

## ユーザビリティ評価実験のデザインに関する考察

菅原 千穂子<sup>†</sup>, 高田 眞吾<sup>†</sup>, 中小路 久美代<sup>‡</sup>, 松本 健一<sup>†</sup>, 鳥居 宏次<sup>†</sup>

<sup>†</sup>奈良先端科学技術大学院大学 情報科学研究科  
〒630-01 奈良県生駒市高山町 8916-5

<sup>‡</sup>(株) SRA ソフトウェア工学研究所  
〒160 東京都新宿区四ツ谷 3-12

{chihok-s, michigan, kumiyo, matumoto, torii}@is.aist-nara.ac.jp

あらまし 本稿では、ユーザビリティ評価実験のデザインを系統的、かつ、容易に行うことのできる具体的手法の確立を目指し、その準備段階として行った3つのケーススタディについて報告する。具体的には、まず、デザインを構成する5つの基本作業を定義する。次に、ケーススタディを観測、分析した結果を、事実や分析結果を、基本機能毎に分類、整理する。

キーワード ユーザビリティ評価実験, ケーススタディ, 実験デザインのガイドライン

## Toward Better Usability Testing Design

Chihoko Sugahara <sup>†</sup>, Shingo Takada <sup>†</sup>, Kumiyo Nakakoji <sup>‡</sup>,  
Ken-ichi Matumoto <sup>†</sup>, and Koji Torii <sup>†</sup>

<sup>†</sup>Graduate School of Information Science, Nara Institute of Science and Technology

<sup>‡</sup>Software Engineering Laboratory, Software Research Associates, Inc.

{chihok-s, michigan, kumiyo, matumoto, torii}@is.aist-nara.ac.jp

**Abstract** While the importance of usability testing has long been recognized, difficulty of usability testing design has been underestimated. Little has been done to identify guidance and know-how for designing a usability testing. This paper reports results of our analysis of three usability tests as a case study to identify such guidance and know-how.

**keywords** Usability Testing Design, Case Study, Design Guideline

### 1 はじめに

ユーザビリティ評価実験とは、ソフトウェアの機能がユーザに正しく理解され、かつ、ユーザにとって使いやすいものかどうかを、被験者(ユーザ)を使って評価する作業である。機能が正しく実装されているかどうかを調べる機能テスト等と共に、ソフトウェア品質を保証する重要な作業の一つと位置づけられている。

ユーザビリティ評価実験では、評価目的に合致した実験を「デザイン」することが重要である。ユーザビリティ評価実験は、被験者を使っ

た文字どおり「実験」であり、機能テスト等と比較すると、実施のコストや期間が多くかかり、やり直しも容易でない。デザインにおいて評価の目的、実験の実施要領、収集データの分析方法、更には、実験で使用、作成するドキュメントの形式をも定義しておかなければ、効果的な実験を行うことは困難である [5][3]。

本稿では、ユーザビリティ評価実験のデザインを系統的、かつ、容易に行うことのできる具体的手法の確立を目指し、その準備段階として行った3つのケーススタディについて報告する。

具体的には、デザインを構成する5つの基本作業を定義し、ケーススタディで観測された事実や分析結果を基本機能毎に分類、整理する。

## 2 ユーザビリティ評価実験

システムの使いやすさや使いにくさ（使用感）を評価する手法には、(1) Heuristic evaluation with usability guidelines, (2) Cognitive walkthroughs, (3) Usability Testing, (4) Usability Engineering, (5) Controlled Experiment の5つがある [1]。このうち、(1)(2)は、専門家の直感に基づく評価である。一方、(3)(4)(5)は、被験者を用いた実験を含む。但し、(4)ではメトリクスを用いた定量的な実験が、(5)ではより厳密に統制された実験が、それぞれ行われる。

本稿では、(3)(4)(5)に含まれる、被験者を用いた実験を総称して「ユーザビリティ評価実験」と呼ぶことにする。ユーザビリティ評価実験では、次の3点を明確にする必要がある。

(1) 評価コンテキスト：誰にとっても、また、どのような状況下でも絶対的に使いやすい、あるいは、使いにくいシステムは存在し得ない。誰にとっての、どのような状況（システムやタスク）下での使用感を実験で評価するのかを予め明確にしておく必要がある。

(2) 評価の基準：ソフトウェアシステムの使用感には、ユーザ（被験者）の主観に基づいて評価される場合が多い。主観的評価の基準を定義することは容易でないが、評価の観点、対象、収集データや分析結果等をガイドラインとして準備することは、評価の信頼性を高めることになる。

(3) 実験の観点：実験の観点には、「仮説検証」と「事実発見」の2つがある。仮説検証とは、専門家の直感や過去の実験結果に基づいて予め設定された仮説が成り立つかどうかを確認する作業である。一方、事実発見とは、実験で観測された具体的な事象や分析結果から、ユーザビリティに関わる新たな事実を抽出する作業である。

## 3 評価実験のデザイン

ユーザビリティ評価実験は「デザイン」、「実施」、「分析」、「評価」という4つのフェーズから成る。この中でもデザインフェーズは、実験の設定を決定する重要なフェーズである。効

果的な評価実験は、実験デザインに依存しているといっても過言ではない。

本稿では、ユーザビリティ評価実験のデザインフェーズを、さらに、(1) 目的設定、(2) 問題設定、(3) 実験設定、(4) 分析設定、(5) パイロットテスト、という5つの基本作業によって定義する（図1参照）。

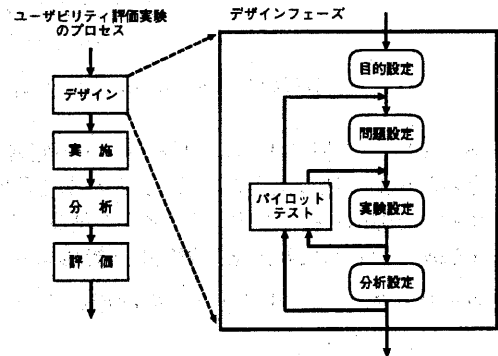


図1: 評価実験のデザインフェーズ

### (1) 目的設定 (図2参照)

ユーザビリティ評価実験をなぜ行いたいのか、実験結果をどう利用するのか、といった評価の目的を明文化する。目的の設定は、評価対象システムの特徴、及び、評価依頼者の意図に基づいて行う。目的設定を疎かにすると、評価依頼者の意図が評価者に伝わらず、実験や分析の方法変更や、やり直しを招くことになる。また、目的を出来る限り詳細に設定することで、他の基本作業が行いやすくなる。特に、分析の粒度の決定が容易になる。

### (2) 問題設定 (図3参照)

システムの特徴と評価目的に基づいて、評価実験で明らかにすべき問題（実験目的）を明文化する。問題は、評価対象と仮説から構成される。評価対象とは、システムの機能、ユーザ、及び、タスクであり、システムの特徴に基づいて決定する。仮説とは、評価者の想定するシステムの使いやすさ、ユーザがシステムを使ってタスクを実行する際に生じると予測される問題点、等である。仮説検証型実験の場合、仮説は出来るだけ詳細に設定し、実験で明らかにすべき点

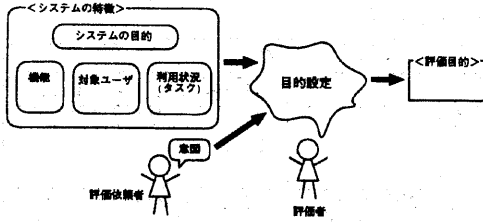


図 2: 目的設定

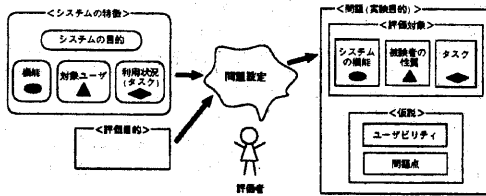


図 3: 問題設定

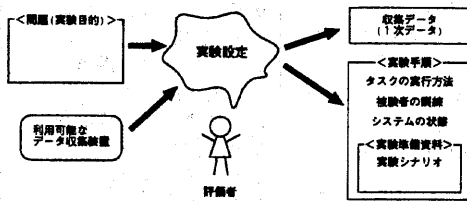


図 4: 実験設定

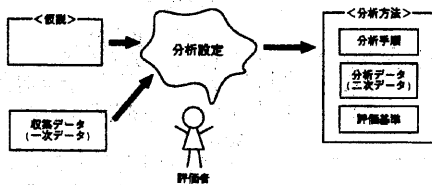


図 5: 分析設定

に絞り込むことが望ましい。

### (3) 実験設定 (図 4 参照)

問題設定作業で設定した問題 (実験目的) を明らかにするための実験の方法を明文化する。「実験のデザイン」を狭義に解釈した場合に対応する作業である。まず、実験手順の大きな流れを決定する。具体的には、タスクの実行方法、被験者への訓練、システムの状態、実験シナリオなどを決定する [5]。更に、収集すべきデータと利用可能なデータ収集装置 (ビデオ、履歴収集ツール等) に基づいて、実際に収集するデータ (一次データ) を決定する。

### (4) 分析設定 (図 5 参照)

収集データ (一次データ) に対する分析方法を明文化する。具体的には、実験結果を想定して、分析手順、分析データ (二次データ)、評価基準等を決定する。様々な分析方法が提案されているので、評価目的に合った方法を設定することが重要である。また、仮説検証型実験の場合、評価基準を出来るだけ詳細に決定しておく必要がある。

### (5) パイロットテスト

実験手順の確認やデザインの改良を目的とした実験であり、本実験の前に行われる。パイロットテストで実験実施上の不都合が発見された場合には、実験デザインの変更を検討する。特に、仮説検証型実験の場合、検証すべき仮説の再設定や分析手法の変更を検討する。

## 4 ケーススタディ

ユーザビリティ評価実験のデザインに関する具体的なノウハウを発見する目的で、3つのユーザビリティ評価実験において実験デザインがどのように行われたかを調査した。3つの実験はいずれも、奈良先端科学技術大学院大学の実習講義において行われたものである。評価依頼者、評価者 (実験毎に3名)、及び、被験者は、全て同大学の大学院生である。なお、評価対象となるシステムは任意に与えられたものであり、実験毎に異なる。また、評価者のユーザビリティ評価実験に関する知識、及び、評価の動機は、実験間で差はないものとする。

各実験におけるデザインの概要を表 1 にまとめる。また、各実験における評価対象システム、及び、評価目的は次の通りである。

### Case-1

評価対象システム：個人情報管理ソフトウェア

市販のソフトウェアであり、既に出荷され保守段階にある。対象ユーザは特に限定されていないが、ビジネスマンが多いと考えられる。

評価目的：

- (1) システムを初めて使うユーザが感じる使用感を調べる。
- (2) システムに関する情報 (ヘルプ、マニュアル、デモ、質問) の有無による、システムの使いこなし方の差を調べる。

表 1: 実験デザインのとまとめ

	Case-1	Case-2	Case-3
システムの機能	スケジュールの入力, 表示	全般	eMMaC(eMMa-Critics)
被験者の対象	ソフトを初めて利用する人, Windows 使用経験のある人	特に設定なし	タスクに必要な操作を行える人.
タスク	決められた3つの異なるスケジュールを入力する.	印刷された文書を読む(事前タスク). それを理解しやすい文書にする. 時間制限のみ 一人 10分	画面上から画像(イメージ)を切りとり色の評価を行う. ×5回 評価の終了 5分程度
<完了基準> <完了時間>	被験者の判断 最大 20分		
仮説 ユーザビリティ 問題点	GUI からの情報のみでタスクを達成できるか. 繰り返しスケジュールの入力に対してシステムの機能を使いこなせるか. 便利だと思われる機能に気付くことができるか.	より良いシステムとするために必要な要件, 及び, システムの問題点を明らかにする. ツールを利用することで, 作業の継承がツールがない時よりうまくいく.	被験者が展開する window に対してどのように対応するかを観察する. window が次々に展開するため, 結果として作業効率が低下する.
実験手順	1 実験の説明 2 事前インタビュー 3 情報なし: タスク 4 中間インタビュー 5 紹介アモ 6 情報有り: タスク 7 事後インタビュー	1 実験の説明 2 システムの操作の説明 3 事前タスク 4 改良方針のインタビュー 5 タスク 6 記入式アンケート × 3人	0 初めて使う人への訓練 1 実験の説明 2 アイカメラの調整 3 高解像度: タスク 4 アイカメラの再調整 5 低解像度: タスク 6 インタビュー
タスクの 実行方法	1 情報(マニュアル, ヘルプ, 質問)なし 2 情報有り think-aloud protocol	think-aloud protocol	1 高解像度 (1024×768pixel) 2 低解像度 (800×600pixel)
被験者の訓練	無	有(簡単な操作説明のみ)	有(30分弱の訓練)
実験準備資料	実験シナリオ, タスク説明書(配布)	実験シナリオ, 実験説明書(配布), 記入式アンケート	実験シナリオ, 実験説明書(配布), 被験者行動観察記録用紙
実験回数	2回+パイロットテスト	1回	2回+パイロットテスト

### Case-2

評価対象システム: 非同期の協調作業を前提としたドキュメント構築支援ツール

研究目的で開発中のツールである. システム設計が行われる前のプロトタイプで, ユーザによって使用されたのはこの実験が初めてであった. なお, 実験のデザインが行われた時点では, 動作確認のできる状態にはなかった.

評価目的:

- (1) 理解しにくいテキストを, 構造化された理解しやすい文書になるよう改良してもらい, テキストの改良の流れを観察, 分析する.
- (2) システムを用いて改良した文書を評価することによりシステムを評価する.

(3) 複数人による非同期作業を観察, 分析する.

### Case-3

評価対象システム: マルチメディアオーサリング支援システム (eMMa システム [2])

研究目的で開発されたシステムである. 完成度の高いプロトタイプで, 研究のための評価が既に何回も行われている.

評価目的:

ユーザインタフェースの問題点の検証し, その改善内容を明らかにする.

### 5 実験デザインに関する考察

4章で紹介したケーススタディを観察, 分析した結果を, 3章で定義した実験デザインの5つの基本作業に毎に考察する. なお, 分析には,

実験者に対するアンケートやインタビューへの回答、実験者による発表資料（デザイン終了時、実験終了時、分析終了時）、及び、最終レポートを利用した。

### (1) 目的設定に関して

評価依頼者と評価者が異なる場合、目標設定は特に重要である。Case-2では、システムの開発者が評価者に一人含まれていたが、他の評価者は、「任意に与えられたシステムのユーザビリティを評価する」という曖昧な目的しか持っていなかった。そのため、システムの特徴の理解に多くの時間を費やし、評価目的の再設定も必要となり、実験のデザインは困難であった。

### (2) 問題設定に関して

#### <システムの機能の設定>

システムの全機能を短時間で評価ことは難しく、評価対象とする機能を絞り込む必要がある。Case-1では、利用できるシステムの機能を実験時に制限することはなかったが、基本操作と思われる「スケジュールの入力」「表示機能」のみに着目した。Case-3では、「ウィンドウが作業の妨げになる」という仮説を検証する目的で、ウィンドウを表示する機能のみを評価対象とした。

#### <被験者の性質の設定>

実験に対する被験者の慣れ（think-aloud、緊張等）、及び、コンピュータに対する慣れと好みを実験結果に与える影響をできるだけ小さくする必要がある。被験者の性質が実験に及ぼす影響については、心理学等で研究がなされている[4]。今回の実験で見られた主な影響は次の通り。

(1) ウィンドウ操作に慣れた被験者であったため「ウィンドウが作業の妨げになる」という仮説は検証できなかった。(Case-3)

(2) マウス操作（ダブルクリックや右ボタン操作等）に対する慣れが、システムに対する操作に影響した。(Case-1)

(3) 入力方法（マウスとキーボード、メニューとGUI、利用OS等）に対する好み、システムに期待するユーザビリティに影響した。(Case-1)

#### <タスクの設定>

タスクの設定では、典型的タスクや問題点の生じそうなタスク等を設定し、完了基準、及び、

完了時間を決定する必要がある。タスクの達成時間を測定する場合は、タスクの開始と終了の目安となる操作、発話、あるいは、動作を決めておくといよい。

決定したタスクの教示（媒体と表現）には細心の注意を払う必要がある。媒体としては、紙、ディスプレイ、口頭等があるが、いずれも一長一短である。なお、被験者にタスクを口頭で説明する場合、被験者に与える情報を同一にするため、説明内容の明文化が必要である。

タスクの表現が適当でない、利用すべき機能を被験者が誤解する恐れがある。Case-1では、「スケジュールの入力」というタスク表現が、「プラン機能」ではなく「スケジュール機能」を使うことを示唆していると誤解され、被験者の行動に影響が見られた。

#### <仮説の設定>

仮説検証であるのか事実発見であるのか、実験の観点を明確にすることで、問題（実験目的）をより詳細化できる。今回の実験では、実験の観点が明確に区別されていなかった。また、ユーザビリティに関しても、評価者の視点と被験者の視点の区別が明確ではなかった。そのために、問題の詳細化がうまく行えなかった。

### (3) 実験設定に関して

#### <実験シナリオ>

実験シナリオは、実験を効果的、かつ、効率よく行う上で必要不可欠である。今回の実験で見られた実験シナリオの主な効果は次の通り。

- 実験の実施手順の予測、確認、把握
- 複数の評価者がいる場合の役割分担の徹底
- 実験中の手順ミスの記録
- 被験者に与える情報の同一化

#### <事前訓練>

被験者の性質の設定によっては、事前訓練の有無とその方法について考慮する必要がある。Case-1では、評価対象ソフトウェアを初めて利用する人を対象としていたため、事前訓練は必要なかった。Case-2では、ツールの簡単な説明のみを行った。Case-3では、システムの利用経験のない被験者にのみ、30分程度の事前訓練を行った。

訓練において、本実験で検証する予定の仮説

に関する情報を与えてしまわない様に注意する必要がある。Case-3の評価者は、評価終了後のアンケートの中で、評価実験がうまく機能しなかった主な原因の一つとして、この点を挙げていた。

#### <システムの状態>

システムの初期状態の明確な記録は、実験の再現性を確保するために必要である。また、データ収集装置を評価対象システム上で稼働させる場合には、システムの負荷を監視する必要がある。

#### <収集データ>

データ収集装置や評価対象システムの不調、あるいは、収集データのバックアップの不備によるデータの消失に備えて、ビデオカメラ等を用いたデータの補助記録を行うべきである。Case-1とCase-2では、被験者の操作画面や操作履歴の記録にミスがあったが、ビデオにより操作画面を撮影していたため、分析を行うことが出来た。

#### (4) 分析設定に関して

分析方法は、収集データとの関係を明確にし、出来るだけ詳細に決定しておく必要がある。今回の実験では、収集すべきデータとデータの分析方法の関係をあまり明確にしていなかった。そのため、分析フェーズにおいて仮説検証が行えず、収集データに合わせる形で、様々な分析方法を後から検討する結果となった。

#### (5) パイロットテストに関して

パイロットテストは、実験デザインの見直しと改良に役立つ。特に、評価実験の経験の少ない評価者にとっては、出来る限り行うべき作業である。今回の実験では、Case-1とCase-2でパイロットテストが行われた。特に、Case-1では、パイロットテストの結果に基づき、次の4つの修正が行われた。但し、分析設定の見直しまでは行われなかった。

- インタビュー内容の明文化  
(被験者の助言や実験の際の自由な質問より)
- 実験シナリオへの時間記入欄の作成  
(ビデオの解析の容易化のため)
- 被験者への実験、タスクの説明の明文化  
(被験者に与える情報の均一化のため)

- 実験時の評価者の役割分担の強化  
(実験をスムーズに行うため)

## 6 まとめ

本稿では、ユーザビリティ評価実験のデザインを構成する5つの基本作業を定義した上で、3件の評価実験をケーススタディとして観測、分析した結果について述べた。前提知識や時間の不足により、観測、分析した実験におけるデザインは満足なものではなかったが、デザインに関するノウハウをある程度得ることが出来た。また、実験デザインに関する次の4点の重要性を再認識することが出来た。

- 評価目的、問題(実験目的)、実験結果の利用法を明文化する。
- 仮説検証型の実験では、仮説を詳細化し、前もって評価基準を定める。
- 事実発見型の実験では、実験結果の利用法を明文化する。
- 評価者の視点からのユーザビリティとユーザの視点からのユーザビリティを区別する。

謝辞：実験を行った実習の参加者、および被験者として実験に協力して頂いた方々、さらに、実験ツールの作成、実験へのアドバイスを下さった方々に心から感謝します。

## 参考文献

- [1] T. M. Backer, J. Grudin, W. A. S. Buxton, S. Greenberg, Readings in Human-Computer Interaction: Toward the Year 2000, 2nd ed, Morgan Kaufmann Publishers, pp.73-91, 1995.
- [2] K. Nakakoji, A. Aoki, B.N. Reeves, "Knowledge-based cognitive support for multimedia information design," Information and Software Technology, 38, pp.77-84, 1996.
- [3] J. Nielsen, Usability Engineering, Academic Press, 1993.
- [4] 利島保生和秀敏(編), 心理学のための実験マニュアル: 入門から基礎・発展へ, 北大路書房, 1993.
- [5] J. Rubin, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, John Wiley & Sons, 1994.