

情報システムの有効性評価
量的評価のガイドライン

(解説編)

第 1.1 版

2012 年 11 月 13 日

情報システムと社会環境研究会
情報システム有効性評価手法研究分科会

情報システム有効性評価手法研究分科会 メンバ (50 音順)

新目真紀 青山学院大学総合研究所
市川照久 静岡大学情報学部
片岡信弘 東海大学情報理工学部
金田重郎 同志社大学理工学部
鎌田真由美 日本アイ・ビー・エム
神沼靖子
栗山 敏 日本アイ・ビー・エム株式会社
児玉公信 情報システム総研 (幹事)
島本栄光 じぶん銀行
高橋尚子 國學院大学経済学部
刀川 眞 室蘭工業大学
中鉢欣秀 産業技術大学院大学
辻 秀一 東海大学
戸沢義夫 産業技術大学院大学 (幹事)
永井好和 山口大学大学情報機構
畑山満則 京都大学防災研
平田昌信 日経 BP
鷺崎早雄 静岡産業大学

目次

1. はじめに	1
1.1 情報システム研究の態度.....	1
1.2 ケース研究としての情報システム学.....	1
1.3 ガイドラインの目的.....	1
2. 有効性をどう見るか	1
2.1 違いを捉える.....	2
2.2 効果の指標.....	2
2.3 心理量を扱う場合の注意点.....	2
2.4 少数データの扱い.....	3
2.5 分析の方法.....	4
3. 効果を見るポイント	4
3.1 これまでの効果の指標.....	4
3.2 新しい効果の指標.....	5
3.3 局所最適と全体最適.....	8
3.4 投資対効果.....	8
4. 情報システム論文のまとめ方	9
4.1 新規性, 有効性, 信頼性.....	9
4.2 対象領域についての説明.....	9
4.3 研究設問.....	10
4.4 有効性の評価.....	10
4.5 参加者(被験者)の統制.....	10
4.6 予備実験.....	11
4.7 論文の基本的な構成例.....	11
4.8 エビデンスの提示.....	12
5. まとめ	12
参考文献	12
APPENDIX 有効性評価の例	1
A.1 企業情報システムにおける有効性の扱い.....	1

1. はじめに

1.1 情報システム研究の態度

情報システムは、企図する者（施主）が一定の意図を持って、組織または個人の活動に変化をもたらそうとする個別一回性をもった働きかけ（介入）である。その働きかけは、名目上、改善あるいは改革と呼ばれる。

組織にはたいてい、情報を流し、処理するための仕組み（IT を伴わないものも含めて）が存在し、それを使って組織の活動が成り立っている。そうした現状のシステムに対して、新たな情報システムを適用することは、介入そのものである。そこには当然、抵抗や非協力的な態度があるし、新しいものを受け入れるまでの学習や調整の時間やコストが必要である。介入によって現れる変化も、予測したところに表出するとは限らないダイナミック性とシステムミック性をもっている。

情報システムの研究者は、これを前提として、ある特定の組織のコンテキストにおいて、介入が、どのような状況で、誰にとって、どのようによい変化が得られたかを観察し、その法則性を議論する。

1.2 ケース研究としての情報システム学

このように、個別一回性の強い情報システムを研究対象とするためには、さまざまな介入の事例を収集し、それらを総合して、一定の法則性を見だし、仮説を立て、検証するというケース研究を、研究者のつながりの中で、ひたすら繰り返して積み重ねていくことになる。

1.3 ガイドラインの目的

介入事例を報告し、共有するためには、ケースとして必要な情報が含まれていなければならない。介入の有効性について正当に判定できないし、事例をまたがる法則性の判断を誤ることになる。

本ガイドラインは、介入としての個々の情報システムの有効性を評価する際の基本的な枠組みを提供し、その事例報告の形式を定めて、ケースを適正に共有することを目的とする。

なお、情報システムの有効性を示す方法として、量的な評価、質的な評価、そして量的評価と質的な評価の統合的な評価のアプローチがあるが、本ガイドラインは、このうち量的評価に限定して枠組みを示す。効果の量的評価であって、その意図の良し悪しを評価するものではない。

2. 有効性をどう見るか

介入の有効性は、それによって組織活動にどのような変化（違い）がもたらされたかをとらえ、それを評価することによって得られる。変化をとらえないで、絶対的な有効性を唱えるには、形式的に証明する必要がある。それは不可能であろう。

実際（actual）の組織活動の変化は、情報システム以外の介入も同時に発生するだけでなく、それらと情報システムとの交互作用も起こりうる。実際の環境で、それらを隔離することはできないし、変化の観測値から取り除くことはできないので、むしろ他要因の存在を認めて、それを漏らさず記述することを奨励する。

2.1 違いを捉える

組織活動の違いを捉えるには、介入の前後の状況を比較（前後比較）するか、複数の疑似的環境を作って、それぞれにおける活動の違いを捉える（クラス比較）のが一般的である。

前後比較の場合、事前状態を前もって測っておくことになるが、後戻りできないし、変化が予測していないところに表れることもある。組織や活動に対して、広く網を張っておく必要がある。

2.2 効果の指標

違いの指標として、量的評価においては、物理量、精神物理量、心理量がある。

(1) 物理量

処理時間、処理速度、処理データ量、処理件数、待ち時間、エラー数、消費資源量（メモリ、工数）、生産資源量、人数

計測手法を明記すること。

複数の指標を組み合わせて導出される指標として、生産性、エラー率などもあるが、これらのベースデータを必ず表示すること。

(2) 精神物理量

疲労度（フリッカー値など）、視力、色覚、聴覚、視点移動、脈拍、血圧、呼吸数、発汗、事象関連電位、脳波、表情計測など、計量可能なヒトの感覚である。

通常、官能検査で得られる。

(3) 心理量

満足度、好悪、態度（やる気、高揚感など）

量的評価法¹として、質問紙、調査票（アンケート）、SD（semantic differential）法などによる評定値を求める方法がある。精神物理量および心理量の計測では、馴化²の影響や個人差があるので工夫が必要である。

2.3 心理量を扱う場合の注意点

情報システムの評価に関わる心理量は、調査票などを用いて得られる調査対象者の“感覚”の評定（rating）である。

調査票は手軽な方法に見えて、実は問題が多い。質問の相手、用語の使い方（wording）、答え（評定値）の設定（5件法など）、嘘回答への対応など、考慮すべき点は多い。何度か繰り返して、安定した質問紙を用意しておくことが必要である。

(1) 調査票の問題

問題の一つ目は調査対象者選択の適切性に関する問題である。調査対象者が評価する情報システムをどの程度知っており、どのように関わっているのかで評価は変わるとの報告もある。それによって調査票の内容も変わってくる。

二つ目の問題は、順序尺度であることである。“感覚”は、「とても〇〇だ」、「やや〇〇だ」、「どちらでもない」といった1軸上の点との対応で表現される（3件法、5件法、7件法と呼ばれる）。これで得られる結果は名義尺度または順序尺度である。これらに対しては通常の統計処理ができない。そのため、第一に結果の度数分布を示すこと。データの特性によっては、数量化の手法を

¹ 質的評価法として、アクションリサーチ、GTA、エスノメソドロジなどの方法があるが、これについては別途述べる。

² 検査に慣れて、差を感じなくなること。

用いることができるのであり，安易に統計処理をしない。

三つ目は対象が人であることである。感覚には，常に個人差と馴化がつきまとう。同じ質問に対しても人によって理解の幅や深さが異なる。質問者の意図を汲んで迎合した回答を選ぶ，あるいは反発して嫌がる回答をわざと選ぶこともある。その状況に慣れてくると感覚が鈍くなってしまふ。

個人差に対しては統計的な誤差として扱うようにする。迎合や反発に対しては，質問文を無機質なものに変える，意図を推し量られないように，関係のない質問を加えるなどの対策をとる。慣れに対しては同じ参加者を何度も使わない，あるいは測定間隔を十分に確保する。

日本人の極端な評定を嫌う傾向にも考慮する。5 件法の評価が中央に寄ってしまい，差が抽出しにくいことがよくある。7 件法は，両端が選ばれないことをむしろ前提として作る（まれに選ばれたときは，一つ内側の評定に含める）。4 件法は，「どちらでもない」を回答させないための方法でもある。また，選択肢だけでなく，自由記述欄を設けておくと，数値には表れない質的データも得られ，質問の誤解が判明することもある。調査票は，予備実験を繰り返して，質問の仕方，ワーディング，回答の与え方を改善する。安定した調査票と評価指標を得るには相当の時間と労力が必要となる。

(2) 無機質な質問の例

提供した情報システムによる効果を直接聞くのではなく，前後比較として，かつてどのような状態にあったか，介入後，どのような状態にあるかを問う。質問も，一定の回答を誘導するような表現は避け，できるだけ無機質な問い方をする。

評価点を回答してもらうだけではなく，そう評価した理由も記述してもらうと良い。この記述内容によっては，質問に対する誤解が明らかになり誤りデータとして除外するかどうかを決める。また，質的データとしても利用できる。

図 1 に調査票の例を示す。Q1 は無機質な質問の例であり，Q2 は前後比較を行わない前提の上に，回答を誘導するような悪い質問の例である。

Q1. 日次の予定在庫数が実在庫数と一致する

よく当てはまる	当てはまる	ときどき当てはまる	あまり当てはまらない	当てはまらない
---------	-------	-----------	------------	---------

なぜそう思いますか(具体的に述べてください)。

Q2. 日次の予定在庫数が実在庫数とよく合うようになった ←×

図 1 調査票の質問の例

2.4 少数データの扱い

少数のデータ（サンプル）を取り扱う場合には，特別な注意が必要である。

まず，想定する母集団が狭められることである。これはどうしようもないことなので，標本の

特性および標本を選択した条件を明らかにすることをもって、得られた結論を外延しないようにする。次に、得られた標本が正規分布しない可能性である。変数変換などによって、正規分布に近づけられるかどうか検討し、それが難しければ、基本的な統計処理は諦めること。

データ数が少ない場合、1, 2件の異常値が統計処理をゆがめることがある。この場合、棄却検定法などを用いることで、“良心的に”異常値を捨てることができる。ただし、その旨およびその理由を明示すること。

2.5 分析の方法

情報システムの有効性評価でよく使われる統計手法を表 1, 2 に示す³。ここにはない統計手法を用いるときは、研究設問に対してその手法を用いることの妥当性についての説明と、当該手法の概要説明および文献の提示が必要となる。

表 1 比較でよく使われる統計的手法

目的	統計手法	前提
度数の差の検定	直接確率計算	
度数の適合性の検定	カイ二乗検定	
度数の独立性の検定	カイ二乗検定	
平均値の差の検定	t 検定	等分散, 正規分布
	ウェルチの t 検定	不等分散, 正規分布
等分散の検定	分散分析	
順位の設定	一対比較法(サーストン)	
	一対比較法(シエッフエ)	
階層的順位の設定	AHP(階層分析法)	一対比較で得られた順序尺度を間隔尺度と見なす

表 2 分析でよく使われる統計的手法

目的	多変量解析手法	尺度
相関の強さを求める	スピアマンの順位相関	順序尺度
	相関分析	間隔尺度
因果関係の推定	共分散構造分析	
変数値の線形推定	回帰分析	間隔尺度
非間隔尺度の回帰分析	数量化Ⅰ類	非間隔尺度
変量の構成要素の抽出	因子分析	間隔尺度
非間隔尺度の因子分析	数量化Ⅲ類	非間隔尺度
類似性の階層化	クラスタ分析	

3. 効果を見るポイント

一口に有効性と言っても、それが意味するものは幅広く、解釈する人によって異なり、かつ曖昧である。それも“情報システムの有効性”となると、さらにわかりにくい。ここでは、有効であるとは「効果があること」であり、システム理論で言うシステムに対して何らかの意図を持って働きかけを行った結果、意図に適合と認められた状況が得られたことを指す。

3.1 これまでの効果の指標

これまでに考えられてきた効果の指標とされるものについて、いったん整理してみる。

³ 学生向けの統計学のテキストとして次を推薦する。向後千春, 富永敦子: 統計学が分かる, 技術評論社, 2007, 向後千春, 富永敦子: 統計学が分かる (回帰分析・因子分析編), 技術評論社, 2009.

(1) QCDSE

非公式ながらも、製品がもつ機能の価値（広義の「品質」）を表すと考えられてきたキーワードは、“QCD”（狭義の「quality, 品質」, 「cost, コスト」, 「delivery, 納期」）である。狭義の「品質」とは、製品に欠陥がないこと、いわゆる「当たり前品質」を指す。最近では、これに“SE”（「safety, 安全性」, 「environmental consideration, 環境への配慮」）を加えることもある。有効性の文脈では、情報システムの利用者がそれを使用することを通して、顧客に対して“QCD(SE)”という切り口で価値を提供すると捉えることができる。

(2) ソフトウェアの品質特性

ISO/IEC 9126-1:2001 のソフトウェアの品質特性を効果の指標とすることも可能である。この国際標準では、製品の品質特性を 6 つの特性（「functionality, 機能性」, 「reliability, 信頼性」, 「usability, 使用性」, 「efficiency, 効率性」, 「maintainability, 保守性」, 「portability, 移植性」）に分類する。

また、ISO/IEC TR 9126-4:2004 はソフトウェア利用時の品質特性を規定している。この技術報告では、4 つの特性（「effectiveness, 有効性」, 「productivity, 生産性」, 「safety, 安全性」, 「satisfaction, 満足性」）に分けてとらえることとしている。

ソフトウェアの品質特性に関する国際標準は、上に述べた ISO/IEC 9126 のほかに、ISO/IEC 14598 の品質メトリックスの規定があり、内容が錯綜してきたため、これらをリファクタリングして置き換える形で、ISO/IEC 25000:2005 シリーズが策定されている。

(3) EEE

情報システムの効果を直接的にとらえるということでは、これも非公式なキーワードながら、“EEE”（「efficacy, 適正性」, 「effectiveness, 有効性」, 「efficiency, 効率性」）が挙げられる。“QCD”が製品（もの）の品質であったのに対して、“EEE”はサービス（こと）の品質とみることができ。“EEE”を、たとえば「リードタイムの短縮」, 「誤りの防止」といった具体的な指標に展開して計測できれば、有効性の指標になる。

(4) 顧客満足度

顧客に提供する価値を総称して「顧客満足度」と呼ぶことがある。ものやサービスから受け取る価値を、顧客の側からとらえようとしている点が強調される。ただし、会社対会社の取引にしても、顧客は多様であり、顧客によって異なる価値観に基づく心理量を一律に測ることは難しい。アンケートが現実的な計測法であり、インタビューは質的評価に用いられる。

3.2 新しい効果の指標

情報システムの効果に関する従来の指標だけでは、不十分と感じられるようになった理由は、情報システムがこれまでのような企業活動の道具としてではなく、企業活動そのものに同化してしまったためであろう。それは、情報システムへの投資判断が、かつては、投資額が、想定される効果でペイアウトできる期間を基準としていたものが、現在はその基準が意味を持たなくなったこととも関連している。

このことは、次のようにも言い換えられる。すなわち、かつて「想定される効果」とは自社内のコスト削減が主であった。現在では、自社の事業継続のためであったり、マーケットでの優位性を維持するためであったり、顧客の利益を確保することで自社の利益につなげようとするためであったりと、効果をこれまでよりも広いシステムの中で、多角的、重層的にとらえるようにな

ってきた。

(1) 効果の分類

このような観点から、情報システムが提供する機能の効果の指標となり得るものを列挙し、整理したものが図2である。

図2の横軸は、その効果が直接的であるか間接的であるかを分類している。直接的効果とは、利用者がそれを直接享受できることを意味し、たとえば「性能向上」は明らかにそれと分かる。間接的効果は、たとえば情報システムの「拡張容易性」のように利用者がそれと分かることはないが、システムの変更コストを総合的に減少することにつながっている。直接効果は比較的容易に定量化でき、間接効果は質的に扱われることが多い。

図2の縦軸は、その効果が施主から見て組織の外部に現れるか、内部にとどまるかで分類している。たとえば、「ビジネスコミットメント」とは、顧客のビジネスプロセスに自社の製品やサービスの提供が組み込まれる状況を指すが、これは明らかに外部効果である。一方、「意識づけ」は社員の業務に対する理解や、改善意識を指す。これは明らかに内部効果である。外部効果は観察によって計測できるが、内部効果の計測のためには、それが効果的であることのメカニズムの分析が前提となる。

列挙された効果の候補を、このような直接-間接、外部-内部の2軸で形成された4つの象限に配置した。各象限に命名すると、第1象限から順に、「収益」の獲得に関わる効果、「機会」の獲得に関わる効果、「能力」の獲得に関わる効果、「性能」の獲得に関わる効果となる。なお、第3象限にはみ出して「組織」と命名した効果は、個別の情報システムの効果というよりも、企業全体に及ぼす効果であり、これを識別しておいた。企業全体の効果と個別業務における効果については、後に議論する。

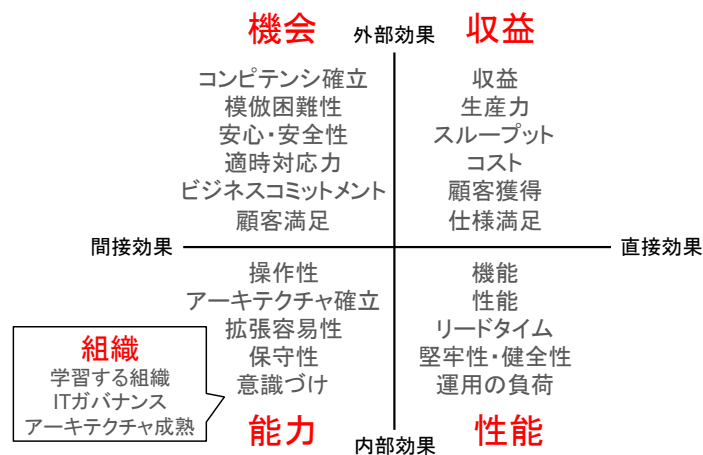


図2 効果の分類例

(2) 目的と効果の関係

情報システムは、意図した目的を達成するために開発され、日々運用される。その目的が、実際に達成されたかどうかを確かめるために、目的に対応して定められた指標値を常に観測（モニタ）し、その目標値（objectives）と比較する。この機能は情報システムに内蔵され、目標値と観測値の乖離が大きい場合は、その旨施主に警告するとよい。これには顧客満足度も含まれるべきである。

常に観測する理由は、その効果の価値が時間とともに変化するためである。多くは陳腐化の方

向にすすみ、情報システムの寿命の終わりを予見する。分科会の議論では、情報システムの効果とは、システムが生まれてから廃棄されるまでの間に、総合的にどれほどの効果を残したか（積分値）で測るべきであるとする意見もある。

目的の達成に貢献しない機能は、他の観点からは効果的であるとしても、有効であるとは言えない。逆に、当初想定していた以外の効果が目的の達成に貢献していると分かった場合、それは副次的効果として算入される。

(3) 誰にとっての効果なのか

効果が価値であるとする、価値を受ける人の価値観がその評価には反映する。たとえば、分科会のあるメンバが生産管理システムの構築プロジェクトで施主と要求定義をしていたときのことである。そこで、ある ERP 製品が使いにくいという話が出た。このような悪評は、その製品を導入しているほとんどの製造業で聞くことなので、だったらやめればいいのにと言いかけたとき、その経営トップが明確に言った。「これは自分のためのツールなのだ。君たちにはその価値は分からない」と。この一言はとても印象に残ったという。

(4) 管理階層と業務相

逆に、情報システムが大きな一枚岩として、経営から末端の担当者のすべてに渡って等しく有効な効果を提供すべきとするのは誤りである。使用者（アクタ）の目的や特性に応じてサブシステムを設定し、それぞれに必要な機能を提供すべきである。

図 3 は分科会のあるメンバが採用している生産管理システムの層別化アーキテクチャの例である。これは、生産管理の国際標準 IEC 62264-1²²⁾に準拠している。階層を管理レベルに基づき、事業、案件、執行、制御に分離する。層内の機能コンポーネントは密結合し、層間では疎結合する。

レベルごとに管理の主体（施主の候補）が異なり、関心事、すなわち管理対象と時間粒度が異なる。この例で管理対象は、利益、注文、製品、ワーク（加工対象物）とし、各サブシステムはそれぞれのライフサイクルを記録し、追跡する。

各サブシステムを大きく 3 つの業務相に区分し、機能コンポーネントを形成する。すなわち、「知識」「計画・実績」「報告」である。たとえば、案件レベルの「計画・実績」機能が注文を受けたとする。「計画・実績」機能は、「知識」機能に注文内容を渡し、製品完成までの作業計画を受け取り、製造指示として該当する工程の執行レベルの「計画・実績」機能に送信する。製造が終了すると、案件レベルの「計画・実績」機能は、その実施報告を受け取って、計画と実績の差異をみて、必要ならばアラームする。無事に製品ができれば、「注文」は完了し、そのライフサイクルを終える。「報告」機能は、日単位や月単位の報告サイクルで実績データを取り出し、日計表や財務諸表の形式で活動をモニタする。

階層ごとに管理主体が異なり、管理目的、管理対象、管理手法が異なることで、価値観が異なり、効果の評価基準も異なることになる。

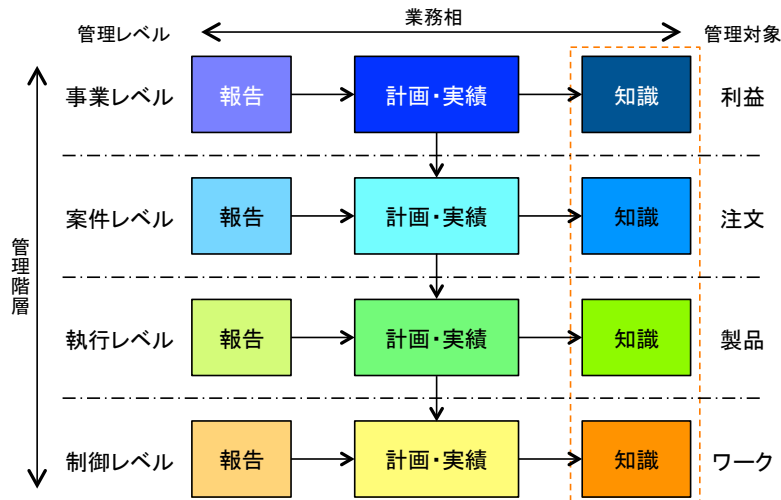


図3 管理階層と機能階層

3.3 局所最適と全体最適

管理階層とは別に、情報システムの構築と運用において、もう一つ、大きな価値観の違いがある。それは、上でも触れた、企業全体としての価値観と個別業務ごとの価値観である。それぞれを全体最適、局所最適と呼ぶこともできる。これらは必ずしも整合しない。

企業情報システムでは、施主は、手っ取り早く自分にとって最適な情報システムを形成しようとする。他のシステムの最適性まで考慮する必要はない。しかし、これが積み重なると、全体の情報システムの観点では、情報資源が重複し、散在し、データの品質が劣化し、技術要素が干渉し合うようになる。そこで、企業情報システムの継続的な全体最適化を図るために、Enterprise Architecture (EA)²³⁾と呼ぶガバナンスの仕組みが提唱されている。業務システムごとの局所最適化を進めるのは施主であり、全体最適の見地からガバナンスを発揮するのは、CIO (Chief Information Officer) である。

図4は、情報システムの構築プロセスと業務の執行プロセスにおける情報システムの利用のイメージを描いている。CIOはEAの長中期計画に基づいて開発案件を制御し、業務システムの運用を通して事業ごとの効果を提供しようとする。個々の業務システムの開発では、提示されたEAのガイドラインに従ってアーキテクチャを選択することによって、局所最適と全体最適が整合される。ただし、技術動向やビジネス動向によって、EAの内容も業務システムも変化するので、CIOも事業の長も、常に互いの乖離を気にしていなければならない。

3.4 投資対効果

情報システム学の研究として、投資対効果が重要な場合もある。しかし、研究内容によっては、投資対効果の評価が必要ない場合もある。投資が適切であるかどうかの判断は、経営者に託された課題であると考えられる。

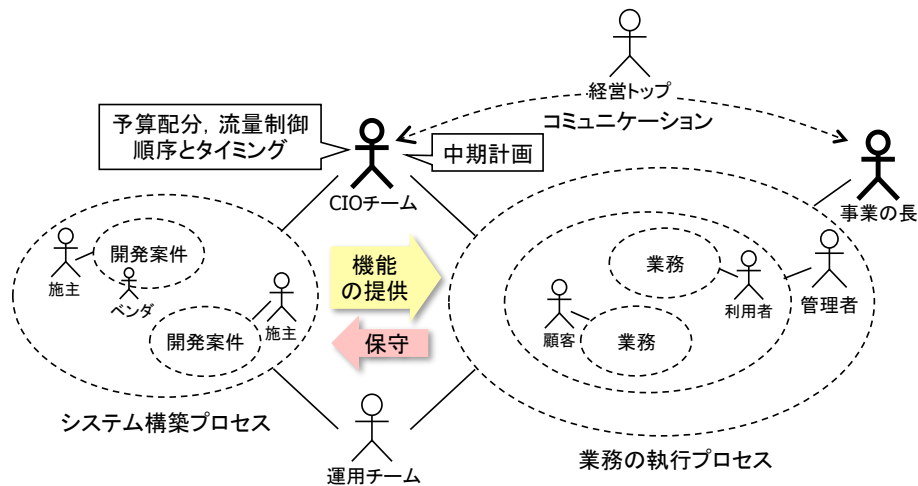


図4 全体最適評価と局所最適評価の調停

4. 情報システム論文のまとめ方

情報システムという組織への介入によって見いだされた変化を、論文として報告することは、ケース研究を中核とする情報システム学にとって貴重な貢献となる。情報システム論文を執筆する際の注意事項を述べる。

4.1 新規性，有効性，信頼性

論文採録の基準は、新規性、有効性、信頼性が一定の水準にあることを査読者が認めることである。情報システム論文においては、要素技術の新規性が必ずしも要求されない。要素技術の組合せ方、適用領域、評価の観点などから、先行研究、関連研究に対する新規性を主張すればよい。ただし、先行研究、関連研究の内容を正しく理解しておくこと。

有効性には、情報システムの有効性と得られた結論や知見の有効性という二面がある。前者については、これまでに述べてきたように、変化を（定量的に）記述して、有効性を客観的に示す（質的評価については別に述べる）。変化を述べるには、介入前の状態、介入後の状態のほかに、その組織が置かれている個別事情（コンテキスト）や介入の必要性または企図を、読者が理解できる程度の詳しさを述べる必要がある。効果の観点については、上に述べたとおりである。後者については、研究者として、得られた結論（知見）の価値を明示する。

信頼性には、研究そのものの信頼性と記述の信頼性という二面がある。前者に対しては、研究設問（Research Question）の妥当性から、計測、評価の妥当性、結論（知見）の妥当性、考察の妥当性までの展開を筋道立てて訴える。後者については、論文の構成、正しい日本語（英語）、文章表現を心がける。

これらを総合して、論文作成のガイドラインを示す。

4.2 対象領域についての説明

情報システムが解決しようとしている問題が存在している対象領域（ドメイン）の知識については、査読者および読者が、研究設問の意味を理解できる程度に述べること。特に、読者はその対象領域の専門家ではないと考えて、特別な用語については解説を施す。ただし、対象領域について詳細に記述する紙幅はないはずなので、適切な文献を示すことで代用する。

4.3 研究設問

事例報告であっても、論文を書く動機としての研究設問（何を明らかにしたいか）または仮説を明示すること。研究設問を述べる際には、当該組織の個別事情（コンテキスト）を可能な限り具体的に示し、企図する者の立場、企図した目的状態、および観察者（論文の著者）の立場を、**目標規定文**として示すこと。研究設問の妥当性を、先行研究、関連研究と対比して述べること。介入の手続きを、追試可能な程度の詳しさを示すこと。結論は、研究設問に対する検証結果として明確に示すこと。

4.4 有効性の評価

結果がよかったから報告するのではなく、結果にかかわらず、実践報告として残して欲しい。

(1) 有効性を違いとして取り出す

情報システムの有効性を、介入前後で観察された活動の変化前と変化後の違いとして示すこと。変化の観点を定めるために、実験計画は必須である。変化は、組織内部だけでなく外部にも表出したり、変化が連鎖したりする可能性がある。変化が組織内で一律でない場合は、観察対象の特性に基づき、群化して見る工夫もする。

その変化が効果的であるかは、観察者ではなく、施主の価値判断に基づく。その判断基準も示すこと。

(2) 計測手段を明示する

第三者が、著者の思考を追跡し、追試できる程度に計測方法について詳しく述べること。実験の方法や要因の設定がうまくいっていなかった可能性があるし、同様の研究をほかの誰かに確認してもらえることは、大きな信頼につながる。

質問紙を用いて心理量を計測した場合は、質問自体を付録などで掲載することで、その妥当性を示すこと。計測環境、方法のほかに、評価参加者の当該情報システムへの関わり方、位置づけを明確にすることも必要である。

(3) 統計的検定

評価実験を行うに当たっては、実験計画を立てて、冗長な計測を避けると同時に、統計的に検定できる枠組みを作る。特に、操作する要因（変数）の定義と、それ以外の要因の変動を統制すること。

心理量を計測するときは、参加者（被験者、サンプル）の統制が重要である。母集団の想定と、そこからの抽出の無作為性を確保すること。参加者の統制はきわめて重要なので、次項にまとめる。

4.5 参加者（被験者）の統制

参加者の統制は、想定する母集団とそこからの無作為抽出（ランダムサンプリング）が基本である。無作為性を担保するために、参加者の採用に関し、次の点に気をつけること。

- 狭い範囲で参加者を募らない
- 一人の参加者を何度も採用しない
- 参加者の層別化を考慮する

大学の研究では、とかく近くにいる学生に参加者をお願いすることが多い。一般化するには母集団として偏っている上に、類似の実験に何度も参加するうちに熟練度が上がってしまい、統計

上のアーチファクト⁴になってしまうおそれがある。

また、参加者を公募する場合でも、一人の参加者に複数の実施条件に参加してもらうことも問題である。参加者は同じ課題を行ううちに、(システムではなく) 課題についての学習が成立してしまうからである。実施条件ごとに異なる参加者を当てるか、学習が成立しにくい作業であれば、同じ参加者でも統計的に学習効果が相殺されるように実施順序を「カウンターバランス」⁵する。

参加者は一律ではない。たとえば、当該システムに対してノービスな参加者と熟練した参加者が存在しうる。熟練度や情報システムに対する知識が評価を左右することがある。参加者を一律に扱ってよいかどうかは、分析結果に依存する。実験計画段階では、参加者の情報システムに関する態度やプロフィールを、補足的実験などによって測定し、記録しておくことが必要である。

4.6 予備実験

評価実験の前に、何度かの予備実験が必要である。予備実験を通して、実験環境の統制、変数の統制、計測手順の洗練を行う。特に、調査票の精度を上げるためには必須である。

4.7 論文の基本的な構成例

参考として、論文の構成例を次に示す。

1. はじめに (研究の背景, 動機)
2. 研究の目的
2.1 導入
対象領域 (ドメイン) の概説, 用語の説明
問題認識への導入
2.2 関連研究, 先行研究 (先行事例)
2.3 問題認識
対象領域についての踏み込んだ説明
利害関係者, 顧客の顧客の存在
問題の状況
2.4 問題解決
問題が解消された状況の定義
2.5 研究設問
Research Question を明示する
3. 提案するシステム/機能/方法/理論
3.1 設計
設計内容についてモデル (クラス図) などで説明
設計判断について根拠を沿えて説明
3.2 実装
実装方式をモデル (配置図) などで説明
4. 評価実験 (観察)
4.1 評価実験の概要
比較対象
用いる指標
4.2 評価方法
4.2.1 環境
組織の特徴, 時期, 体制
4.2.2 参加者 (被験者)
4.2.3 実験の手順

4 Artifact. 統計分析では、実験の統制不良により入り込む攪乱要因および結果を指す。

5 Counter-balance. たとえば、同じ課題を A, B の条件で行う場合、参加者を実験順序 A→B で行う群と B→A で行う群で同数になるようにランダムに分ける。これで、統計的には実験順序の効果を打ち消すことになるが、念のために実験順序を要因に入れて分析し、効果がないことを確認する。

4.2.4	使用する分析手法
4.2.5	評価方法の正当性 正当性への配慮
5.	結果
5.1	結果の表示
5.1.1	統計値, 検定値
5.1.2	補助的統計 (下位検定)
5.2	結果の評価 結果の分析 直接的解釈 導出される結果
6.	考察 (結論) Research Question に対する答え 結果が意味すること 他研究との関連づけ 他の研究との対比
7.	まとめ (おわりに, 結語)
	参考文献
	付録 エビデンスなどの提示

4.8 エビデンスの提示

調査票を行う場合は、その内容 (少なくともその一部)、回答例、および結果のコーディング (採点) 基準を付録に添付すること。報告書にすべてを掲示しない場合でも、研究者間で調査票を共有できるようにしてもらえると、情報システム研究の水準が高まっていくと期待される。

5. まとめ

学会の投稿論文や修士論文の査読をしていると、情報システムの評価方法の杜撰さを痛感する。それは、情報システムなるものを試作して、「調査票で有効性を評価してもらったら『有効である』という回答が 50%を超えたので、この情報システムは有効である」とするたぐいである。そもそも情報システムとは何なのか。その理解も浅いまま、実際の現場に適用することもなく、怪しい調査票を安易に使って、安易な結論を出す。これでは効果も有効性も信頼性もない。

情報システム論文の執筆に当たっての態度や査読に関する最初のガイドライン¹⁾は、2001年に書かれ、それ以来、情報システム論文の特集号が生まれ、論文執筆に関するワークショップが開かれてきた。2007年にも、情報システム論文の特性と評価に関する見解²⁾が出されている。これらの貢献によって特集論文では、幾分品質の改善が見られるものの、修士論文まではこの基準の精神が行き渡っていないと感じる。

本ガイドラインでは、すでに刊行された情報システム論文に関するガイドライン¹⁾²⁾をより具体化することを心がけた。記述のいたらない点は改めていくので、ご活用いただき、ご意見をいただけるとありがたい。質的評価については別途まとめるつもりである。

参考文献

- 1) 永田守男：「情報システム論文の書き方と査読基準の提案」, 情報処理学会研究報告, IS-77, No.4, 2001/6/26

- 2) 神沼靖子：「情報システム論文の特質と評価」, 情報処理学会論文誌, Vol.48(3), 970-975, 2007
- 3) 児玉公信, 新目真紀：「情報システムの有効性評価と統計手法の適用における問題点について」, 情報処理学会研究報告, Vol.2011-IS-115, No.15, 2011/3/15
- 4) 児玉公信：「情報システムの有効性評価のガイドラインについて（中間報告）」, 情報処理学会研究報告, Vol.2011-IS-117, No.13, 2011/9/5
- 5) Edwards, N. P.: “On The Evaluation of the Cost-Effectiveness of Command and Control Systems,” *Proceedings of AFIPS spring joint computer conference*, 211-218, ACM, 1964.
- 6) 力 利則, 藤野喜一：「顧客満足度計測モデルと計測手法についての研究」, 情報処理学会論文誌, Vol.38(4), 891-903, 1997.
- 7) 小幡孝一郎：「IS 研究のコア特性を巡る議論（その 1）」, 情報システム学会誌, Vo.1(1), 18-23, 2006.
- 8) 高橋善文, 牛島和夫：「計算機マニュアルのわかりやすさの定量的評価方法」, 情報処理学会論文誌, Vol.32(4),460-469, 1991.
- 9) Mathieson, K., and Ryan, T.: “The Effect of Definitional Variations on Users' Evaluations of Information Systems,” *SIGMIS Database*, ACM, Vol. 25(2), 35-48, 1994.
- 10) 田中 敏・中野博幸：「実践データ解析法 十秒でできるクイックデータアナリシス」, 新曜社, 2004.
(<http://www.kisnet.or.jp/nappa/software/star/info/new.htm>)
- 11) 増山元三郎：「少数例のまとめ方」, 現代応用数学双書, 竹内書店, 1964.
(<http://www.sci.kagoshima-u.ac.jp/~ebsa/masuyama01/index.html>)
- 12) 豊田秀樹ほか：「原因をさぐる統計学」, 講談社, 1992.
- 13) 中村 義作：「よくわかる実験計画法」, 近代科学社, 1997
- 14) 盛山和夫：「社会調査法入門」, 有斐閣ブックス, 有斐閣, pp.79-90, 2004.
- 15) ISO/IEC 9126-1:2001 “Software engineering - Product quality - Part1: Quality model” (JIS X0129-1: 2003 「ソフトウェア製品の品質－第 1 部：品質モデル」)
- 16) ISO/IEC TR 9126-4:2004 “Software engineering -- Product quality -- Part 4: Quality in use metrics.”
- 17) ISO/IEC 14598:1998 “Software Engineering - Product Evaluation - Part 1: General Overview,” (JIS X0133: 1999 「ソフトウェア製品の評価」)
- 18) ISO/IEC 25000:2005 “Software Engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Guide to SQuaRE.”
- 19) IPA SEC, 「システム及びソフトウェア品質の見える化、確保及び向上のためのガイド」,
http://www.meti.go.jp/policy/it_policy/softseibi/metrics/product_metrics.pdf (2011/8/5 アクセス)
- 20) Fornell, C., et al, “Customer Satisfaction and Stock Prices: High Returns, Low Risk,” *Journal of Marketing*, Vol. 70, January, pp.3-14, 2006.
- 21) 松島桂樹, 「IT 投資マネジメントの発展」, 白桃書房, 2007.
- 22) IEC 62264-1:2003, Enterprise - Control System Integration Part 1: Models and Terminology.
- 23) The Open Group, “TOGAF version 9,” <http://www.opengroup.or.jp/togaf.html> (2011/8/5 アクセス)

Appendix 有効性評価の例

事業領域別に、情報システムの有効性について論点を例示する。

A.1 企業情報システムにおける有効性の扱い

企業活動と企業情報システムとは一体であると考えて、まず企業活動を簡単にモデル化し、そのモデルからみた効果について議論する。

(1) 製造業の例

図 A.1 は、製造業の企業活動のモデルである。管理階層は図 3 でいう案件レベルである。このモデルでは、顧客 (Customer) からの注文に基づいて、作業展開および部品展開を行い、自社 (Organization) において内部資源の引き当て (在庫の引き当て、設備能力の引き当て) を行った後、正味所要量を求めて、取引先 (Supplier) に材料を発注する。同時に生産を開始して材料の到着タイミングと同期させながら粛々と生産し、完成したら納入し、それぞれに対価を得る。この間、知識 (部品表などの製品定義情報)、資源 (材料の在庫や中間品)、資金を環流させて効率化を図るものである。他の管理階層でもこれと同様の構造がある。これは、非製造業でもほぼ同様と考えられる。

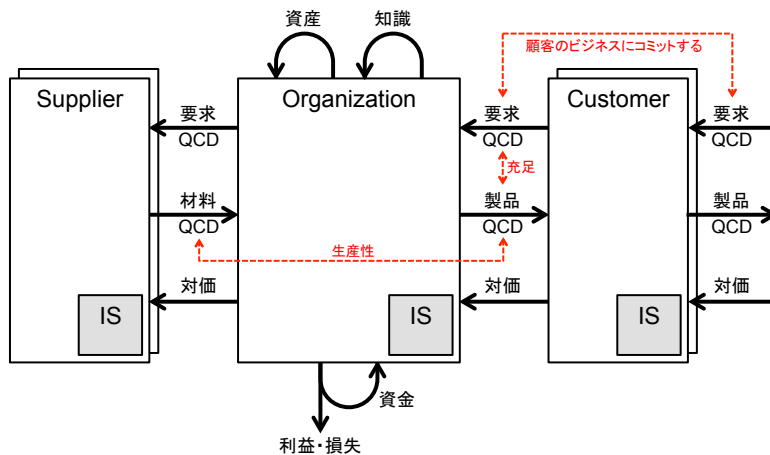


図 A.1 企業システムの基本構造と情報システム

この図に、これまで述べた効果の指標のごく一部を定位して赤字で示した。「生産性」、「コスト」、「収益」の指標は、手持ち資源の制約と材料の入力に対してどれほどの製品が産出できたか、どれほどのコストがかかったか、結果として収益はそれほどあったかを表す。「仕様の充足」は、要求された“QCD”に対して実現された“QCD”が収まっているかを表す。「ビジネスコミットメント」は、真の要求が Customer のさらに Customer の要求にあることを知り、暗黙の“QCD”を付度することを表す。Customer との密接な関係によって囲い込みを狙う。そのほかの指標についての定位は省略するが、内部効果は、概して、この図に表現されていない内部構造に関わる。

(2) 製造業における情報システムの効果

企業は、情報システムの変更も含め、企業活動を変更することによって価値を確保しようとする。効果とは、組織、仕事への介入がもたらす価値の変化であり、図 2 で整理したように、自社の収益が向上する (第 1 象限)、ビジネスの機会が増える (第 2 象限)、自社の潜在能力が高まる

(第3象限), 生産性が上がる(第4象限), そして組織の力が強まる(第3象限のはみ出し部分)などの観点から, 変化を取り出して見ることになる。例として, その様子を図 A.2 に示した。図中の「Organization (O)」は自社の旧組織, 「ISo」は旧の情報システム, 「Organization (N)」は同じく新組織, 「ISN」は新の情報システムを意味する。ここでは, 第1象限と第2象限の効果に着目して前後の差を取り出そうとしている。実際には, Customer や Supplier の要求や仕事の変化が自社の仕事の変化を誘発することもあるが, 一般に自社の変化だけをとらえることでよい。

分科会で行った情報システムの有効性評価に関する論文調査研究³⁾においても, 効果を, 変化前と変化後の前後差や, 変化前の一般的な指標値としての基準値との比較としてとらえていた。あるいは, 複数の異なる方法間での変化の違いを見るクラス比較もあった。効果の議論においては, こうした変化としてのとらえ方と, それを期待する者の立場, あるいは価値観を示す必要がある。

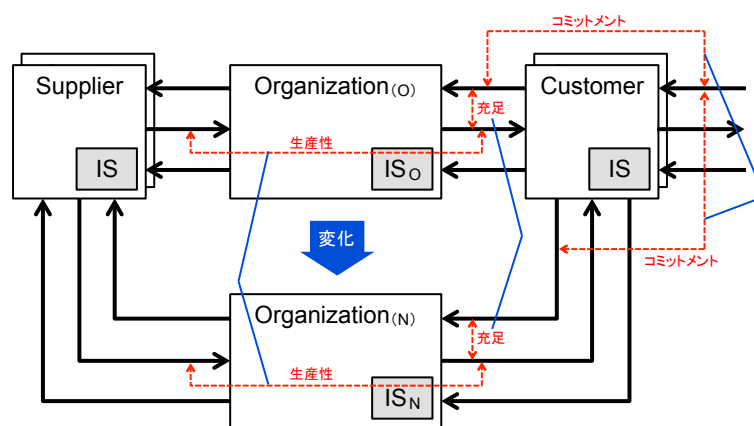


図 A.2 もたらされた変化を効果と見る

ここで, 有効性とは, 利害関係者間で合意された, またはその変化を企図する者(施主)が定めた, システム変更の目的または期待に合致した効果が得られることと定義できる。したがって, 有効性を主張するためには, システムに加える変更の内容だけでなく, その目的とそれがもたらす期待効果を指標値で具体的に示し, それは誰にとっての価値であるか, さらにその背景を示した上で, 達成状況を示す必要がある。