HAL open science

# Do Teams Achieve Usability Goals? Evaluating Goal Achievement with Usability Goals Setting Tool

Anirudha Joshi, N. L. Sarda

**To cite this version:**

Anirudha Joshi, N. L. Sarda. Do Teams Achieve Usability Goals? Evaluating Goal Achievement with Usability Goals Setting Tool. 13th International Conference on Human-Computer Interaction (INTERACT), Sep 2011, Lisbon, Portugal. pp.313-330, 10.1007/978-3-642-23774-4_26 . hal-01590579

**HAL Id: hal-01590579**

**https://inria.hal.science/hal-01590579**

Submitted on 19 Sep 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

# Do Teams Achieve Usability Goals? Evaluating Goal Achievement with Usability Goals Setting Tool

Anirudha Joshi[1], NL Sarda[1]

[1] Indian Institute of Technology, Bombay
anirudha@iitb.ac.in, nls@iitb.ac.in

**Abstract.** Do teams achieve important usability goals most of the time? Further, is goal achievement uniform or are practitioners more mindful of some goals than others? This paper presents an empirical study on usability goal achievement in industry projects. We used Usability Goal setting Tool (UGT), a recommender system that helps teams set, prioritize, and evaluate usability goals. The practitioner creates profiles for the product and its users. Based on these inputs, UGT helps the practitioner break down high-level usability goals into more specific goal parameters and provides recommendations, examples, and guidelines to assign weights to these parameters. UGT suggests strategies to evaluate goal parameters after the design is ready and assign them scores. UGT was used to collect data from 65 projects in the Indian software industry in which participants assigned weights and scores to the goal parameters. The 30 goal parameters suggested by UGT were found to be internally reliable, and having acceptable granularity and coverage. It was observed that goal parameter weights and scores correlated, but only moderately. Another interesting observation was that more than a third of the important goal parameters did not score well. We identify eight goal parameters that are typically high-weighted but have poor weight-score correlations. We call these *"latent but important"* goal parameters. Design teams will do well to pay closer attention to these goal parameters during projects.

**Keywords:** Usability goals achievement, usability goal parameters, latent goals, design tools, methods

## 1. Introduction

Setting goals is an important step early in the design process. Setting goals before design gives the team a target to achieve. Goals help guide the design process, make the design activity tangible, and help evaluate the designs. In the field of human-computer interaction (HCI), often multi-disciplinary teams are involved; hence, setting goals early and getting an agreement from all stakeholders is important.

Goal have been discussed extensively in HCI literature [1-9]. To help teams set goals, we developed a Usability Goal Setting Tool (UGT) [10], [11]. UGT maintains a repository of profiles of past projects, their users, goals, and scores. Based on this experience, UGT provides guidance to HCI practitioners to set and evaluate usability

goals in new projects. Is such a tool necessary or is the current knowledge, skills and practices sufficient to let HCI practitioners set usability goals and evaluate products against them? During our interactions with software development teams in the industry, we observed that HCI practitioners are often unclear about the usability goals of the product they are designing. In some cases, the HCI practitioners are clear, but there is no explicit agreement on these by the other stakeholders in the team.

The questions we investigate in this paper are: To what extent do teams achieve their usability goals? Do teams achieve the important usability goals most of the time? Further, is goal achievement uniform or are practitioners more mindful of some usability goals than other?

We collected retrospective data with the help of UGT from projects in the Indian industry to evaluate goal achievement. We explored the data to establish the coverage, relevance, granularity, and internal reliability of the goal parameters.

In section 2, we review literature related to goal setting in design and in HCI literature. Section 3 gives an overview of UGT. In section 4, we present the data collected from industry projects and investigate the coverage, relevance, granularity, and internal validity of goals suggested in UGT. Section 5 investigates the goal achievement in these projects. Section 6 talks about conclusions and future work.

## 2. Goals in Design and HCI

The importance of goal-driven design has been stressed in literature for a long time. Design for a 'need' has been a part of traditional industrial design thinking. Charles Eames reportedly said, *"Design is a plan for arranging elements in such a way as best to accomplish a particular purpose"* [12]. Archer explains that since design is necessarily associated with change, identifying goals means *"defining the needs and pressures, which constitute the driving force for change"*. The first step is to determine the goals of the design effort together with *"the essential criteria by which a 'good' solution will be distinguished from a 'not so good' solution"* [13].

The closely related fields of HCI, usability, interaction design, and information architecture emphasise the importance of **usability goals**. ISO 9241 defines usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [2]. Shneiderman and Nielsen agree on five high-level usability attributes of a product – learnability, speed of use, error-free use, retention over time, and subjective satisfaction [7], [1]. Mayhew categorizes usability goals as qualitative and quantitative, and as performance goals, preference goals, and satisfaction goals [3]. ISO 9126-1 describes usability in terms of understandability, learnability, operability, and attractiveness [4]. Cooper and Reimann emphasise the importance of goals in usability and interaction design process [6]. Hornbæk et al. suggest a technique to integrate business goals in usability evaluations [9]. Bevan summarises several other ways of organising usability measures, which could be looked upon as goals [8].

One purpose of defining these usability goal-sets was to reduce the ambiguity associated with the terms *"usability"* or *"user-friendliness"*, and define them in terms of measurable components. Usability has been recognised to be a multi-dimensional

property that needs to be broken down into measurable attributes such as learnability, speed of use etc. [1].

However, these attributes themselves are still at too broad to be interpreted consistently in the context of a given project. For example, one can interpret *learnability* as *"users should take less time to learn to use the product"* or as *"users should be able to learn to use the product on their own"*. *Error-free use* could be either interpreted as *"product should not induce errors"*, *"product should tolerate user's errors"*, or *"product should help the user recover from errors"*. These interpretations could lead to communication gaps between team members and could make it difficult to evaluate the design. As summarised below, our approach in UGT is to split high-level usability goals into more granular, measurable, and less ambiguous goal parameters.

Relatively recent research interest in human emotions has broadened the traditional focus of researchers from usability to **user experience**. However, approaches to user experience goals are not universal. Unlike in usability, where high-level goals have been more or less agreed upon, when it comes to user experience design, there seems to be much less agreement. After a survey of user experience professionals, Law et al concluded that the concept of user experience is dynamic, context-dependent, and subjective [14]. There is no single accepted definition of what user experience is, and some researchers question whether the user experience could be designed at all [15].

While there may be no explicit agreement on a list of user experience goals, most practitioners favour a *"goal-driven"* approach in design of interactive products, e.g. [3], [5], [6]. When a practitioner sets goals in a particular project, he may not explicitly differentiate between whether a particular goal is related to usability of the product or its broader user experience.

In a typical industrial context, it may not be possible to meet all goals and it is necessary to **prioritize**. Depending on the context, users, and platform, some goals may be more important for a project, while others might be irrelevant. The need of prioritising goals has been acknowledged for a while. For example, Archer talks about *"rank ordering"* sub-problems as a method of prioritizing goals and resolving conflicts [13]. More recently, Cross talks about an objective tree method – organizing objectives into a hierarchy of higher and lower level objectives [16]. Shneiderman states, *"a clever design for one community of users may be inappropriate for another community"*, and *"an efficient design for one class of tasks may be inefficient for another class"* [7]. However, other than UGT (which we summarise below), there has been no guidance to prioritize usability goals systematically.

In our earlier work, we found that goal achievement is affected by the process. When HCI activities are better integrated with software development, teams following Agile processes achieve their usability goals better than teams following the Waterfall process [17]. We also found that unless they integrate HCI activities with software development rigorously, companies providing software development services do worse in usability goals achievement than companies involved in software product development. Elsewhere, we report the relative contributions of different HCI activities on goal achievement [18].

Despite the extensive literature on usability goals, there have been no empirical studies on the **extent of usability goals achievement**. We also do not know if some usability goals are achieved more frequently than others are. Further, it is not clear

whether the current understanding of goals among practitioners is sufficient, or tools such as the UGT are necessary to improve goal achievement. In this paper, we present one such study.

## 3. UGT Overview

The primary purpose of UGT is to help set usability goals, given a context of user goals and business goals [10]. This section provides an overview of UGT. A more detailed version of UGT including systematic instructions is available online [11].

UGT is envisioned as a recommender system (**Fig. 1**). Based on prior experience of similar product and user profiles, UGT recommends priorities for usability goals. UGT also recommends usability evaluation guidelines. At all times, the practitioner has the freedom to add goals, edit, or re-word the suggested goals, to regroup goals, or to evaluate them differently than suggested in UGT. These decisions are collected and fed back into UGT to improve its recommendations in future, making UGT flexible and extensible.

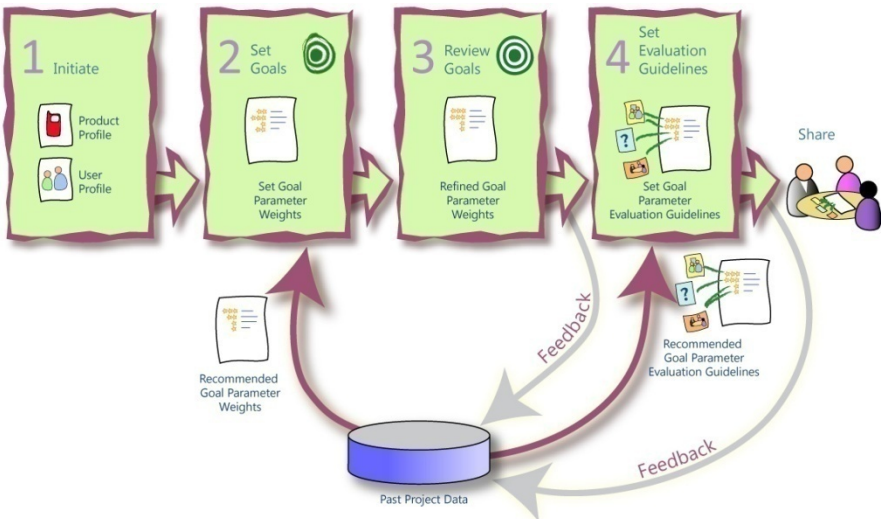



**Fig. 1**: An overview of UGT usage workflow.

### 3.1 Goal Parameters in UGT

UGT helps break down high-level usability goals into 30 concrete, specific, and measurable goal parameters (**Table 1**). These were derived from the usability attributes suggested by Shneiderman [7] and Nielsen [1], and elaborated through brainstorming and formative evaluations as discussed in [10]. Our attempt is to keep

these goal parameters granular, concrete, and yet relevant to a large number of projects.

The HCI practitioner determines the importance of each goal parameter in the context of his project by assigning it a suitable weight. Weights express the relative importance of goal parameters in the context of a project. For example, a product meant to be used several times a day is likely to have higher weight for a speed related goal parameter such as goal parameter 8, *"user must be able to do primary task / frequent tasks quickly, easily, at all times"*. A one-time use product might give higher weight to learnability and ease of use goal parameters such as 1, *"findability: options / data / information should be visible / easy to find"* and 17, *"intuitiveness: user should be able to predict the next step / task"*. A life-critical product is likely to rate highly error-free use goal parameters such as 25, *"product should not induce errors"* and may sacrifice a learnability goal parameter such as 3, *"users should be able to learn on their own"*.

We suggest the following scale for assigning weights: Extremely important / unique selling proposition (5), very important (4), important (3), usual relevance / hygiene factor (2), somewhat relevant (1), irrelevant (0). Based on experience of prior projects, UGT recommends weights for the goal parameters on this scale [11].

We do not claim that the goal parameters and weights recommended by UGT cover all aspects of user experience in every situation. They are suggestive of the kind of goals that the practitioner *could* set. An experienced practitioner may assign different weights, add, edit, or re-word the suggested goal parameters, or group goal parameters differently if preferred. Yet, as we report below, the current list of goal parameters seems to have wide coverage, reasonable relevance, and sufficient granularity.

Goal-setters should be aware that while it may be tempting to set a high weight to each goal parameter, it might not be necessary, practical, or even possible to achieve such a design. The weights should reflect the priorities of the business, the stakeholders, and the users. The weights would help prioritize usability evaluation activity as the highest rated goals and parameters would be evaluated more rigorously.

After the evaluation, practitioners can assign a score to each goal parameter. We suggest the following scale for assigning scores: Outstanding, exceptional, best possible user experience against this goal parameter (100); acceptable, good enough, though not exceptional (75), undecided, neither good nor bad (50), bad, but not the worst (25), very bad, worst possible user experience for this goal parameter (0).

**Table 1**: Goals and goal parameters in UGT.

| | Learnability |
|---|---|
| 1 | **Findability**: options / data / information should be visible / easy to find |
| 2 | User should take less **time to learn**: (e.g. in < 10 minutes, in < 2 hours practice) |
| 3 | Users should be able to **learn on their own** |
| 4 | Product should be **internally consistent** |
| 5 | Product should be consistent with **other products**, older **methods** / past **habits** |
| 6 | Product should be consistent with **earlier version** |
| 7 | User should remember / **retain** critical, but infrequent tasks |

| | |
|---|---|
| | Speed of use |
| 8 | User must be able to do **primary** task / **frequent tasks** quickly, easily, at all times |
| 9 | User should be able to **navigate** quickly and easily |
| 10 | Product should not load user's **memory** / product should not put **cognitive load** |
| 11 | **Flexibility**: User should **control** the sequence of tasks |
| 12 | User should be able to complete tasks in specific **time** / no. of **steps** / in less **efforts** |
| 13 | Product should be **personalised** for the user automatically |
| 14 | Product should be **localised** for specific market segments |
| 15 | User should be able to **customise** the product for himself |
| | Ease of use |
| 16 | Interface should clearly communicate the **conceptual model** |
| 17 | **Intuitiveness**: User should be able to **predict** the next step / task |
| 18 | No **entry barrier**: user must be able to complete **critical first tasks** |
| 19 | Product should require no **unnecessary tasks** |
| 20 | Product should **automate** routine tasks / minimise user **task load** |
| 21 | Product should be **always on**, always accessible |
| | Ease of Communication |
| 22 | **Information architecture**: well **aggregated, categorised, presented** |
| 23 | **Communication** should be clear / user should easily **understand** text, visuals |
| | Error-free use |
| 24 | Product should give good **feedback** / display its current **status** |
| 25 | Product should **not induce errors** |
| 26 | Product should tolerate user's errors / **forgiving** interface / should prevent errors |
| 27 | Product should help user **recover from errors** / help users **troubleshoot** problems |
| | Subjective Satisfaction |
| 28 | User should **feel in control** of the product / **behavioural** appeal |
| 29 | User should feel **emotionally engaged** / brand / fun / reflective appeal / trust |
| 30 | User should find the product **aesthetically** appealing / **visceral** appeal |
| | Other goals |
| | Add goal parameters specific to your project here. |

## 3.2 Using UGT

The HCI practitioner approaches UGT when he is clear about the product brief and has sufficient understanding about the domain, the problems, the context, and the users. Using UGT involves four steps: initiate, set goals, review goals, and set evaluation guidelines.

To begin with, UGT asks the practitioner to specify the **product profile** and one or more **user profiles**. The product profile captures information such as the industry domain, the work-practice domain, the expected cost to the user, the platforms, and the number of user profiles. Each user profile captures further information such as the

age, the expected level of tech-savvyness of the user, the frequency of use, the product complexity, and the motivation of the user to use the product.

The product profile and the user profile form the input to UGT. Based on these, UGT provides recommendations for **assigning weights** to the goal parameters. Detailed recommendations for all the 30 goal parameters are available online [11]. These recommendations have been derived from experience of projects for which UGT has been used so far.

After the practitioner has done one pass of assigning weights to goal parameters, the list of goal parameters is re-presented, this time sorted by their weights. Goal parameters with weights that deviate substantially from the recommended range are highlighted. The practitioner **reviews** the weights and tweaks them. This step was introduced because it was observed during evaluations that practitioners over-assigned weights in the first pass, but preferred to tone them down during a review.

UGT recommends **evaluation guidelines** of each goal parameter. It recommends at least one possibility for a user-based test and one for a review-based evaluation. The evaluation guidelines for the 30 goal parameters are available online [11]. These are based on the formative evaluation reported in [10], and further data from projects where UGT has been applied so far.

As shown in **Fig. 1**, UGT has self-learning capabilities. The practitioner may use UGT recommendations or choose to override them. These decisions are captured and fed back to enrich UGT further. As UGT collects data from more projects, it gives more accurate and relevant recommendations.

## 4. Data and Analyses

Using UGT, we collected data about goal parameter weights and scores from industrial projects. First, we evaluate whether UGT goal parameters have sufficient coverage, relevance, and granularity, and whether the weights and scores used in UGT are internally reliable. Next, we determined the extent to which the HCI practitioners achieved goals of their projects, and whether practitioners are likely to be equally mindful of all goals parameters.

HCI practitioners from the Indian IT industry were trained on UGT. Then they were invited to participate in the study by contributing data from industry projects that they had recently completed. For each project, following data was collected:

- Product profiles and user profiles
- Goal parameter weights (0-5)
- Goal parameter scores (0-100). Participants were encouraged to use information from usability tests, reviews, and user feedback where available while giving their scores.

44 HCI practitioners participated and contributed data from 65 projects. Some participants contributed data from more than one project. The participants came from a wide variety of companies including four large (25,000+ employees) companies engaged in developing software on contract, a few relatively smaller companies, some multi-national companies with large product development centres in India, and smaller product development companies. The projects represented a wide variety as

well. These included different platforms (desktop, web and mobiles), industry domains (finance, telecom, entertainment etc.), targeted users (call centre agents, sales persons, farmers etc.), types of business models (contracted software development companies and product companies), and process model used (waterfall and agile).

**Table 2** lists the number of projects that assigned a particular weight (from 0 to 5) and a particular score (0-25-50-75-100) to each goal parameter. It also lists the means and standard deviations of weights and scores of each goal parameter.

**Table 2:** Number of projects that assigned a particular weight to a goal parameter, goal parameter weight means and standard deviations, number of projects that assigned a particular score to a goal parameter, and goal parameter score means and standard deviations. # refers to the goal parameter numbers in Table 1. (n = 65)

| # | No. of projects out of 65 with weights | | | | | | Weights | | No. of projects out of 65 with scores | | | | | Scores | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | Mean | SD | 0 | 25 | 50 | 75 | 100 | Mean | SD |
| 1 | 0 | 0 | 3 | 13 | 30 | 19 | 4.00 | 0.83 | 0 | 1 | 16 | 38 | 10 | 71.92 | 16.83 |
| 2 | 1 | 5 | 12 | 24 | 17 | 6 | 3.06 | 1.13 | 0 | 7 | 22 | 31 | 5 | 63.08 | 19.82 |
| 3 | 5 | 6 | 15 | 14 | 19 | 6 | 2.83 | 1.40 | 5 | 5 | 16 | 36 | 3 | 60.38 | 24.56 |
| 4 | 1 | 4 | 17 | 19 | 17 | 7 | 3.05 | 1.16 | 2 | 4 | 18 | 28 | 13 | 67.69 | 24.09 |
| 5 | 5 | 5 | 19 | 18 | 8 | 10 | 2.75 | 1.41 | 4 | 7 | 31 | 17 | 6 | 55.38 | 24.40 |
| 6 | 37 | 3 | 8 | 6 | 6 | 5 | 1.32 | 1.75 | 34 | 1 | 14 | 14 | 2 | 30.38 | 34.09 |
| 7 | 10 | 12 | 11 | 17 | 10 | 5 | 2.31 | 1.52 | 8 | 7 | 26 | 22 | 2 | 51.15 | 25.93 |
| 8 | 3 | 3 | 5 | 14 | 28 | 12 | 3.49 | 1.28 | 4 | 2 | 10 | 33 | 16 | 71.15 | 25.86 |
| 9 | 1 | 1 | 12 | 11 | 30 | 10 | 3.51 | 1.11 | 0 | 7 | 11 | 32 | 15 | 71.15 | 22.63 |
| 10 | 1 | 2 | 23 | 14 | 17 | 8 | 3.05 | 1.18 | 4 | 10 | 15 | 33 | 3 | 58.08 | 25.04 |
| 11 | 7 | 8 | 16 | 19 | 11 | 4 | 2.48 | 1.37 | 6 | 9 | 22 | 25 | 3 | 53.85 | 25.86 |
| 12 | 4 | 1 | 13 | 20 | 15 | 12 | 3.18 | 1.33 | 2 | 3 | 22 | 31 | 7 | 64.62 | 21.60 |
| 13 | 15 | 10 | 12 | 6 | 17 | 5 | 2.23 | 1.70 | 19 | 9 | 12 | 21 | 4 | 43.08 | 33.80 |
| 14 | 17 | 12 | 8 | 10 | 9 | 9 | 2.14 | 1.79 | 20 | 5 | 18 | 13 | 9 | 44.62 | 35.77 |
| 15 | 29 | 9 | 6 | 8 | 9 | 4 | 1.55 | 1.73 | 33 | 6 | 9 | 13 | 4 | 30.38 | 35.21 |
| 16 | 0 | 5 | 8 | 18 | 21 | 13 | 3.45 | 1.17 | 0 | 4 | 20 | 37 | 4 | 65.77 | 17.44 |
| 17 | 1 | 4 | 11 | 25 | 18 | 6 | 3.12 | 1.10 | 0 | 3 | 27 | 27 | 8 | 65.38 | 19.11 |
| 18 | 7 | 6 | 10 | 19 | 16 | 7 | 2.80 | 1.47 | 6 | 8 | 20 | 25 | 6 | 56.54 | 27.34 |
| 19 | 3 | 2 | 19 | 26 | 12 | 3 | 2.78 | 1.08 | 1 | 9 | 16 | 32 | 7 | 63.46 | 23.00 |
| 20 | 10 | 7 | 12 | 16 | 16 | 4 | 2.51 | 1.51 | 9 | 8 | 22 | 23 | 3 | 51.15 | 27.75 |
| 21 | 10 | 6 | 16 | 10 | 10 | 13 | 2.66 | 1.70 | 13 | 6 | 11 | 26 | 9 | 54.62 | 33.92 |
| 22 | 0 | 1 | 3 | 24 | 18 | 19 | 3.78 | 0.98 | 0 | 3 | 15 | 34 | 13 | 71.92 | 19.52 |
| 23 | 0 | 0 | 5 | 14 | 29 | 17 | 3.89 | 0.89 | 0 | 3 | 17 | 31 | 14 | 71.54 | 20.19 |
| 24 | 0 | 0 | 10 | 21 | 22 | 12 | 3.55 | 0.97 | 1 | 5 | 19 | 27 | 13 | 67.69 | 23.27 |
| 25 | 3 | 0 | 17 | 24 | 14 | 7 | 3.03 | 1.16 | 3 | 6 | 22 | 30 | 4 | 60.00 | 22.88 |
| 26 | 3 | 6 | 20 | 24 | 10 | 2 | 2.58 | 1.10 | 4 | 7 | 28 | 20 | 6 | 56.54 | 24.72 |
| 27 | 2 | 5 | 19 | 18 | 16 | 5 | 2.86 | 1.20 | 3 | 7 | 24 | 24 | 7 | 59.62 | 24.48 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 1 | 2 | 9 | 28 | 19 | 6 | 3.23 | 1.01 | 3 | 2 | 15 | 40 | 5 | 66.15 | 21.39 |
| 29 | 8 | 12 | 10 | 15 | 12 | 8 | 2.54 | 1.58 | 11 | 7 | 24 | 17 | 6 | 50.00 | 29.97 |
| 30 | 2 | 6 | 18 | 21 | 14 | 4 | 2.78 | 1.17 | 2 | 9 | 20 | 26 | 8 | 61.15 | 24.62 |

### 4.1 Goal Parameter Coverage, Relevance, Granularity and Reliability

While recommending goal parameters in UGT, our aim was to achieve a reasonable level of granularity so that the goal parameters are concrete, specific, measurable, and interpreted unambiguously. At the same time, we wanted the goal parameters to be widely applicable, relevant to a large number of projects, and internally reliable. We evaluated the data from projects to verify if this was indeed the case.

**Coverage** is the extent to which the goal parameters suggested by UGT suffice the goal-setting needs of a project. One method of evaluating if UGT has enough coverage was to look at how many new goal parameters were added by practitioners. Throughout the study, the participants had the freedom to add, edit, merge, or regroup goal parameters if the suggested ones did not suit their needs. This was indicated to them at the beginning of the study. To remind the participants, the UGT form always had blank lines under each group of goal parameters and at the bottom as shown in **Table 1**. After assigning goal weights, the participants were encouraged to express goals relevant to their projects beyond the ones listed in UGT.

A majority of the participants did not specify additional goal parameters. The goal parameters currently recommended in UGT could express goals in 59 out of 65 projects to the satisfaction of the practitioners in those projects. Practitioners of only 6 projects added goal parameters. Among these, 4 specified one additional goal parameter each, and 2 specified 2 additional goal parameters each. We can conclude that currently UGT has a reasonable coverage and is applicable to a majority of projects in the industry.

**Relevance** is the number of goal parameters that could be removed from UGT without hampering coverage significantly. **Table 2** lists the number of projects that assigned a particular weight to each goal parameter. Out of the (65 projects x 30 goal parameters=) 1,950 cases, a weight greater than 0 is assigned to 1,757 (90%) cases and in only 193 (10%) of the 1,950 of the cases, the weight was set to 0 (irrelevant). As we can see in **Table 2**, weight 0 is not restricted to a few goal parameters, but spread across several. Further, even goal parameters weighted irrelevant by a large number of projects (goal parameter 6 and 15, for instance) have been weighted important (3 or more) by many other projects. We can conclude that the goal parameters suggested in UGT are relevant to projects.

**Granularity** is the extent to which UGT helps break down high-level goals into concrete goal parameters. Our aim was to achieve a good balance of granularity. We wanted the goal parameters to have sufficient granularity to allow practitioners to express themselves precisely, and yet not so fine that they cannot differentiate between adjacent elements. If the list of suggested goal parameters were too long, it could hamper the usability of UGT itself.

We could consider the granularity of UGT in two ways; the granularity of the goal parameters and the granularity of the scale (weights 0-5).

During the study, the participants seemed comfortable with the scale. Assigning a specific meaning to each point on the scale seems to have helped. Once they got used to it, the participants could easily differentiate between the adjacent points on the scale and could justify why one goal parameter would have a weight of 2 while another would have a weight of 3.

Granularity of the goal parameters is harder to evaluate. In some cases, a few of the goal parameters could be split up further and made more granular, e.g. goal parameter 29 *"user should feel emotionally engaged / brand / fun / reflective appeal / trust"*. On the other hand, since none of the participants felt the need for doing so for their projects though they had a choice, it is perhaps not necessary.

Our objective behind splitting high-level goals into goal parameters is to make goal setting less ambiguous and consistently interpretable. During the study, participants could interpret the goal parameters unambiguously. Occasionally, they needed to refer to the UGT documentation and go through examples for some of the goal parameters. Nevertheless, once they understood the meaning, they could unambiguously interpret the goal parameter in the context of their project, determine its importance, and think of ways to evaluate that goal parameter. We believe that we have achieved a right balance of granularity in UGT goal parameters. Of course, in case a particular project needs more granular goal parameters, UGT always allows for the flexibility.

When variables are used as components of a larger construct, their internal consistency reliability is important. **Internal consistency reliability** is the extent to which individual variables are measuring the same underlying construct of interest [19]. Cronbach's alpha evaluates the variation accounted for by the true score of the underlying construct. Cronbach's alpha above 0.7 is considered an acceptable measure of internal reliability. To test if any particular component varies differently than the other components, and therefore does not measure the same underlying construct, that component is dropped and alpha is calculated again with remaining components. If the resulting alpha increases substantially, then that component is not measuring the same underlying construct and it must be deleted.

The overall internal reliability of the 30 goal parameters in the 65 projects using Cronbach's alpha was 0.7870 for the weights, and 0.8733 for the scores. Both values indicate an acceptable level of internal reliability. Each of the 30 goal parameters was deleted by turns and Cronbach's alpha was re-calculated for the remaining 29 goal parameters. The resulting alphas did not vary much – from 0.7720 to 0.7969 for weights and from 0.8629 to 0.8780 for scores. The alpha resulting from dropping any particular goal parameter did not increase by much (maximum increase < 0.01). We can conclude that the 30 goal parameters are internally consistent and essential to measure the same construct.

## 5 Goal Achievement

We analysed extent to which the HCI practitioners achieve goals of their projects. In an ideal situation, design teams would like to score well against all goal parameters. In practice, if resources were limited, design teams would still like to score well on high-weighted goal parameters. This analysis was done to test the following

hypothesis: **Design teams naturally concentrate on achieving the high-priority goals; goal weights and goal scores are positively correlated.**

This hypothesis can be tested in two different ways. We can evaluate the correlation between the mean weights and mean scores of the 30 goal parameters reported in **Table 2**. However, mean weights and scores can disguise the divergence between individual weight-score pairs. Therefore, we can also evaluate the correlations between individual weight-score pairs (65 projects x 30 goal parameters = 1,950 pairs).

Pearson's correlation r is used to evaluate correlations between parametric variables, where $r^2$ indicates the extent of variation in the dependent variable because of the independent variable [20]. Pearson's correlation of mean weights and mean scores of the 30 goal parameters (**Table 2**, n = 30) was calculated. A significant positive and strong correlation resulted ($r = 0.957$, $n = 30$, $p < 0.0005$, $r^2 = 0.916$). We can conclude that the mean weights of goal parameters strongly correlate with mean scores of goal parameters, and mean weight of goal parameters predicts 91.6% variation in mean scores.

Next, a crosstabs summary was generated for the 1,950 weight-score pairs (**Table 3**). This shows how many goal parameters achieved each weight-score pair. For example, among the goal parameters that were set a weight of 5, 52 achieved a score of 100, while 97 achieved a score of 75. We can see that out of 1,950 goal parameters 757 (43%) scored 50 or less (undecided or worse). A matter of particular concern is that goal parameters with weights 3, 4 and 5 (the important goal parameters) scored 50 or less in 39%, 36% and 33% cases respectively (the gray cells in **Table 3**).

**Table 3:** SPSS crosstabs output for 30 goal parameter weights and scores for 65 projects. The highlighted cells (important, but low-scoring goal parameters) are a matter of concern.

| | | Goal score | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 0 | 25 | 50 | 75 | 100 | |
| Goal weight | 0 | 144 (75%) | 9 (5%) | 19 (10%) | 19 (10%) | 2 (1%) | 193 (100%) |
| | 1 | 21 (14%) | 35 (24%) | 44 (30%) | 37 (25%) | 10 (7%) | 147 (100%) |
| | 2 | 17 (5%) | 41 (11%) | 142 (39%) | 146 (40%) | 23 (6%) | 369 (100%) |
| | 3 | 10 (2%) | 37 (7%) | 157 (30%) | 254 (49%) | 60 (12%) | 518 (100%) |
| | 4 | 11 (2%) | 28 (6%) | 138 (28%) | 256 (51%) | 65 (13%) | 498 (100%) |
| | 5 | 1 (0%) | 18 (8%) | 57 (25%) | 97 (43%) | 52 (23%) | 225 (100%) |
| Total | | 204 (11%) | 168 (9%) | 557 (29%) | 809 (42%) | 212 (11%) | 1,950 (100%) |

Spearman's rho is used to evaluate correlations between ordinal variables and the square of the Spearman's rho indicates the extent of variation in the dependent variable because of the independent variable [20]. As the goal parameter weights and scores are ordinal variables, a Spearman's rho was calculated between the 1,950 pairs of weights and goal parameter scores reports. A statistically significant correlation emerged but the value of the rho was quite small ($rho = 0.391$, n = 1,950, $p < 0.0005$). A correlation value between 0.25 and 0.5 is considered a moderate association. The $rho^2 = 0.153$. This indicates that only 15.3% of the variation in the goal parameter scores is explained by the variation in the goal parameter weights.

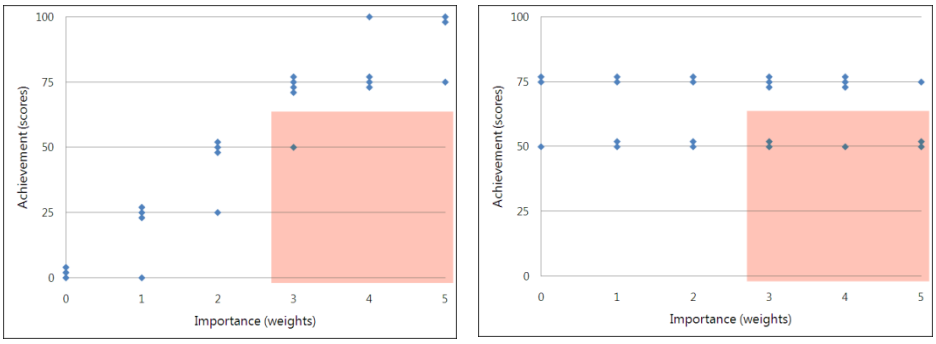
While mean weights and mean scores of goal parameters correlate very well, individual goal parameter weight-score pairs correlate only moderately, and as many as 36.8% of the important goals are not achieved. This could have happened possibly because the practitioners did not set usability goals systematically during the project, or because they did not track goal achievement of important goals explicitly at the end of the project. As a result, they may have not explicitly realised the importance of some of the goal parameters for their projects until they participated in our study.

A limitation of this study was that data was collected retrospectively, i.e. after the projects were completed, and this may have introduced some biases in the data. However, this may not be significantly affecting our main finding that a large proportion of important usability goals are not achieved. It is anecdotally observed that practitioners set more ambitious goals in the beginning of the project than at the end. If at all, it could be possible that in reality an even larger proportion of important usability goals are not achieved.

### 5.1 Latent Goals and Explicit Goals

If the practitioner knew that a goal parameter is important in the context of a project, he would strive to achieve it well. In our study, he would assign that goal parameter a high weight. If he were successful in achieving that goal parameter, he would also assign it a high score.

A high correlation between the weights and scores of a particular goal parameter across projects (**Fig. 2a**) indicates that largely, practitioners put in efforts to achieve high scores against that goal parameter in projects where it was important. Conversely, a low correlation (**Fig. 2b**) indicates that scores of that goal parameter are independent of its weights. This could happen if either the goal parameter is either particularly easy or difficult to achieve (and hence uniformly high or low scores), or if the practitioners were not mindful of that goal parameter during projects.




**(a)** **(b)**

**Fig. 2:** Importance / achievement scatter plot for (a) a highly correlated goal parameter, and (b) a non-correlated goal parameter (hypothetical data). Each dot represents data from one project. The pink area represents the projects that gave weight of 3 or more, but scored 50 or less.

If practitioners were equally mindful of all goal parameters during development, there would not be a large variation between weight-score correlations across goal parameters. On the other hand, if some goals tended to be *"latent"* (i.e. the practitioners were less mindful of them during development), these would tend to have lower weight-score correlations. This analysis was done to test the following hypothesis: **Design teams are equally mindful of all goal parameters; weight-score correlations do not vary much across goal parameters.**

Spearman's rhos were calculated to explore the relationship between weights and scores for individual goal parameter across the 65 projects. **Table 4** lists the rhos and significances for the 30 goal parameters, along with mean weights and mean scores. The statistically significant rhos ($p < 0.05$) are highlighted in grey. The table has been sorted in the descending order of mean weights assigned to each goal parameter.

**Table 4:** Spearman's rhos for weights and scores, rho significances of individual goal parameters (n = 65 projects each). The mean weights and mean scores are reproduced here from Table 2 for convenience.

| | Goal parameters (see Table 1 for longer names) | Spearman's rho for weight – score | Significance of rho $p$ <= | Mean weights (0-5) | Mean scores (0-100) |
|---|---|---|---|---|---|
| 1 | Findability | 0.17 | 0.186 | 4.00 | 71.92 |
| 23 | Clear communication | -0.19 | 0.138 | 3.89 | 71.54 |
| 22 | Info architecture | -0.12 | 0.335 | 3.78 | 71.92 |
| 24 | Feedback | -0.02 | 0.901 | 3.55 | 67.69 |
| 9 | Quick and easy navigation | 0.09 | 0.482 | 3.51 | 71.15 |
| 8 | Do primary tasks quickly | 0.34 | 0.006 | 3.49 | 71.15 |
| 16 | Conceptual model | 0.03 | 0.834 | 3.45 | 65.77 |
| 28 | Users feel in control | 0.17 | 0.181 | 3.23 | 66.15 |
| 12 | Complete tasks in time | 0.17 | 0.186 | 3.18 | 64.62 |
| 17 | Intuitiveness | 0.23 | 0.070 | 3.12 | 65.38 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | Less time to learn | 0.22 | 0.074 | 3.06 | 63.08 |
| 4 | Consistent: internally | 0.39 | 0.001 | 3.05 | 67.69 |
| 10 | Memory / cognitive load | 0.02 | 0.897 | 3.05 | 58.08 |
| 25 | Not induce errors | -0.04 | 0.725 | 3.03 | 60.00 |
| 27 | Help error recovery | 0.21 | 0.092 | 2.86 | 59.62 |
| 3 | Learn on their own | 0.54 | 0.0005 | 2.83 | 60.38 |
| 18 | No entry barrier | 0.28 | 0.021 | 2.80 | 56.54 |
| 19 | No unnecessary tasks | 0.07 | 0.558 | 2.78 | 63.46 |
| 30 | Aesthetic appeal | 0.33 | 0.007 | 2.78 | 61.15 |
| 5 | Consistent: other prods. | 0.38 | 0.002 | 2.75 | 55.38 |
| 21 | Always on | 0.55 | 0.0005 | 2.66 | 54.62 |
| 26 | Forgiving interface | 0.35 | 0.004 | 2.58 | 56.54 |
| 29 | Emotional engagement | 0.53 | 0.0005 | 2.54 | 50.00 |
| 20 | Minimise task load | 0.48 | 0.0005 | 2.51 | 51.15 |
| 11 | Flexibility / user control | 0.33 | 0.007 | 2.48 | 53.85 |
| 7 | Retain infrequent tasks | 0.24 | 0.059 | 2.31 | 51.15 |
| 13 | Auto-personalised | 0.71 | 0.0005 | 2.23 | 43.08 |
| 14 | Localised | 0.64 | 0.0005 | 2.14 | 44.62 |
| 15 | User can customise | 0.74 | 0.0005 | 1.55 | 30.38 |
| 6 | Consistent: earlier version | 0.77 | 0.0005 | 1.32 | 30.38 |

As reported (**Table 2**), the mean scores of goal parameters are between 30 and 72 and standard deviations of scores are greater than 16. This indicates that none of the goal parameters was particularly easy or particularly difficult to achieve. Yet, the weight-score rhos are not significant for half the goal parameters (**Table 4**) in spite of having a reasonable sample size of projects (n = 65). We can conclude that practitioners were not mindful of these goal parameters during projects, but gave them a higher weight during our study.

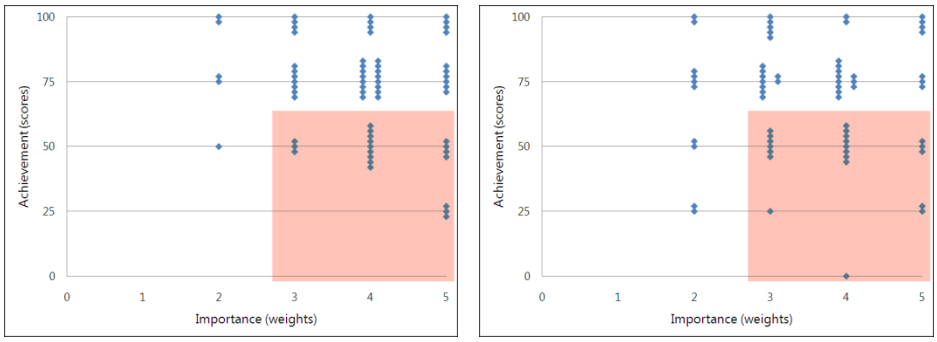
Interestingly, goal parameters with high mean weights tend to have less significant weight-score rhos (the upper half of **Table 4**). Conversely, the goal parameters with low mean weights tend to have higher and statistically significant weight-score rhos (the lower half of **Table 4**). The Pearson's correlation between the goal parameter weight means and the weight-score Spearman's rhos is strongly negative and statistically significant ($r = -0.815$, $n = 30$, $p < 0.0005$).

We can conclude that weights and scores are better correlated for low-weighted goal parameters, but not so well correlated on higher weighted goal parameters. Design teams achieve better scores when a typically low-weighted goal parameter gets an occasional higher weight, but do not achieve similar better scores when a typically high-weighted goal parameter gets an even higher weight.

A possible interpretation is that the typically high-weighted goals are **latent but important**. These goals are perhaps not explicitly asked for by the stakeholders and hence are perhaps not explicitly evaluated for during usability evaluations. The practitioners gave higher weights to these goal parameters when they saw them listed in UGT during the study, but they were not very mindful of them during the projects.

The following goal parameters have particularly low weight-score Spearman's correlations (rho < 0.15) and are potentially latent but important. Two of these have been plotted in **Fig. 3** as an illustration. These 8 goal parameters have a mean weight of 3.40.:

- 23 - Communication should be clear: rho = -0.19 ($p$ = 0.138) (**Fig. 3a**)
- 22 - Information architecture should be well aggregated, categorised, presented: rho = -0.12 ($p$ = 0.335)
- 25 - Product should not induce errors: rho = -0.04 ($p$ = 0.725)
- 24 - Product should give good feedback / display its current status: rho = -0.02 ($p$ = 0.901) (**Fig. 3b**)
- 10 - Product should not load user's memory / put cognitive load: rho = 0.02 ($p$ = 0.897)
- 16 - Interface should clearly communicate the conceptual model: rho = 0.03 ($p$ = 0.834)
- 19 - Product should require no unnecessary tasks: rho = 0.07 ($p$ = 0.558)
- 9 - Users should be able to navigate quickly and easily: rho = 0.09 ($p$ = 0.482)

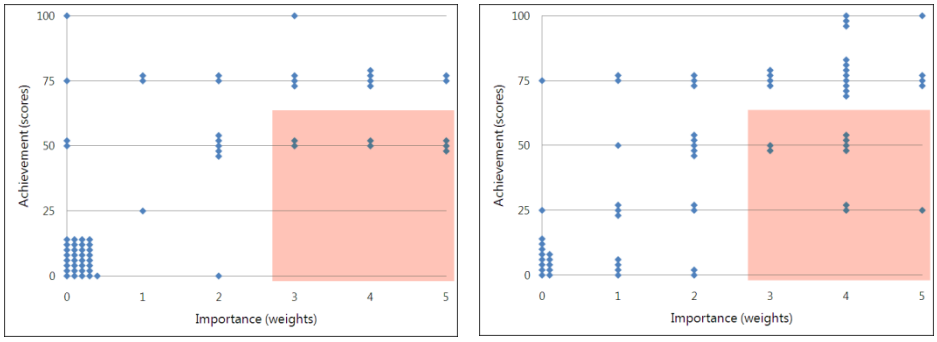



**(a)** **(b)**

**Fig. 3:** Importance / achievement scatter plot for (a) Goal parameter 23 "*communication should be clear*", mean weight = 3.89 (SD = 0.89), mean score = 71.54 (SD = 20.19), rho = -0.19 ($p$ = 0.138), and (b) Goal parameter 24 "*product should give good feedback*", mean weight = 3.55 (SD = 0.97), mean score = 67.69 (SD = 23.27), rho = -0.02 ($p$ = 0.901) (n = 65 in each case). Each dot represents data from one project. The pink area represents the projects that gave weight of 3 or more, but scored 50 or less.

A corollary to the above would be that low-weighted, high correlation goal parameters are **explicit**. These are more readily expressed by the stakeholders (such as clients or product managers) and are specifically addressed during usability evaluations. The following goal parameters have high and significant weight-score Spearman's correlations (rho > 0.50, $p$ < 0.05) and are potentially explicit. These 7 goal parameters have a mean weight of 2.21, which is significantly less than the weights of latent goals above (p < 0.0005). Two of these have been plotted in **Fig. 4** as an illustration:

- 6 - Product should be consistent with earlier version: rho = 0.77 ($p$ = 0.0005) (**Fig. 4a**)
- 15 - User should be able to customise the product for himself: rho = 0.74 ($p$ = 0.0005)
- 13 - Product should be personalised for the user automatically: rho = 0.71 ($p$ = 0.0005) (**Fig. 4b**)
- 14 - Product should be localised for specific market segment: rho = 0.64 ($p$ = 0.0005)
- 21 - Product should be always on, always accessible: rho = 0.55 ($p$ = 0.0005)
- 3 - Users should be able to learn on their own: rho = 0.54 ($p$ = 0.0005)
- 29 - User should feel emotionally engaged with product / brand / product should be fun / reflective appeal / trust: rho = 0.53 ($p$ = 0.0005)




**(a)** **(b)**

**Fig. 4:** Fig: Importance / achievement scatter plot for (a) Goal parameter 6 *"consistent with earlier version"*, mean weight = 1.32 (SD = 1.75), mean score = 30.38 (SD = 34.09), rho = 0.77 ($p$ = 0.0005), and (b) Goal parameter 13 *"product should be personalised for the user"*, mean weight = 2.23 (SD = 1.70), mean score = 43.08 (SD = 33.80), rho = 0.71 ($p$ = 0.0005), (n = 65 in each case). Each dot represents data from one project. The pink area represents the projects that gave weight of 3 or more, but scored 50 or less.

## 6. Conclusions and Future Work

We knew that there are differences in goal achievement in projects following waterfall and agile process models and between product and services companies [17]. We also knew that some HCI activities contribute to goal achievement more than others [18]. In this paper, we presented an empirical study that looks at goal achievement in 65 industry projects with the help of a tool for usability goal setting (UGT). In these projects, the goal parameter weights and scores correlated, though only moderately. HCI practitioners seem to be achieving important usability goals only moderately better than other usability goals. A particular matter of concern is that more than a third of the important usability goals scored undecided or worse. This clearly establishes the need to pay more attention to explicitly setting usability goals

early on in the project, tracking the achievement of these goals during the project, and evaluating against these goals at the end.

We also identified the differences in goal achievement patterns between usability goals. Projects seem to score well on a typically less important usability goal when it occasionally becomes more important, but do not achieve better scores when a typically important usability goal becomes even more important. We identified these usability goals that could be interpreted as *"important but latent"*. Practitioners gave a higher weight to these goals when they saw them during the study, but judging by the scores of these goal parameters, we can only conclude that they were not very mindful of these higher weights during the projects.

Further research is needed in goal setting, tracking, and achievement to determine the extent to which goal achievement varies based on variables such as organisational maturity, domain, platform, and experience of practitioners.

The study established the usefulness of a tool such as UGT. In an earlier survey, we had reported that practitioners thought that UGT was useful, it helped them understand the context of the project better, and made them think about goals that they had not considered earlier [10]. This study independently corroborates this result. It points to a need for a tool such as UGT that will help practitioners set and track goals during their projects systematically and provides guidance in prioritisation and evaluation of goals. As a long-term activity, organizations can use such a tool to keep a history of past goals, usability evaluations, practices, and costs, and use the repository as a knowledge base for identification and prioritization of usability goals in future projects. More studies will be required to determine if tools such as UGT could help improve usability goal achievement in practice. We also plan to explore the relationships between the user profiles, the product profiles, and the goal parameter weights with the aim of improving the recommendations for the goal parameter weights.

The study had some limitations that we acknowledge. The study comprised of projects from the Indian IT industry only. This could have introduced certain biases. A majority of the projects used the waterfall process model of software development, and a few used agile process models. No projects reported the use of Rational Unified Process models. This proportion may be different in other countries / contexts. The maturity of the usability, the mix of business models (product vs. service) and nature of domains addressed may also be different in other countries. Further, as discussed above, data was collected retrospectively. Thirdly, for most projects, only one representative participated in the study. This happened due to practical constraints. In reality, the data could vary somewhat if we get inputs from multiple stakeholders in each project as UGT is meant to take.

## Acknowledgements

We thank all the participants who provided data and helped us evaluate the UGT. We thank Dr. Sanjay Tripathi and Prof. SV Sabnis for their help and feedback in the statistical analysis. We thank Prof. UA Athavankar, Prof. Umesh Bellur, and Prof. S Sudarshan of IIT Bombay for their suggestions in this project.

## References


1. Nielsen, J.: Usability Engineering. Morgan Kaufmann, 26-37 (1993)
2. International Organization for Standardization: ISO 9241-1:1997 Ergonomic Requirements for Office Work with Visual Display Terminals. (1997)
3. Mayhew, D.: Usability Engineering Lifecycle. Morgan Kaufmann, 123-145 (1999)
4. International Organization for Standardization: ISO/IEC 9126-1:2001 Software Engineering - Product Quality. (2001)
5. Preece, J., Rogers, Y., Sharp, H.: Interaction Design, Beyond Human-Computer Interaction, 1st edition. John Wiley & Sons, 13-20 (2002)
6. Cooper, A., Reimann, R.: About Face 2.0. Wiley, 64-67 (2003)
7. Shneiderman, B.: Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition). Addison Wesley, 14 (2004)
8. Bevan, N.: Classifying and Selecting UX and Usability Measures. In : International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (2008)
9. Hornbæk, K., Frøkjær, E.: Making Use of Business Goals in Usability Evaluation: An Experiment with Novice Evaluators. In : CHI, Florence (2008)
10. Joshi, A.: Usability Goals Setting Tool. In : 4th Workshop on Software and Usability Engineering Cross-Pollination: Usability Evaluation of Advanced Interfaces, Uppsala (2009)
11. Joshi, A.: Usability Goals Setting Tool - Online Version. (Accessed 2009) Available at: http://www.idc.iitb.ac.in/~anirudha/ugt.htm
12. Eames, C.: Charles Eames Quotes. In: BrainyQuote.com. Available at: http://www.brainyquote.com/quotes/quotes/c/charleseam169187.html
13. Archer, B.: Systematic Method for Designers. Council of Industrial Design (1965)
14. Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J.: Understanding, Scoping and Defining User eXperience: A Survey Approach. In : CHI 2009 (2009)
15. McCarthy, J., Wright, P.: Technology as Experience. The MIT Press, 49-129 (2004)
16. Cross, N.: Engineering Design Methods 3rd ed. John Wiley & Sons, 61-76 (2000)
17. Joshi, A., Sarda, N., Tripathi, S.: Measuring Effectiveness of HCI Integration in Software Development Processes. Journal of Software Systems, (2010) DOI: 10.1016/j.jss.2010.03.078.
18. Joshi, A., Sarda, N.: Evaluating Relative Contributions of Various HCI Activities to Usability. In : Human-Centred Software Engineering, Reykjavik (2010)
19. Hatcher, L.: A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling. SAS Publishing, 131-140 (1994)
20. Kurtz, N.: Introduction to Social Statistics. McGraw Hill Book Company, 270-275 (1983)