]Original statement in Portuguese by I1: *"Em qualquer método de avaliação – e PARTICULARMENTE nos métodos de inspeção – a experiência do avaliador é tudo."*

[15]Original statement in Portuguese by I4: *"os alunos também sentem dificuldades para atribuir uma heurística a um problema que identificam."*

[16]Original statement in Portuguese by I2: *"O MIS, de certa forma, por ser um método profundamente interpretativo, acaba instigando mais isto."*

In addition, the importance of the theory knowledge was emphasized by the authors. Author I2 mentioned that the elaboration of a broader and competent opinion about the meta-communication depends on how much the evaluator knows the theory. *"If you do not know [the theory], you will be able to do it, but will be more at the lowest level of abstraction"*[17]. That is, the more one knows about Semiotic Engineering the richer the analysis of the results will be. According to I4, *"students who have a superficial knowledge about Semiotic Engineering (which is the reality of the teaching context) can apply the method, but the results are superficial. They cannot analyze the consequences of problems"*[18].

Costs/Disadvantages and Benefits/Advantages We also asked the authors what were the costs/disadvantages and benefits/advantages of SIM. Regarding costs and disadvantages of SIM the authors I1, I2 and I3 mentioned that SIM requires evaluators to have skills and abilities that are not trivial. According to I1, SIM requires the steps of segmented analysis to be followed strictly as proposed by the method. He/she believes, for instance, that it is very difficult to be strict and not include dynamic signs in the static signs analysis step. Also it requires accuracy in the consolidation of the segmented analysis into a single diagnosis about the communicability of the meta-message. Therefore, to perform an application of SIM the authors believe it requires the evaluator to have a good abstract and interpretative reasoning. As mentioned by I2 and I3 this is a more specific cost of SIM. According to I3, Heuristic Evaluation, for instance, has a recipe to be followed and students can follow and apply it without problems. Whereas, although SIM has well-defined and chained steps, students often have difficulties in applying it; they cannot make the necessary abstractions, and end up applying it in the wrong way.

A disadvantage mentioned by I2 is that SIM does not involve users, but he/she emphasizes that this is a characteristic and disadvantage of all inspection-based methods (i.e., it is not an exclusive feature of SIM). The same author also mentioned as a disadvantage the fact that SIM requires a functional prototype to be used. In his/her view there should have a formative version of SIM, in other words, that supports the evaluation of interfaces early in the process.

In addition, I3 mentioned as a disadvantage the fact that SIM does not have a support tool. This drawback may be associated with what was mentioned by I4 that

[17] Original statement in Portuguese by I2: *"Se você não souber, será capaz de fazer, mas ficará mais no nível mais baixo de abstração."*

[18] Original statement in Portuguese by I2: *"Alunos que tem um conhecimento superficial sobre Engenharia Semiótica (que é a grande realidade do contexto de ensino) até conseguem aplicar o método, mas os resultados são superficiais. Eles não conseguem analisar as consequências dos problemas."*

applying SIM demands time. The authors believe that this effort to apply it is not necessarily a disadvantage of the method. They argue that the effort associated to its application is related to a cost that should be considered when choosing the method to evaluate a system. This characteristic of SIM (i.e., that it demands time) is justified by the fact that it is a qualitative method, because *"... since it is a qualitative method, in order to have richer results more time is needed."*[19] (I4).

Although the authors agree that SIM demands time, all of them mentioned a great advantaged of SIM that offsets this cost: SIM allows obtaining richer and deeper results. According to I4, SIM allows the evaluator to obtain *"... a thorough understanding of the system to be evaluated."*[20]. Author I1's view is in line with this statement and he/she goes further:

> *"The main advantage (which in my opinion beats all the costs) is to be able to learn a wealth of opportunities, resources and facets of human-computer interaction. Nobody comes out of an application of SIM without learning lots of things about HCI. Not even me. :) That's why I think that SIM is the most valuable tool of Semiotic Engineering today. As Don Norman, himself, has already drawn attention in more than one of his articles, the great advantage of Semiotic Engineering, which exemplary crystallizes in SIM, is the integrated view of how many and so many things that are involved in an HCI project. This contrasts very strongly with other methods that have a more sporadic or fragmented, or superficial performance."*[21]

6.2.3 Discussion

We could observe that there was a consensus among novices and authors for most of the SIM characteristics raised. They perceive as the costs of SIM the high time/effort demanded, which can be related to the fact that the method is deeply qualitative and interpretive, and, therefore, demands time and effort to obtain richer results. In

[19] Original statement in Portuguese by I4: *"Como é um método qualitativo, para que o avaliador tenha resultados mais ricos é necessário mais tempo."*

[20] Original statement in Portuguese by I4: *"...um conhecimento aprofundado sobre o sistema a ser avaliado."*

[21] Original statement in Portuguese by I1: *"A principal vantagem (que a meu ver bate todos os custos) é APRENDER a riqueza de oportunidades, de recursos e de facetas da interação humano-computador. Ninguém sai de uma aplicação do MIS sem ter aprendido MONTES de coisas sobre IHC. Nem eu. ;) Por isto, acho que o MIS é a ferramenta mais valiosa da EngSem atualmente. Como o Don Norman, mesmo, já chamou a atenção em mais de um artigo dele, a grande vantagem da EngSem, que se cristaliza exemplarmente no MIS, é a visão integradora de quantas e tantas coisas estão envolvidas no projeto de IHC. Isto contrasta muito fortemente com outros métodos que têm uma atuação mais pontual ou fragmentária, ou superficial."*

addition, the knowledge and skills needed to use it were also mentioned as a cost. First, there is a cost associated to learning the method and its underlying theory. According to I4, *"the cost is high because you need to understand well Semiotic Engineering."*[22]. Semiotic Engineering knowledge is seen as essential for the successful application of the method, and, therefore, it's a cost to be considered when choosing the method. Secondly, there are costs related to skills and abilities that SIM requires of the evaluator that are not easily taught, as for example, *"... how to think, reflect, interpret and abstract"* [23].

The difficulty related to learning the method has also been identified in the work of Bim [2009] in which the author notes that the cause of these difficulties could be considered due to the need to break a widely used way of thinking in the computer science field, which is more predictive, accurate and repeatable. She also pointed out that these difficulties are not exclusive of SIM and CEM. Interpreting, abstracting and building a global vision has been identified as the cause for serious difficulties in the teaching and usage of other methods, for example, in programming, interface design and usability engineering. Thus, we conclude that SIM does require an effort to be learned due exclusively to the fact of being a theory-based method and requiring its knowledge to apply it. In addition, the application of SIM requires certain skills that are not easily taught and, probably, learned.

Regarding the effectiveness, SIM was considered by novice evaluators as an effective method since it allows identifying, in most cases, relevant problems and formalizing perceived problems. In the interview the authors complemented mentioning that SIM allows the evaluator to obtain a thorough understanding of the system being evaluated, generating richer and detailed results about the system's communicability. The authors also pointed out that this is the main advantage of SIM and that it beats all the costs. According to I2, *"it is costly, but worthwhile"* [24].

Novices and authors agree that SIM allows a thorough analysis of the system, supporting the evaluator in formalizing, arguing and explaining the problems encountered. It also produces rich and deep results about the system's communicability and has a good cost-effective ratio. In addition, the novice evaluators' perception about SIM's applicability is aligned with the authors view. The advantage related to the fact that the method can be applied to different domains without adaptations was mentioned by the novice evaluators and reinforced by the authors. The fact that the

[22] Original statement in Portuguese by I4: *"O custo é alto pois é preciso compreender bem a Engenharia Semiótica."*

[23] Original statement in Portuguese by I3: *"... ensinar a pensar, refletir, interpretar e abstrair."*

[24] Original statement in Portuguese by I2: *"É custoso, mas vantajoso."*

method allows identifying the impact of problems related to other system properties (e.g., accessibility, sociability) was also a consensus.

Being a theory-based method was contradictory among novice evaluators (some mentioned as a disadvantage the fact of being based on a complex theory). We argue that this could probably be related to the cost mentioned previously (i.e., that SIM is a difficult method with a high learning curve). Again this reinforces the cost regarding knowledge and skills necessary to apply the method. The author strengthen the view that being a theory-based method is an advantage of SIM. According to I1, *"methods based on theory are naturally more powerful in the sense that their results talk to concepts that are beyond the specific situation or context of the application"* [25]. Author I2 reinforces this view stating that the theory *"brings results integrated and consolidated into a chained reasoning"*[26]. Although some novice evaluators have mentioned the theory as an advantage, some of them did not include any advantages related to the theory when asked about the advantages of the method. As shown, the authors of SIM have a clearer view of the benefits related to the theory. This clearer view can be justified by the fact that they do not only have a larger experience with the method and Semiotic Engineering, but also with HCI in general.

Regarding the disadvantages, the lack of tools to support the application of the method was mentioned by both novices and authors. Although mentioned as a disadvantage, we do not characterize as a drawback because this can be easily solved by creating a support tool. Given the results obtained in this study, we now think that the effort is worth it and we believe that it could be an interesting future work to be conducted (which is already being pursued in our research group).

Finally, it is interesting to note that both the novice evaluators and the authors of the method have a similar view in relation to the costs and benefits of MIS. It is recognized by the authors and perceived by novices that SIM demands time to apply and learn it, and requires certain knowledge and skills of the evaluators. Despite these costs, SIM is perceived as a method that allows evaluators to obtain richer and more detailed results about the system's interface. Furthermore, we may argue that the cost of learning the theory and method is fixed, and not associated to every application of the method. Thus, SIM is perceived as a method with good cost-benefit.

As we can see, we identified some advantages, disadvantages, costs and benefits of the method, which can be useful to assist not only in deeper investigations of SIM,

[25] Original statement in Portuguese by I1: *"Métodos baseado em teoria são naturalmente mais potentes no sentido de que seus resultados conversam com conceitos que estão além da situação ou contexto específico da aplicação."*

[26] Original statement in Portuguese by I2: *"Traz resultados integrados e consolidados em um raciocínio encadeado."*

but also to identify ways to improve SIM's use in the HCI field. The reason of having performed this step by obtaining information from the evaluators' perspective is due to the fact that the evaluators are responsible for deciding which method to use for an interactive system evaluation. Then gather information about the costs and benefits from the evaluators' perspective may be important in defining (or revising) a strategy to present and/or teach the method, and also conduct deeper assessments of the method. In particular, the disadvantage mentioned by the evaluators related to the lack of educational material points to the need of developing support materials so that the method can potentially be better understood and more widely used.

The assessment presented in this Chapter has some limitations. As already mentioned, the main limitation is the number of respondents: only 25 participants and 4 experts (i.e., the authors) answered, respectively, the questionnaire and interview. This number is not statistically significant considering the international HCI community. However, we considered a good number of responses since SIM is a relatively recent method, and it is not widely used or taught (yet).

In addition, the qualitative analysis of the participants' responses and generation of the advantages and disadvantages lists were done only by the author of this research, which could be a threat to the validity of the results. However, to avoid bias in the analysis the results were discussed with the advisor and a review process was also conducted.

Chapter 7

Empirical Assessment

The third study to evaluate the Semiotic Inspection Method (SIM) aimed at identifying its capabilities and limitations through an empirical assessment. One of the goals of this empirical assessment was to collect data regarding some of the contradictory statements made by some evaluators (i.e., if the effort needed is high or low and if the method has a high or low cost-effectiveness) and also identify other features of the method.

In this assessment we compared SIM to other Semiotic Engineering evaluation method. SIM is the only inspection-based method of the Semiotic Engineering theory. As mentioned in the Chapter 2, the other existing methods of the Semiotic Engineering are ISIM and CEM. ISIM is based on SIM and it has a special focus on evaluating the system interface from other stakeholder's perspective (e.g., a teacher's perspective on a learning support system). CEM, in turn, is the Semiotic Engineering user-based evaluation method. To carry out this comparative assessment we chose CEM. The reason for this choice is because CEM was the first Semiotic Engineering theory method proposed, and also because, together with SIM, they are the most established methods of the theory. The comparison of these methods could provide contributions not only to the HCI field as a whole, but also to the theory itself.

Considering that inspection-based and user-based methods are different (i.e., produce different results), our goal was not to treat SIM and CEM as interchangeable. Rather, the methods are complementary, and then the purpose of this study was to identify SIM's strengths, weaknesses, differences and similarities compared to CEM. It is noteworthy that the research aims at evaluating SIM, and, therefore, we present the analysis focusing on it. Before showing how this assessment was conducted and the results obtained, we next present CEM.

7.1 Communicability Evaluation Method

Communicability Evaluation Method (CEM), unlike SIM, is a method that involves users in the evaluation and the goal is to identify communication breakdowns from the observation of the meta-communication reception (i.e., from user's perspective) [Prates et al., 2000; Prates and Barbosa, 2007; de Souza and Leitão, 2009]. CEM has three basic steps, as shown in Figure 7.1 [Prates et al., 2000; de Souza, 2005]: (i) preparation, (ii) application, and (iii) data analysis.

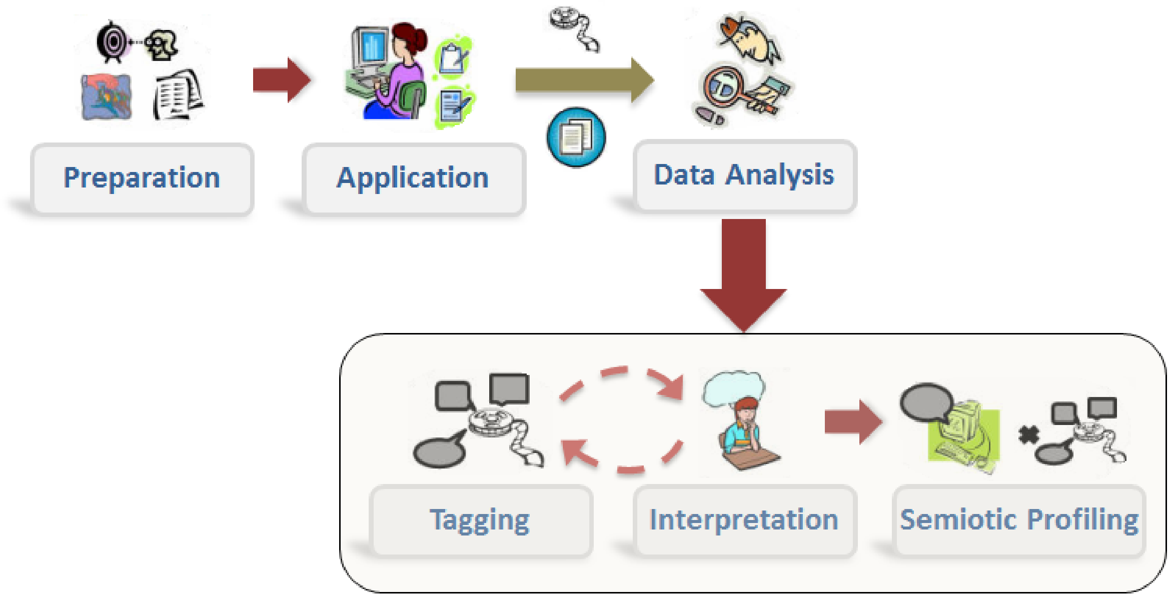

Figure 7.1. CEM Steps [Adapted from Bim et al. [2009]].

Steps (i) and (ii) are similar to other methods that involve users' participation. In the preparation step the evaluator must perform an inspection of the artifact to be evaluated to identify the appropriate focus of the investigation. It is also necessary to define the criteria for selecting participants. Next, the evaluator proceeds to elaborate the test scenario, the consent form, the script of a pre-test and post-test interviews and/or questionnaires (if necessary), as well as the necessary hardware and software infrastructure for the test. Finally, the evaluator does a pilot test to verify if the test application needs further adjustments [de Souza and Leitão, 2009; Bim, 2009]. After the preparation step of the test, the evaluator applies it.

Step (iii) is the central step of CEM and is divided into three other steps: (1) tagging, (2) interpretation, and (3) semiotic profiling. In step (1) the evaluator observes the video recorded of each user and identifies sequences of actions that indicate a communication breakdown. Each identified breakdown is tagged with one of 13 specific expressions, which represents the researcher's interpretation of user behavior in the context in which the breakdown occurs. The 13 specific expressions published by de Souza and Leitão [2009] are presented in Table 7.1.

Table 7.1: Description of CEM 13 specific expressions [de Souza and Leitão, 2009, p.38]

"I give up."	This utterance is used to tag interaction where the user explicitly admits her inability to achieve her goal. The general symptom of this breakdown is that the user interrupts her activity without having accomplished all of the proposed task(s). This may occur at any time during the test and is always associated to other breakdowns in the reception of meta-communication.
"Looks fine to me."	This tag is applied when the user is convinced that she has achieved her goal but, in fact, has not. The symptom of this breakdown is that the user terminates the test falling short of achieving all the tasks described in the test scenario. When asked if all tasks have been achieved, the user will say that they have.
"Thanks, but no, thanks."	This utterance is used when the user is aware of the designer's deputy's meta-communication regarding the types of conversations that are expected to lead to a particular effect, but chooses to do something different than is expected. Knowing what is "expected" is the result of careful examination of explicit manifestations of the designer regarding how certain tasks and operations are achieved. This is typically included in help material. Because the user gives the researcher evidence that she knows what the designer is saying, but decides to follow a different interactive path, she declines the designer's invitation to engage in that particular kind of communication. Hence, there is breakdown, even if, from a cognitive point of view, there is evidence that the user is in full control of interaction.

"I can do otherwise."	This tag is used when the user is not aware of the designer's deputy's meta-communication regarding the types of conversations that are expected to lead to a particular effect. She then chooses to do something different than is expected, but achieves the same effect. This situation is slightly but critically different from the previous one, where the researcher should use the "Thanks, but no, thanks." tag. The breakdown tagged with "I can do otherwise." is in some respect more severe than the previous one because now the user reveals that she has not received the designer's message about how the system should be used in the context where she is.
"Where is it?"	This tag is used when the user expects to see a certain sign that corresponds to a particular element of her strategy, but cannot find it among the signs expressed by the designer's deputy. The user must be convinced that the sign she is looking for is the one she needs to express her current goal (otherwise, the problem is associated to another kind of breakdown).
"What happened?"	This utterance is used to tag interaction where the user repeats an operation because she cannot see or understand the evidence of the effects caused by her actions. The typical symptom of "What happened?" is the user's repeated activation of a function whose feedback is either absent or not perceived.
"What now?"	This tag is used when the user is temporarily clueless about what to do next because none of the designer's deputy's signs mean anything to her. The typical symptoms of "What now?" is when the user is following a random path in interaction. No connection can be traced between one interactive step and the next. The difference between a "What now?" tag and a "Where is it?" lies in the user's knowing the content she wants to express (the case of "Where is it?") or not having any notion (the case of "What now?"). This kind of breakdown can turn into a severe case of miscommunication if, during random interaction, the user cannot find a sign that will spark interpretations that will bring her back into communication with the designer's deputy and eventually lead her out of the breakdown situation.

"Where am I?"	This tag is used when the user is interpreting (and potentially using) signs that belong to the designer's deputy's vocabulary, but doing so in the wrong context of communication. The main problem in this breakdown is the signification of context, which confuses the user.
"Oops!"	This tag is used when the user momentarily makes a mistake and immediately corrects it. She sees that she has made a wrong step and usually activates the "undo" function immediately. However, if the attempt to correct her mistake develops into a long search for a way to cancel the effects of a slip, then it indicates a very serious communication problem.
"I can't do it this way."	This utterance is used to tag interaction where the user abandons a path of interaction (composed of many steps) because she thinks it is not leading her towards her goal. The typical symptom of an "I can't do it this way." is when the user suddenly interrupts an activity she is engaged in and takes a totally different direction.
"What is this?"	This tag is used when the user expects to see an explanatory tip or any other cue to what a particular interface sign means.
"Help!"	This tag is used when the user explicitly resorts to metalinguistic meta-communication in order to restore productive interaction. She may deliberately call a help function by pressing F1 or read documentation material offline. Although used less frequently than one might expect, online help is certainly a privileged communicative resource for designers.
"Why doesn't it?"	This utterance is used to tag interaction where the user is trying to make sense of the designer's deputy's message by repeating the steps of previous unsuccessful communication in order to find out what went wrong. She does not know how to express her intent, but suspects that the sign she is currently examining is the one to be used for achieving the intended goal. In other words, the user is using experimentation to make sense of how the system works.

Having tagged all the passages in which an evidence of a communicability breakdown occurs, in step (2), the evaluator begins to interpret the meaning of the set of tags. This interpretation is based on the presence or absence of each of the 13 tags, their frequency and distribution. The expressions can also be classified according to the type of failure, as shown in Table 7.2. Then, finally, in step (3), the semiotic profile is generated which concludes the whole semiotic process with a detailed characterization of the meta-communication reception.

Table 7.2. CEM's Failure Type [de Souza, 2005, p.138]

Category Type	Subcategory	Distinctive Feature	Utterance
(I) Complete Failures		(a) User is conscious of failure.	"I give up."
		(b) User is unconscious of failure.	"Looks fine to me."
(II) Temporary Failures	a. User's semiosis is temporarily halted	(1) because he cannot find the appropriate expression for his illocution	"Where is it?"
		(2) because he does not perceive or understand the designer's deputy's illocution	"What happened?"
		(3) because he cannot find an appropriate intent for illocution	"What now?"
	b. User realizes his illocution is wrong	(1) because it is uttered in the wrong context	"Where am I?"
		(2) because the expression in illocution is wrong	"Oops!"
		(3) because a many-step conversation has not caused the desired effects	"I can't do it this way."
	c. User seeks to clarify the designers' deputy's illocution	(1) through implicit meta-communication	"What is this?"
		(2) through explicit meta-communication	"Help!"
		(3) through autonomous sense making	"Why doesn't it?"
(III) Partial Failures		(a) User does not understand the design solution.	"Thanks, but no, thanks."
		(b) User understands the design solution.	"I can do otherwise."

7.2 The Empirical Assessment

As shown in Figure 7.2, the empirical assessment was conducted in four major steps: (1) target system definition; (2) preparation; (3) execution; and (4) comparison. Next we describe each step.

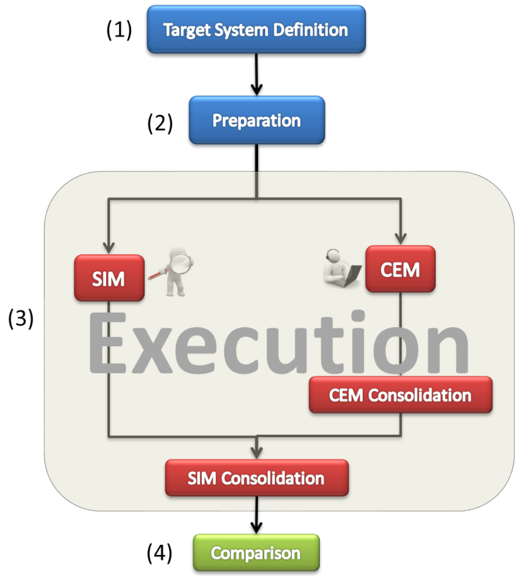

Figure 7.2. Experiment Steps.

7.2.1 Target System Definition

According to de Jong and Schellens [2000] an important factor that has been neglected in comparative assessments so far is the initial quality of the target system. It is important to make sure that the target system has enough problems that allow the assessment proposed. It is not sensible to explore the benefits of an evaluation method using systems that are clearly below normal standards (i.e., systems that could be easily improved without the help of an evaluation method) [de Jong and Schellens, 2000]. In order to choose the target system, we pre-defined three candidate systems:

- **todoist (http://todoist.com/):** an online task manager;
- **Granatum (http://www.granatum.com.br/):** an online financial control system; and
- **1DayLater (http://1daylater.com/):** an online system to track time, expenses and business miles.

To define the target system we evaluated them applying SIM. The criteria to choose the target system were: the system should have a small set of functionalities (to make the assessment viable), it should provide all the three types of sign classes (i.e., metalinguistic, static and dynamic), it should have communicability problems and an apparently nice and easy to use interface (i.e., not clearly below normal standards). So after doing SIM on the three candidate systems the results were discussed between the author and her advisor, and both selected **todoist** as the target system. Next we present the system main features[1].

7.2.1.1 Todoist

Todoist is an online task manager and the designer message to users is that it is *"useful, fast and easy to use"*. The system provides on the homepage (before login on the system) a short introduction video (Figure 7.3 – item 1), information about security, privacy and backups (Figure 7.3 – item 2), and a summary of its main features (Figure 7.3 – item 3). After logging into the system, the system help appears on the start page (Figure 7.4). On the same page users have access to the same introduction video presented in the homepage and more information about the system's features (Figure 7.4 – item 1). On the left side the features "Add project" and "Filter tasks" are presented (Figure 7.4 – item 2).

The system help disappears when the user creates his first project or clicks on the title of one of the existing projects, and the system help is replaced by the main tasks pages as shown on Figure 7.5. In order to add a new project the user has to click on the "Add project" link (Figure 7.5 – item 1). When doing this, a dialog box is opened and the user must inform the project name and click on "Add project" button (Figure 7.6 – item 1). To make a subproject the user must use the left and right arrows to indent the project accordingly (Figure 7.6 – item 2). The project setting is available clicking on the gear sign, which only appears when hovering over the project title. In the project settings the user can create a new project above/below, edit or delete and change its color (Figure 7.7). The user can also reorder the project clicking on the link "Reorder" (Figure 7.5 – item 1).

After adding a project the system provides a way of adding tasks. Tasks always have to be assigned to a project. The tasks page is shown on the right/center side of the page (Figure 7.5 – item 2), which has as its title the name of the project, an email sign where the user can create tasks by sending an e-mail and a gear sign that represents some tasks actions (i.e., sort the tasks by date or priority). The tasks list is presented right below with a checkbox on the right side, where the user can mark the task as completed. The link "Show completed tasks" allows the user to see all the tasks that have been marked as completed, and the user may also delete all completed tasks by clicking on the "Delete completed" link.

[1] The assessments occurred from September to November 2011. It is worth noting that there are differences in the current interface version of the system compared to the one used in the empirical assessment.

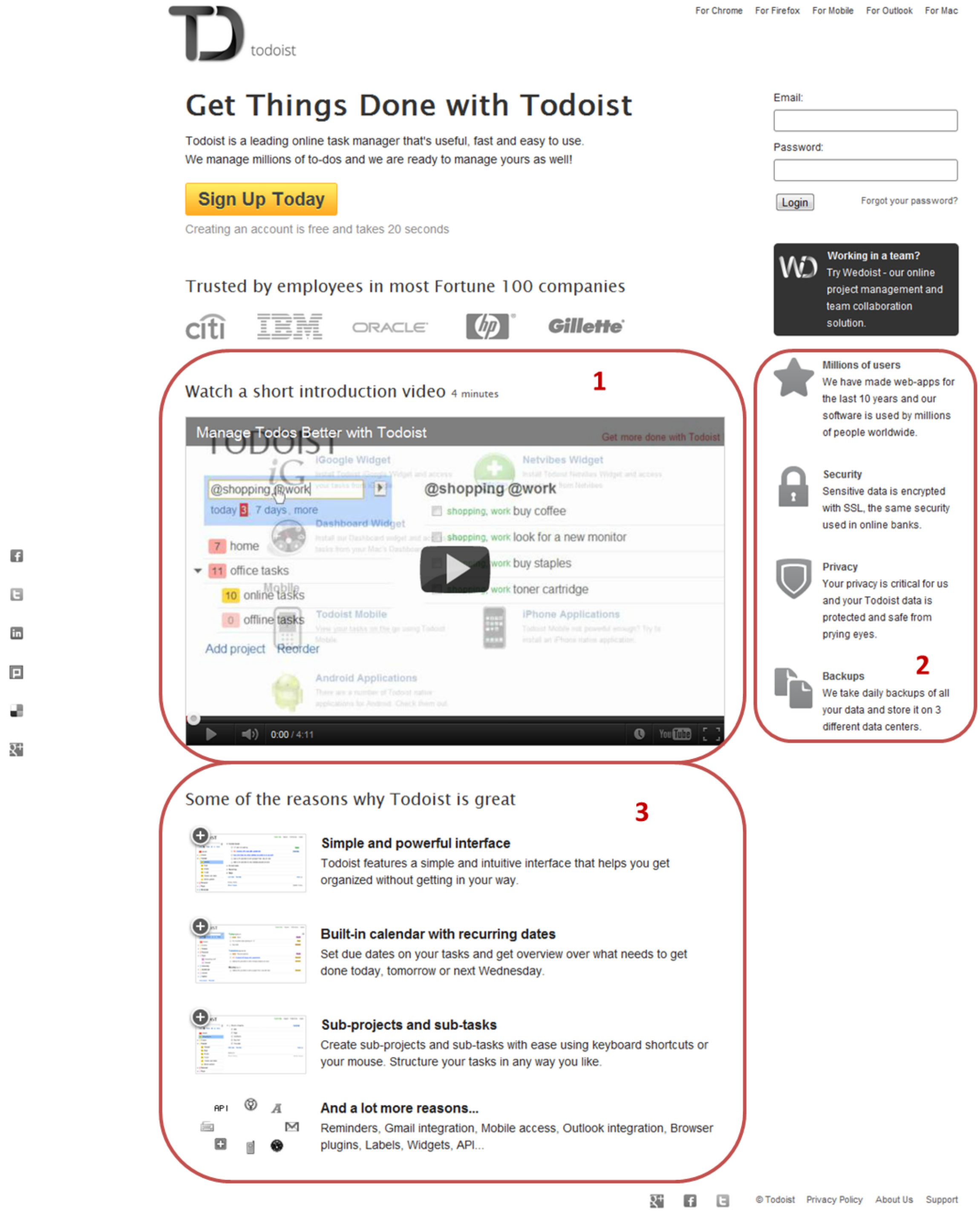

Figure 7.3. Todoist – Homepage (before login on the system).

To add a new task the user must click on the "Add task" link. When doing this a dialog box is opened and the user must inform the task description and, optionally, the task due date (in the field below the task description to the right) and click on "Add task" button (Figure 7.8 – item 1). Every time the user adds a task a new dialog box

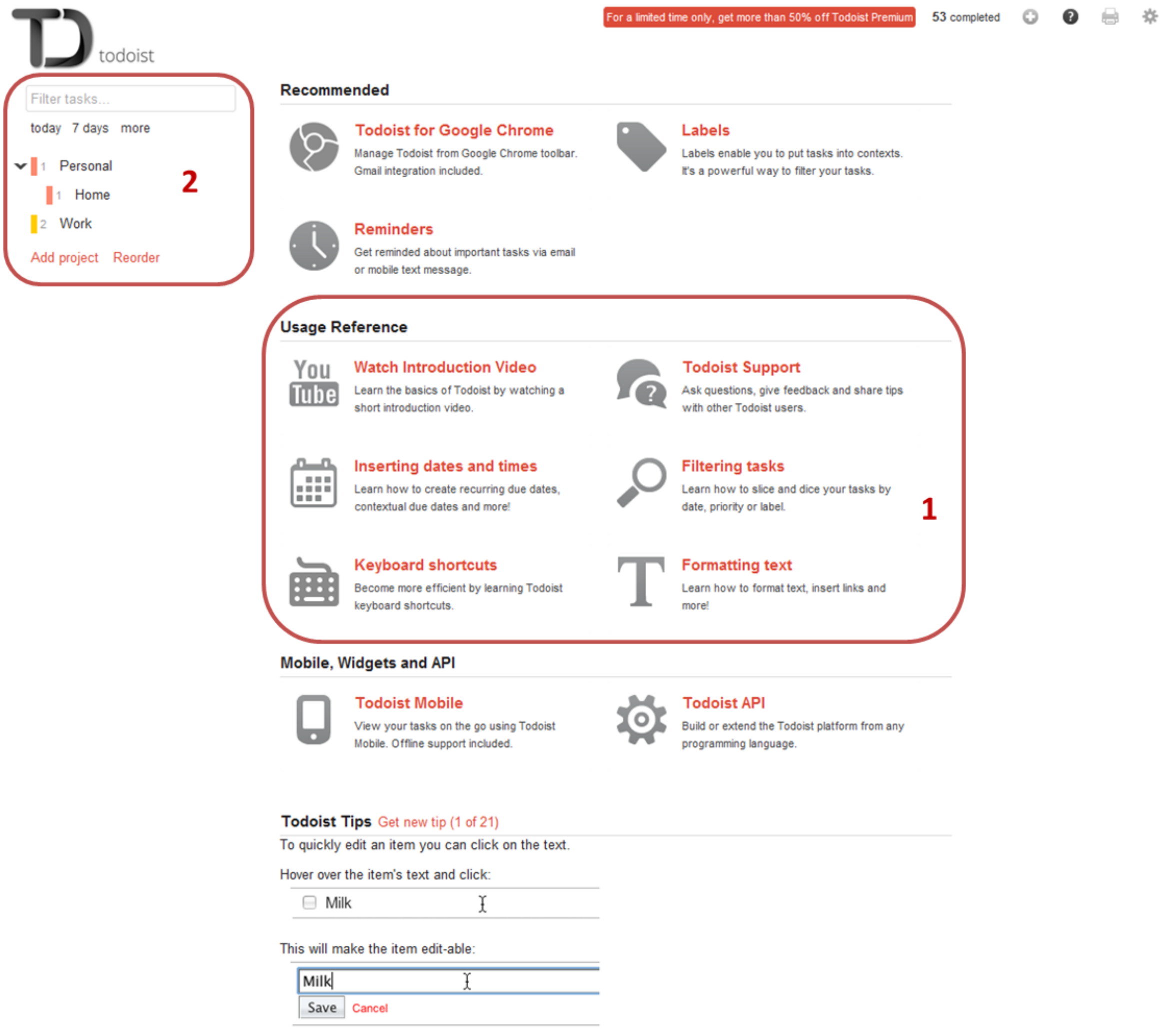

Figure 7.4. Todoist – Start Page (after login on the system).

is opened, so the user can continue adding tasks. To close the dialog box the user must click on "I'm done adding tasks." (Figure 7.8 – item 2). The designer also provides the option to create subtasks clicking on the left and right arrows to indent the task accordingly (Figure 7.8 – item 3).

The task setting is available clicking on the gear sign, which only appears when hovering over the task title. In the task settings the user can create a new task above/below, postpone the task in one day, edit or delete it, set the task priority, set reminder (available only for premium accounts), move the task to history (which is the same as mark the task as complete) and move the task to another project (Figure 7.9). The user can also reorder the task clicking on the link "Reorder" (Figure 7.5 – item 2).

Users can also search for tasks created using the feature "Filter tasks" (Figure

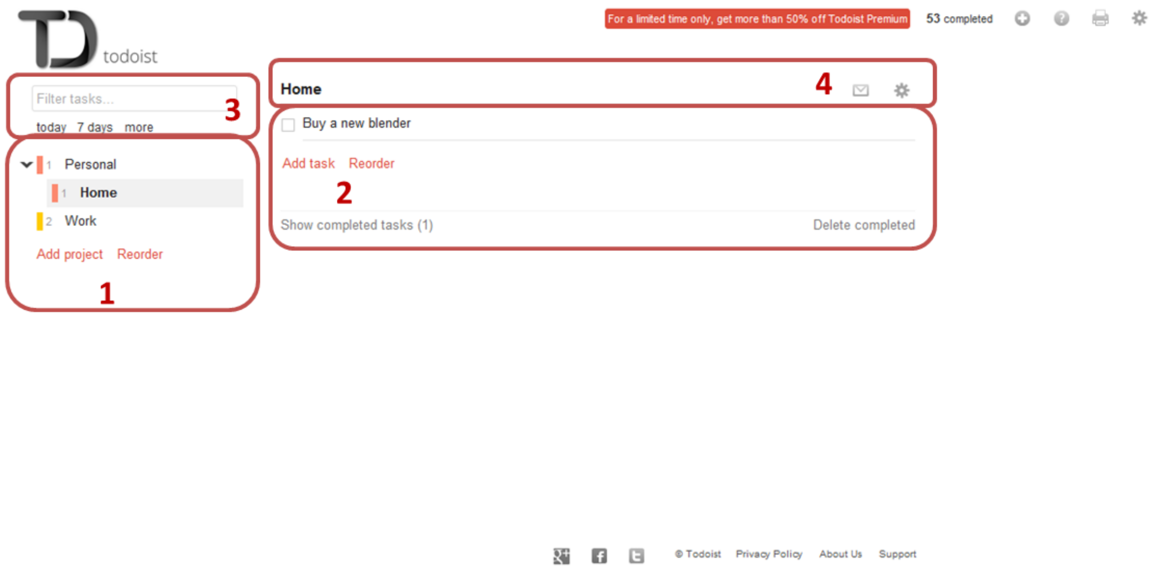

Figure 7.5. Todoist – Features page.

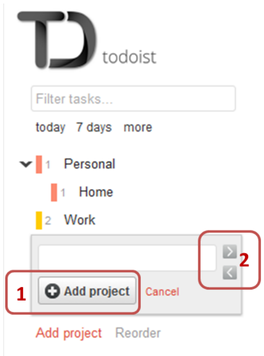

Figure 7.6. Todoist – Add project feature.

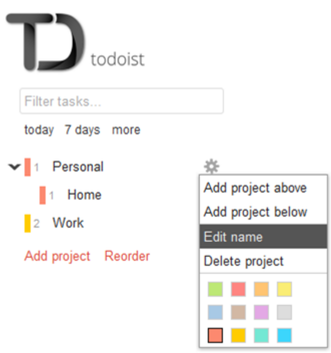

Figure 7.7. Todoist – Project Settings.

7.5 – item 3). There are some options right below the search field where the user can easily search for all the tasks that are due today, in the next seven days, or access other

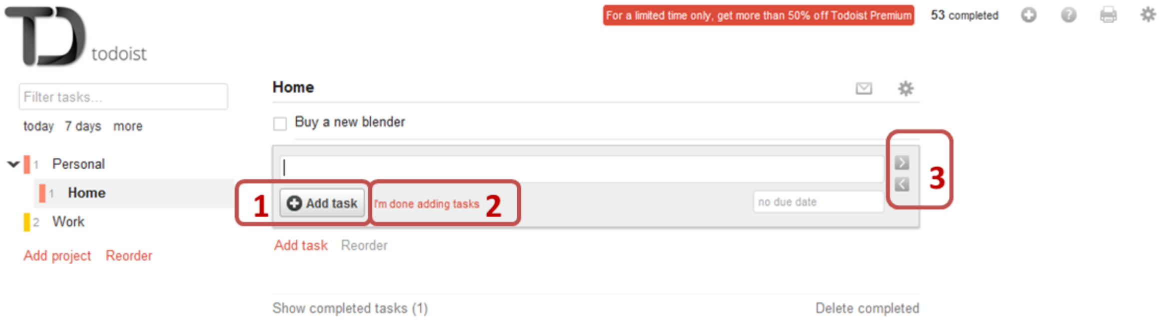

Figure 7.8. Todoist – Add Task.

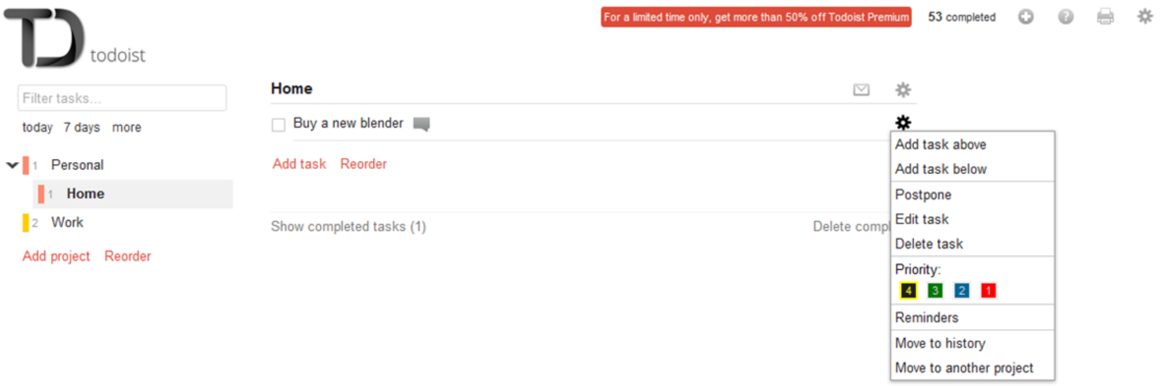

Figure 7.9. Todoist – Task Settings.

filters (Figure 7.10). The user can also type his search on the field "Filter tasks...".

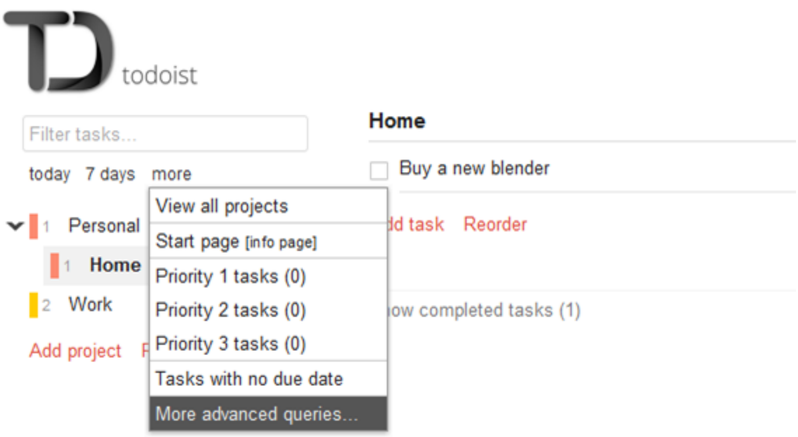

Figure 7.10. Todoist – Filter tasks.

7.2.2 Preparation

In this step we defined how the empirical assessment would be conducted and prepared all the material necessary. In preparation for SIM, we elaborated the interaction scenario to provide the contextual structure necessary for the communication analysis. Since we did a first evaluation of the system using SIM in the "Target System Definition" step, we used the same scenario for the evaluations:

> *Joana is an entrepreneur who operates in the sector of Human Resources. Her company offers various services, including: human resources consulting, organizational climate research, performance appraisal and recruitment and selection. She is responsible for coordinating a team of 5 employees. Besides being a great professional, she is mother of two children: Gabriel, 6 years old, and Valeria, 3 years old. Joana has a busy routine, which besides following her staff activities, involves contacting clients, evaluating reports generated by her team and also performing personal activities (e.g., children activities, husband and home). Because of the difficulty on managing and remembering all the activities she must do during the day, she decided to look for a system that supports this task. Joana is a novice user of technology and therefore needs an easy to use system. In a Google search, she found the **todoist** (`http://todoist.com/`), a task management system, and decided to explore it to see if it meets her needs.*

After defining the scenario, a guideline was elaborated in order to give instructions about the assessment (Appendix C). In the same document, a consent form was provided. As both the invitation and the guidelines were all made and distributed by e-mail, the participants did not sign the consent form. However, when sending back the results of their evaluation, participants were demonstrating to be in accordance with the consent form. A template was also given in order to facilitate the analysis in a later step. Material about how to conduct the method was made available in case it was needed.

In preparation for CEM, the evaluator made an inspection of the artifact to be evaluated to identify the appropriate focus of investigation. This inspection was done when we were selecting the target system. Therefore, we used the results of this inspection to prepare the material necessary for CEM. A concise version of the scenario used in SIM inspection was used for CEM evaluation so that users were able to get the essential information regarding context. We defined 11 tasks that were related to the main functionalities of the system and also related to the main potential problems identified with the initial SIM application (Appendix E). We also elaborated the consent form (Appendix D), the pre-test questionnaire (Appendix F) and post-test interview (Appendix G). Finally, we did a pilot test to verify if the test needed adjustments.

7.2.3 Execution

We invited several evaluators to conduct SIM sending all the material prepared in the previous step. Although we invited more than 15 evaluators, only 7 participated in the empirical assessment. Table 7.3 presents the evaluators' profile. As we can see, we selected evaluators from computer science field that had different levels of experience with SIM which were classified as expert (used SIM at least 6 times) and novice (used SIM at most 3 times). The evaluators were asked to return the template (prepared in the previous step) filled out with the list of problems encountered.

Table 7.3. Evaluators' Profile

ID	Gender	Education	English Knowledge	Experience with SIM (# of applications)	Novice or Expert?
E1	Female	Master Incomplete	Intermediate	6 or more	Expert
E2	Female	Master Complete	Intermediate	6 or more	Expert
E3	Female	Master Incomplete	Advanced	6 or more	Expert
E4	Male	Master Incomplete	Advanced	3	Novice
E5	Male	Undergraduation Incomplete	Basic	2	Novice
E6	Female	Doctorate Incomplete	Intermediate	2	Novice
E7	Female	Undergraduation Complete	Intermediate	0	Novice

After inviting evaluators to conduct SIM, we started the evaluation with CEM. We invited 9 users and we present their profile in Table 7.4. As we can see, we select users with different occupations, gender and ages. Since the assessment was carried out in Brazil, which the official language is Portuguese and the target system is in English, we selected users that had English knowledge. Most of the users had at least intermediate English level and only one user had basic English level. The user with basic knowledge mentioned not to have had problems related to English while using the system. All the users selected were familiar with computers and internet, but most considered themselves basic or intermediate users. We mixed users that had some experience with other similar systems that help organizing activities and tasks (e.g., Google Calendar and Outlook) with the ones that had no experience at all. None of them had ever used **todoist** before.

Table 7.4. Users' Profile

ID	Gender	Age	Education	English Knowledge	Used Any Other Similar System?
U1	Female	38	Economy	Advanced	No.
U2	Female	35	Psychology	Intermediate	Yes, Google.
U3	Female	33	Administration	Intermediate	No.
U4	Male	26	Sociology	Intermediate	Yes, Google and Outlook.
U5	Male	26	Mechanical Engineering	Intermediate	Yes, Outlook.
U6	Female	26	Social Communication	Advanced	Yes, Google and Outlook.
U7	Female	27	Social Communication	Basic	Yes, Outlook.
U8	Female	25	Administration	Intermediate	Yes, Outlook.
U9	Male	37	Mechanical Engineering	Advanced	Yes, Outlook.

In the test section the users were first asked to read and sign (in accordance) the consent form. Then they answered the pre-test questionnaire (about their profile) and read the scenario to get contextualized. We gave a sheet of paper with the 11 tasks previously elaborated, to be done one at a time. Users were told they could express their thoughts during the evaluation. The user-system interaction and audio were recorded for later analysis. During the evaluation, the users were observed by the author of this study which took notes regarding their interaction. After the user interaction, a post-test interview was done in order to better understand the user experience while interacting with the system.

After conducting all the tests, CEM was analyzed and consolidated by the author of this study. For each user the following steps were performed: (1) video tagging; (2) interview transcription; and (3) registration of problems found by the user in a spreadsheet. A total of **132** problems were found using CEM (considering the sum of the problems encountered by each user). After the analysis, all the problems were consolidated in a list of **30** unique problems.

It is noteworthy that CEM analysis was done before SIM to reduce bias in the results (i.e., so the author would not be influenced by the problems encountered with SIM). Although the author had conducted an inspection using SIM in the first step of the empirical assessment (i.e., "Target System Definition"), we believe that performing the analysis of CEM previously would minimize bias in the results since in the analysis of SIM who points the problem is the evaluator, then the author, analyzing the problems encountered, could not add problems that were not described by the evaluators. However, in CEM the evaluator is responsible for identifying and characterizing the problems experienced by users. Thus, it could be influenced by previous knowledge of existing problems by the evaluator.

Table 7.5 presents the CEM problems consolidated and their severity level. According to de Souza [2005], there are three main communicative categories in communicability evaluation and they correspond to different levels of severity: complete failures (high/major); temporary failures (medium); and partial failures (low/minor). As described in Section 7.1, each tag is associated to one of these categories. Then, since some problems generated different breakdowns (as described on the "Evidence" column on Table 7.5), the problems' severity was defined based on the breakdown generated most frequently by the problem. Since we will compare CEM and SIM results, we marked with a gray background the problems not identified by SIM (their analysis will presented in the Section 7.2.4).

Table 7.5: CEM Problem Set.

ID	**Problem**	**Evidence**	**Severity**	**# of Users Affected**
1	The system presents as the home screen, whenever the user logs into the system, the help system in the central part (as shown in Figure 7.4). The system features are available only in the left column. At this moment users became lost in relation to what should be done and what that information represented.	There is a high occurrence of the expressions "What is this?", "Where is it?" and "What now?" at the beginning of users' interaction until the moment they really find out where to start. Some users also mentioned the problem in the interview.	Temporary	9

Continued on Next Page...

ID	Problem	Evidence	Severity	# of Users Affected
2	The system presents a concept in which the tasks must be added within projects (there is no option to add tasks without associating them to a project). This concept is not in accordance with users' expectations. Users commented that they believed it would be more natural to think first about the tasks and then organize them into projects.	All users started adding tasks as projects. Some of them believed, initially, that they were doing the correct steps to include a task. After a while they realized that it was not the correct way. At this point they started exploring the interface to find the best way to accomplish the task (many occurrences of the expression "Where is it?") and some visited the help system (occurrence of the expression "Help!"). Some users were unable to execute the task, and found alternative ways to perform it, such as adding the task as project and including its time in its description. However, they mentioned they were aware that was not the expected way to do it (the occurrence of "I give up!" [2]).	Complete	9
3	The system does not make clear to users the ability to add a task at a specific time, and nor how to add recurring tasks (i.e., that occur every day). Users must access the help system to understand that these features exist. Another factor that complicates the users' understanding is the title of this feature in the help "Help on advanced dates? Recurring dates, Contextual dates and more...". It can be observed that the text does not refer to the time feature.	Some users did not identify that the same field in which the due date for a task is defined can also be used to set its time. Thus, they were looking for this feature in the options available at the task menu (the occurrence of "Where is it?"). Many found the feature "Reminder" and believed that this feature would allow them to set the time. Clicking on this option users noticed that it would not work (the occurrence of "I can't do it this way"). Users then went back to the date field and found the help they needed to set the time (occurrence of "Help!"). It is noteworthy that some users were unable to complete this task (the occurrence of "I give up.").	Complete	9

Continued on Next Page...

[2] Although the user's intent was to solve the problem differently from what was expected (would be an occurrence of "I can do otherwise") he/she did not achieve the same effect.

ID	Problem	Evidence	Severity	# of Users Affected
4	Users had trouble creating subprojects. The problem was that the arrow (left and right arrows beside the projects title field – as shown in Figure 7.6 – item 2) icons associated to this feature were not intuitive to users. Furthermore, they had no metalinguistic message (i.e. tooltip or instruction) to explain them. Even when interacting with the icons it was not evident to users what they referred to. Moreover, the help system shows step by step how to create a subtask and does not explain how to create a subproject.	Before even performing the task of creating a subproject users hover over the arrows and click on them in order to understand their functionality (the occurrence of "What is this?"). One participant even commented *"This is unnecessary isn't it?"*[3] (U3). When performing the task of creating a subproject some users got confused with the option "Add project below" that enables them to add another project below (the occurrence of "I can't do it this way."). Others even accessed the help system (the occurrence of "Help!") but still were unable to understand how to create a subproject. They then decided to organize themselves by creating a subtask aware that it was not what had been requested, but was the only way to get organized (occurrence of "I give up."[4]).	Complete	9
5	The system designer uses the Windows Explorer metaphor in which the page is separated into two frames: the left contains projects and on the right the tasks associated to a project (similar to the structure of folders on the left and files on the right in windows explorer).	This metaphor was not completely built and users tried to drag and drop tasks between projects and use the right mouse click (occurrence of "Why doesn't it?") which is available in the windows explorer and not available in the system.	Temporary	9

Continued on Next Page…

[3]Original text in Portuguese: *"Isso aqui é desnecessário, né?"*

[4]Although the user's intent has been to make an "I can do otherwise" (trying to solve the problem differently from what was expected) he/she did not achieve the same effect.

ID	Problem	Evidence	Severity	# of Users Affected
6	Users became confused about how the cancel option worked while adding a new task. In the project insertion window there is an option "Cancel" next to the Add Project button (as shown in Figure 7.6), which is a clear and users did not present difficulties related to it. However, in the tasks insertion window the command is "I'm done adding tasks" (as shown in Figure 7.8 – item 2). Some users had difficulty understanding that this command would cancel the tasks entry window.	After inserting tasks, some users looked for the Cancel button (the occurrence of "Where is it?"), and, as they did not find it, they clicked again on the title of the project to refresh the screen and thus close the window (occurrence of "I can do otherwise."). Other users expected that the "I'm done adding tasks" (as shown in Figure 7.8 – item 2) command would save the task and close the task insertion window. However, when users filled out the fields and clicked at "I'm done adding tasks" option the window was closed but the task was not saved (the occurrence of "What happened?").	Temporary	8
7	There is a breakdown in the users' semiosis when a task was marked as completed. Once users discovered the option represented by a gear containing the tasks options (as shown in Figure 7.9), they used it whenever they wanted to perform an action with the task. However, when a task was marked as completed the gear no longer appeared.	When the user wanted to perform an action with a completed task they hovered over the region where the gear was shown before (occurrence of "Where is it?"). When they realized that the gear was not displayed anymore they started wandering on the screen (the occurrence of "What now?"), and some of them even tried to find the gear again ("Why doesn't it?").	Temporary	8
8	The lack of tooltips in some signs made users click on them trying to understand what they meant.	Users hovered over the signs in an attempt to understand some of the interface elements ("What is this?"). The signs that had a tooltip were less clicked than the ones that did not have it.	Temporary	8

Continued on Next Page…

ID	Problem	Evidence	Severity	# of Users Affected
9	The system provided a format for entering date and time of tasks quite flexibly. The user could enter the date separated by slashes (i.e., 09/09/2011), textual (i.e., 10 Oct 2011), using a calendar that is displayed by clicking on it or even by entering keywords in English that refer to time (i.e., tomorrow, today). With respect to hours, the user must enter it after entering the date and use the keyword "at" or @ (i.e., tomorrow @ 10). Because the format is neither explicit nor intuitive (at least to users whose first language is not English), there is a communicability problem, since the user does not understand the format that the system accepts.	All the users went to the help system to understand how to enter the time of a task (the occurrence of "Help"). Some users entered the time in a way not intended by the designer (the help system does not mentioned this format), but the system interpreted correctly (occurrence of "I can do otherwise."). Users ended up finding that the format with the word hour written in Portuguese was correctly understood by the system Example: "10/10/2011 @ 10 **horas**".	Partial	7
10	The options menu associated to a project is accessible via a gear sign that only appears when the user passes the mouse over the title of the project. The same happens with the options menu associated to tasks. Users had difficulty finding these options menu, since it is only represented by a dynamic sign.	This problem occurred when users wanted to edit or delete the project and were looking for these functionalities (the occurrence of "Where is it?").	Temporary	5

Continued on Next Page...

ID	Problem	Evidence	Severity	# of Users Affected
11	The use of the same sign to represent different functionalities confused users. The gear sign is used to represent the options menu of projects and tasks, as well as to represent the tasks actions menu (as shown in Figure 7.5 – item 4), preferences and logout functionalities.	This problem is, somewhat, related to the previous problem. This gear sign may refer to setting options, and, therefore, since the options menu of projects and tasks were hidden, users believed that the gear sign that represented the tasks' actions, which was not hidden, (see Figure 7.5 – item 4) would be a way to access the options menu (occurrence of "Where is it?"). For instance, the users were trying to find the project options menu and tried to click on the gear sign available close to the title of the project. Since the sign was close to the project title and with the same gear sign, the user though it would be the projects option. However, it was not what they wanted. One user even tried to click on other gear option available at the top right part of the system but, again, it was not the project options menu. This user U9 also mentioned in the interview: *"The symbol is in the same side with three complete different meanings."*.	Temporary	5
12	Users had difficulty understanding how to mark/unmark a task as completed. The checkbox used was not clear to users, and some chose to follow a different and longer path that would get to the same result, using the "Move to history" feature.	Some users typed the word "completed" in the task field (occurrence of "Where am I?" because the system used keywords for actions, but this action has no keyword – and the users got confused). Other users went in the task menu and clicked on "Move to history", which is correct, but it's a longer way (occurrence of "I can do otherwise"). One user did not find any of the options offered by the designer and gave up completing the task ("I give up.").	Complete	5

Continued on Next Page...

ID	Problem	Evidence	Severity	# of Users Affected
13	Users who understood how to mark a task as completed (using the checkbox) reported another difficulty: assimilating the concept of the functionality "Move to history." They had expected it to be associated to a different functionality.	When asked what the "Move to history" feature was some users believed it would remove a not completed task to a history of pending issues (that could be recovered later).	Complete	5
14	The system's search field did not make clear to users what the expected search format was. Since the field was an input text and users were asked to search for the tasks that were due "tomorrow", they got confused with what patterns could be used.	Users hovered and clicked on the pre–defined search options (the occurrence of "What is this?"). One user asked if he should put the date in English or Portuguese (U3). Other users went to the help and found out how to do it (occurrence of the expression "Help!").	Temporary	4
15	In the insertion window of a project or task the system displays two buttons "Add project" (see Figure 7.6) and two buttons "Add task" (see Figure 7.6), respectively. This duplicity of buttons confused users when adding items.	Users were not sure which button to click to actually add the project or task. Users hovered over both buttons and then chose one (occurrence of "What now?").	Temporary	3
16	The system's feedback about the recurring task is not clear. First, when a recurring task is created, the system only creates one item (not one for each recurring instance). When users perform a search, for example, for the tasks of the next 7 days, only one item is shown (even if it is a task to be executed every day in this period). Second, when users mark the task as completed, the recurring task does not leave the task list, only the date is postponed to the next day.	After users created a recurring task they did not understand if it was created properly and tried to make sure of that. Additionally, when users marked the task as completed they did not understand why the task continued to be in the list. Some of them clicked more than once to verify that the date was postponed (occurrence of "Why doesn't it?").	Temporary	3

Continued on Next Page...

ID	Problem	Evidence	Severity	# of Users Affected
17	The sign used to reorder items is not clear. Users, at first, believed that the interaction would be through clicks on the arrows to move them up or down. However, the reordering only occurs through the interaction style "drag and drop."	Users clicked on the arrows to move the items within the list, only to realize that nothing happened (the occurrence of "What happened?").	Temporary	3
18	When adding a new task in the system it is automatically created as priority 4 (i.e., black color). This is not communicated to the user, who has no option to change the priority during the creation of the task, but only by editing it after it is created. In addition, in the edition menu the current priority of the task was marked with a subtle yellow shade in the tasks option menu (see Figure 7.9), which turned out to be imperceptible to the user.	Even with the task classified as priority 4 (shown on options menu with a yellow shade – see Figure 7.9) the user clicked on it again. That is, the user was not able to realize that the yellow shadow in the task options menu corresponded to the current priority of the task (occurrence of "Why doesn't it?").	Temporary	3
19	Clicking on the system's logo updates the current page of the system. Despite having an associated tooltip that explains it, users clicked on the logo with the intent to return to the system's initial page (in this case, they wanted to access the help system), which is the most common function associated to logo's on web systems (i.e., the logo takes back to homepage).	Some users clicked on the logo repeatedly believing that the system was malfunctioning (occurrence of "Why doesn't it?"). One user commented *"Just for your information, if I could get back to the initial page I would watch the video again."*[5] (U8).	Partial	3
20	Users had difficulties in understanding what the command "Delete completed" is. This command deletes all the tasks marked as completed. The difficulty probably because the command was ambiguous. Even participants who had advanced knowledge in English had this difficulty.	In the tasks in which users had to delete completed tasks they chose to unmark the task and delete it through the menu options (occurrence of "I can do otherwise.").	Partial	3

Continued on Next Page...

[5] Original text in Portuguese by U8: *"Só para a sua informação, se eu conseguisse entrar na página inicial eu iria assistir o vídeo de novo."*

ID	Problem	Evidence	Severity	# of Users Affected
21	Even when users created the subproject correctly, the communication of the relationship between the project and subproject was not clear to them.	When reordering the projects that have a subproject, users realized that the subprojects associated to a project did not move along with the project and got confused (occurrence of "What happened?"). Some even commented *"... it does not make sense to me that a subproject stays in a different order in the hierarchy in relation to the project."*[6] (U4).	Temporary	2
22	Users had trouble memorizing the colors of each priority level.	When users were sorting the tasks by priority they returned to the menu to remember the colors associated to each level.	Partial	2
23	Help system was inconsistent with interface. Thus, users who accessed the help system to watch the video in order to understand how to perform a task (e.g., subproject or menu options) could not identify the interface signs shown. Thus, they watched it again before they tried other options.	Some users had to go to the help system at least twice before they could figure out how to do an intended action.	Temporary	2
24	The system allows users to use shortcuts. However these shortcuts are explained only in the help system, and are not easily memorized and confused users.	A user watched the introduction video, but became confused with the shortcuts mentioned and did not memorize the shortcut he should use. He then verbalized: *"Is it Ctrl + Shift? What is it? [The shortcut] to Save I did not memorize."*[7] (U3). The user then decided to click on "Enter" to save the task ("I can do otherwise").	Partial	1

Continued on Next Page...

[6] Original text in Portuguese by U4: *"...para mim não faz sentido um subprojeto estar numa outra ordem que não seja a de hierarquia em relação ao projeto."*

[7] Original text in Portuguese by U3: *"É Ctrl+Shift? Como é que é? Para salvar que eu não gravei"*

ID	Problem	Evidence	Severity	# of Users Affected
25	The system has four priority levels numbered from 1 to 4. However, there is no explanation in the interface about which one is the highest priority (only the introductory video presents this information).	A user commented *"I do not see anywhere what the definition of priority 1, 2 ... The natural order of things is to think that the most urgent is 1."*[8] (U7). The user then established that 1 would be the most urgent ("I can do otherwise."[9]).	Partial	1
26	Users had difficulty in understanding what the number after the word "Today" in the search referred to. This number indicated the number of tasks for today, but it was unclear to users.	A user commented *"What is this number two here?"*[10] (U6) (occurrence of "What is this?").	Temporary	1
27	The functionality priority was imperceptible to some users.	When asked to define priority, a user chose to reorder the tasks instead of marking their priority levels (occurrence of "Looks fine to me.").	Complete	1
28	When a user registers a task with a date before the current date the system changes it automatically to the following year. The problem is that no message is shown and the user may not notice the change.	During the test the user selected in the calendar a date before the current date and saved the task. The system than automatically changed the date to the following year. For example, the user selected the date "5 Dec 2012" and when he/she saved the task the system changed it to "5 Dec 2013". When the user searched for the tasks he noticed that they were not displayed. He/she thought it strange (occurrence of "What happened?"), but did not try to investigate what had happened.	Temporary	1

Continued on Next Page...

[8]Original text in Portuguese by U7: *"Eu não vejo em nenhum lugar o que me defina se é prioridade 1, 2... A ordem natural das coisas é a gente pensar o que é mais urgente é 1."*

[9]The user could not understand the designer intent and set a rule that made sense to him/her. Although the user was able to do as it was expected, it was not intentional (i.e., it was by chance). For this reason we defined this problem as "I can do otherwise".

[10]Original text in Portuguese by U6: *"Esse dois aqui é o que?"*

ID	Problem	Evidence	Severity	# of Users Affected
29	It was not clear for users whom the system was meant to.	User U9 asked during the interview to whom the system had been developed and said that he believed the system was not designed for people working in industry.	Partial	1
30	When users clicked on "Todoist for Chrome" they became confused with the "Download Now" that appeared.	The user thought that he/she should install the software in the browser to use it. However, this installation is optional and is only for those who want to manage their tasks using a plugin rather than entering in the system web page.	Temporary	1

SIM was analyzed and consolidated by the same author who made CEM analysis in order to maintain the same judgment criteria. The following steps were conducted: (1) analysis of each problem to ensure that each item actually corresponds to a single problem; (2) elimination of problems that were not in the scope of the analysis (i.e., suggestions, general problems, problems related to English understanding and accessibility assessment – shown on Table 7.6); (3) labeling of each item with a term that summarized the problem described; and (4) consolidation of the problems set by label.

In total there were **103** problems reported by evaluators using SIM and they were consolidated into a list of **47** unique problems. Table 7.7 presents the SIM problems consolidated. We also categorized each problem according to its severity. Although SIM does not have a pre-defined way to categorize the problems in respect to its severity, we used the same categories used for CEM to enable the comparison of the problems. In this case, the problems that were also found by CEM were categorized with the same severity level since it was possible to observe what happens when the user faces the problem. The categorization of the other problems was done by the author of this study.

We matched the problems that both SIM and CEM found and then we marked with a gray line the problems that were identified by SIM and not by CEM. We provide an identifier of the corresponded CEM problem, in the column "CEM ID", for those problems that were also found through CEM evaluation. We also present if the problem

Table 7.6. SIM Problems Eliminated.

ID	Problem	Justification for Elimination
1	In order to guide users in using Todoist, the designer provides the help system with a video of the main system features and how they can be used. However, the fact that this information is available only in a video format can make them inaccessible to deaf users.	This problem was eliminated because the deaf user's profile was not the inspection focus of this analysis, and known by the evaluator (he mentioned in the report that this problem is beyond the scope of evaluation). For real assessments this would be relevant, but, in the case of this experiment, it was eliminated for comparison purposes.
2	Todoist does not support the world's most used browser. Firefox and Chrome are cited, but Internet Explorer is not.	The evaluator mentioned in the report that this problem is not a communication breakdown, but a decision of the designer. That is, the evaluator is aware that the problem is beyond the scope of evaluation.
3	In addition to the existing filter options, it would be interesting if designer provided a calendar as a filter option to search for tasks. Thus, the user could select the desired date from the calendar.	Redesign suggestion.
4	A possible suggestion to circumvent the communicability breakdown that may arise from features that are available only for Premium accounts would be not to display these signs in the free account or keep them disabled with a tooltip explanation associated to them.	Redesign suggestion.
5	The project offers the possibility to create a hierarchy among the projects. But while on the left side of the interface the designer displays a tree that describes this hierarchy, it would be interesting to reinforce this information (e.g., with the use of "breadcrumbs") in the header preceding the tasks´ lists of a subproject. Currently the designer chose to display only the name of the subproject. Such a decision may confuse the user, even momentarily.	Redesign suggestion. Although we eliminated the evaluator suggestion of redesign, the problem reported in the last sentence of this excerpt (i.e., "Currently the designer chose to display only the name of the subproject. Such a decision may confuse the user, even momentarily.") was considered as a problem of the system.
6	The phrase *"Create subprojects and subtasks with ease using keyboard shortcuts or your mouse. Structure your tasks in any way you like."* may not be clear to the user because the phrase refers to keyboard shortcuts and the user may think this is obvious: "Of course I can edit with the keyboard, I cannot type with my mouse.".	In this problem was noticed that the evaluator had problems interpreting English. It can be seen clearly in the phrase that the user has two options for interaction: keyboard shortcuts or mouse clicking.
7	The meta-message reconstructed from metalinguistic signs indicates that the tool was built for users who want to manage their tasks easily and efficiently. However, when looking at the meta-message reconstructed based on static signs it is possible to observe that the user can sometimes get lost, without knowing the meaning of terms used in the interface. By analyzing the meta-message reconstructed from the dynamic signs it is possible to realize that sometimes the system can take the user to perform unwanted actions.	The evaluator did a general appraisal about the system communicability. He/she did not describe a specific communicability problem.

was found by experts or novices, respectively, in the columns "Experts" and "Novices" where we put "X" if the problem was found and "-" if not. In addition, the total number of evaluators who found the problems is also shown in the column "Total # of Evaluators" (considering experts and novices).

Table 7.7: SIM Problem Set.

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
1	The option of defining a project/task as subproject/subtask (right and left arrow that appear on the right of the box – shown in Figures 7.6 – item 2 and 7.8 – item 3) can only be understood by the user through "trial and error" or if he sees the introduction video, which may not occur. Even interacting with the system the functionality may not be perceived by users, because they may not realize that this movement is equivalent to changing the level of the project (making a subproject).	Complete	4	X	X	6
2	At the time of a task creation the designer offers the user the possibility to assign a due date to the task. However, in his message, he does not convey to users the information about how to assign a specific time to the task. The same occurs with the functionality of adding recurring tasks (i.e., that occur every day). The information about these features can only be obtained through the introduction video or clicking on "Help on advanced dates?". Even in this case, as can be seen, this link refers to "Advanced Dates" and does not refer to the possibility setting a time.	Complete	3	X	X	6

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
3	In the instruction video users are informed that to register they should click on "Sign Up for Free", referring to the idea that the system is free. On the home page the button to get registered is "Sign Up Today", leaving it unclear whether the system is free or not. After login into the system, on the main screen, users can see that there is a Premium account (i.e., paid account). This uncertainty about the types of accounts and what features are offered in each type of account may create false expectations to users, causing them to give up using the system.	Temporary	-	X	X	6
4	The search filter is quite poor of static and dynamic signs. Based only on the static signs it is very difficult to understand how it works. Even during the interaction it is difficult to know how the filter can be used, especially because there are many keywords that can be used and there is no information about how to use them. Communication is also deficient in metalinguistic signs because it does not say anywhere in the system that the search only returns tasks that were not completed and that a comma is used to concatenate filters.	Temporary	14	X	X	5
5	Some signs only appear when hovering over a project or task description (i.e., gear). These signs are hardly perceived by users because they do not have a static representation, and are only seen during the interaction. This lack of communication worsens because the sign represents the project/task configuration which is a very important feature of the system. In addition, the color of the sign resembles disabled items and this may lead the user to think that the option is disabled, when in fact it is not.	Temporary	10	X	X	5
6	In both the addition of project and task, the action of adding an item appears duplicated, either being displayed as button or as a link with the same name ("Add project" and "Add Task"). Despite the fact that the dynamic signs demonstrated that both represent the same function, this may confuse the user.	Temporary	15	X	X	4

Continued on Next Page…

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
7	The introductory video presents some interface elements that are different from the ones used on the interface as, for example, the settings option (this option in the video is presented as a down arrow and on the interface is presented as a gear) and others (e.g., the names "projects" and "tasks" shown on the interface in the settings option are shown in the video only with the name "item"). This inconsistency can confuse the users.	Temporary	23	X	X	3
8	Each task has a checkbox in front of the description, but no information regarding what it is related to is given. Only by marking one of them can users understand that they cause the task to go to the list of completed tasks. The user must use the strategy of "trial and error" to understand this feature.	Complete	12	X	X	3
9	The same symbol of a gear is used to represent different functions in the system. In one place (in the right corner of a task or project) it represents a task or project settings, in the other (near the printer symbol) it represents ways to sort the task list. If the user is familiar with the symbol representing the settings of an item, he may not realize that there is a gear symbol representing ways of organizing the task list. He may, for example, want to sort a list by due date and do it manually by choosing "Reorder", unaware that there is an item for this ("Sort by date").	Temporary	11	X	X	3
10	Some signs used to filter tasks are clearly presented to users. For example, the sign "@" and "q:" do not inform users what type of refinement can be done through these options. Only when the user accesses the help will he understand that the "@" means to show all tasks labeled and that the sign "q:", in turn, allows him to search by task description. In addition, the filter "Today" could be interpreted as a filter to retrieve all tasks created today, when in fact it will retrieve the tasks due today.	Temporary	-	X	X	3

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
11	The sign "@" and "q" available below the text box "Filter Tasks" suggest that the user can apply them to refine his search for tasks. However, if the user goes to the Todoist introduction video, he will realize that these options are not available in the free version. Only accessing the help he will understand that these options are only available for users who have a Premium account.	Temporary	-	X	X	3
12	The designer uses different types of buttons/links: red background, black background, black link, red link, blue link, red link, gray link and conventional button. It was not possible to reach any conclusions regarding any coded use of colors. The lack of a standard can cause problems in user interaction. For instance, through the exploitation of static signs, users could assume that the gray links meant that they were disabled. When exploring the dynamic signs it was revealed that some fields depicted in gray were really disabled, whereas others were enabled.	Partial	-	X	X	3
13	If the user selects a date in the past (i.e., prior to the current date) the system always interprets this date as a future date, registering a date with the same day and month, but the following year. The system does not invoke users attention to what was registered (the task is registered one year after the due date). The user will only notice that the date was recorded for the following year if he pays close attention to the date when registering. However, he may not find the task afterwards and not understand what went wrong.	Temporary	28	X	X	2
14	To change the priority of a task users must go to item settings (only available after the task creation). There is no way to assign a priority during the insertion of a task. In other words, it is not communicated to the user that this option exists and he will only learn it if he gets curious about seeing the task settings.	Complete	27	X	X	2

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
15	When the user associates a priority to a task the system changes the color of the task description to show that it has a different priority. This requires the user to memorize/remember that the tasks with the red text have priority 1, blue have priority 2, green have priority 3 and black have priority 4.	Partial	22	X	X	2
16	By checking on the checkbox the task is marked as completed, removed from the tasks list and displayed in the completed list. The user can also use the "Move to history" option in the menu to mark the task as completed. This concept that when a task is completed it is moved to a history is not clear in the system and thus users may find it difficult to understand the feature "Move to history". Users may need to read the help system to understand how it works.	Complete	13	X	-	2
17	There are no standards among metalinguistic signs (i.e., tooltips) used to explain some elements of the interface. For example, the symbols "@" and "q:", located near the filter bar, have an explanation but the element "more" does not. In addition, tooltips are not always very informative.	Temporary	8	-	X	2
18	The interface does not make clear to users the definition of projects or tasks. Thus, users may assign different meanings to them. The only explanation offered through metalinguistic signs is that tasks are created within projects. However, that information may go unnoticed if users do not read the help system.	Complete	2	X	X	2
19	The designer communicates in the help system that it is possible to format the description of the tasks (e.g., italics, bold, link) using a special syntax. However, this syntax is only communicated through the help system. Thus, if users do not access the help system they may never learn this possibility.	Temporary	-	X	-	2

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
20	The designer defined that the "@" character followed by a word (i.e., @test) allows the user to create labels that can be assigned to tasks. However, this information is only conveyed through the introductory video, and the user may never become aware of it.	Temporary	-	X	X	2
21	The system works differently when adding/deleting projects or tasks. To add multiple projects users should add one at a time, repeating the steps "Add project", inform the project name and click "Add project" to save. In the case of adding tasks users click on "Add task", enter the desired description, and click on "Add task" to save the task and the system automatically displays a new add task field allowing users to enter all the intended tasks. Once users are finished they must click on "I'm done adding tasks.". When deleting the system displays a confirmation screen only when users are deleting a project. This does not occur when excluding tasks.	Partial	-	X	X	2
22	When searching for a word that was in the description of one or more tasks the system displays a message "data error". However, what should be informed to users is that this function is only available in the Premium version.	Temporary	-	-	X	2
23	The designer uses terms that may not be very common to users as, for example, "plugin" and "data center" (even for English speakers), since these terms are technical terms and may not be known to users who are in the technology field. Once Todoist is designed for any users who want to manage their tasks, there may be a communication breakdown.	Partial	-	X	X	2
24	The help mentions that Todoist can be used in Firefox, Chrome, MAC, etc. The information is presented in a way that the user may find that the system must be installed before starting to use. However, this is not the case because it is an online web system. What the designer probably intended to communicate was that there are plugins that can be added and used in addition to the online version.	Temporary	30	-	X	1

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
25	Some evidences show that Todoist was meant to manage professional projects. This evidence is highlighted on the system's initial page. The system is also meant for personal use, to see this information you must click on "About us". This evidence can cause a communication breakdown because as soon as the user enters the system he is informed that the system is used by employees in several companies while the information that the software is also for personal use has a much lower visibility. This breakdown can lead users to think that the system is not for them and give up on trying it.	Partial	29	-	X	1
26	The static sign chosen to represent the function of creating tasks through email may not be very intuitive. Only by clicking on the email option it is possible to understand that the user can register tasks using e-mail. Initially, when looking to the static sign, the user may think that the feature is meant to send tasks list to his e-mail, since it uses a graphical representation usually associated to sending email.	Partial	–	-	X	1
27	The designer offers users the possibility to prioritize his tasks into four levels, from 1 to 4. However, there is no communication to users which of these is meant as highest priority. He can only identify that the first priority depicted (priority 4) is the highest one through metalinguistic signs.	Partial	25	X	-	1
28	The decision of the designer to provide keyboard shortcuts to access some features of Todoist is interesting (i.e., CTRL + M to newline). However, this information is only available in the help system and is not available through other metalinguistic signs (e.g., tooltips) or even through static signs. So if users do not access the help system they may never learn about the existence of this feature.	Partial	24	X	-	1

Continued on Next Page…

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
29	The system offers the possibility to create a hierarchy among projects. On the left side of the interface the system displays a tree that describes this hierarchy. Despite this communication, in some points of the system this link between project and subproject is not clear. For example, when selecting a subproject in the tree the page title becomes the name of the selected subproject, not indicating to which project it belongs to. The same occurs when the user performs a search for all the tasks, receiving as a result the list of tasks organized by projects but does not differentiate projects and subprojects. These design decisions can hinder the user's understanding about the relationship between projects and subprojects.	Temporary	21	X	-	1
30	The "Delete completed" link provided by the designer aims at deleting completed tasks can have an ambiguous interpretation: the action of deleting something that is complete (which is the right interpretation) or a feedback message stating that the requested deletion was performed.	Partial	20	-	X	1
31	Usually the logo allows the user to return to the home screen. However, the system displays the logo as a way to update the current page which can be confusing to users. To access the help system again (which is displayed at the initial page of the system) the user must click the help sign.	Partial	19	X	-	1

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
32	Although the system explains that it can be used to manage professional or personal tasks, it does not clearly convey who it was designed for. The designer says that it is used by millions of people around the world which suggests the idea of a system used in several countries and therefore different cultures, and can be used by people who do not have English as a first language. However, the shortcuts used by the system may hinder the user interaction, since some of them require knowledge of words contractions (which are usually only known to people with advanced English knowledge). Example: "ev weekday", where "ev" means "every." Another similar example is the use of "tues" to represent "Tuesday". The system provides instructions on these shortcuts. However, the point is that they cannot be intuitive for some users as the system is meant to be.	Partial	9	X	-	1
33	The Cancel option available when creating a task may not be clear to users. The designer clearly offers the user the possibility to cancel the creation of a project through the link "Cancel". However, he uses different codes for projects ("Cancel") and tasks ("I'm done adding tasks") This may cause users to have doubts whether there is or not a possibility to cancel.	Temporary	6	X	-	1
34	When you save a change in the description of a task, a message appears stating that it was "Moved to <date>" even if no change has been done to its due date. This can induce users to think that she inadvertently changed the date or that it was changed by the system.	Temporary	-	-	X	1
35	The designer shows to users the total number of tasks that they have concluded. However, even temporarily, the user can interpret that the information is contextualized by project and not in general. Over time the user can understand that this information refers to all the tasks (including deleted tasks), but it may still cause an initial problem.	Partial	-	X	-	1

Continued on Next Page...

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
36	When users delete their completed tasks within a project the system resets the number that appears in brackets to the right of the label "Show completed tasks". However, when the user selects another project and then returns to the first, the system displays the total number of completed tasks, including those that were deleted. It also allows users to click on the link "Delete completed", which can confuse them even more.	Partial	-	-	X	1
37	Todoist does not allow selecting multiple tasks so actions like "delete" cannot be applied simultaneously to more than one task in the same project. The designer decision may hinder the work of a user who wants to, for example, delete multiple tasks simultaneously.	Partial	-	X	-	1
38	When users click on the menu item "Move to another project" (on the task settings) the system displays a dialog box showing the projects users can move the task to. However, the system lists among the projects the project to which the task already belongs to.	Partial	-	-	X	1
39	The system displays static signs for filtering tasks when the user has not yet added any tasks or projects.	Partial	-	-	X	1
40	It is unclear whether the functionality "Print" would print the current view (full screen) or just the tasks currently displayed.	Temporary	-	-	X	1
41	Both the initial page of the tool, which is displayed before the user logs in, and in the help page, conveys the possibility of integrating Todoist with Gmail and Outlook. However, signs were not identified as of how such possibility could be put into practice.	Temporary	-	-	X	1
42	When the user clicks on the link "Show completed tasks" the system displays the list of completed tasks of the selected project, but it does not allow hiding the list again.	Temporary	-	-	X	1

Continued on Next Page…

ID	Problem	Severity	CEM ID	Experts	Novices	Total # of Evaluators
43	The error message when reporting an invalid password does not follow the pattern of error messages when users change the password (displayed in a box). The lack of pattern may cause confusion to the user.	Temporary	-	-	X	1
44	Todoist does not provide any warning if users create two or more projects with the same name or two or more tasks with the same name associated to the same project.	Partial	-	X	-	1
45	There is no explanation of what the designers have implemented in the system to ensure privacy.	Partial	-	X	-	1
46	The metalinguistic signs do not explain how the user can become more productive using the system.	Partial	-	X	-	1
47	It is possible to organize tasks by creating subprojects and subtasks. Besides the term subprojects and subtasks, the designer uses terms such as group of projects, group of tasks and tasks hierarchy that seems to be the same thing. Using several different terms, even in the same sentence, it may cause confusion to the user.	Partial	-	-	X	1

7.2.4 Comparison

After the consolidation of the two sets of problems we did the comparison between the methods. As previously mentioned, the purpose of this comparison is to identify SIM and CEM strengths, weaknesses, differences and similarities.

The first step was defining how the methods could be compared. As shown in the Related Works (Chapter 3), some researchers criticized the methodology of previous works that aimed at assessing usability methods [Gray and Salzman, 1998]. Moreover, researchers [Sears, 1997; Hartson et al., 2003] proposed measures to assess usability evaluation methods (e.g., validity, thoroughness, effectiveness). However, the HCI field still needs further studies related to methodologies to support researchers in conducting assessments of new methods that have emerged recently, as well as those that take into account properties other than usability.

Semiotic Engineering methods have been proposed recently and criteria applied to comparing usability evaluation methods may not be the most appropriate for communicability evaluation methods [Salgado et al., 2006]. Further investigation on the applicability of these measure proposed by [Sears, 1997; Hartson et al., 2003] to the context of communicability evaluation methods are still needed. Therefore, we decided to carry out an assessment of SIM aiming at identifying its characteristics regarding some of the topics discussed in the SIM's assessment from evaluator's perspective discussed in the previous Chapter and other topics that we thought would be interesting to discuss. Next we will analyze SIM in the following topics:

- **Types of problems identified by each method:** a qualitative analysis to find the differences between the nature of the problems identified by SIM and CEM;

- **A severity analysis of the problems encountered:** an analysis and contrast of the severity of problems identified by both methods;

- **Evaluators influence in SIM application:** we present an assessment of the evaluators expertise influence during SIM evaluation and an average proportion of problems found as a function of number of evaluators; and

- **Time/Effort needed to apply SIM:** we present the time and effort need to apply SIM and CEM, and an analysis of the cost-effectiveness of SIM in relation to CEM.

7.2.4.1 Types of Problems Identified by Each Method

Raw counts of problems have limited reliability and validity as an indicator of the quality of a method [Gray and Salzman, 1998; Hornbæk, 2010]. Hornbæk [2010] suggests that the counting of problems is accompanied by an analysis, called by de Jong and Schellens [2000] as congruent validity, which checks the similarity and differences between the problems found by different methods.

Therefore, we did a qualitative analysis to find the differences between SIM and CEM. In order to do so, we associated each SIM problem to a CEM problem and we encountered **23** equal problems (Figure 7.11). In other words, from **47** problems identified by SIM **24** were not identified by CEM and from **30** problems identified by CEM **7** were not identified by SIM. Next we describe the differences between SIM and CEM.

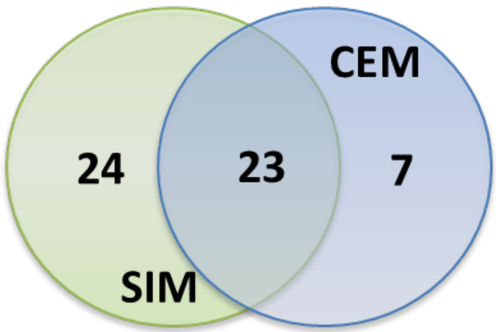

Figure 7.11. SIM and CEM intersection.

Table 7.8 presents the types of problems encountered by SIM and not encountered by CEM. As we can see problems were classified in four different types: (1) metalinguistic signs, which are problems related to the system's documentation, instructions and messages; (2) bugs and potential use problems; (3) problems encountered because SIM allowed a broader assessment; and (4) minor problems. Next we describe each type of problem.

Table 7.8. Types of problems encountered by SIM and not by CEM.

SIM and not by CEM	# of problems	SIM ID
Metalinguistic Signs	6	19, 20, 23, 45, 46 and 47
Bugs and Potential Use Problems	10	11, 22, 34, 35, 36, 37, 38, 39, 42 and 44
Broader Scope	4	3, 10, 40 and 41
Minor Problems	4	12, 21, 26 and 43

Metalinguistic Signs SIM allowed finding problems related to: (1) hidden features (problems 19 and 20); (2) confusing terms (problems 23 and 47); (3) inconsistencies between help and the system (problems 45 and 46). The first type of problems – hidden features – is related to features offered by the system that are described in the metalinguistic signs, but appear at the interface only as dynamic signs, and which do not have any static sign associated to them. Since, these features do not have a static sign associated to them, they may be difficult to be perceived by the user, or may not even be perceived at all. For instance, the designer communicates in the help that it is possible to format the description of tasks (e.g., italic, bold, link), but while inserting or editing a task there is no static sign that inform the user about this feature. In the second case (i.e., confusing terms) terms used by the designer may not be understood by users, such as "plugin", which could be a common term to developer but not to users.

Finally, in the case of the inconsistencies between help and the system the help system announces specific properties of the system which are not offered by the system. In **todoist** the help system states that the system will increase users' productivity and provide privacy, but it is not clear how the system provides these properties. For instance, users may want to know how their information is being stored and protected (i.e., privacy) and this information is not clear. This could also lead users to give up using the system, if this property is important to him.

Bugs and Potential Use Problems Although SIM focuses on communicability, it allowed evaluators to find bugs and problems regarding how the designer expects the system to be used. In the case of bugs, the evaluators were able to find functional problems in the system such as counting of tasks and a link that did not work (problems 35, 36 and 42) and incorrect (information) error messages (problems 22 and 34). SIM allowed finding problems that could be due to a designer option (problems 38 and 44) such as the case in which the designer allows the user to add project or tasks with the same name and information redundancy, which would characterize it as a bug. In the case of potential use problems, SIM allowed to find problems where the user must take more steps to realize that he cannot use a feature (problem 11 and 39) and of non-existing features that could help improve the user interaction (problem 37). In the former the communication could be done more directly in order to reduce the users' effort in discovering features that he/she could not use. This happened in **todoist** when the designer provides static signs related to features that are available only for premium accounts (i.e., paid accounts). Considering only the interface quality of use we characterize this as a problem of **todoist**. However, it is important to mention that this problem may be also a designer option, since, from marketing point of view, it is interesting to let users know which features of the system he could have if he/she opted to use a premium account. In both cases the problems are related the communicability efficiency. As mentioned by de Souza [2005], one of the points for an efficient communication is to present only the necessary amount of information for a speech to be successful. As we can observe, these problems encountered in **todoist** provides more information than needed and, therefore, may damage user's interaction with the system affecting directly their productivity. One possible explanation for SIM having allowed these types of problems to be identified can be related to the fact that the method does not limit the evaluator's semiosis. Although there is a scenario to guide evaluators, we cannot limit their interpretation in relation to the signs inspected.

Broader Scope Although the same contextualization scenario was used for the evaluation with both methods, evaluation with SIM generated broader results. The reason for that was that users were guided by the general scenario and given specific tasks to be executed. SIM evaluators, on the other hand, often explored more than the main tasks and, thus, encountered problems related to tasks that users had not performed during CEM evaluation (problems 3, 10, 40 and 41). One could argue that users could be asked to perform a larger number of tasks, but that could be unfeasible in terms of the time users would be willing to dedicate do the evaluation, especially when they do not have a direct benefit from the system. For instance, in our CEM evaluation the user session lasted on average one hour, and if it had to be any longer, it could be too tiring for users or difficult to find participants.

Minor Problems SIM evaluators identified problems regarding consistency of interface signs (problems 12, 21, 26 and 43) that did not represent a breakdown for the user. For instance, in **todoist** the user explicitly had to click on “Add project” to include a new project, and repeat this action for the next projects she wanted to include. When including tasks, after one task was included the system automatically opened the dialog box to include another one, and the user should cancel when finished. One may argue that if users did not experience the problem, it may not be worth fixing it. However, we know that these problems may have an impact on learning and remembering the interface. Thus, even if they were not experienced in one controlled test, it could be a problem for the user in the long run.

Table 7.9 presents the types of problems encountered by CEM and not encountered by SIM. As we can see there were two types of problems: (1) problems related to user previous knowledge and experience; and (2) problems intrinsically related the users' experience. Next we describe the types of problems CEM was able to find and SIM was not.

Table 7.9. Types of problems encountered by CEM and not by SIM.

CEM and not by SIM	# of problems	CEM ID
User Interaction Style	2	5 and 17
User Experience	5	1, 7, 16, 18 and 26

User Interaction Style CEM allowed the identification of problems regarding the previous knowledge and interaction experience users were expected to have (problems 5 and 17). For instance, the **todoist** designer used a metaphor to represent the way in which projects and tasks (which can only be created inside a project) are structured and presented. He made an analogy to the folders (projects) and files (tasks) structure used at Windows Explorer. This analogy made users expect the same kind of interaction: through drag and drop and right mouse click. However, the metaphor was not completely built and users had problems (e.g., they tried to drag and drop tasks between projects). Evaluators did not identify this problem through SIM. One possible explanation was that evaluators may have considered that users would be aware that these interactive behaviors are not usually available at web interfaces. However, even though users browsed the web frequently they did not have this previous knowledge regarding web interfaces.

User Experience Some problems experienced by users were not foreseen by SIM evaluators (problems 1, 7, 16, 18 and 26). For instance, the word "Today" was followed by a number that indicated the number of tasks due Today. Some users were not able to understand what this number meant.

The results obtained indicate that evaluators using SIM were able to identify most of the problems found with the application of CEM (approximately 77%). However, some relevant problems experienced by users were only found by CEM. Thus, this shows that ideally both methods should be applied when evaluating a system. Applying SIM before CEM would allow evaluators to identify relevant problems and also focusing CEM evaluation on verifying the real impact of the problems considered to potentially have a negative influence on user experience.

Another interesting aspect is that although the scope defined for both systems was the same, SIM evaluation examined a larger part of the system and, therefore, allowed finding problems beyond the scope considered in CEM. This could indicate that the use of SIM, before the user evaluation, could allow the designer to eliminate a great number of the problems by applying the method in order to fix the major problems. Then, after fixing the problems encountered with SIM, the designer would be able to use CEM in order to encounter more specific issues.

In addition, SIM also detected problems that did not represent a breakdown for users, but could represent a problem in the longer run. Also, it analyzed the help system which is the fallback artifact whenever users have relevant breakdowns and is hardly ever evaluated through user observation.

7.2.4.2 Severity Analysis

According to de Jong and Schellens [2000], without the indication of problem severity, the number of problems detected is not very informative. Therefore, after classifying the problems in relation to severity level, we analyzed the percentage of problems found by SIM in each level. In this analysis we generate the following data (shown in Table 7.10):

- $SIM \cap CEM$ – the number of problems encountered by both methods by severity;
- $SIMnotCEM$ – the number of problems found with SIM only by severity;
- $CEMnotSIM$ – the number of problems found with CEM only by severity;
- $SIM \cup CEM$ – the number of problems found in the total by severity.

Table 7.10. Problems by Severity.

	Complete	**Temporary**	**Partial**
$SIM \cap CEM$	6	10	7
$SIMnotCEM$	0	11	13
$CEMnotSIM$	0	7	0
$SIM \cup CEM$	6	28	20

Analyzing the results, we note that the SIM allowed finding all the "Complete" and "Partial" problems that users had while interacting with the system. In relation to "Temporary" problems, SIM allowed finding 10 out of 17 (61%). Now considering all the problems encountered by both methods (i.e., $SIM \cup CEM$ – 54 problems) we can observe, in Table 7.11, that SIM finds a large number of minor problems. In this assessment we observed that most of the problems encountered with SIM that users did not experienced were "Partial" problems (65%).

Table 7.11. Problems encountered with SIM and not experienced by users.

	Complete	**Temporary**	**Partial**
$\frac{SIMnotCEM}{SIM \cup CEM}$	0%	39%	65%

The results indicate that although SIM allows finding most of the high (major) and medium severity problems experienced by users (which is an advantage of the method), a high number of low severity problems not experienced by users are also detected. One may argue that this might be a disadvantage of the method, however we argue that these low severity problems can be experienced by the user in the longer run. Therefore, although less important, these problems may help designers improve the system's interface to a broader range of users.

7.2.4.3 Evaluators Influence

According to Prates and Barbosa [2007] from the Semiotic Engineering point of view each evaluator generates a set of possible interpretations on the meta-communication. Therefore, it is not necessary to apply SIM with more than one evaluator, since all possible interpretation paths generated by one or more evaluators are plausible. However, an analysis by more than one evaluator allows enriching the results with different views. For this reason, we conducted an evaluation of the evaluators influence in the use of SIM.

The analysis of evaluators influence will be made considering two groups of evaluators: experts and novices. We considered in the experts group the evaluators who had conducted at least six evaluations using SIM and the evaluators considered novices did at most 3 evaluations. In our empirical assessment 3 participants were experts and 4 were novices. Next we present an analysis made inspired on the study done by Nielsen [1992] when evaluating the Heuristic Evaluation method.

In this study, Nielsen [1992] made an empirical assessment conducted by three groups of evaluators with different levels of expertise. The aim was to understand the factors that may influence the probability of finding usability problems. Several factors were considered in his assessment, they are: expertise of the evaluator, severity of usability problems, the individual heuristics, and the activities needed to identify the problems.

Our assessment of the impact of the evaluator's expertise on SIM's evaluation was inspired on Nielsen [1992] method. It is important to note that the number of participants in our study is not statistically significant [11] and, therefore, do not allows us to generate absolute truths about the evaluators influence in the evaluations. However, this is an initial indicator and may be further investigated.

Table 7.12 presents the results separated by groups of evaluators. As we can see no single evaluator was able to find all the problems. It also indicates that experts are better than novices at finding problems with SIM. However, some novices were able to find problems that experts did not find.

As shown, the average performance of individual evaluators is not very good. However, the picture changes if we consider the performance of multiple evaluators. Figure 7.12 shows the average proportion of problems that would be found by aggregating the sets of problems found by several evaluators. We created groups 1, 2, 3 and 4 (the latter was only for novices). For each group size we measured the average

[11] However, since SIM is a relatively recent method, and it is not widely used or taught (yet), we considered that it was worth the analysis.

Table 7.12. The proportion of evaluators who found each of the 47 problems.

Complete

SIM ID	Novices	Experts
1	75%	100%
2	75%	100%
8	25%	67%
14	25%	33%
16	0%	67%
18	25%	33%
32	0%	33%
Average	**32%**	**62%**

Temporary

SIM ID	Novices	Experts
3	75%	100%
4	75%	67%
5	75%	67%
6	50%	67%
7	50%	33%
9	25%	67%
10	50%	33%
11	50%	33%
13	25%	33%
17	50%	0%
19	0%	67%
20	25%	33%
22	50%	0%
24	25%	0%
29	0%	33%
33	0%	33%
34	25%	0%
40	25%	0%
41	0%	33%
42	25%	0%
43	25%	0%
Average	**35%**	**33%**

Partial

SIM ID	Novices	Experts
12	25%	67%
15	25%	33%
21	25%	33%
23	25%	33%
25	25%	0%
26	25%	0%
27	0%	33%
28	0%	33%
30	25%	0%
31	0%	33%
35	0%	33%
36	25%	0%
37	25%	0%
38	25%	0%
39	0%	33%
44	25%	0%
45	0%	33%
46	0%	33%
47	25%	0%
Average	**16%**	**21%**

number of problems found – a problem was considered found if at least one member of the group had found it. As we can see, experts performed better than novices. In addition, we could also observe that 3 expert evaluators (or more) are possibly the best number to carry out an evaluation with SIM since they can find approximately 70% of the problems. In the case of novices, to find the same amount of problems, it would be needed at least 4 evaluators.

7.2.4.4 Time/Effort Needed

We present now the effort spent to conduct the empirical assessment. Unfortunately some records were lost as, for example, time spent preparing the MAC environment and time spent during the introduction of the test (i.e., reading the consent term, answering questionnaire). However, these values were estimated and presented in Table 7.13 and Table 7.14.

Table 7.13 presents the time spent by each evaluator to conduct SIM independently. In total, seven evaluators together spent 131 hours and 7 minutes, which gives an average of 18 hours and 43 minutes (σ 7 hours and 25 minutes) for each one to inspect the system individually. Adding the time spent by the author of this assessment to prepare the necessary materials and to consolidate the problems encountered, in total 151 hours and 7 minutes were spent, which gives an average of 18 hours and

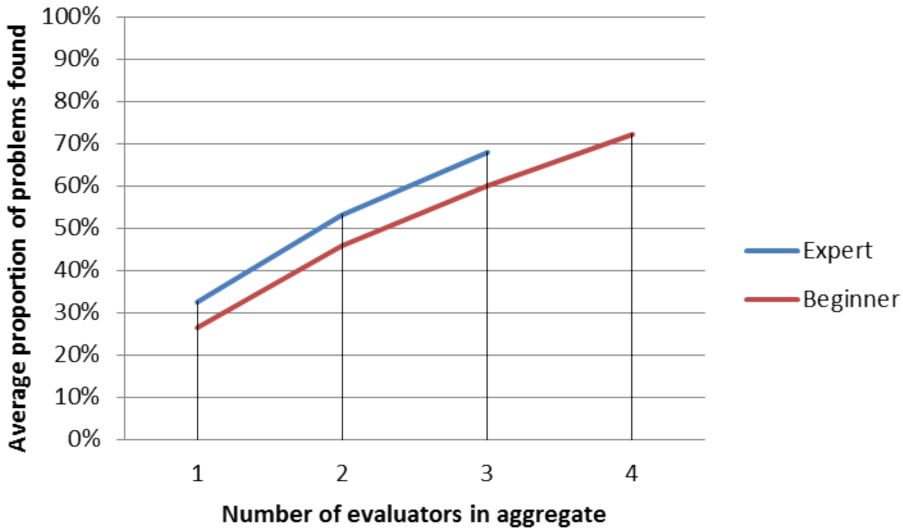

Figure 7.12. Average proportion of problems found as a function of number of evaluators in a group performing SIM.

53 minutes per person involved in the SIM evaluation.

Table 7.13. The effort spent to conduct SIM.

SIM	**Time**
Expert Evaluator 1	28:05:00
Expert Evaluator 2	8:32:00
Expert Evaluator 3	11:30:00
Novice Evaluator 1	16:00:00
Novice Evaluator 2	22:00:00
Novice Evaluator 3	18:00:00
Novice Evaluator 4	27:00:00
Total	**131:07:00**
Average	**18:43:51**
Standard Deviation	**7:25:31**
Preparation: scenario, template and instructions.	4:00:00
Consolidation	16:00:00
Total	**151:07:00**
Average	**18:53:22**

Table 7.14 presents the effort spent to conduct CEM. We can observe that the test execution with all users together with the evaluator time during the test and the tagging and transcription time was done in 70 hours and 14 minutes, which gave an average of 7 hours and 48 minutes (σ 1 hour and 35 minutes) of effort per user. Adding the time spent by the author of this assessment to prepare the necessary materials,

conduct the pilot test and to consolidate the problems encountered, in total 119 hours and 14 minutes were spent, which gives an average of 11 hours and 55 minutes per person involved in the CEM evaluation.

Table 7.14. The effort spent to conduct CEM.

CEM	**Environment Preparation**	**Introduction and Pre-test**	**Test**	**Post-test Interview**	**Evaluator's Time**	**Tagging and Transcription**	**Total**
User 1	0:10:00	0:15:00	0:30:03	0:17:23	1:12:26	4:44:36	6:59:28
User 2	0:10:00	0:15:00	0:42:09	0:14:26	1:21:35	5:39:30	8:12:40
User 3	0:10:00	0:15:00	0:35:55	0:14:02	1:14:57	4:59:42	7:19:36
User 4	0:10:00	0:15:00	0:52:20	0:25:41	1:43:01	7:48:06	11:04:08
User 5	0:10:00	0:15:00	0:38:15	0:10:39	1:13:54	4:53:24	7:11:12
User 6	0:10:00	0:15:00	0:25:56	0:15:19	1:06:15	4:07:30	6:10:00
User 7	0:10:00	0:15:00	0:35:13	0:13:40	1:13:53	4:53:18	7:11:04
User 8	0:10:00	0:15:00	0:33:47	0:09:51	1:08:38	4:21:48	6:29:04
User 9	0:10:00	0:15:00	0:48:23	0:18:45	1:32:08	6:42:48	9:37:04
Total							**70:14:16**
Average							**7:48:15**
Standard Deviation							**1:35:19**
Preparation: scenario, SIM execution, tasks, consent form, pretest and posttest.							16:00:00
Pilot Test and Adjustments							6:00:00
Interpretation and Consolidation							27:00:00
Total							**119:14:16**
Average							**11:55:26**

As we can observe, on average SIM consumed more effort than CEM. A possible explanation for this is that SIM is an exploratory method where we cannot control the evaluator exploration or limit him in the interpretation of signs. This can be noticed by looking at the standard deviation in SIM evaluations. As we can see there is an inconsistency between the time/effort spent by evaluators to conduct the method (e.g., the Expert Evaluator 1 spent 3 times more than the time spent by the Expert Evaluator 2). Besides, as explained in the analysis of types of problems each method is able to find, SIM is usually carried out in a broader scope than CEM, and as a consequence, we would expect to spend more time when using SIM.

Nevertheless, the effort demanded to identify a problem is equivalent in both methods. As we can observe from the previous analysis, SIM allowed finding a greater number of problems demonstrating to be a very complete method as regards the scope of analysis. So analyzing the time spent by each method to find each problem, we can observe in Table 7.15 that the time/effort required to find one problem using SIM is smaller than the CEM.

Table 7.15. The effort spent to find one problem by method.

	Time/Effort	**# of Problems**	**Time/Effort by Problem**
SIM	151:07:00	47	3:12:55
CEM	119:14:16	31	3:50:47

The results indicate that although SIM, on average, spent more time/effort per person that CEM, SIM has a greater cost-effectiveness compared to CEM as the time/effort to find each problem is lower.

7.3 Limitations of this Assessment

Although since the beginning of the empirical assessment planning we were concerned about the validity of the results, this study has some limitations. Regarding the sample of evaluators and users that, respectively, inspect and perform tests on the system they were not the sample recommended to perform empirical assessment Gray and Salzman [1998] argue that standards sample size for user testing (i.e., as mentioned by Nielsen and Landauer [1993] and Nielsen [2000] that 3 to 5 users are enough to evaluate an interface) is not appropriate for empirical researches; empirical assessments must use large number of participants. However, the fact that SIM and CEM are relatively recent methods, we had a great difficulty in finding the sample needed (as the ones Nielsen used to evaluate the Heuristic Evaluation method in Nielsen [1992] – 31, 19 and 14 evaluators). At any rate, although not statistically significant, according to Hartson et al. [2003] the results contribute and are valuable measures within a development project.

Another limitation of our research is that all the assessment was conducted and analyzed by only one person. Some researchers argue that a single expert review cannot be considered a reliable evaluation approach [de Jong and Schellens, 2000], so in order to minimize this limitation, a review process was carried out by a second person. The reviewer (i.e., advisor) revised all the decisions made, and also analyzed the consolidated problems list and the problems categorization.

Also, unfortunately, during **todoist** evaluation process we observed that some changes were performed in the system interface. Aware of this, we checked if the changes made would impact the results obtained. For this analysis we compared the problems encountered by users before the changes and the problems encountered after the changes and we found that the results were similar. That is, despite the changes made, the users continued to have the same kinds of problems. For this reason we believe that the modifications do not invalidate the results of our assessment, because even with the initiative to improve the interface, most of the identified problems persisted. The changes performed on **todoist** are shown in Appendix H.

Finally, another limitation is related to time registration in the SIM evaluation. Since SIM was not conducted by the evaluators in a controlled environment (i.e., they conducted in their homes or other places unknown to the authors of the empirical assessment) is not possible to know how reliable the values are. For instance, it is unclear whether the evaluators made breaks during the evaluation (either for quick conversations or answering the phone) and if these breaks were registered or not. It was very difficult to find volunteers with the necessary knowledge to participate in this assessment in an uncontrolled format, and we believe it would have been even harder to find people willing to perform it in a controlled environment. However, even then, as future work, it would be interesting to be able to repeat this assessment in a more controlled format.

Chapter 8

Final Discussion

After carrying out all phases of the proposed assessments, this Chapter aims at presenting a final discussion and consolidation of the results obtained in each one of the assessments. In this final discussion we have used not only the results presented in the previous Chapters, but also the literature (e.g., books, papers, thesis). The goal was to reach a conclusion about the cost-effectiveness, advantages and disadvantages of SIM.

The first analysis of our research aimed at investigating the applicability of SIM. Through this analysis we could see great benefits of the method, including: (1) SIM is a domain and technology independent method (considering the scope of analysis performed), and in the domains applied there was no need for adaptations in the method; (2) the method allows identifying domain-specific breakdowns; and (3) besides communicability, it allows the investigation of other properties.

The applicability of SIM was already expected by the authors of the method who raised this hypothesis when they formalized the method [de Souza et al., 2006], however, the hypothesis had not yet been confirmed. Through the analysis of studies that applied SIM (i.e., through the SLR) and the analysis of the questionnaire and interview responses (second step of our assessment) where the evaluators mentioned the applicability of the method as a benefit, we could confirm this hypothesis. It is noteworthy that we can confirm this hypothesis only for the scope of analysis performed with the method so far (i.e., to the domains so far investigated and presented in this study).

Another point also identified in the SLR that was also pointed by both novice evaluators and authors of the method is the possibility to investigate, besides communicability, other properties. SIM allows evaluators to expand their analysis beyond the scope of communicability, allowing the identification of breakdowns related to, for example, sociability and accessibility. It was also possible to identify this aspect during the empirical assessment, where SIM allowed finding potential use problems in the system evaluated, and, more specifically, problems that could impact the user's productivity. As already shown, one of the authors of SIM explained that this is possible because SIM focuses on communication, and the communication process is the support for other process such as, cognitive and productive.

Although this feature was identified as a benefit, the author I2 said that this feature should not be highlighted. The author's justification is that the method should be good at doing what it is proposed to do and declaring that SIM allows finding usability problems, for instance, may frustrate the evaluator, because there are specific evaluation methods to evaluate usability and they are sure to be more powerful than SIM in evaluating this property. Therefore, we note that the evaluator, when choosing SIM as an evaluation method, must not raise high expectations if he/she intends to evaluate other properties.

Regarding the possibility of identifying domain-specific issues some authors think it is possible and mentioned in the interview to have identified issues specific to the domain in their analysis with SIM. As evidenced in the assessment conducted through an SLR, SIM enabled to identify specific issues, for instance, in the domain of collaborative and educational systems, then confirming this possibility.

By contrasting the results identified in the questionnaire and interview (in the evaluators' perception) with the other results we could confirm some features of SIM that by analyzing only the evaluators' responses it would not have been possible. As we mentioned in Chapter 6, some features have been raised by evaluators as both advantages and disadvantages of the method (i.e., contradictions between the evaluators) and had not been clarified, they are: (1) easy to learn x difficult to learn; (2) high effort x low effort; (3) low cost-benefit x high cost-benefit; and (4) theory advantage x theory disadvantage. By contrasting these contradictions with the results obtained in other stages of the research, we could reinforce some features of SIM.

In relation to learning the method, as mentioned in the Chapter 6, the novice evaluators and authors believe that there is a high cost to learn it, which is usually related to the Semiotic Engineering theory. Besides the theory, SIM requires some characteristics and abilities of the evaluator that are not easily taught such as thinking, reflecting, interpreting and abstracting. These difficulties have also been identified in the work of Bim [2009] in which the author notes that the cause of these difficulties could be considered due to the need to break a widely used way of thinking in the computer science filed, which is more predictive, accurate and repeatable. She also pointed out that these difficulties are not exclusive of the methods that were under investigation (i.e., SIM and CEM). Interpreting, abstracting and building a global vision have been shown to be serious difficulties in the teaching and usage of other methods, for example, in programming, interface design and usability engineering. Thus, we conclude that SIM does require an effort to be learned due exclusively to the fact of being a theory-based method and requiring its knowledge to apply it. In addition, the application of SIM requires certain skills that are not easily taught and, probably, learned.

In relation to the effort of applying the method, we could observe this feature not only in the empirical assessment, but also in the assessment from evaluators' perspective where the evaluators mentioned to believe that SIM requires a high effort of application. However, in the interview the authors mentioned that this effort is not necessarily a disadvantage of the method. They argue that the effort of application is related to a cost that should be considered when choosing the method for evaluating a system. A disadvantage mentioned by both novice evaluators and the authors that could decrease this cost is the fact that SIM does not have a support tool. We believe that this would be an interesting future work to be investigated and conducted.

Analyzing the cost-effectiveness of the method, the evaluators reported how beneficial it is to perform an application with SIM. In the survey novice evaluators pointed out that SIM allows to make a thorough analysis of the system under evaluation. The authors also agree with this view and complemented saying that *"the main advantage (which in my opinion beats all the costs) is to be able to LEARN a wealth of opportunities, resources and facets of human-computer interaction ... the great advantage of Semiotic Engineering, which exemplary crystallizes in SIM, is the integrated view of how many and so many things that are involved in an HCI project."*[1].

This cost-effectiveness of SIM was also observed during the empirical assessment, where we noticed that the method allowed finding problems beyond the scope of CEM assessment. SIM also allowed finding approximately 77% of the problems that users faced when interacting with the system under evaluation, and 87% of the problems identified in general (by both SIM and CEM). Moreover, the time/effort required to find each problem was lower compared to CEM, confirming the high cost-effectiveness of SIM in this case.

Finally, the evaluators agree that being a theory-based method can be considered as an advantage and also a cost of SIM. On the one hand, if the evaluators know little about the theory it will require more time to learn the method (I4) and consequently increase the cost of applying it. On the other hand, according to I1, theory-based

[1]Original quote by I1 in Portuguese: *"A principal vantagem (que a meu ver bate todos os custos) é APRENDER a riqueza de oportunidades, de recursos e de facetas da interação humano-computador ... a grande vantagem da EngSem, que se cristaliza exemplarmente no MIS, é a visão integradora de quantas e tantas coisas estão envolvidas no projeto de IHC."*

methods are *"naturally more powerful in the sense that their results talk with concepts that are beyond the specific situation or context of the application"* [2] and also *"brings results integrated and consolidated into a chained reasoning"* [3] (I2). Therefore, once the evaluators know the theory it is no longer a cost (of learning) and starts to be a benefit of the method.

With regard to the disadvantages the evaluators mentioned: (1) the influence of the evaluators experience on the results; and (2) the lack of a support tool. The first was also identified during the empirical assessment, where we noticed that the experience of the evaluator influenced the results obtained with SIM. It was pointed out that a greater number of novice evaluators are needed to find about the same amount of problems encountered by experienced evaluators. However, as mentioned during the interview, the influence of the evaluator's experience on the method's application is not necessarily a drawback of the method, because it does not differ from other existing methods. According to the authors, the evaluator experience is everything in any inspection method. It is true that an inspection with expert evaluators will provide better results, and that the evaluators experience influences not only the results obtained with SIM, but probably any other inspection-based method, as well. However, being a deeply interpretive method, SIM requires more of the evaluator experience than other methods as, for example, Heuristic Evaluation.

As already mentioned, contradiction occurred between novice and expert evaluators. Evaluators mentioned that one advantage of SIM would be the possibility of using it early in the software development process. However, one author said that SIM requires a functional prototype to be used and then this was considered as a disadvantage of the method. We argue that this disadvantage may be related to the fact that SIM depends on a functional prototype to be applied, and not actually related to its applicability in formative assessment. It is possible to apply SIM when a functional prototype of the system exists, and its applicability to functional prototypes characterizes a formative assessment.

Some characteristics of SIM identified by the evaluators were also contrasted with the literature. Bim [2009] presents a study on obstacles to teaching the Semiotic Engineering evaluation methods, including SIM. In this study the author highlights the lack of material developed specifically for teaching, handling both educational descriptions about the methods and the availability of examples of how to apply them. This

[2] Original statement in Portuguese by I1: *"naturalmente mais potentes no sentido de que seus resultados conversam com conceitos que estão além da situação ou contexto específico da aplicação."*

[3] Original statement in Portuguese by I2: *"Traz resultados integrados e consolidados em um raciocínio encadeado."*

lack of educational materials was also mentioned as a disadvantage by the evaluators, which reinforce the need for such material. In additions, an advantage of SIM that was mentioned by the evaluators and can be confirmed by the literature is the possibility to generate new knowledge through its scientific application. This possibility has been also demonstrated by the works of de Souza and Leitão [2009] and de Souza et al. [2010].

The analysis carried out through questionnaires and interview (from evaluators perspective) allowed us to identify costs, benefits, advantages and disadvantages of SIM that were not identified in the other two assessments (i.e., SLR and empirical assessment), and, thus, could not be contrasted with the other assessments. All of them were argued in the Chapter 6 and, therefore, we will not make the same argumentations once again in this chapter. Although the characteristics could not be contrasted, we still understand them as aspects of the method.

We did not contrast the results with respect to the advantages and disadvantages mentioned by the evaluators in comparing SIM to other methods (presented in Chapter 6). As already mentioned, in most of the comparisons made by the participants they did not mention the method they were comparing SIM to. Therefore, it was very difficult to generate conclusions about their views regarding SIM in relation to other evaluation methods. Besides, there is no way to know whether participants had any data to support their statements, or whether these statements resulted from feelings they had based on their experiences. Thus, these issues raised about comparing the methods are not taken as final results about the method, but rather as aspects that could be interesting to further investigate. To confirm or contradict what was mentioned by evaluators it would be necessary to perform other empirical assessments with SIM in order to compare it to other evaluation methods.

Finally, we generated Table 8.1 that consolidates the characteristics of SIM identified in this research and in accordance with the analysis presented in this Chapter.

Table 8.1. Advantages and disadvantages consolidated.

Advantages	SIM allows identifying domain-specific breakdowns.
	Identifies problems related to communicability.
	Allows the analysis of the impact of each sign class (i.e., metalinguistic, static and dynamic) independently.
	Allows formalizing, arguing and explaining the problems encountered while interacting with the system.
	It has a good cost-benefit relation (i.e., good results at a relatively low cost).
	Only one evaluator is needed to perform an evaluation.
	It is a theory-based method.
	It is a method that produces good results and allows finding important problems.
	Allows a thorough analysis of the system and identification of its problems.
	Allows generating new knowledge using a scientific application of SIM.
	It does not require an expert evaluator to understand problems reported.
	It is a method that, unlike other interface evaluation methods, performs an assessment of the help and system's documentation.
	Allows identifying the impact of problems related to other properties of the system (i.e., accessibility, sociability).
	Allows evaluating the system from evaluators' point of view (i.e., meta-message emission).
	It can be applied to different domains and technologies without adaptations.
	It is a good guide to find problems (i.e., provide procedures that guide the evaluator during the evaluation).
	Can be used to evaluate system's interface in the beginning of the development process.
Disadvantages	It is based on a complex theory.
	It is a difficult method to apply with a high learning curve.
	The experience of the evaluator is very important to get good results. Novices may have difficulties and may not generate good results.
	Demands a lot of time and effort to apply it.
	It is a laborious and repetitive method, and is tiring at times. This may cause the evaluator to overlook some problems.
	There is no tool available to support the evaluation.
	It is a method which requires the generation of an extensive and detailed documentation, which increases the evaluation time.
	Lack of material to support learning, such as case-study examples on how to apply the method.
	Lack of experts, which makes difficult to arrange a team to apply the method.

Chapter 9

Conclusion

This study presented an assessment of SIM in different aspects. The first assessment was conducted in order to investigate SIM's applicability, and it was done through a systematic literature review. The motivation for this analysis started from the hypothesis raised in 2006, when the method was officially proposed [de Souza et al., 2006], that SIM is a domain and technology independent method. Through the SLR it was possible to generate indicators to support this claim. Then we continued SIM's assessment conducting a survey through a questionnaire answered by 25 participants (novices' evaluators – students and researchers) and an interview answered by 4 authors of SIM (experts' evaluators). The analysis performed was of a qualitative nature and the main goal was not to draw any final conclusions about the cost-effectiveness ratio of SIM. Rather the goal was to outline what is perceived as advantages, disadvantages, costs and benefits from the evaluators' perspective and, therefore, raise the main issues regarding the method to be further investigated using deeper approaches (e.g., empirical assessments). The analysis was successful to raise interesting insights and characteristics of the method.

In the assessment carried out from the evaluators' perspective we identified some contradictory views regarding some of the method's features, in special considering their cost and benefits (i.e., high/low effort needed and high/low cost-effectiveness). Thus, we conducted an empirical assessment of SIM through a case study performed comparing SIM and CEM methods, aiming at better defining the costs and effectiveness of SIM. The purpose of this assessment was not to define which method is the best, but rather to show their differences and similarities. In this assessment we could perceive new characteristics of SIM that were not uncovered in the previous assessments, and also collect more evidences about issues evaluators disagreed on. Finally, we made a final discussion and consolidated the results obtained in the assessments previously mentioned.

Throughout the study and in the previous Chapter we discussed a number of advantages, disadvantages, costs and effectiveness of SIM. We present here only the main findings. A great benefit of SIM is its technology and domain independency. Thus, although evaluators have to learn the underlying theory and concepts in order to learn the method – that has a high learning curve – once the evaluator has learned it, it can be applied broadly without any adaptations or "add-ons". Besides, SIM is a method that allows a thorough understanding of the system being evaluated, providing rich results about its communicability quality (which is a property that has also been identified as important according to evaluators). Therefore, SIM's cost is offset by the results obtained and could be considered a cost-effective method. In comparison to CEM, SIM allowed finding about 77% of the problems that users faced when interacting with the system under evaluation. Moreover, although the time spent to conduct SIM was in general higher than CEM, the time/effort required to find each problem was lower, confirming the high cost-effectiveness of SIM. However, we also realize that SIM and CEM, as expected, are complementary methods, since they allow finding different types of problems.

The results generated by this research contribute to the Semiotic Engineering Theory research, since it provides data regarding how a theory-based method is being used and perceived by evaluators and its characteristics in an empirical application compared to CEM. This study is relevant not only for Semiotic Engineering research, but also to the HCI field as a whole, since it has already been identified the need to research new methods [Greenberg and Buxton, 2008], as well as HCI theories and methods based on it [Shneiderman et al., 2002; Carroll, 2003; de Souza, 2005]. Therefore, the contributions of this research to the HCI field are: (1) it provides an overview of the main advantages/disadvantages of a theory-based method; (2) the main advantages/disadvantages found are important to support professionals and researchers in assessing whether or when to use SIM; and (3) the results may assist future assessments of SIM, specifically in the definition of criteria to be used in assessing evaluation methods that focus on properties, other than usability. Such contributions may be strengthened by the acceptance of papers presenting some results shown here at important HCI conferences in Brazil and abroad [Reis and Prates, 2011, 2012].

We also suggest further and deeper investigations of the method. First we suggest that the applicability of the method continues to be investigated in domains that the method has not been used yet. As already mentioned, the findings presented here regarding SIM's applicability cannot be generalized since we cannot guarantee that we selected all the studies in the area, and we also cannot affirm that the domains we found in this study in which SIM was applied are enough to generalize the results. For instance, no studies have investigated SIM's applicability to mobile systems domain, augmented reality systems and tangible interfaces.

Second we suggest a broader application of the questionnaire since, although we distributed the invitation to answer the questionnaire in different channels, we could not reach a large number of participants. One of the reasons may be the fact that the method is recent, there may not be a very large number of evaluators who have used the method and also the period in which the questionnaire was distributed, which was the end of the term and beginning of winter break, in which students and professors are busy and on vacation. At any rate, a larger sample of participants could make it possible to validate the results obtained in this study and draw more definite conclusions regarding on evaluators' perspective of SIM's advantages and disadvantages.

Another possible future investigation that would be interesting is comparative assessments carried out with SIM and other evaluation methods. The empirical assessment can be done again comparing SIM and CEM in order to provide information to be triangulated with the results obtained in this study, as well as evaluate SIM in relation to other inspection-based methods as, for example, Heuristic Evaluation. As discussed before a comparison between SIM and Heuristic Evaluation should take into account that they evaluate different properties. However, since there is an overlap of problems that can be identified with both properties [de Souza, 2005], it could be interesting and useful in giving a better understanding of costs and benefits of SIM, since most HCI researchers and practitioners have a good knowledge and understanding of Heuristic Evaluation. It could also be interesting to better illustrate how the different properties change the focus of the evaluation and analysis of the system's interface.

From the data analyzed we could identify future works not only related to a deeper evaluation of the method but also related to supporting the learning and application of the method and decreasing its associated cost. We suggest the development of a tool to support the application of SIM, since it could reduce the effort required to apply the method which was mentioned by participants as a disadvantage of the method. Another suggestion would be the development of more educational material about the method and theory since the participants reported difficulties not only in learning the method, but also finding educational material related to it. Reducing these costs could create favorable conditions for a broader application of the method, and consequently, providing more information for a deeper analysis of the method's use.

Finally, although we argue that HCI field still needs further studies related to methodologies to support researchers to conduct assessments of the methods, we presented a way of assessing an evaluation method from different perspectives and in a non-comparative way. Even in the empirical assessment carried out, we presented it in a way where we identified similarities and differences between the methods without intending to identify one as better than the other. We suggest as future works the proposal of methodologies to assess interface evaluation methods, considering their differences regarding type of evaluation (i.e., inspection-based and user-based) and properties (i.e., usability and communicability). This investigation is important because the existing methods focus on usability and comparative assessments (e.g., [Jeffries et al., 1991; Desurvire et al., 1992; Karat et al., 1992]). However, nowadays new technologies (e.g., tangible interfaces) and properties (e.g., sociability, playability) may require the proposal of new methods [Greenberg and Buxton, 2008]. Thus, being able to assess these new methods becomes necessary.

Bibliography

Abreu, P. M. (2010). Recomendações para projetos de tics para apoio a alfabetização com libras. Master's thesis, MSc Dissertation in Computer Science - Universidade Federal de Minas Gerais. 91p.

Baker, K., Greenberg, S., and Gutwin, C. (2001). Heuristic evaluation of groupware based on the mechanics of collaboration. In *Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction*, EHCI '01, pages 123–140, London, UK. Springer-Verlag.

Barbosa, G. A. R., Corrêa, L. P. D., and Prates, R. O. (2011a). Análise da sociabilidade de comunidades online para os usuários surdos: Um estudo de caso do orkut. In *Proceedings of X Brazilian Symposium on Human Factors in Computing Systems and V Latin American Conference on Human Computer Interaction*, IHC '11 and CLIHC '11, pages 237–246, New York, NY, USA. ACM.

Barbosa, G. A. R., Silva, I. S., Gonçalves, G., Prates, R. O., Benevenuto, F., and Almeida, V. (2011b). Characterizing interactions among members of deaf communities in orkut. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part III*, INTERACT'11, pages 280–287, Berlin, Heidelberg. Springer-Verlag.

Barbosa, S. D. J. and Silva, B. S. (2010). *Interação Humano-Computador.* Rio de Janeiro: Série SBC.

Bento, L. F. H. (2010). Desenvolvimento de interfaces em diferentes tecnologias para teleoperação de robôs móveis. Master's thesis, MSc Dissertation in Computer Science - Universidade Federal de Minas Gerais. 90p.

Bento, L. F. H., Prates, R. O., and Chaimowicz, L. (2009). Using semiotic inspection method to evaluate a human-robot interface. In *Proceedings of the 2009 Latin*

American Web Congress, LA-WEB '09, pages 77–84, Washington, DC, USA. IEEE Computer Society.

Bertini, E., Gabrielli, S., and Kimani, S. (2006). Appropriating and assessing heuristics for mobile computing. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pages 119–126, New York, NY, USA. ACM.

Bim, S. A. (2009). *Obstáculos ao ensino dos métodos de avaliação da Engenharia Semiótica*. PhD thesis, Ph.D Thesis on Informatics - Pontifícia Universidade Católica do Rio de Janeiro. 181p.

Bim, S. A., Leitão, C. F., and de Souza, C. S. (2009). The challenge of teaching hci qualitative evaluation methods: a case study on the communicability evaluation method. Technical report, Technical Report no. 32/07 - Pontifícia Universidade Católica do Rio de Janeiro. 19p.

Brandão, A., Trevisan, D. G., Brandão, L., Moreira, B., Nascimento, G., Vasconcelos, C. N., Clua, E., and Mourão, P. (2010). Semiotic inspection of a game for children with down syndrome. In *Proceedings of the 2010 Brazilian Symposium on Games and Digital Entertainment*, SBGAMES '10, pages 199–210, Washington, DC, USA. IEEE Computer Society.

Capelão, L., Coutinho, F. R. d. S., Pereira, K., and Prates, R. O. (2011). Communicability evaluation of moodle to deaf and hearing users. In *Companion Proceedings of X Brazilian Symposium on Human Factors in ComputingSystems and V Latin American Conference on Human Computer Interaction*, IHC '11 and CLIHC '11, New York, NY, USA. ACM. Evaluation Competition.

Carroll, J. M. (2003). *HCI models, theories, and frameworks: toward a multidisciplinary science*. Morgan Kaufmann, 1 edition.

Carvalho, D. B. F., Nasser, R. B., and de Souza, C. S. (2010). Um estudo sobre a utilização de programas com interface baseada em mapas. Technical report, Technical Report no. 17/10 - Pontifícia Universidade Católica do Rio de Janeiro. 28p.

Castro, T. and Fuks, H. (2009). Inspeção semiótica do colabweb: proposta de adaptações para o contexto da aprendizagem de programação. *Revista Brasileira de Informática na Educação*, 17(1):71–81.

Coopamootoo, P. and Ashenden, D. (2011). A systematic evaluation of the communicability of online privacy mechanisms with respect to communication privacy management. In Marcus, A., editor, *Design, User Experience, and Usability. Theory,* *Methods, Tools and Practice*, volume 6770 of *Lecture Notes in Computer Science*, pages 384–393. Springer Berlin / Heidelberg.

Coutinho, F. R. d. S., Chaimowicz, L., and Prates, R. O. (2011). An analysis of information conveyed through audio in an fps game and its impact on deaf players experience. In *Proceedings of the 2011 Brazilian Symposium on Games and Digital Entertainment*, SBGAMES '11.

da Silva, E. J., de Souza, C. S., Prates, R. O., and Nicolaci-da Costa, A. M. (2003). What they want and what they get: a study of light-weight technologies for online communities. In *Proceedings of the Latin American conference on Human-computer interaction*, CLIHC '03, pages 135–146, New York, NY, USA. ACM.

da Silva, R. F. and Prates, R. O. (2008). Avaliação da manas na identificação de problemas de impacto social: um estudo de caso. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '08, pages 70–79, Porto Alegre, RS, Brazil. Sociedade Brasileira de Computação.

de Jong, M. and Schellens, P. J. (2000). Toward a document evaluation methodology: what does research tell us about the validity and reliability of evaluation methods? *IEEE Transactions on Professional Communication*, 43(3):242–260.

de Souza, C. S. (2005). *The Semiotic Engineering of Human-Computer Interaction*. MIT Press.

de Souza, C. S. and Cypher, A. (2008). Semiotic engineering in practice: redesigning the coscripter interface. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '08, pages 165–172, New York, NY, USA. ACM.

de Souza, C. S., da Costa, A. M. N., da Silva, E. J., and Prates, R. O. (2004). Compulsory institutionalization: investigating the paradox of computer-supported informal social processes. *Interacting with Computers*, 16(4):635–656.

de Souza, C. S. and Leitão, C. F. (2009). *Semiotic engineering methods for scientific research in HCI*. Morgan & Claypool.

de Souza, C. S., Leitão, C. F., Prates, R. O., Bim, S. A., and da Silva, E. J. (2010). Can inspection methods generate valid new knowledge in hci? the case of semiotic inspection. *International Journal Human-Computer Studies*, 68:22–40.

de Souza, C. S., Leitão, C. F., Prates, R. O., and da Silva, E. J. (2006). The semiotic inspection method. In *Proceedings of VII Brazilian symposium on Human factors in computing systems*, IHC '06, pages 148–157, New York, NY, USA. ACM.

Desurvire, H., Kondziela, J., and Atwood, M. E. (1992). What is gained and lost when using methods other than empirical testing. In *Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems*, CHI '92, pages 125–126, New York, NY, USA. ACM.

Desurvire, H., Lawrence, D., and Atwood, M. (1991). Empiricism versus judgement: Comparing user interface evaluation methods on a new telephone-based interface. *SIGCHI Bull.*, 23:58–59.

dos Santos, R. L. and Prates, R. O. (2010). Estratégias para comunicar qualidade na wikipedia. In *Proceedings of the IX Symposium on Human Factors in Computing Systems*, IHC '10, pages 71–80, Porto Alegre, Brazil, Brazil. Brazilian Computer Society.

dos Santos Pereira, R. L. (2011). Qualidade de artigos na wikipedia para seus usuários - análise e proposta de interação. Master's thesis, MSc Dissertation in Computer Science - Universidade Federal de Minas Gerais. 93p.

Egan, T. M. (2002). Grounded theory research and theory building. *Advances in Developing Human Resources*, 4(3):277–295.

Engström, E., Runeson, P., and Skoglund, M. (2010). A systematic review on regression test selection techniques. *Inf. Softw. Technol.*, 52:14–30.

Fernandes, E., . M. A. (2002). *Grounded theory*. E. Fernandes & L. Almeida (Eds.), Modelos e técnicas de avaliação: Novos contributos para a prática e investigação psicológicas.

Ferreira, J. J. and de Souza, C. S. (2011). Agentes no agentsheets®: como o agentsheets® comunica o conceito de agentes. Technical report, Technical Report no. 09/11 - Pontifícia Universidade Católica do Rio de Janeiro. 19p.

Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, Aldine.

Goulding, C. (2001). Grounded theory: A magical formula or a potential nightmare. *The Marketing Review*, 2(1):21–34.

Gray, W. D. and Salzman, M. C. (1998). Damaged merchandise? a review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13(3):203–261.

Greenberg, S. and Buxton, B. (2008). Usability evaluation considered harmful (some of the time). In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 111--120, New York, NY, USA. ACM.

Guimarães, F. J. Z. and de Souza, C. S. (2008). Análise de um ambiente de apoio a comunidades de prática utilizando o método de inspeção semiótica. Technical report, Technical Report no. 06/08 - Pontifícia Universidade Católica do Rio de Janeiro. 22p.

Hartson, R. H., Andre, T. S., and Williges, R. C. (2003). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction*, 15(1):145–181.

Hornbæk, K. (2010). Dogmas in the assessment of usability evaluation methods. *Behav. Inf. Technol.*, 29:97–111.

Jeffries, R., Miller, J. R., Wharton, C., and Uyeda, K. (1991). User interface evaluation in the real world: a comparison of four techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, CHI '91, pages 119–124, New York, NY, USA. ACM.

Karat, C.-M. (1990). Cost-benefit analysis of usability engineering techniques. In *Proceedings of the Human Factors Society*, pages 839–843, Orlando, FL.

Karat, C.-M., Campbell, R., and Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92, pages 397–404, New York, NY, USA. ACM.

Khajouei, R., Hasman, A., and Jaspers, M. W. M. (2011). Determination of the effectiveness of two methods for usability evaluation using a cpoe medication ordering system. *International Journal of Medical Informatics*, 80(5):341–350.

Kitchenham, B. A. (2004). Procedures for undertaking systematic reviews. Technical report, Technical Report - Computer Science Department of Keele University (TR/SE-0401) and National ICT Australia Ltd (0400011T.1).

Kitchenham, B. A., Dyba, T., and Jorgensen, M. (2004). Evidence-based software engineering. In *Proceedings of the 26th International Conference on Software Engineering*, ICSE '04, pages 273–281, Washington, DC, USA. IEEE Computer Society.

Koutsabasis, P., Spyrou, T., and Darzentas, J. (2007). Evaluating usability evaluation methods: criteria, method and a case study. In *Proceedings of the 12th international conference on Human-computer interaction: interaction design and usability*, HCI'07, pages 569–578, Berlin, Heidelberg. Springer-Verlag.

Leitão, C. F., de Souza, C. S., and Barbosa, C. M. d. A. (2007). Face-to-face sociability signs made explicit in cmc. In *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction*, INTERACT'07, pages 5–18, Berlin, Heidelberg. Springer-Verlag.

Lindgaard, G. (2006). Notions of thoroughness, efficiency, and validity: Are they valid in hci practice? *International Journal of Industrial Ergonomics*, 36(12):1069–1074.

Mattos, B. A. M., Santos, R. L., and Prates, R. O. (2009). Investigating the applicability of the semiotic inspection method to collaborative systems. In *Simpósio Brasileiro de Sistemas Colaborativos (SBSC)*, pages 53 –60.

Moura, J. R. F., Santos, R. L., Oliveira, V. C., Silva, R. F. d., and Prates, R. O. (2008). Avaliação da comunicabilidade e possíveis impactos sociais do orkut. In *Companion Proceedings of Brazilian Symposium on Human Factors in Computing Systems*, IHC '08, New York, NY, USA. ACM. Evaluation Competition.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '92, pages 373–380, New York, NY, USA. ACM.

Nielsen, J. (2000). Why you only need to test with 5 users. Available at: http://www.useit.com/alertbox/20000319.html. Accessed in January 12th, 2012.

Nielsen, J. and Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, CHI '93, pages 206–213, New York, NY, USA. ACM.

Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, CHI '90, pages 249–256, New York, NY, USA. ACM.

Norman, D. A. and Draper, S. W. (1986). *User Centered System Design: New Perspectives on Human-computer Interaction*. CRC Press; 1 edition.

Oliveira, E. R. d. (2010). Investigação sobre a aplicabilidade dos métodos de avaliação de comunicabilidade ao domínio educacional. Master's thesis, MSc Dissertation in Computer Science - Universidade Federal de Minas Gerais. 206p.

Oliveira, E. R. d., Luz, L. C. S., and Prates, R. O. (2008). Aplicação semi-estruturada do método de inspeção semiótica: estudo de caso para o domínio educacional. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '08, pages 50–59, Porto Alegre, Brazil, Brazil. Sociedade Brasileira de Computação.

Peirce, C. S. (1992). *The essential Peirce*. Indiana University Press.

Peixoto, D. C. C., Prates, R. O., and Resende, R. F. (2010). Semiotic inspection method in the context of educational simulation games. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 1207–1212, New York, NY, USA. ACM.

Polson, P. G., Lewis, C., Rieman, J., and Wharton, C. (1992). Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *Int. J. Man-Mach. Stud.*, 36:741–773.

Prates, R. O. and Barbosa, S. D. J. (2003). Avaliação de interfaces de usuário - conceitos e métodos. In *Jornadas de Atualização em Informática (JAI/SBC)*.

Prates, R. O. and Barbosa, S. D. J. (2007). Introdução à teoria e prática da interação humano computador fundamentada na engenharia semiótica. In *Jornadas de Atualização em Informática (JAI/SBC)*, pages 263–326.

Prates, R. O., de Souza, C. S., and Barbosa, S. D. J. (2000). Methods and tools: a method for evaluating the communicability of user interfaces. *interactions*, 7:31–38.

Preece, J., Rogers, Y., and Sharp, H. (2007). *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley, 2 edition.

Reis, S. d. S. and Prates, R. O. (2011). Applicability of the semiotic inspection method: a systematic literature review. In *Proceedings of the Brazilian Symposium on Human Factors in Computing Systems*, IHC '11, pages 177–186, New York, NY, USA. ACM.

Reis, S. d. S. and Prates, R. O. (2012). An initial analysis of communicability evaluation methods through a case study. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, page (to appear), New York, NY, USA. ACM.

Salgado, L. C. d. C., Bim, S. A., and de Souza, C. S. (2006). Comparação entre os métodos de avaliação de base cognitiva e semiótica. In *Proceedings of VII Brazilian symposium on Human factors in computing systems*, IHC '06, pages 158–167, New York, NY, USA. ACM.

Salgado, L. C. d. C., de Souza, C. S., and Leitão, C. F. (2009). A semiotic inspection of icdl. Technical report, Technical Report no. 31/09 - Pontifícia Universidade Católica do Rio de Janeiro. 14p.

Sears, A. (1997). Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction*, 9(3):213–234.

Seixas, M. L. A. (2004). *Um método de avaliação para interfaces baseadas em mapas*. PhD thesis, Ph.D Thesis on Informatics - Pontifícia Universidade Católica do Rio de Janeiro. 113p.

Shneiderman, B., Card, S., Norman, D. A., Tremaine, M., and Waldrop, M. M. (2002). Chi@20: fighting our way from marginality to power. In *CHI '02 extended abstracts on Human factors in computing systems*, CHI EA '02, pages 688–691, New York, NY, USA. ACM.

Souza, G. S. (2010). Sobre a engenharia semiótica da interação com sistemas de monitoração. Master's thesis, MSc Dissertation in Informatics - Pontifícia Universidade Católica do Rio de Janeiro. 108p.

Steves, M. P., Morse, E., Gutwin, C., and Greenberg, S. (2001). A comparison of usage evaluation and inspection methods for assessing groupware usability. In *Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '01, pages 125–134, New York, NY, USA. ACM.

Strauss, A. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. SAGE.

Villela, M. L. B., Prates, R. O., and Moura, M. A. (2011). Qualidade em redes sociais online: Um estudo de caso contrastando perspectivas das áreas de ihc e ciência da informação. In *IADIS Ibero Americana WWW/INTERNET*, pages 43–50.

Yen, P.-Y. Y. and Bakken, S. (2009). A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol - an example from a web-based communication tool for nurse scheduling. In *Proceedings of the Symposium AMIA*, pages 714–718.

Appendix A

Questionnaire

Introduction

Dear Participant,

The aim of this questionnaire is to collect data in order to identify advantages and disadvantages associated with the Semiotic Inspection Method.

Your participation is voluntary and is not part and will not be considered as an assessment in this course you are attending this semester. The completion of this questionnaire should take approximately 20 minutes.

In answering this questionnaire you are automatically accepting that the data provided may be used and disclosed for research. We guarantee that the disclosure of data will be reported anonymously, and personal data reported here will be kept confidential.

This research is being conducted by Soraia Reis, a Computer Science Masters Student at the Computer Science Department of the Federal University of Minas Gerais (DCC / UFMG), under the guidance of Professor Raquel O. Prates.

For more information, please contact through the e-mail soraiareis@dcc.ufmg.br.

*** Required**

Participant Identification

* Gender

- Female
- Male

* Age

[ ]

Higher Level Education/Training

* What is your current profession?

- [ ] Student
- [ ] Researcher
- [ ] Professor
- [ ] IT Professional
- [ ] Professional in other area

* Highest level of education

- Incomplete Undergraduate Degree
- Complete Undergraduate Degree
- Incomplete Master Degree
- Complete Master Degree
- Incomplete Doctorate Degree
- Complete Doctorate Degree

*** Required**

* **Estimated Completition (mm/yyyy)** Shown only if the level of education is **Incomplete**.

* **Completition Year (yyyy)** Shown only if the level of education is **Complete**.

* **What is your research area?** Shown only if the level of education is **Master** or **Doctorate**.

* **Course**

- Computer Science
- Information Systems
- Other

* **University**

Professional Experience

* **Do you have internship or work experience in the Human-Computer Interaction (HCI) area?**

- Yes
- No

* **How many years do you do internship or work in the HCI area?** Shown only if the participant have experience in HCI area.

*** Required**

Identification of the course in which you learned SIM

* Enter in the form below the name of the course and year/semester (yyyy-s) when you had contact with the Semiotic Inspection Method (SIM). At least one course must be informed.

Course	Year/Semester yyyy-1 or yyyy-2	Course of: Undergraduation	Post-graduation
		☐	☐
		☐	☐
		☐	☐
		☐	☐

If you have other courses to inform (to complement the above question) report it below. Example: Human-Computer Interaction - 2008-1 (Undergraduate and/or Postgraduate)

Experience with SIM

* How many times have you evaluated a system using the Semiotic Inspection Method (SIM)? Skip the entire section if the answer to this question is None.

- None
- 1
- 2
- 3
- 4 or more

*** Required**

* Which were the systems evaluated?

* How many evaluators were involved, on average, in each assessment?

- 1
- 2
- 3 or more

* How did you obtain information to apply the SIM? Check as many options as you want.

- ☐ Lessons
- ☐ Educational Material
- ☐ Scientific Papers
- ☐ Other

*** Required**

* In the experience you had with the application of SIM, how much do you think is important to know the Semiotic Engineering theory?

- Low - the theory knowledge is not important.
- Reasonable - the theory knowledge may help but is not essential.
- High - the theory knowledge is crucial to achieve good results.

* In the experience you had with the application of SIM, do you consider that:

- [ ] SIM helped you to perceive problems that had no impact on the system use.
- [ ] SIM helped you to formalize problems that had already been noticed before application of the method.
- [ ] SIM helped you to perceive relevant issues in the system.

* How do you categorize the difficulty of applying the method?

- Very Easy
- Easy
- Medium
- Difficult
- Very Difficult

*** Required**

* Select below which step of the method that you have more difficulty.

- I had no difficulty.
- Step 1 - Analysis of Metalinguistic Signs
- Step 2 - Analysis of Static Signs
- Step 3 - Analysis of Dinamic Signs
- Step 4 - Comparison of meta-communication messages
- Step 5 - Assessment of the system communicability

* Select below which activity of the method that you have more difficulty.

- I had no difficulty.
- Difficulty in identifying what each sign represents in overall designer meta-communication.
- Difficulty in abstracting problems encountered in classes of problems.
- Difficulty in consolidating and making the interface analysis to have a global view (i.e., reconstruction of the meta-message).
- Other

*** Required**

* Select which of the options below reflects your greatest difficulty in applying the method.

- None. I had no difficulty.
- Difficulty in learning the method.
- Difficulty of the theory - the Semiotic Engineering is a complex theory with a lengthy content.
- The method requires time and concentration to perform a good evaluation.
- Other

* Would you consider applying the SIM in future evaluations in which you are responsible for defining the evaluation method to be used?

- Yes
- No
- Perhaps

Comment your previous answer.

In your experience what are the main advantages of SIM?

*** Required**

In your experience what are the main disadvantages of SIM?

Feel free to make any further comment about your experience with the method.

Knowledge/Experience with Evaluation Methods in General

* Have you ever applied any other evaluation method (other than SIM)? Skip the entire section if the answer to this question is No.

- Yes
- No

*** Required**

* Check the context in which you have already applied some interface evaluation method.

	Never Applied	Practical work in a course	Internship / Work (i.e., professional experience)	Scientific Research	Other
Heuristic Evaluation	☐	☐	☐	☐	☐
Usability Test	☐	☐	☐	☐	☐
SIM (Semiotic Inspection Method)	☐	☐	☐	☐	☐
CEM (Communicability Evaluation Method)	☐	☐	☐	☐	☐
Cognitive Walkthrough	☐	☐	☐	☐	☐

* Considering the methods that you have already applied, what is your experience in applying it?

	Never Applied	Once	Applied 2 to 3 times	Applied 4 to 5 times	Applied 6 or more
Heuristic Evaluation	●	●	●	●	●
Usability Test	●	●	●	●	●
SIM (Semiotic Inspection Method)	●	●	●	●	●
CEM (Communicability Evaluation Method)	●	●	●	●	●
Cognitive Walkthrough	●	●	●	●	●

*** Required**

How do you compare SIM in terms of costs and benefits over other methods?

Contact

* An interview step is expected in this research. If you wish to further contribute to this research and allow us to contact you to check your availability to participate in the interview please provide your name and e-mail (this information will be used only for this purpose). Skip the entire section if the answer to this question is **I do not authorize contact**.

- I authorize contact
- I do not authorize contact

* Full Name

* E-mail

*** Required**

---------- original version in Portuguese ----------

Introdução

Caro Participante,

O objetivo deste questionário é coletar dados que permitam identificar vantagens e desvantagens associadas ao Método de Inspeção Semiótica.

A sua participação é voluntária e não faz parte e nem será considerada na sua avaliação na disciplina em que esta cursando neste semestre. O preenchimento deste questionário deve levar aproximadamente 20 minutos.

Ao responder este questionário você esta automaticamente aceitando que os dados informados poderão ser utilizados para pesquisa e divulgados. Garantimos que a divulgação dos dados informados será de forma anônima, e os dados pessoais aqui informados serão mantidos em sigilo.

Esta pesquisa esta sendo realizada por Soraia Reis, aluna do mestrado em Ciência da Computação no Departamento de Ciência da Computação da Universidade Federal de Minas Gerais (DCC/UFMG), sob a orientação da Professora Raquel O. Prates.

Para maiores informações, favor entrar em contato através do e-mail soraiareis@dcc.ufmg.br.

Identificação do Participante

* Sexo

● Feminino ● Masculino

* Idade

[ ]

*** Obrigatório**

Nível Mais Alto de Formação

* Qual a sua profissão atual?

- ☐ Estudante
- ☐ Pesquisador
- ☐ Professor
- ☐ Profissional na área de TI
- ☐ Profissional que não é da área de TI

* Nível de formação mais alto

- ● Graduação Completo
- ● Graduação Incompleto
- ● Mestrado Incompleto
- ● Mestrado Completo
- ● Doutorado Incompleto
- ● Doutorado Completo

* **Previsão de Formatura (mm/aaaa)** Exibida apenas se o nível de formação for **Incompleto**.

\[\]

* **Ano de Conclusão (aaaa)** Exibida apenas se o nível de formação for **Completo**.

\[\]

*** Obrigatório**

*** Qual a sua área de pesquisa?** Exibida apenas se o nível de formação for **Mestrado** ou **Doutorado**.

[ ]

* Curso

- Ciência da Computação
- Sistemas de Informação
- Outro

  [ ]

* Universidade

[ ]

Experiência Profissional

* Tem experiência com estágio ou trabalho na área de Interação Humano-Computador (IHC)?

- Sim
- Não

*** Há quantos anos faz estágio ou trabalha na área de IHC?** Exibida apenas se tiver experiência na área de IHC.

[ ]

*** Obrigatório**

Identificação da Disciplina que Aprendeu o MIS

* Informe abaixo o nome da disciplina e semestre/ano (mm/aaaa) em que você teve contato com o Método de Inspeção Semiótica (MIS). Pelo menos uma disciplina deve ser informada.

Disciplina	Ano/Semestre aaaa-1 ou aaaa-2	Disciplina de: Graduação	Pós-graduação
[ ]	[ ]	☐	☐
[ ]	[ ]	☐	☐
[ ]	[ ]	☐	☐
[ ]	[ ]	☐	☐

Caso tenha outras disciplinas a informar (para complementar a questão anterior) informe abaixo. Exemplo: Interação Humano-Computador - 2008-1 (Graduação e Pós-graduação)

[ ]

*** Obrigatório**

Experiência com o MIS

* **Quantas vezes você já avaliou um sistema utilizando o Método de Inspeção Semiótica (MIS)?** Pular toda a seção se a resposta desta questão for **Nenhuma**.

- Nenhuma
- 1
- 2
- 3
- 4 ou mais

* **Qual(is) foi(ram) o(s) sistema(s) avaliado(s)?**

* **Quantos avaliadores estavam envolvidos, em média, em cada avaliação?**

- 1
- 2
- 3 ou mais

*** Obrigatório**

* Como obteve informações para aplicar o MIS? Marque quantas opções desejar.

- ☐ Aulas em disciplina
- ☐ Material didático
- ☐ Artigos científicos
- ☐ Outro

* Na experiência que você teve com a aplicação do MIS, o quanto você acredita ser importante o conhecimento da teoria da Engenharia Semiótica?

- ● Baixo - o conhecimento da teoria não é importante.
- ● Razoável - o conhecimento da teoria pode ajudar, mas não é fundamental.
- ● Alto - o conhecimento da teoria é fundamental para se realizar uma boa avaliação.

* Na experiência que você teve com a aplicação do MIS, você considera que:

- ☐ O MIS te ajudou a perceber problemas que não tem impacto no uso do sistema.
- ☐ O MIS te ajudou a formalizar problemas que já tinha percebido antes da aplicação do método.
- ☐ O MIS te ajudou a perceber problemas relevantes no sistema.

*** Obrigatório**

* Como você categoriza a dificuldade de aplicação do método?

- Muito Fácil
- Fácil
- Médio
- Difícil
- Muito Difícil

* Selecione abaixo qual a etapa do método que você tem mais dificuldade.

- Não tive dificuldades.
- Passo 1 – Análise dos Signos metalinguísticos
- Passo 2 – Análise dos Signos estáticos
- Passo 3 – Análise dos Signos dinâmicos
- Passo 4 – Comparação das mensagens de metacomunicação
- Passo 5 – Avaliação da comunicabilidade do sistema

*** Obrigatório**

* Selecione abaixo qual a atividade do método que você tem mais dificuldade.

- Não tive dificuldades.
- Dificuldade em identificar o que cada signo representa na metacomunicação geral do designer.
- Dificuldade em abstrair os problemas encontrados em classes de problemas.
- Dificuldade em consolidar e fazer a análise da interface para ter uma visão global (i.e., reconstrução da metamensagem).
- Outro

* Selecione qual das opções abaixo reflete a sua maior dificuldade em aplicar o método.

- Nenhuma. Não teve dificuldade.
- Dificuldade de aprender o método.
- Dificuldade da teoria – a Engenharia Semiótica é uma teoria complexa com um conteúdo extenso.
- O método demanda tempo e concentração para que uma boa avaliação seja realizada.
- Outro

*** Obrigatório**

* Você consideraria aplicar o MIS em avaliações futuras em que você seja responsável pela definição do método de avaliação a ser utilizado?

- Sim
- Não
- Talvez

Comente sua resposta anterior.

Na sua experiência quais são as principais vantagens do MIS?

Na sua experiência quais são as principais desvantagens do MIS?

Fique à vontade para fazer outros comentários sobre sua experiência com o método.

*** Obrigatório**

Conhecimento/Experiência de Métodos de Avaliação em Geral

*** Você já aplicou algum outro método de avaliação (diferente do MIS)?** Pular toda a seção se a resposta desta questão for **Não**.

- Sim
- Não

*** Marque o(s) contexto(s) no(s) qual(is) você já aplicou algum método de avaliação de interfaces.**

	Nunca aplicou	Trabalho Prático em uma disciplina	Estágio/Trabalho (i.e., experiência profissional)	Pesquisa Científica	Outro
Avaliação Heurística	☐	☐	☐	☐	☐
Teste de Usabilidade	☐	☐	☐	☐	☐
MIS (Método de Inspeção Semiótica)	☐	☐	☐	☐	☐
MAC (Método de Avaliação de Comunicabilidade)	☐	☐	☐	☐	☐
Percurso cognitivo	☐	☐	☐	☐	☐

*** Considerando os métodos que já aplicou, qual sua experiência na sua aplicação:**

	Nunca aplicou	Aplicou 1 vez	Aplicou de 2 a 3 vezes	Aplicou de 4 a 5 vezes	Aplicou 6 ou mais vezes
Avaliação Heurística	●	●	●	●	●
Teste de Usabilidade	●	●	●	●	●
MIS (Método de Inspeção Semiótica)	●	●	●	●	●
MAC (Método de Avaliação de Comunicabilidade)	●	●	●	●	●
Percurso cognitivo	●	●	●	●	●

*** Obrigatório**

Como você compara o MIS em termos de custos e benefícios a outros métodos?

Contato

* Esta pesquisa prevê uma etapa de entrevista. Caso deseje contribuir ainda mais para esta pesquisa e nos autorize a lhe contactar para verificar sua disponibilidade para participar na época da entrevista favor fornecer seu nome e e-mail (estas informações serão utilizadas apenas para este fim). Pular toda a seção se a resposta desta questão for Não Autorizo Contato.

- Autorizo Contato
- Não Autorizo Contato

* Nome Completo

* E-mail

*** Obrigatório**

Appendix B

Interview Script

Interview Script

Thematic Block	Key Points
Acknowledgment	First I would like to thank you for your willingness to participate in the interview late this year that I know is very busy: Christmas and New Year.
Contact	Given that we are doing this interview over the Internet, there is always the possibility of having problems during the interview (i.e., stop the light or lose the connection). For this reason, I would like to leave my contact is (31) XXXX-XXXX if anything happens. If possible I would like you to pass me your contact if necessary.
Introduction	As I said, this research is part of a work I'm doing for my masters and aims to raise advantages and disadvantages about the Semiotic Inspection Method. I'll do some questions about the method and would like you to respond by giving as much detail as possible about it. Before we begin, I will send you a consent form that I would like you to read and respond if you agree or not to participate in the interview.
Consent Form	**Consent Form** ----------------------------- This interview is being conducted through an online instant messaging. All communication will be recorded for research purposes. The purpose of this interview is to determine, through the information collected, the advantages and disadvantages Semiotic Inspection Method (SIM). The data collected during the interview will be used for research and may be published. Any data submitted for publication will be used to ensure your anonymity. If you wish, you may request a copy of the data generated by you via e-mail soraiareis@dcc.ufmg.br. You are free to stop the interview at any time. Your participation in this survey is voluntary and will not be offered any compensation. If you wish, you can specify new conditions that must be met for you to participate in this interview. Just specify them before the interview. **Free and Clarified Consent (Voluntary Agreement)** The above text describing the research objectives and procedures of Assessment of the Semiotic Inspection Method (SIM) has been read and understood. I had the opportunity to ask questions about the survey, which were answered satisfactorily. If you agree, type "I agree to participate as a volunteer.".

Profile Summary	Before starting with the questions, I raised some information about your profile [Lattes] that I would like you to confirm if I'm correct. <Participant profile regarding his background, current profession, and if a professor or researcher in which department/university, how long, acting in undergraduate and/or post-graduation courses and which research area.>
Interviewee Profile	▪ In your research related to Semiotic Engineering, what are your focuses? ▪ Do you minister disciplines related to HCI? Which one(s)? o Are they undergraduate or graduate subjects? o Do you teach SIM in these disciplines? o Do you only teaches the theoretical part or do you also give practical works?
Experience with HCI	▪ What is your experience in the HCI field? ▪ How long do you run interface evaluations in general? ▪ What HCI methods have you applied? o In which contexts have you used these methods? (e.g., teaching, research and/or extension projects/industry). o Do you consider yourself an expert in all of them? o Which ones you do not? Why?
Experience with SIM	Application of SIM ▪ Are you involved with SIM from the beginning (around 2003)? o If not, when did you get involved? ▪ In which contexts (e.g., teaching, research and extension projects/industry) and domains (i.e., class of systems) have you used SIM? o Was there a need for adjustments in the method? Which one(s)? o Was it possible to identify specific issues in the domain applied? o Although SIM focus on communicability, do you think it was possible to identify issues related to other qualities of use? ▪ What type of application of the method that you do most often (technical or scientific)? ▪ Did the applications occur individually or in groups? o If in group: How was the consolidation of the results? Industry ▪ Have you applied SIM in the industry? o Why did you choose to use SIM in the project? ▪ Example: (1) because you are expert in the method, (2) because you wanted to generate data for research, (3) because you thought it was more appropriate, (4) other. o Did you have any project that you decided to not apply SIM? Why?

Advantages X Disadvantages	I will now ask some questions related to your perception about SIM.
	Knowledge of the Theory of Semiotic Engineering
	▪ In your opinion, what is the relevance of the knowledge of the Semiotic Engineering Theory in applying the method?
	▪ In your opinion, how important is the evaluator experience in the application of SIM?
	  o Do you think this is different to other evaluation methods?
	Advantages X Disadvantages
	▪ What are the main costs/disadvantages of SIM?
	▪ What are the main benefits/advantages of SIM?
	  o Do you think there is a cost associated with the application of the method? Which?
	    ▪ In your experience, which step has the greatest complexity in SIM?
	  o What do you think about the time of application of SIM?
	  o In your opinion, what is the cost of learning SIM?
	    ▪ For those who are new, in your opinion (or according to a report of students) what is the most complex step of SIM?
	  o What are the costs/benefits that the theory brings to the method?
	▪ How do you evaluate the cost-effectiveness of SIM over other methods that you know?

System	▪ Do you think that a system to support the application of SIM would be interesting?
	▪ Have ever used one? Which?
	  o What did you find interesting about the system? What did not?
	▪ What do you think would be the main points in which a system could assist in the evaluation?
	  o How? Why?
	▪ If you could choose to have three features in the system which would they be?
	▪ Do you think it would be interesting a collaborative system, allowing the evaluation of more than one evaluator?

---------- original version in Portuguese ----------

Roteiro da Entrevista

Bloco Temático	Principais Pontos
Agradecimento	Primeiramente gostaria de agradecer pela sua disponibilidade em participar da entrevista neste final de ano que eu sei que é muito corrido: natal e ano novo.
Contato	Como estamos realizando esta entrevista pela internet, existe sempre a possibilidade de termos problemas durante a entrevista (i.e., acabar a luz ou perder a conexão). Por este motivo, eu gostaria de deixar meu contato que é (31) XXXX-XXXX caso aconteça alguma coisa. Se possível gostaria que me passasse o seu contato caso seja necessário.
Introdução	Como lhe falei, esta pesquisa é parte do trabalho que estou fazendo para o meu mestrado e tem por objetivo levantar vantagens e desvantagens a respeito do Método de Inspeção Semiótica. Eu vou realizar algumas perguntas a respeito do método e gostaria que respondesse dando o maior detalhe possível a respeito. Antes de começar, eu vou lhe enviar um termo de consentimento que eu gostaria que você lesse e respondesse se concorda ou não em participar da entrevista.
Termo de Consentimento	**Termo de Consentimento** ---------------------------- Esta entrevista está sendo realizada online através de um sistema de mensagem instantânea. Toda a comunicação será gravada para fins de pesquisa. O objetivo desta entrevista é apreciar, através das informações coletadas, as vantagens e desvantagens do Método de Inspeção Semiótica (MIS). Os dados coletados durante a entrevista serão utilizados para pesquisa e poderão ser publicados. Quaisquer dados utilizados para publicação serão apresentados de forma a garantir o seu anonimato. Caso deseje, poderá solicitar uma cópia dos dados gerados por você através do e-mail soraiareis@dcc.ufmg.br. Você é livre para interromper a entrevista a qualquer momento. A sua participação nesta pesquisa é voluntária, e não será oferecida nenhuma remuneração. Caso deseje, você pode especificar novas condições que devem ser atendidas para que você participe desta entrevista. Basta especificá-las antes do início da entrevista. **Consentimento Livre e Esclarecido (Acordo Voluntário)** O texto acima descrevendo objetivos e procedimentos da pesquisa Avaliação do Método de Inspeção Semiótica (MIS) foi lido e entendido. Eu tive a oportunidade de fazer perguntas sobre a pesquisa, que foram respondidas satisfatoriamente. Caso você concorde, digite "Eu estou de acordo em participar como voluntário(a).".
Resumo do Perfil	Antes de iniciar com as perguntas, eu levantei algumas informações sobre seu perfil [Lattes] que gostaria que confirmasse se estou correta. <Perfil dos participantes a respeito de seus conhecimentos, profissão atual, e se um professor ou pesquisador em qual departamento/universidade, quanto tempo, se atua nos cursos de graduação e/ou pós-graduação e que área de pesquisa.>

Perfil do Entrevistado	▪ Na sua pesquisa relacionada à Engenharia Semiótica, quais são seus focos?
	▪ Você ministra disciplinas relacionadas à IHC? Quais?
	    o As disciplinas são de graduação ou pós-graduação?
	    o Você ensina o MIS nestas disciplinas?
	    o Você só ensina a parte teórica ou também dá trabalhos práticos?

<table>
<tr>
<td>Experiência com IHC</td>
<td>

Qual é a sua experiência na área de IHC?
Há quanto tempo executa avaliações de interface em geral?
Quais métodos de IHC você já aplicou?

Em quais contextos você utilizou este(s) método(s)? (e.g., ensino, pesquisa, projetos de extensão/indústria).
Você se considera especialista em todos eles?
Em quais que não? Por quê?

</td>
</tr>
<tr>
<td>Experiência com o MIS</td>
<td>
Aplicação do MIS

Você esta envolvida com o MIS desde o início (por volta de 2003)?

Se não, quando você se envolveu?

Em quais contextos (e.g., ensino, pesquisa, projetos de extensão/indústria) e domínios (i.e., classe de sistemas) você já utilizou o MIS?

Houve necessidade de adaptações no método? Quais?
Foi possível identificar questões específicas do domínio?
Embora o MIS foque em comunicabilidade, você acha que foi possível identificar questões relacionadas a outras qualidades de uso?

Qual o tipo de aplicação do método que você faz mais comumente (técnica ou científica)?
As aplicações ocorreram individualmente ou em grupo?

Se em grupo: Como ocorreu a consolidação dos resultados?

Indústria

Você já aplicou o MIS na indústria?

Por quê você escolheu o MIS para aplicar no projeto?

Exemplo: (1) porque você é especialista no método; (2) porque gostaria de gerar dados para pesquisa; (3) porque achou que era o mais adequado; (4) outro.

Teve algum projeto que você deixou de aplicar o MIS? Por quê?

</td>
</tr>
</table>

Vantagens X Desvantagens	Vou fazer algumas perguntas agora relacionadas a sua percepção sobre o MIS.
	Conhecimento da Teoria da Engenharia Semiótica
	▪ Em sua opinião, qual a relevância do conhecimento da Teoria da Engenharia Semiótica para aplicação do método?
	▪ Em sua opinião, qual a importância da experiência do avaliador na aplicação do MIS?
	    o Você acha que isso é diferente para outros métodos de avaliação?
	Vantagens X Desvantagens
	▪ Quais os principais custos/desvantagens do MIS?
	▪ Quais os principais benefícios/vantagens do MIS?
	    o Você acha que existe algum custo relacionado a aplicação do método? Qual?
	        ▪ Na sua experiência qual o passo de maior complexidade do MIS?
	    o O que você acha a respeito do tempo de aplicação do MIS?
	    o Na sua opinião, qual é o custo de aprendizagem do MIS?
	        ▪ Para quem é iniciante, na sua opinião (ou de acordo com relato de alunos) qual é o passo de maior complexidade do MIS?
	    o Quais os custos/benefícios que a teoria traz para o método?
	▪ Como você avalia o custo-benefício do MIS em relação aos outros métodos que você conhece?

Sistema	▪ Você acha que um sistema de apoio à aplicação do MIS seria interessante?
	▪ Já utilizou algum? Qual?
	    o O que achou interessante sobre o sistema? O que não achou?
	▪ Que você acha que seria os principais pontos em que um sistema poderia apoiar na avaliação?
	    o Como? Por que?
	▪ Se pudesse escolher 3 funcionalidades para ter no sistema quais seriam?
	▪ Você acha que seria interessante que ele fosse colaborativo, permitindo a avaliação por mais de um avaliador?

Appendix C

SIM Guideline

Dear Participant,

This document aims to inform you about the experiment that you are participating.

This experiment is part of the research methodology of my dissertation, which has the title "Assessment of the Semiotic Inspection Method" and is being advised by the Professor Raquel O. Prates. This experimental phase of the research aims to evaluate the effectiveness of the Semiotic Inspection Method (SIM). For this we choose a system so that several participants make an evaluation of its communicability using SIM. A parallel evaluation will be conducted by me using the Communicability Evaluation Method (CEM). From the results obtained in both steps we will determine the efficacy of SIM in finding real problems encountered by users. Some metrics have been selected not only to measure this variable (i.e., effectiveness), but also intend to analyze the reliability of the results generated by the method. These are summarized information about the assessment, but if you want more information on how the evaluation will be conducted, please contact me.

We would like to mention that the participation in the experiment is voluntary. By submitting the results of the assessment I understand that you agree with the participation and accepting that the data generated by you may be used for research and disseminated. We guarantee that the disclosure of the data generated by you will be anonymous, and personal information provided will be kept confidential. Read the consent form that follows for more information about participating in the experiment.

Below is a consent form and instructions on how the evaluation should be performed.

Soraia de Souza Reis

soraiareis@dcc.ufmg.br

Consent Form for Participation

Title: Experimental Evaluation of Semiotic Inspection Method (SIM)

Date: August/2011

Institution: DCC/UFMG

Responsible for Research: Soraia Reis (soraiareis@dcc.ufmg.br)

Introduction: This Consent Form contains information on the experimental evaluation of SIM. To ensure you are informed about your participation in this research, we ask that you make a thorough reading of this Consent Form. If you have any questions, feel free to ask the responsible for this research. We will understand that upon receiving the results of your assessment you are in accordance with this Consent Form.

Objective of the assessment: The purpose of this assessment is to assess the effectiveness of SIM in detecting problems that affect users when using the system.

General information about the assessment: You will be asked to perform the evaluation of a system using the Semiotic Inspection Method. This assessment should be documented according to the instructions that will be passed. After the assessment you must send the document to the person responsible for the research. For identification of the participants' profile, you will also be asked to answer a questionnaire.

Use of data collected: The data collected by you during the system assessment will be used to evaluate SIM as mentioned in the assessment objective. Any data submitted for publication will be used to ensure the anonymity of the participants in this experiment.

Privacy: Information that can identify the participants of the evaluation will not be disclosed. Your name will not appear in any reports. If you wish, you may request copies of the reports generated by this research.

If you decide not to participate in the evaluation: You are free to decide, at any time, whether or not to participate in this experiment.

Compensation: The participation in this experiment is voluntary and will not be offered any compensation to participants.

If you have any problems or have any other questions: If you have a problem you think may be related to your participation in this experiment, or if you have any questions about it, you may contact the person responsible for the research at any time through the e-mail soraiareis@dcc.ufmg.br.

New conditions: If you wish, you can specify new conditions that must be met for you to participate in this evaluation. To do this, please contact the person responsible for the research through the e-mail and specify the conditions for your participation.

Instructions for Evaluation

Before performing the assessment, please answer a questionnaire that we did in order to collect information about the profile of evaluators who are participating in the experiment. Follow the link to this questionnaire.

https://qtrial.qualtrics.com/SE/?SID=SV_4V3slkHrIYCcPl2

Target System	TODOIST (http://todoist.com)
Method to be used	Semiotic Inspection Method
Scenario	Joana is an entrepreneur who works in the Human Resources area. Her company offers several services including: human resources consulting, organizational climate research, performance evaluation and recruitment and selection. She is responsible for coordinating a team of 5 employees. Besides being a great professional, she is mother of two kids, Gabriel with 6 years old and Valeria with 3 years old. Joan has an agitated routine, which besides following her team activities, she has to make contact with clients, evaluate reports generated by her team and also perform personal activities (e.g., participate in your children activities and home). Due to the difficulty in managing and remembering the several activities that she must do during the day, she decided to seek for a system that supports this task. Joana is a novice user of technology and therefore need a system easy to use. In a Google search, she found TODOIST (http://todoist.com), a task management system, and decided to explore it to see if it suits your needs.
Template	A template to be filled out was sent to you along with this instruction sheet with the name **template_mis.doc**. **Attention:** We ask that you fill out this document with as much information as possible, recording all the details you can. **Default:** If this template is not what you're normally accustomed to use, we let it free that you add new topics if you need it, but we ask that the existing topics are not changed. **Hours Appropriation:** In the template you are asked to fill in the date, start time, end time and duration. We would like to clarify that it is not our intention to control how you evaluated the system. This item aims to get an idea of how long it took to you to perform the inspection in the system. So if you are not comfortable letting us know the actual day and time that the assessments occurred, feel free to not inform them, but please at least tell us the total time spent on evaluation. **Step 4 of SIM:** The fourth step of the evaluation is extremely important for our experiment. For this reason we ask special attention while completing this item. We need you to list in detail the communicability problems found in the system; including evidences (e.g., print screens). For each problem listed is very important that you

	classify it. Below, under "Classification of Communication Breakdowns", we mention how the classification should be performed.
	Follow links to materials that you may help about the evaluation with SIM: • **Article about SIM:** The semiotic inspection method.pdf • **Scientific Paper about SIM:** Semiotic Engineering Methods for Scientific Research in HCI.pdf[1] The use of auxiliary material is not mandatory, but you should consult them if you have questions regarding the application of the method.
Contact	Any doubt about the experiment, the material available and about SIM can be sent to the responsible for this research through the e-mail soraiareis@dcc.ufmg.br.

Classification of Communication Breakdowns

Complete: The user is unable to understand the message the designer consciously or not. Can occur when (1) the user cannot understand the message sent by the designer and give up to accomplish the task, or (2) the user performs the task in a different way the intention of the designer and believes that doing it correctly.

Partial: They are associated with unexpected paths chosen by the interactive user. Can occur when (1) the user understand what the designer is communicating, but he prefers to use other means to achieve his goals, or (2) the user does not understand the message the designer and consciously try to do the job otherwise.

Temporary: These are situations where the user can temporarily do not understand what happened, not knowing what to do to proceed, not knowing in what context he is or it may be momentarily confused.

[1] Although this book demonstrate the use of scientific SIM, it may be useful to understand how to perform the steps of the method since every scientific evaluation involves a technical evaluation.

---------- original version in Portuguese ----------

Caro (a) Participante,

Este documento visa informá-lo sobre o experimento que você esta participando.

Este experimento faz parte da metodologia de pesquisa da minha dissertação, que possui o título "Avaliação do Método de Inspeção Semiótica" e esta sendo orientada pela Professora Raquel O. Prates. Esta etapa experimental da pesquisa visa avaliar a eficácia o Método de Inspeção Semiótica (MIS). Para isso escolhemos um sistema para que vários participantes fizessem uma avaliação de sua comunicabilidade utilizando o MIS. Uma avaliação em paralelo a esta será realizada por mim utilizando o Método de Avaliação de Comunicabilidade (MAC). A partir dos resultados obtidos em ambos os passos iremos verificar a eficácia do MIS em encontrar problemas reais encontrados pelos usuários. Algumas métricas já foram selecionadas para medir não só esta variável (i.e., eficácia), mas também pretendemos analisar a confiança dos resultados gerados pelo método. Estas são informações resumidas da avaliação, mas caso você queira saber mais informações de como a avaliação será conduzida, entre em contato comigo.

Gostaríamos de mencionar que a participação no experimento é voluntária. Ao enviar os resultados da avaliação entendemos que você está de acordo com a forma de sua participação e aceitando que os dados gerados por você poderão ser utilizados para pesquisa e divulgados. Garantimos que a divulgação dos dados gerados por você será de forma anônima, e os dados pessoais informados serão mantidos em sigilo. Leia o termo de consentimento que segue para mais informações sobre a participação no experimento.

Abaixo apresentamos o Termo de Consentimento e instruções de como a avaliação deve ser realizada.

Soraia de Souza Reis

soraiareis@dcc.ufmg.br

Termo de Consentimento de Participação

Título: Avaliação Experimental do Método de Inspeção Semiótica (MIS)

Data: Agosto de 2011

Instituição: DCC/UFMG

Responsável pela Pesquisa: Soraia Reis (soraiareis@dcc.ufmg.br)

Introdução: Este Termo de Consentimento contém informações sobre a avaliação experimental do MIS. Para assegurar que você esteja informado sobre a sua participação nesta pesquisa, pedimos que faça a leitura completa deste Termo de Consentimento. Caso tenha alguma dúvida, não hesite em perguntar ao avaliador responsável. Entenderemos que ao receber os resultados de sua avaliação você esta de acordo com este Termo de Consentimento.

Objetivo da avaliação: O objetivo desta avaliação é avaliar a eficácia do MIS em detectar problemas que afetam os usuários durante o uso do sistema.

Informação geral sobre a avaliação: Você será solicitado a realizar a avaliação de um sistema utilizando o Método de Inspeção Semiótica. Esta avaliação deverá ser documentada conforme instruções que serão passadas. Ao fim da avaliação você deverá encaminhar o documento ao responsável pela pesquisa. Para fins de identificação do perfil dos participantes, você também será solicitado a responder um questionário.

Utilização dos dados coletados: Os dados coletados por você durante a avaliação do sistema serão utilizados para a avaliação do MIS conforme objetivo da pesquisa. Quaisquer dados utilizados para publicação serão apresentados de forma a garantir o anonimato dos participantes deste experimento.

Privacidade: Informações que possam identificar os participantes da avaliação não serão divulgadas. O seu nome não aparecerá em nenhum relatório. Caso deseje, poderá solicitar uma cópia dos relatórios gerados por esta pesquisa.

Se você decidir não participar na avaliação: Você é livre para decidir, a qualquer momento, se quer participar ou não neste experimento.

Compensação: A participação neste experimento é voluntária, e não será oferecida nenhuma remuneração aos seus participantes.

Se tiver algum problema ou se tiver outras perguntas: Se você tiver algum problema que pensa que pode estar relacionado com sua participação neste experimento, ou se tiver qualquer pergunta sobre a mesma, poderá entrar em contato com o responsável pela pesquisa a qualquer momento pelo e-mail soraiareis@dcc.ufmg.br.

Novas condições: Caso deseje, você pode especificar novas condições que devem ser atendidas para que você participe desta avaliação. Para isso, entre em contato com o responsável pela pesquisa através do e-mail e especifique as condições para sua participação.

Instruções sobre a Avaliação

Antes de realizar a avaliação, favor responder a um questionário que fizemos com o objetivo de coletar informações sobre o perfil dos avaliadores que estão participando do experimento. Segue o link de acesso a este questionário.

https://qtrial.qualtrics.com/SE/?SID=SV_4V3slkHrIYCcPl2

Sistema a ser avaliado	TODOIST (http://todoist.com)
Método a ser utilizado	Método de Inspeção Semiótica
Cenário	Joana é uma empresária que atua no setor de Recursos Humanos. A sua empresa oferece diversos serviços, dentre eles: consultoria em recursos humanos, pesquisa de clima organizacional, avaliação de desempenho e recrutamento e seleção. Ela é responsável por coordenar uma equipe de 5 funcionários. Além de ser uma grande profissional, ela é mãe de dois filhos: Gabriel de 6 anos e Valéria de 3 anos. Joana tem uma rotina agitada, onde além de acompanhar a atividades de sua equipe, ela tem que realizar contato com clientes, avaliar os relatórios gerados pela sua equipe e também realizar atividades pessoais (e.g., atividade dos filhos, marido e casa). Devido à dificuldade encontrada por Joana em administrar e lembrar as diversas atividades que ela deve realizar durante o dia, ela resolveu procurar um sistema que apoiasse nesta tarefa. Joana é uma usuária iniciante de tecnologia e por isto precisa de um sistema fácil de utilizar. Em uma pesquisa no Google, Joana encontrou o TODOIST (http://todoist.com/), um sistema de gerenciamento de tarefas, e resolveu explorá-lo para verificar se o mesmo atendia as suas necessidades.
Documento Modelo	Um modelo do documento a ser preenchido foi enviado a você juntamente com este documento de instruções com o nome ***template_mis.doc***. **Atenção:** Pedimos que você preencha este documento com a maior quantidade possível de informações, registrando todos os detalhes que puder. **Padrão:** Caso este modelo não seja o que você esta geralmente acostumado(a) a preencher, deixamos livre que você adicione novos tópicos se achar necessário, mas pedimos que os tópicos existentes não sejam alterados. **Apropriação de Horas:** No modelo você é solicitado a preencher a data, hora de início, hora de término e duração. Gostaríamos de esclarecer que estas informações não tem como obtivo controlar a forma como você avaliou o sistema. Este item tem como objetivo ter uma ideia de quanto tempo você levou para realizar a inspeção do sistema. Portanto, se não estiver confortável em nos informar o dia real e a hora em que realizou as

	avaliações fique a vontade em não informá-las, mas pedimos por gentileza que pelo menos nos informe o tempo total gasto em sua avaliação. **Passo 4 do MIS:** O passo 4 da avaliação é extremamente importante para o nosso experimento. Por este motivo pedimos atenção especial no preenchimento deste item. Precisamos que você liste com detalhes os problemas de comunicabilidade encontrados no sistema, inclusive com evidências (e.g., *print screen* das telas). Para cada problema listado é muito importante que você o classifique de acordo. Abaixo, no item "Classificação das Rupturas de Comunicação", mencionamos como a classificação deve ser realizada.
	Segue links de materiais que você poderá consultar sobre a avaliação com o MIS: • **Artigo sobre o MIS:** The semiotic inspection method.pdf • **Livro sobre o MIS científico:** Semiotic Engineering Methods for Scientific Research in HCI.pdf[2] O uso de material auxiliar não é obrigatório, mas você deverá consultá-lo caso tenha dúvidas em relação a aplicação do método.
Contato	Qualquer dúvida em relação ao experimento, ao material disponibilizado e ao MIS poderá ser enviada ao responsável pela pesquisa através do e-mail soraiareis@dcc.ufmg.br.

Classificação das Rupturas de Comunicação

Completa: O usuário é incapaz de compreender a mensagem do projetista conscientemente ou não. Pode ocorrer quando (1) o usuário não consegue entender a mensagem enviada pelo projetista e desiste de realizar a tarefa; ou (2) o usuário realiza a tarefa de um jeito diferente da intenção do projetista e acredita que esta fazendo corretamente.

Parcial: Estão associados a inesperados caminhos interativos escolhidos pelo usuário. Pode ocorrer quando (1) o usuário entende o que o projetista esta comunicando, mas ele prefere utilizar outros meios para atingir seus objetivos; ou (2) o usuário não entende a mensagem do projetista e conscientemente tenta realizar a tarefa de outra forma.

Temporária: São as situações em que o usuário poderá temporariamente não entender o que aconteceu, não entender o que fazer para prosseguir, não saber em qual contexto esta ou poderá momentaneamente se confundir.

[2] Apesar deste livro demonstrar o uso do MIS científico, ele poderá ser útil para entender como realizar os passos dos MIS visto que toda avaliação científica envolve uma avaliação técnica.

Appendix D

CEM Consent Form

Consent Form

Title: Communicability Evaluation of TODOIST

Date: ____ / ____ / 2011

Institution: DCC/UFMG

Responsible for Research: Soraia Reis (soraiareis@dcc.ufmg.br)

Introduction: This Consent Form contains information on the communicability assessment of TODOIST that you are participating. To ensure that you are informed about your participation in this research, we ask that you make a thorough reading of this document. If you have any questions, feel free to ask the responsible for this research. After reading and agreeing to participate you must sign this form.

Objective of the evaluation: The purpose of this evaluation, which is part of the dissertation of the responsible researcher, is to identify *communicability* problems in the TODOIST system that hinders user-system interaction.

General information about the evaluation: You will be asked to perform some simple tasks using the system. Your interaction with the system while doing the task will be recorded for later analysis by the evaluator. At the end of tasks execution, there will be an interview about your experience with the system.

Use of data collected: The data collected during the evaluation will be used for the analysis of TODOIST as the research objective. Any data submitted for publication will be used to ensure the anonymity of the participants.

Privacy: Information that can identify the participants of the evaluation will not be disclosed. Your name will not appear in any reports. If you wish, you may request a copy of the data generated by you.

If you decide not to participate in the evaluation: You are free to decide, at any time, whether or not to participate in this evaluation and may even stop it if necessary.

Compensation: Participation in this evaluation is voluntary and will not be offered any remuneration/compensation to participants.

If you have any problems or have any other questions: If you have a problem you think may be related to your participation in this evaluation, or if you have any questions about it, you can contact the evaluator at any time by e-mail soraiareis@dcc.ufmg.br.

New conditions: If you wish, you can specify (in the field below) new conditions that must be met for your participation in this evaluation.

Free and Clarified Consent (Voluntary Agreement)

The document describing the above conditions of participation in the "Communicability Evaluation of TODOIST" was explained. I had the opportunity to ask questions about the evaluation, which were answered satisfactorily. I agree to participate as a volunteer.

Participant		Researcher	
Signature:	______________________	Signature:	______________________
Name:	______________________	Name:	______________________

---------- original version in Portuguese ----------

Termo de Consentimento

Título: Avaliação da Comunicabilidade do TODOIST

Data: ____ / ____ / 2011

Instituição: DCC/UFMG

Avaliador Responsável: Soraia Reis (soraiareis@dcc.ufmg.br)

Introdução: Este Termo de Consentimento contém informações sobre a avaliação da comunicabilidade do sistema TODOIST que você esta participando. Para assegurar que você esteja informado sobre a sua participação nesta pesquisa, pedimos que faça a leitura completa deste documento. Caso tenha alguma dúvida, não hesite em perguntar ao avaliador responsável. Após a leitura e se concordar em participar você deverá assinar este termo.

Objetivo da avaliação: O objetivo desta avaliação, que faz parte da pesquisa de dissertação do avaliador responsável, é identificar problemas de *comunicabilidade* no sistema TODOIST que dificultam a interação usuário-sistema.

Informação geral sobre a avaliação: Você será solicitado a realizar algumas tarefas simples utilizando o sistema. A realização dessa tarefa será gravada para posterior análise pelo avaliador. Ao fim da execução da tarefa, será realizada uma entrevista sobre sua experiência com o sistema.

Utilização dos dados coletados: Os dados coletados durante a avaliação serão utilizados para a análise do TODOIST conforme objetivo da pesquisa. Quaisquer dados utilizados para publicação serão apresentados de forma a garantir o anonimato dos participantes da avaliação.

Privacidade: Informações que possam identificar os participantes da avaliação não serão divulgadas. O seu nome não aparecerá em nenhum relatório. Caso deseje, poderá solicitar uma cópia dos dados gerados por você.

Se você decidir não participar na avaliação: Você é livre para decidir, a qualquer momento, se quer participar ou não nesta avaliação, podendo inclusive interrompê-la se achar necessário.

Compensação: A participação nesta avaliação é voluntária, e não será oferecida nenhuma remuneração/compensação aos seus participantes.

Se tiver algum problema ou se tiver outras perguntas: Se você tiver algum problema que pensa que pode estar relacionado com sua participação nesta avaliação, ou se tiver qualquer pergunta sobre a mesma, poderá entrar em contato com o avaliador a qualquer momento pelo e-mail soraiareis@dcc.ufmg.br.

Novas condições: Caso deseje, você pode especificar (no campo abaixo) novas condições que devem ser atendidas para que você participe desta avaliação.

Consentimento Livre e Esclarecido (Acordo Voluntário)

O documento mencionado acima descrevendo as condições de participação da "Avaliação da Comunicabilidade do TODOIST" foi explicado. Eu tive a oportunidade de fazer perguntas sobre a avaliação, que foram respondidas satisfatoriamente. Eu estou de acordo em participar como voluntário.

Participante

Assinatura: ___________________________

Nome: ___________________________

Pesquisador

Assinatura: ______________________________

Nome: ______________________________

Appendix E

CEM Scenario and Tasks

Scenario

You are an entrepreneur and your company offers several services including: consulting and training. You have a team of 5 people. In addition, you are married and have several activities to be performed on your day-to-day life. Your routine is agitated, which besides following your team activities, you have to make contact with clients, evaluate the reports generated by your team and also perform personal activities (e.g., participate in your children activities and home). Due to the difficulty in managing and remembering the several activities that must be done during the day, you decided to seek for a system that supports this task. In a Google search, you found the TODOIST (http://todoist.com), a task management system, and decided to explore it to see if it suits your needs. Perform the following tasks in the system and evaluate whether it meets your needs.

Tasks

Task 1: Suppose that you want to register several activities in the system, some personal and other professional activities. I would like you to register in the system the following activities separating it into personal and professional activities.

Activity/Task	Personal or Professional?	Due Date
Pay cellphone bill	Personal	20th December 2011
Call to schedule a dentist	Personal	Today
Buy a blender	Personal	No due date
Medical visit	Personal	Tomorrow at 8h
Send the final project A report	Professional	18th December 2011
Prepare training X	Professional	19th December 2011
Call to schedule a meeting with John Doe	Professional	Tomorrow
Meeting with Beltran	Professional	Tomorrow at 14h
Interview Sara	Professional	Today at 10h

Task 2: Now that you have your activities/tasks registered, suppose that you want to differentiate projects by color for easy identification. Thus, your task is to change the display color of the projects (Personal and Professional) to the color you want.

Task 3: You may notice that on your personal task list one task is related to a home task (i.e., Buy a blender). Now suppose that you want to define better the Personal tasks separating the tasks related to your home in a subproject called "Home". Your task now is to create within the project Personal the subproject Home and then move the task "Buy a blender" to the subproject "Home".

Task 4: Create a new project called "Academic". You will notice that it will be created below the existing projects. Suppose that you prefer that the projects are displayed in alphabetical order. Your task is then to edit the order of the projects (i.e., sort them).

Task 5: Suppose you like to read daily news, but because of the day-to-day rush you end up forgetting to do this. Register the task "Access Globo.com" in the Personal project. Attention, as this task occurs every day you must create this task in order it repeats daily.

Task 6: Suppose that you want to set the priority of tasks registered. Your task is to set the priority of all tasks registered in the Personal project.

Task 7: Suppose that you want to see the tasks registered of a project in the priority order. Go to the Personal project and sort the activities by priority.

Task 8: Suppose that you just called to your dentist scheduling a visit. Mark then the task "Call to schedule a dentist" as completed.

Task 9: Suppose that you want to know which activities you should do tomorrow. Your task is to search for all the activities/tasks that you must complete tomorrow.

Mark all tasks as completed.

Task 10: Suppose that after marking as completed you realized that marked a task wrong and you want to undo what was done. Try doing that in the system.

Task 11: Delete at least one of the tasks/activities completed the Personal project.

---------- original version in Portuguese ----------

Cenário

Você é uma empresária e a sua empresa oferece diversos serviços, dentre eles: consultorias e treinamentos. A sua equipe é de 5 funcionários. Além disso, você é casado(a) e possui diversas atividades a serem realizadas no seu dia-a-dia. Sua rotina é agitada, onde além de acompanhar a atividades de sua equipe, você tem que realizar contato com clientes, avaliar os relatórios gerados pela sua equipe e também realizar atividades pessoais (e.g., atividade dos filhos e casa). Devido à sua dificuldade em administrar e lembrar as diversas atividades que deve realizar durante o dia, você resolveu procurar um sistema que apoiasse nesta tarefa. Em uma pesquisa no Google, você encontrou o TODOIST (http://todoist.com), um sistema de gerenciamento de tarefas, e resolveu explorá-lo para verificar se o mesmo atende às suas necessidades. Execute as seguintes tarefas no sistema e avalie se o mesmo atenderá às suas necessidades.

Tarefas

Tarefa 1: Suponha que você queira cadastrar diversas atividades no sistema, algumas delas atividades pessoais e outras profissionais. Gostaria que você cadastrasse no sistema as seguintes atividades separadas em atividades pessoais e profissionais.

Atividade/Tarefa	Pessoal ou Profissional?	Data para Concluir
Pagar conta de celular	Pessoal	20/12/2011
Ligar para marcar dentista	Pessoal	Hoje
Comprar liquidificador	Pessoal	Sem data para concluir
Consulta médica	Pessoal	Amanhã às 8h
Entregar relatório final do projeto A	Profissional	18/12/2011
Preparar treinamento X	Profissional	19/12/2011
Ligar para marcar reunião com Fulano	Profissional	Amanhã
Reunião com Beltrano	Profissional	Amanhã às 14h
Entrevistar Ciclano	Profissional	Hoje às 10h

Tarefa 2: Agora que você possui suas atividades/tarefas cadastradas, suponha que você deseje diferenciar os projetos por cores para facilitar a identificação. A sua tarefa então é alterar a cor de exibição dos projetos (Pessoal e Profissional) para a cor que você mais desejar.

Tarefa 3: Você pode observar que na lista de tarefas pessoais uma delas esta relacionada a tarefas de casa (i.e., Comprar liquidificador). Suponha então que você queira definir melhor as tarefas do projeto Pessoal separando as tarefas relacionadas à sua casa em um subprojeto chamado "Casa". Sua tarefa agora é criar dentro do projeto Pessoal o subprojeto Casa e em seguida mover a tarefa "Comprar liquidificador" para o subprojeto Casa.

Tarefa 4: Crie um novo projeto chamado "Acadêmico". Você vai observar que ele será criado abaixo dos projetos já existentes. Suponha que você prefira que os projetos estejam em ordem alfabética. A sua tarefa é então fazer modificações na ordem dos projetos (i.e., ordená-los).

Tarefa 5: Suponha que você goste de ler notícias diariamente, mas que com a correria do dia-a-dia você acaba esquecendo de fazer. Cadastre a tarefa "Acessar o Globo.com" no projeto Pessoal. Atenção, como esta tarefa ocorre todos os dias você deve criar esta tarefa de forma que ela seja repetida diariamente.

Tarefa 6: Suponha agora que você gostaria de definir a ordem de prioridade das tarefas cadastradas. A sua tarefa é definir a prioridade de todas as tarefas cadastradas no projeto Pessoal.

Tarefa 7: Suponha que você queira visualizar as tarefas cadastradas de um projeto na ordem de prioridade. Vá até o projeto Pessoal e ordene as atividades por prioridade.

Tarefa 8: Suponha que você acabou de ligar para sua dentista marcando uma consulta. Marque então a tarefa "Ligar para marcar dentista" como concluída.

Tarefa 9: Suponha que você gostaria de saber quais são as atividades que você deve fazer amanhã. Sua tarefa é pesquisar todas as atividades/tarefas que você deve concluir amanhã.

Marque todas as tarefas como concluída.

Tarefa 10: Suponha que após marcar como concluída você percebeu que marcou uma tarefa de forma errada e você queira desfazer o que foi feito. Tente fazer isso no sistema.

Tarefa 11: Exclua pelo menos uma das tarefas/atividades concluídas do projeto Pessoal.

Appendix F

CEM Pre-test Questionnaire

Pre-test Questionnaire

Name: ____ **Date:** ____ / ____ / 2011

Gender: ( ) Female ( ) Male **Age:** ________

Please enter the data requested with the option that best represents your answer to each question.

1. What is your English knowledge level?

( ) Basic

( ) Intermediate

( ) Advanced

2. What is your training level?

( ) Undergraduate ( ) Complete

( ) Post-graduate ( ) Incomplete

( ) Master

( ) Doctor

3. **Course:** ____

4. **How often per week do you use the computer and/or internet?**

( ) 1 or 2 times a week

( ) from 3 to 5 times a week

( ) every day

5. **How often daily do you use the computer and/or internet?**

( ) up to 2 hours a day

( ) from 3 to 5 hours a day

( ) from 6 to 8 hours a day

( ) more than 8 hours a day

6. **List the systems/software you use most.**

7. **How would you rate your computer knowledge?**

( ) Basic

( ) Intermediate

( ) Advanced

8. **Do you use or have used a system to manage your day-to-day activities/ tasks (i.e., register tasks and have a record/manage the ones you already completed or will complete)?**

( ) Yes If you checked Yes, cite which one(s): __________________________________

( ) No If you checked No, you would be interested in using such a system?

( ) Yes ( ) No

9. Do you know the system TODOIST?

( ) Yes

( ) No

---------- original version in Portuguese ----------

Questionário Pré-Teste

Nome: ____ **Data:** ____ / ____ / 2011

Sexo: ( ) Feminino ( ) Masculino **Idade:** _________

Por favor, informe os dados solicitados com a opção que melhor representa a sua resposta a cada pergunta.

1. Qual é o seu nível de conhecimento da língua inglesa?

( ) Básico

( ) Intermediário

( ) Avançado

2. Qual é o seu nível de formação?

( ) Graduação ( ) Completo

( ) Pós-graduação ( ) Incompleto

( ) Mestrado

( ) Doutorado

3. Curso: ____

4. Com qual frequência semanal utiliza o computador/internet?

( ) 1 ou 2 vezes por semana

( ) de 3 a 5 vezes por semana

( ) todos os dias da semana

5. Com qual frequência diária utiliza o computador/internet?

( ) até 2 horas por dia

( ) de 3 a 5 horas por dia

( ) de 6 a 8 horas por dia

( ) mais de 8 horas por dia

6. Liste os sistemas/software que você mais utiliza?

7. Como você classificaria o seu conhecimento em informática?

( ) Básico

( ) Intermediário

( ) Avançado

8. Você utiliza ou já utilizou algum sistema para gerenciar suas atividades/tarefas (i.e., cadastrar tarefas e ter um registro/gerenciar aquelas que você já concluiu ou deverá concluir) do dia-a-dia?

( ) Sim   Caso tenha marcado Sim, cite qual(is): ____________________________________

( ) Não   Caso tenha marcado Não, você teria interesse em utilizar um sistema deste tipo?

( ) Sim   ( ) Não

9. Você conhece o sistema TODOIST?

( ) Sim

( ) Não

Appendix G

CEM Post-test Interview

Post-test Interview – TODOIST

1. In the main page of the system, did you have some initial difficulty to locate and identify where to start?

2. What did you think of how the system provides the functionality to assign dates to the tasks (e.g., common and recurring)?

3. Did you have some initial difficulty to add a project or task?

4. In relation to the project and task setup menu, did you have some trouble finding the features? Were there any that you did not understand?

5. In the task setup menu there is the option "Move to history", did you come to think what would this functionality be? Did you understand what it is for?

6. Is there an icon, button and/or field that you had difficulty in understanding? Cite which and the difficulties encountered?

7. Did you feel any difference in patterns on the system as, for example, use of an icon, button and/or field differently in different contexts?

8. Did you have any difficulty in relation to mark a task as completed, view completed tasks and/or delete completed tasks?

9. What did you think of the search tasks feature?

10. Did you struggle to increase/decrease the indentation of projects and tasks? Which one(s)?

11. Would you like to make additional comments about your experience with the interface of this system?

---------- original version in Portuguese ----------

Entrevista Pós-Teste – TODOIST

1. Na página principal do sistema, você chegou a ter alguma dificuldade inicial para localizar e identificar por onde começar?

2. O que você achou da forma como o sistema disponibiliza a funcionalidade de atribuir datas às suas tarefas (e.g., comum e recorrente)?

3. Você teve alguma dificuldade inicial para inserir algum projeto ou tarefa?

4. Em relação ao menu de configuração do projeto e tarefa, você teve alguma dificuldade para encontrar as funcionalidades? Teve alguma que você não entendeu?

5. No menu de tarefas existe a opção "Move to history", você chegou a pensar o que seria essa funcionalidade? Entendeu para que ela serve?

6. Existe algum ícone, botão e/ou campo que você teve dificuldade em entender? Cite quais e as dificuldades encontradas?

7. Você sentiu alguma diferença de padrão no sistema como, por exemplo, uso de um ícone, botão e/ou campo de forma diferente em diferentes contextos?

8. Você teve alguma dificuldade em relação a marcar tarefa como concluída, visualizar tarefas concluídas e/ou excluir tarefas concluídas?

9. O que você achou da forma de pesquisa de tarefas?

10. Você teve dificuldade para aumentar/diminuir o recuo de projetos e tarefas? Qual(is)?

11. Você gostaria de fazer comentários adicionais sobre a sua experiência com a interface deste sistema?

Appendix H

Todoist Changes

TODOIST Changes

Entering Page

- [Figure 1 – items 1a and 1b] Menu Bar
  - As we can see the menu bar at the top of the system has changed. Before it had four elements: (1) Get it all done with Todoist Premium, which was an option to the user to upgrade his account to a premium version; (2) Todoist help, which was a link to show/hide the system help (the system help is shown in this figure); (3) Preferences, which is the account settings; and (4) Logout. In the new version the first element (i.e., Get it all done with Todoist Premium) was excluded, the Todoist help is now a question image, the third and fourth elements are now available when clicking on the gear sign (see Figure 2). In the new version of Todoist there are two new images: (1) a plus sign, which allows the user to add a new task (see Figure 3); and (2) a printer sign.
- [Figure 1 – items 2a and 2b] System Help Content
  - As shown in Figure 1 the system help content has not changed. The only difference found was related to the images, that are, in the new version, black and white.
- [Figure 1 – items 3a and 3b] Logo and Search Filter
  - As shown in Figure 1 the system logo has changed, but the actions are the same. When clicking in the logo the system refresh the current view. In the filter options the only change was related to the filters "@" and "q:", which were premium features, and is not shown now.
- [Figure 1 – items 4a and 4b] Information About How to Begin
  - The information given by the system about how the user should begin had small changes. Only in the colors and images.

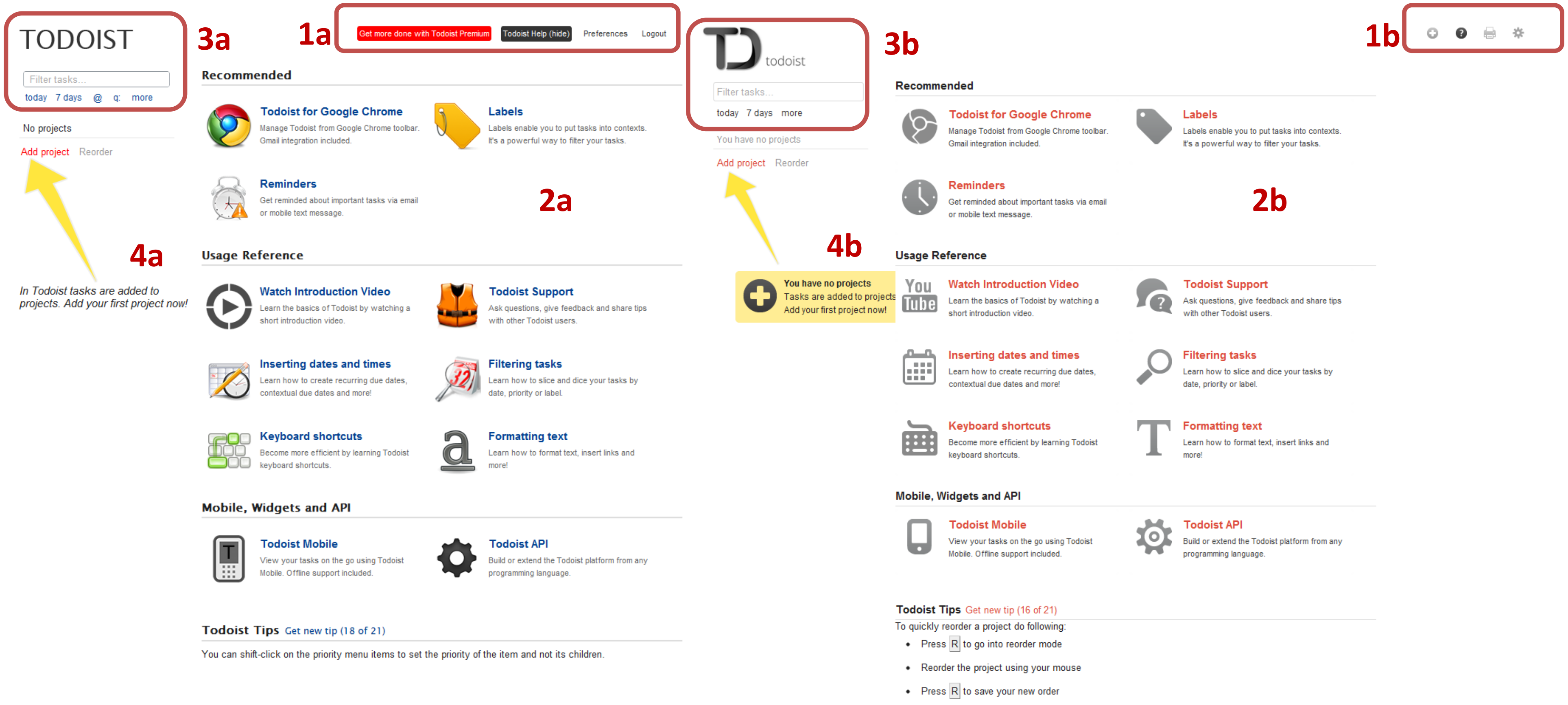

Figure 1. Entering Page Differences

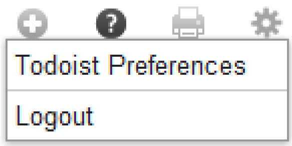

Figure 2. Menu shown when clicking on the gear sign.

Figure 3. Menu shown when clicking on the plus image.

Projects/Tasks Page

- [Figure 4 – items 1a and 1b] Tasks Recording
  - As we can see in Figure 4 the only difference detected is related to the box where the tasks recording number is shown besides the project name. In the new version of the system this box is thin and the number is shown outside of the box, right beside de project name.
- [Figure 4 – items 2a and 2b] Completed tasks
  - The designers of the system also changed the term used to show completed tasks. Before they used the term “Get” and after they used the term “Show”.

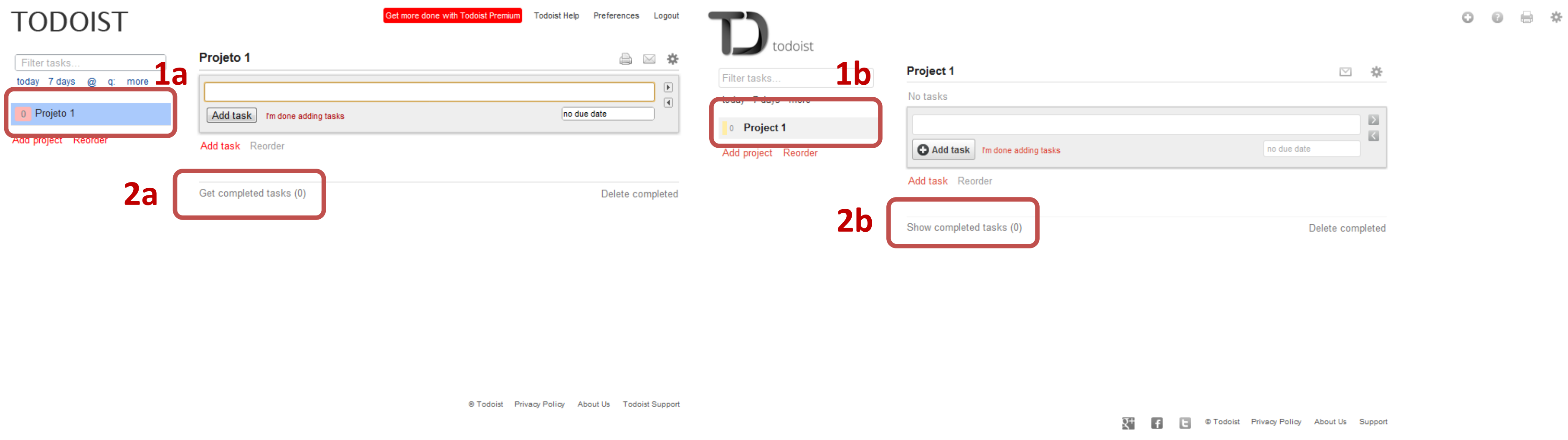

Figure 4. Projects/Tasks Page Differences

Tasks Menu

- We also can see changes in the task menu (Figure 5), where an additional feature was added: “Postpone”. This feature allows the user to postpone the task date in 1 day.

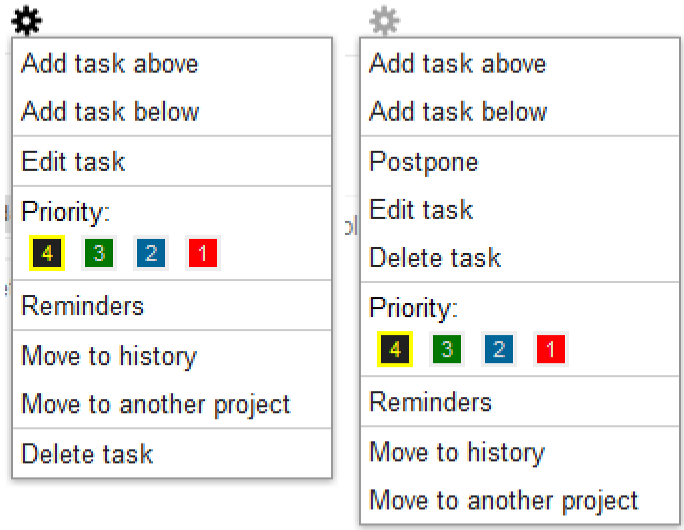

Figure 5. Tasks Menu Differences