使い勝手を定量的に捉えることによるUX評価の改善手法

- 製品利用現場の実態に沿った問題分析・課題認識を容易にして製品改善を牽引 -

Improving UX Evaluation by Quantitative Understanding of Usability

- Driving Product Improvement by Simplifying Identification and Analysis of Problems as They Occur in the Field -

富士通株式会社 共通ソフトウェア開発技術本部 ソフトウェア検証統括部
Software Quality Assurance Division, Software Technologies Unit, FUJITSU LIMITED

○宮下 直也	神野 昌和	高井 直人	野田 梨香レジナ	柳生 幾美
○Naoya Miyashita	Masakazu Jinno	Naoto Takai	Regina Lica Noda	Ikumi Yagyu

Abstract The conventional method of evaluating software usability is by creating and applying checklists based on past experiences. However, there are problems in evaluating accurately the way the product is actually being used, and providing satisfactory quantitative evidence of detected usability problems to designers and developers. This can result in products with poor usability and a drop in user satisfaction. In this paper, we seek to solve these problems by applying refined "Interaction Design Evaluation Method", which is a refinement of the established usability evaluation method Novice Expert ratio Method (NEM). We also present our findings on the validity of our method for the increasingly virtualized and cloud-based computing world.

1. はじめに

我々は第三者検証部門として主にミドルウェア製品の品質評価を実施している。当社のこれまでの製品の使い勝手評価では、過去の経験に基づいたチェックリストの利用を中心にしてきたが、以下の問題があり、お客様にとって使いやすい製品とならず、お客様満足度向上の阻害要因になっていた。

- ✓ 検出した使いにくさの問題を、製品設計・開発者に対して定性的な問題指摘では納得性を与えることができない。結果、問題改善の重要度を設計・開発者に適切に伝えられず、対処が見送られる
  - 例1：操作ガイドの表示位置が不適切でお客様が気づかず、操作ミスを誘発すると指摘。しかし、本当にお客様で発生する根拠が示せず。結果、対処が見送られ、出荷後に問題を誘発する可能性が残存
- ✓ ユーザーテストの実施が望ましいが、限られた開発費の中では全ての評価対象製品で実施することが難しく、チェックリストでの評価をベースとしてきた。しかし、お客様製品利用の実態が日々変動する中、新たな使い勝手の観点をカバーできず、使いにくさの問題を見逃す
  - 例2:仮想化によるサーバー集約で運用管理者一人当たりの監視システムが増大。製品が提供する画面で、運用現場に則した監視業務での効率的な可視化/情報特定手段の観点がカバーできず。運用管理者の手間が増えてしまうなどの問題を見逃す

我々は、前述の問題を解決することを目的として、ユーザビリティ評価手法であるNEM法（Novice Expert ratio Method）[1][2]を応用・拡張した「インタラクションデザイン評価手法」を開発・実践した。

一連の活動を共有すると共に「インタラクションデザイン評価手法」が昨今のトレンドである仮想化技術やクラウド、マウスやキーボードを用いないスマートデバイス等、新たな製品利用実態においても有効であることの可能性、および今後の研究の方向性についても報告する。

富士通株式会社 共通ソフトウェア開発技術本部 ソフトウェア検証統括部
Software Quality Assurance Division, Software Technologies Unit, FUJITSU LIMITED

静岡県静岡市駿河区南町18-1　サウスポット静岡 Tel:054-203-0234
18-1, Minami-cho, Suruga-ku, Shizuoka Southspot Shizuoka
e-mail:miyashita.naoya@jp.fujitsu.com

2. NEM 法の応用・拡張

今回、問題解決を図る足がかりとして、評価結果を定量的に示すことが可能なユーザビリティ評価手法である、NEM 法[1]の応用・拡張を検討した。

2.1 NEM 法から抽出した評価現場への適用への応用・拡張部分

NEM 法の応用・拡張の検討の中で抽出した以下の 3 点の対応により、使いにくさの問題点を定量的、かつ、発生契機をより明確にでき、製品開発部門に対して、修正改善を促すことが可能になると考えた。

(1) 被験者(製品操作者)選択の拡張

NEM 法では Novice を一般利用者、Expert を製品設計・開発者と定義。評価では Novice と Expert のそれぞれのグループ毎に操作時間を平均化し、比較することで評価している。

しかし、実際の製品利用現場では、利用者層として、「一般利用者」と「製品設計・開発者」の二種類だけでは不足要因がある。そこで、Novice を一般利用者にするのではなく、初心者とし、Expert を製品設計・開発者でなく熟練者と定義した。たとえば「該当製品の初心者であるか、熟練者であるか」、「該当業務の初心者であるか、熟練者であるか」といった要素を加えるなどして、それらの被験者間の比較をすることで、製品の問題点を抽出できると考え研究を進めた。

(2) 操作時間の差(NE 比)に加え、作業フェーズ毎の目標時間、限界時間の導入

NEM 法では問題個所を時間差(NE 比)で評価するが、それだけでは使いにくさの問題を見逃す可能性のある以下の点について、NEM 法の拡張が必要と考えた。

- NE 比の値が小さいが、そもそもどちらのグループでも操作時間がかかっているケース

  作業ストーリー（表 1）を NEM 法で評価した結果（図 1）、フェーズ 4 は NE 比≒1.0 である。NE 比≧3.0 を使いにくさ問題とすると数値データだけでは判断ができない。しかし、Novice、Expert 共に操作時間がかかることが、お客様の期待している操作時間でないケースがあり、設計の妥当性を検証する必要があった。

- Novice、Expert 毎のグループで平均を取ることによる特異値が見過ごされるケース

  表 2 のように、フェーズ 3 での Expert3 名の操作時間が計測された場合、Expert-C は Novice 平均を超える操作時間を要している。機能追加により Expert-C のみがこれまでの慣れていた操作感と異なってしまい操作に躓くケースである。図 1 のように平均化されてしまうと、このような重要な改善ポイントが見過ごされることになる。

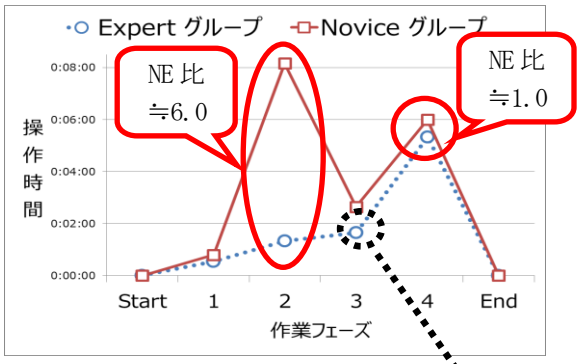

図 1. NEM 法による評価結果

表 1. 作業ストーリーと各フェーズ

【作業ストーリー】 あなたはシステム管理者です。 トラブル発生時のログを採取します。	
フェーズ	期待するユーザーの操作
1	アプリを探し起動する
2	ログ情報出力画面を探す
3	対象のログを選択する
4	問題発生時のログを抜粋する

表 2. フェーズ 3 に置ける Expert の各被験者操作時間

被験者	Expert-A	Expert-B	Expert-C	Novice 平均	Expert 平均
操作時間	49 秒	46 秒	3 分 22 秒	2 分 40 秒	1 分 39 秒

1 NEM 法(Novice Expert ratio Method 法)とは、製品設計・開発者(Expert)が一連の流れを操作した際の時間を計測し、同じ流れを一般利用者(Novice)が操作した際の時間との差を評価するものである。たとえば、設計者と初心者の差が大きいところには何らかの設計の問題があると判断する。

我々はこのような差が小さい箇所や、同一グループ内での操作時間の差についても、操作時間のデータから問題を見過ごさないため、あらかじめ操作時間に閾値を設定し問題個所を特定する施策も必要と考え研究を進めた。

(3) 使いにくさ問題の定量的な重度判定基準の定量化と自動判定

NEM 法では抽出した使いにくさの問題がお客様への使い勝手に対して、どれだけの影響を及ぼすのかの指標がなく「至急修正が必要」「製品利用現場を監視して対処検討」などの問題対処の重度判定は定性的で、設計・開発者の修正対応に差が生じてしまう。

我々は評価者誰でも一定の重度判定基準を用い、使いにくさの問題を定量的に判断できる施策の研究を進めた。

2.2 応用・拡張部分の取り組み詳細

NEM 法の試行から抽出した 3 点の応用・拡張部分の取り組みの詳細を述べる。

なお、本評価手法を「インタラクションデザイン評価手法」(以降、IxD 評価と記載) とし、現場実践に重点に PDCA を多く回せるようにし、研究を進めた。

(1) 評価すべきユーザーモデルの抽出の型決め

NEM 法での Novice(一般利用者)を「初心者」に、Expert (設計・開発者) を「熟練者」と置き換えた。さらに、該当製品、該当業務のそれぞれの側面で初心者と熟練者を定義している。

これらを表 3 のようにユーザーモデルマッピングとしてマトリクスで表現する事により、評価対象の製品特長、提供目的に合わせた充足性の高い被験者の抽出が可能と考えた。

例えば Java 言語の知識が前提となるアプリケーションサーバーの機能追加開発では②④⑥⑧の 4 つのユーザーモデルが選択され、4 名の被験者の用意で網羅的なモデル選択が可能になる。

表 3. ユーザーモデルマッピング

<table>
<tr><td colspan="2">熟練度
導入形態</td><td colspan="2">業務初心者</td><td colspan="2">業務熟練者</td></tr>
<tr><td colspan="2"></td><td>IT スキル初心者</td><td>IT スキル熟練者</td><td>IT スキル初心者</td><td>IT スキル熟練者</td></tr>
<tr><td rowspan="2">自社製品初心者</td><td>新規導入</td><td>①製品を利用して業務習得目指す初心者モデル</td><td>②製品を利用して業務習得目指す初心者モデル</td><td>③初めて製品を利用する業務熟練者モデル</td><td>④初めて製品を利用する業務熟練者モデル</td></tr>
<tr><td>他社製品からの移行</td><td colspan="2">該当モデルなし
(他社製品での業務実施者であるため)</td><td>⑤他製品の癖に慣れた利用者モデル</td><td>⑥他製品の癖に慣れた利用者モデル</td></tr>
<tr><td>自社製品熟練者</td><td>バージョンアップ</td><td colspan="2">該当モデルなし
(自社製品での業務実施者であるため)</td><td>⑦旧製品を業務活用している既存利用者モデル</td><td>⑧旧製品を業務活用している既存利用者モデル</td></tr>
<tr><td colspan="6">IPA 情報処理推進機構で公開されている「IT スキル標準」から利用者が理解しやすい文言としている。
IT スキル標準のレベル 1 を各初心者、レベル 3 以上を各熟練者という考え方で設定している。</td></tr>
</table>

また、一般的なユーザビリティ評価の場合、ニールセン博士が提唱している「各ユーザーモデル 5 人の被験者で 84%の使いにくさの問題を検出可能」[3]という定説に従うことが多い。製品提供先のユーザーモデルが多岐にわたるソフトウェアなどでは、モデル毎に 5 人の被験者、合計 40 人の被験者が必要となり、コストの面からも、適用は難しくなってしまう。そこで、NEM 法の「被験間のギャップから問題点を抽出する」という考え方を適用することで、少ない被験者数でもニールセン博士の「ユーザーモデル毎に 5 人の被験者」のレベルに到達できるかを検証することとした。

(2) 操作時間の目標・限界値、および操作時間重度の判定

- **各フェーズに目標時間、限界時間を設定**

熟練者の操作時間を基に目標時間、限界時間を設定する運用とした。（表 4）

表 4. 各操作時間の定義

製品熟練者時間	該当製品の設計・開発者や製品習熟度が高い評価担当者の操作時間
目標時間	ターゲットとしている全ての被験者に対して、達成させたい操作時間
限界時間	この時間を超えた場合には操作続行不可能と判断する操作時間

我々は便宜的に目標時間を製品熟練者時間の 2 倍、限界時間は 4 倍という指針を設定している。ただし、評価対象製品の特長などを考慮した設定も必要と考えている。（以下、事例）

- スマートデバイス対応製品
  メジャーなアプリの熟練者の操作時間を測定し 1.2 倍を目標時間として設定
- 仮想化技術によるサーバー集約を担う製品
  集約で操作対象が増大するが、集約前のサーバー１台の操作時間と同等の目標値を設定

- **操作時間重度の設定**

操作時間を表 5 の考え方で分割し、各被験者の操作時間がどこに属するのかを判断する。これにより操作時間の重み付けを可能とした。この重み付けを「操作時間重度」と呼ぶ。

表 5. 操作時間重度

	操作時間重度	説明
	3	【限界時間越え】 長時間作業が進まず時間経過し、作業が完了できず
限界時間		
	2	【目標時間～限界時間】 多くの時間を費やしたが、なんとか限界時間内に作業を終了
目標時間		
	1	【製品熟練者時間～目標時間】 操作に躓く部分もあるが目標時間内に作業を終了
製品熟練者時間		
	0	【製品熟練者時間以下】 操作に躓くことなく、製品熟練者時間内で作業を終了

(3) 使いやすさ問題の重度判定

NEM 法のように被験者グループ毎に平均化せず、全被験者の個々の操作時間データに対し、操作時間重度の割合を用い重度判定を行い、使いにくさの問題の特定を試みた。重度判定については 5 段階で結果を自動的判定するようにし、かつ、修正の必要性を設計・開発者に対して分かりやすいように定義した。（表 6）

表 6. 使いやすさの問題の重度判定

重度判定		修正の必要性	重度判定基準(例:4 人)
A	お客様自身での問題回避が難しく、業務続行が不可能に陥る可能性が高い	必須 (機能障害)	2.0 以上
B	お客様自身での問題回避に時間がかかり、業務続行に著しい影響を与える恐れがある	必須 (重大)	1.75 以上 2.0 未満
C	お客様自身で問題回避判断が可能であり業務続行が可能だが使いにくいと感じる	必須 (非重大)	1.5 以上 1.75 未満
D	改善を行うことで、より使い勝手向上に繋がる	要望	1.25 以上 1.5 未満
E	問題無し	―	上記以外
ランク A、B は出荷までの改善を必須としているが、ランク C、D については問題の内容を関係者間で協議し、製品出荷後の状況を監視しながら次版対応ということも可能としている。 重度判定基準の算出：各被験者の操作時間重度の総和÷被験者数(被験者数によって基準は変動)			

表 6 の重度判定方式を適用した出力結果は図 2 のようになる。このグラフでは、「限界時間」を超えている人数を見ると、フェーズ４は全ての被験者が超過しており、既存の製品利用者も含め、大きな不満を与えるものから Rank A と判定され、早急な改善が必要な使いにくさの問題となる。フェーズ 2 では初めての製品利用者に不満を与えるもので Rank B と判定している。

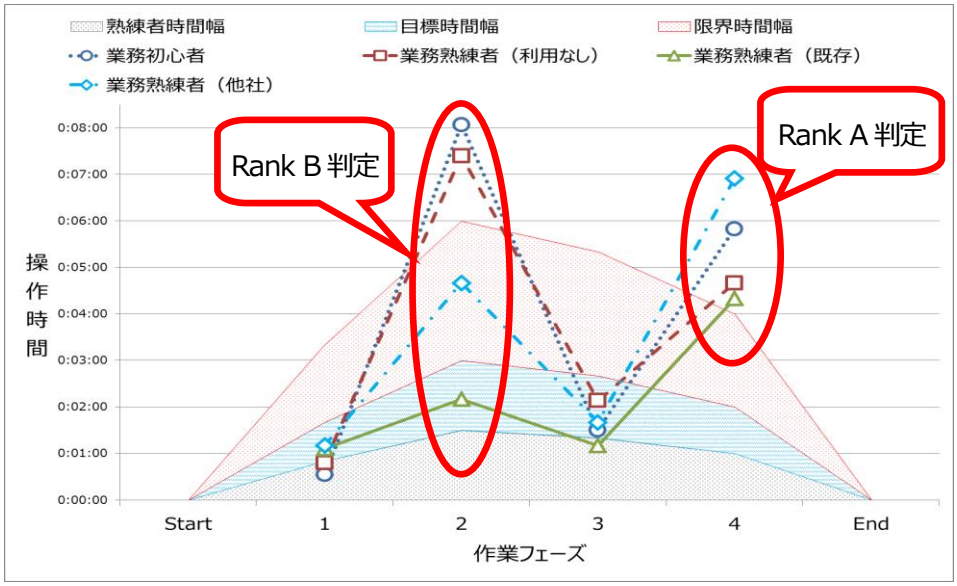

図 2. IxD 評価による判定結果（グラフ表示）

3. IxD 評価の効果

2011 年度から 2015 年度までに弊社のミドルウェア 60 製品に対して IxD 評価を実施した結果を述べる。主要な評価製品は以下に示す通り、様々な種別に対して実践した。

- ビジネスアプリケーション基盤ソフトウェア群（Interstage ファミリー）
- データベースソフトウェア群（Symfoware ファミリー）
- 統合運用管理ソフトウェア群（Systemwalker ファミリー）
- ストレージ基盤ソフトウェア群（ETERNUS ファミリー）
- 仮想化基盤ソフトウェア群（ServerView ファミリー）

評価時期は各製品の開発プロセス（主にウォータフォール型）の結合テスト、システムテストの工程(最終の第三者テスト工程以前)で実施。

3.1 使いにくさの問題の検出率、修正率の向上

「2.2 応用・拡張部分の取り組み詳細」で述べた施策展開により、NE 比だけでは見逃す可能性のある使いにくさの問題、およびこれまでのチェックリストの利用では検出する事が出来なかった問題[2]も検出可能となった。これにより使いにくさの問題の検出率が向上した。(図 3)

IxD評価適用前と1年目を比較すると検出率は3.7倍、修正率は 2.7 倍となり、大きな効果を測定した。

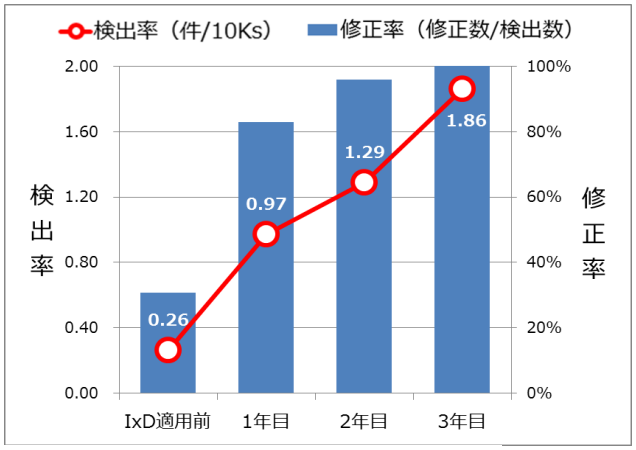

図 3. 検出率と修正率の推移

2 操作可能なボタンと認識できない、メニュー階層が深く対象に気づけない 等

また、「操作者の全員が躓いた」「目標時間を大きく超過した」などを操作時間という定量的なデータで示すことで、開発部門への説得性・納得性が増し、修正率が格段に向上した。(IxD 評価適用前：31% ⇒ 適用後：100% [Rank A/B の修正率])

なお、PDCA を繰り返すことで、作業ストーリーの検討方法の確立、被験者評価時の観察ポイントの整理、被験者評価時のファシリテーションノウハウの整理などにより、年々、検出率、修正率が向上した。結果、適用前と比較して 3 年目では検出率 7 倍、修正率 3.3 倍まで向上している。

3.2 使いやすさの問題検出数の被験者数の相関

これまでに検出した使いやすさの問題数と被験者数の関連を図 4 に示す。

結果、IxD 評価でのユーザーモデルマッピングで定義した複数モデルの被験者数の変化に伴う、問題検出率 [3] の推移は、一つのユーザーモデルに対するニールセン博士の結果を上回る曲線を描けた。

「2.2 応用・拡張部分の取り組み詳細」で述べた事例では 4 モデルにより、4×5 人の 20 人必要となるが、4 人の被験者で同等の検出率を維持でき、16 名分の評価コストが削減できた。

以上から、現在では費用対効果から、検出率が 70%を超える3名以上でのIxD評価を推奨している。

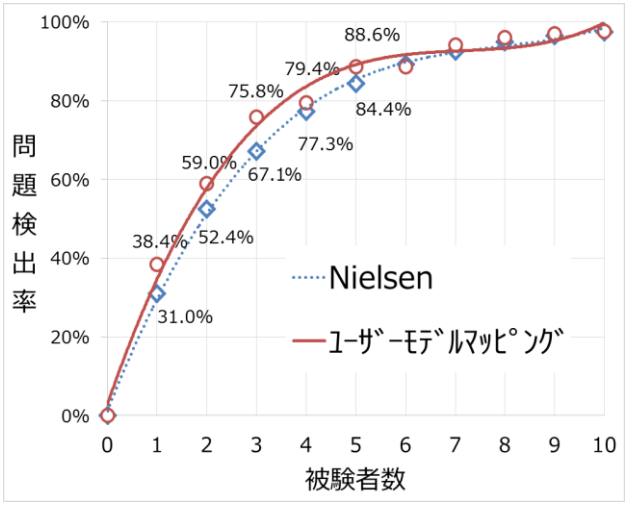

図 4. 問題検出数と被験者数の関係

3.3 評価工数の削減

改善 PDCA を廻したことによる評価工数の推移を図 5 に示す。初年度は研究開始時の想定より、準備、まとめ作業に時間を費やしたが、改善を繰り返したことにより、評価プロセスや、各種評価支援ツール等が充実し、評価工数の削減ができた。

また、本手法適用前に工数を要していた開発者との問題改善の調整時間も、定量的なデータで示すことで大幅に削減(87%)できた。

評価チームの構成人数についても、当初 6 名（ファシリテータ 1 名、タイムキーパー2 名、問題記録者 3 名）だったが、現在では 3 名構成で運用可能となっている。結果として、問題検出数の向上を計りながら、工数削減を実現することができた。

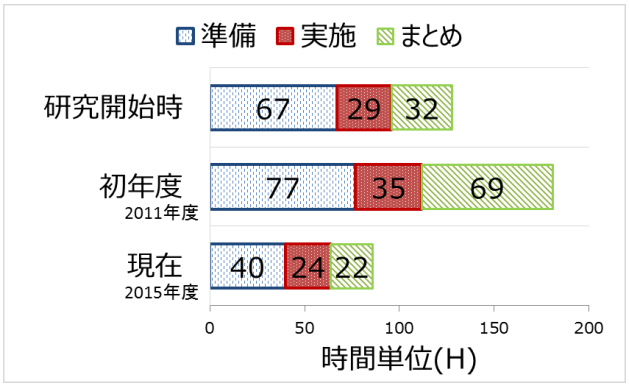

図 5. IxD 評価の工数推移

3.4 効果のまとめ

今回の IxD 評価の研究、実践を通して冒頭に述べた 2 つの問題点の改善を確認できた。

- ✓ 使いにくさ問題を定量的なデータで示すことで、開発元への説得性・納得性が増し、問題の修正率が 3.3 倍に向上した
- ✓ ユーザーモデルマッピンによる少ない被験者数でのユーザーテストが可能となったこと、また、改善を繰り返したことで評価工数の削減ができたことで様々な種別の製品への適用が可能となった。これにより使いにくさ問題の検出率も 7 倍に向上し、これまで検出できなかった問題も検出可能になった。

3 算出式：問題検出率＝IxD 評価検出数／設計～テスト＋お客様先で検出された使いにくさ問題の総数

4. 付帯効果

以降、今回の取り組みにより、得られた付帯効果を紹介する。

4.1 使い勝手の品質マネージメントを実現

IxD 評価の効果測定、および改善 PDCA を廻すことを目的に、使いやすさの問題の目標値を開発規模 10Kstep あたりの Rank A、B に該当する使いにくさの問題数を採用し設定した。

IxD 評価適用の 1 年目の実績値（0.97 件/10Ks）をベースに、2 年目以降、半期ごとに目標値を修正する運用とし、目標値に対して各製品の評価結果を監視する品質マネージメントが可能となった。

また、各製品の問題検出数に対して、目標差との原因分析を実施することにより、IxD 評価のプロセス・手段の改善部分の抽出や、その時の製品利用の実態に則した鮮度の高い評価観点[4]も、即時、反映できるようになった。

これは参考文献[4]に提言されている「利用時の品質の定量的データ監視」に準ずるものであり、その効果は高いということを実証できたものと判断している。

4.2 設計/開発者のモチベーションの向上

使いやすさの問題修正後に再度 IxD 評価を実施するプロセスとすることで、修正の改善効果を定量的に測定することが可能となった（図 6）。

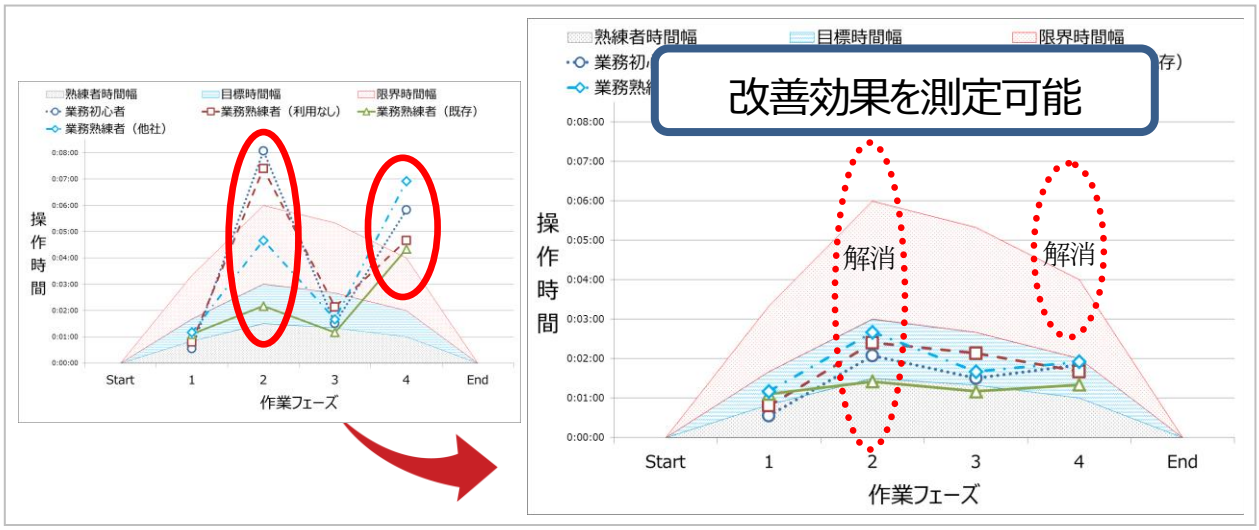

図 6. 問題の改善効果の可視化

また、参考文献[5]を参考にし、競合製品や過去バージョンとの比較評価にも利用することで、操作時間の優劣・バラツキの可視化も可能にし、優位部分はさらに伸ばし、劣位部分は改善優先度を上げる、等の判断もできるようにした。（図 7）

これらにより、これまでの使いにくさの問題抽出・修正では、「どの問題修正を優先すべきかわからない」、「お客様にとって本当に良くなったのかがわからない」という設計/開発者の悩みが解消され、より良い製品を開発するというモチベーションに繋がったというコメントも複数頂いた。

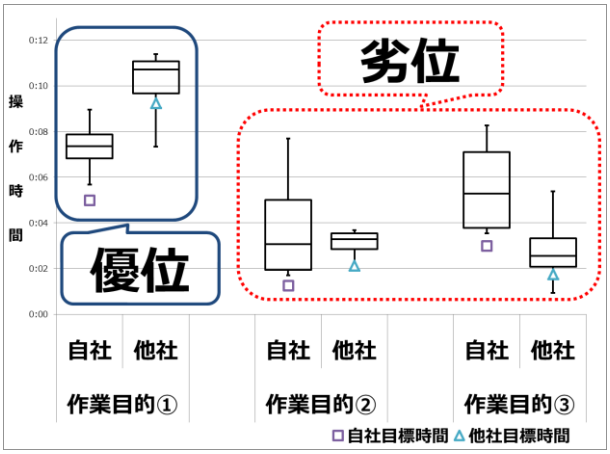

図 7. ベンチマーク評価結果

4 災害対策システム構築・仮想化サーバー集約・運用時、等で新たに抽出した使い勝手の評価観点

4.4　製品提供者の自己満足ではなく、お客様の生の声からも効果を実感

ここまで述べた効果は製品提供側での効果であるが、製品出荷後の効果もお客様の生の声から得られた。

- 本手法適用前の 2011 年度と 2015 年度比較で、お客様からの「製品が使いにくい」というコメント数が 35%減少
- 本手法を適用した製品利用者から「他社と比較して大変良くできていて、富士通の老舗ぶりが伺える。」とのコメントを頂いた。開発・評価部門は、直接的にお客様からお褒めのコメントを頂けることが少ない中、非常にうれしい効果を実感

5. 今後の展開

表 7 に示す通り、使いやすさの品質(ISO25010:利用時の品質モデル)における「有効性」「効率性」は、IxD 評価により定量的な評価が可能になった。

現在は、UX の重要な要因である「満足性」に繋がる「実用性」「信用性」「快感性」「快適性」等の定量的な評価手法の確立に挑戦している。

有効性・効率性についても更なる適用範囲、評価精度向上に向けて、キーボード操作/マウス操作を伴わないスマートデバイス/ウェアラブル等の評価への適用も研究、実践中である。これらについては、コンピューター上の操作の記録だけでなく、人の動きをどう記録するかということも含め、各種先端技術を取り入れ、融合し、定量的な使い勝手の評価を拡張していきたいと考えている。

表 7.　ISO25010:利用時の品質モデルにおける評価の適用範囲

利用時の品質 (ISO 25010)	有効性	ー	インタラクションデザイン評価手法で定量的評価を実現
	効率性	ー	
	満足性	実用性/信用性 快感性/快適性	現在、評価範囲拡大を検討・試行中
	リスク回避性	研究対象外（現時点）	
	利用状況網羅性		

参考文献

[1]鱗原 晴彦, 設計者と初心者ユーザーの操作時間比較によるユーザビリティ評価手法,
ヒューマンインターフェースシンポジウム 1999, 1999, 10.04

[2]鱗原 晴彦, 定量的ユーザビリティ評価手法：NEM による操作性の評価事例およびツール開発の報告,
ヒューマンインターフェースシンポジウム 2001, 2001. 10.02

[3]JAKOB NIELSEN、2000.3.19、「Why You Only Need to Test with 5 Users」,
https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/, 2010.03.21 参照

[4]神田周一, つながるシステムにおける利用時の品質向上にむけた品質要求事項定量化の提案,
13thWOCS2, 2016.01.21

[5]河野哲也, ユーザビリティ評価方法の実践的拡張および適用, JaSST'13 Tokyo, 2013.01.30