

## 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について

児玉公信<sup>†</sup> 新目真紀<sup>††</sup>

情報システムの有効性評価手法分科会では、情報システム研究における適切な評価手法の採用を促すために、これまでの情報システムの評価に関わる文献調査を行って、共通する二つの問題を見いだした。一つの問題は、統計手法が適切に適用されていないこと、もう一つは、実適用される個別の情報システムは一回性が強く、量的研究に持ち込めないことである。前者に対しては評価パターンごとに適切な統計手法と注意点を提示する。後者に対しては、質的研究として注目される GTA (Grounded Theory Approach) の適用事例を紹介する。

## On a Solution to the Problem of Applying Statistical Approach to Effectiveness Evaluation of Information Systems

Kiminobu Kodama<sup>†</sup> Maki Arame<sup>††</sup>

The sub-SIG for Survey and Study of Evaluating Information Systems proposes two approaches for proper evaluation of information systems. Two common issues have been identified by surveying existing papers of information systems evaluation. One is improper use of quantitative methods, and the other is hardness to apply quantitative methods because of inherent nature, uniqueness and individuality, of information system. The SIG provides a set of application patterns of statistical methods to properly evaluate information systems to address the first issue. For the second issue, the SIG introduces several cases of the Ground Theory Approach (GTA) applications as a quantitative analysis method.

<sup>†</sup> 情報システム  
Information Systems Institute, Ltd.  
<sup>††</sup> 青山学院大学  
Aoyama Gakuin University

### 1. はじめに

情報システム研究において、研究対象の情報システムの有効性を客観的に評価することは必須の研究プロセスである。情報システムおよびその構築過程は、発注者とそれを使用する組織の特性を反映した個別一回性の強いものである。このことは、従来の多サンプルによる定量的評価手法、たとえば統計的手法が適切とは言えないことを意味する。

定量的評価のために、情報システム評価の手段としてアンケートがよく使われる。そのこと自体に問題はないが、そうしたアンケートの中には、残念ながら、「このシステムは有効だと思いますか」といった質問がなされることがある。これは本質的に意味をなさない質問であり、研究の品質を下げ、本来議論すべきことを見失わせる。一方で、看護学などでは個別一回的な事象を対象とする質的研究法、たとえば GTA (Grounded Theory Approach) が一定の学問的地位を固めつつある。

情報システムの有効性評価は古くから行われていた。費用効果に関してアクセス可能な最も古い文献は、軍隊の指揮統制システムに関するものである<sup>1)</sup>。そこでは、効果と費用を見積もる要因として、効率の改善、機能数、開発実施レベルを挙げていた。近年では、システム監査の観点から、品質特性や価値特性を体系化して有効性を評価しようとする事例がある<sup>2)</sup>。情報システムの用途がますます拡大し、その開発形態および運営形態が、高度化・複雑化している現在、有効性評価についての新しい枠組みと手法の提供が必要となっている。

#### 1.1 情報システムの「有効性評価手法分科会」の設立

そこで筆者らは、情報システムと社会環境研究会の活動の一環として、情報システムの有効性評価手法について、一定の見解またはガイドラインを出すための研究分科会の設立を提案した。2010年6月5日の運営委員会において研究分科会の設置が承認され、発足の準備、メンバの募集を経て、9月6日に第1回目のミーティングが開催された。そこで、本分科会の名称を「有効性評価手法分科会」とすること、研究テーマを「情報システムの有効性評価手法の調査・研究」とすること、ほぼ隔月でミーティングを開催することなどを決定した。参加メンバ数は、17名に上った。

#### 1.2 活動の目標と方法

現代の情報システムの特性を考慮した有効性評価手法または評価モデルを確立し、情報システム研究自身の有効性と信頼性を高めることを目標とする。

その第一歩として、これまでの情報システムや関連分野の論文で採用された有効性の評価手法について調査し、現状を知ることから始める。ここでは、どのような評価手法を用い、それが妥当であったか、ほかの評価手法が可能でなかったのかを議論する。この議論を通して、可能であれば、より適切な効果の指標、分析手法、モデル化

手法、アンケート形式やインタビュー方法などを最初の成果として例示する。

その後の目標については、前段での成果を受けて検討するが、筆者は、より理想的な情報システム評価モデルを確立し、情報システム研究のアイデンティティ<sup>3)</sup>を確立する方向に持って行ければよいと思っている。

以下では、本研究分科会の中間報告として、第一回の論文調査の結果を報告する。

## 2. 論文の調査

情報システムや関連分野の論文の評価手法を調査した。1990年から2010年までの情報処理学会およびACMのDigital Libraryに収録されている論文誌および研究報告のうち、キーワードとして「情報システム (Information System)」と「評価 (Evaluati\*)」を含むものを抽出し、明らかに情報システム論文でないものを除いたもの邦文29編と、英文11編の合計40編を取り上げた。

### 2.1 調査方法

取り上げた40編を、無作為に1編あたり二人が担当するように分科会のメンバに割り当て、次の観点で評価手法を評価し、評価シートに記入してもらった。一つの論文に二人の担当を割り当てたのは、評価手法に対して主観的評価を述べることから、見解がなるべく偏らないようにするためである。

- 書誌情報 (省略)
- 被評価システムの目的・ねらい
- 評価の枠組み
  - 評価の方略
  - 計測値
  - 計測の環境
  - 評価方法
- 評価の結果
- 担当者の見解
  - 評価法に対する評価
  - 評価法の問題点と改善案
  - 気づき
  - 重要と思われる参考文献

### 2.2 調査結果

担当した40編のうち、実質、情報システムを扱っていないとされた5編を除いた35編について、各担当者が記入した評価シートを基に、全体の傾向、問題などを浮き彫りにする。

#### 2.2.1 評価の方略

評価の方略とは、被評価事例がその優位性を主張するための、いわば「手口」である。ここでは、「前後比較」「クラス比較」「基準値比較」「仮説検証」「動作検証」の5種類に分けた。

「前後比較」とは被評価事例の前後の状態を比較することによって優位性を主張する方略を指す。これは改善効果、学習効果を見る場合によく使われる。「クラス比較」とは類似の他事例との比較により被評価事例の優位性を主張する方略、「基準値比較」とはあらかじめ設定されている基準値、具体的には経営的基準 (たとえばROI)、監査基準、法的基準との比較によって優位性を主張する方略、「仮説検証」は当該論文で仮説的に導出した基準値と実測値を比較する方法、「動作検証」とは、動作したこと自体を評価する方略である。

方略ごとの度数は次のとおり。「前後比較」が11編、「クラス比較」が7編、「基準値比較」が6編、「仮説検証」が2編、「動作検証」が9編であった。優位性が相対的な判定である以上、比較相手が必要となる。一方、「仮説検証」は仮説の妥当性が問われ、「動作検証」は設計の妥当性が問われる。それぞれを証明する手段としてシステムが実装されることになる。しかし、今回の「動作検証」の9編は作ったこと自体を評価しようとしているように感じられる。その指標は、多くはアンケートによる評定であり客観性を担保できていない。

#### 2.2.2 計測値

計測値は、何を測って評価しようとしたかである。ここでは、所要時間やデータ量などの「物理量」、被験者<sup>a</sup>の感覚、感情などの「心理量」、計測者による良否、有無判定の「合否」と、質的データなどの「その他」の4種類で、傾向を見てみる。ただし、「物理量」と「心理量」の両方が計測されている事例もあるので、必ずしも合計が35にならない。

最も多いのは「物理量」で15編、次が「心理量」で13編、次が「合否」で5編、「その他」が5編となっている。自動的に計測でき、客観的データとして扱いやすい物理量が好まれているが、心理量もほぼ同じ程度に採用されている。

#### 2.2.3 計測の環境

計測の環境は、どのような環境で計測したかである。真の効果を測ろうとすれば、実際に被評価事例が使用されている「実環境」で測ることが望ましい。しかし、事業に影響がある、事前に評価したいなどの理由によって、実験室などの「準実環境」で行わざるを得ないこともある。また、コンピュータ上の「模擬的環境」で計測する場合もある。この傾向は次のようである。

<sup>a</sup> 最近では、否定的なニュアンスのある「被験者 (subject)」という語よりも「参加者 (participant)」という語が使われるようになっていく。ここでは、わかりやすさのために「被験者」を用いておく。

最も多いのは「実環境」で18編、次が「準実環境」で10編、次が「模擬的環境」で4編であった。それ以外は、純粋に理論的検証で、環境を用いないものが3編であった。

#### 2.2.4 計測方法

計測方法は、計測値をどのように得たかである。これについてはさまざまな方法があり、明確に分類しにくい。ざっと傾向を見ると、「アンケート」方式が8編、システムが動作中に計測する「自動計測」が7編、計測者による「観測」が6編で、それ以外の「インタビュー」、「質問紙」が合わせて6編で、残りは計測方法が評価シートからは読み取れなかった。

#### 2.2.5 解析手法

計測値について、どのように解析したかである。実はこれが明らかにされている論文は少ない。物理量や心理量が計測されているにもかかわらず、適切な解析手法が使われていることがトレースできない。

その中で、文献調査の担当者が情報システムの論文の一つのあり方として参考になると挙げたものは、こうした解析経過をわかりやすく提示した2編であった。これらについて、次節で紹介する。

### 2.3 参考となる論文

評価シートの「担当者の見解」の内容は、否定的なものも含めて多岐にわたる。ここではそれらのコメントをいちいち挙げないが、概してより良い有効性評価とするために工夫すべき余地がある。調査文献40編のうち、参考となる論文として2編が挙げられたので紹介する。

#### 2.3.1 アンケートによる評価事例

参考とすべき論文の一つ目は、顧客満足度の計測手法の提案<sup>2)</sup>である。

この論文では、先行研究を概括した後、新たに顧客満足度の指標を定義し、実際の情報システム2件についてアンケート調査を行い、満足度指標を評価している。満足度の指標は、JIS X0129:1994「ソフトウェア製品の評価」<sup>b)</sup>の品質特性のほかに、価値特性として「有効性」と「効果性」を加えた8つの要因を挙げ、その評定値を用いる。総合的満足度を算出する際に要因ごとの負荷点を求める。この負荷点は、8つの指標で対比較した結果からAHP（階層的意志決定法）を用いて決定した。対比較のアンケートは、顧客層（役職などの階層）別に集計した。結果は、総合的顧客満足度が回答者の実感に近い値が得られたとする。

本論文の評価手法として参考となるとされた点は、評価シート記入者によると、「8つの尺度による提案手法は効果的であり、現実的である。良くサーベイされており、また実システム評価を複数のシステムで行うなど、きちんとした論文である。顧客満

b 現在は、2003年の改訂版がある。

足度の範囲にかぎれば参考にできるのではないか」。また、「情報システム維持管理部門の情報システムへの満足度の定量化に挑戦している点評価したい」としている。

調査結果を確認する分科会のミーティングにおいても、実際に行ったアンケートの一部が付録に掲載されていることは、アンケート型の評価手法では見習うべき点であるとされた。

#### 2.3.2 多変量解析による評価事例

参考とすべき論文のもう一つは、マニュアルの「わかりやすさ」について多変量解析を用いてモデルを構築し、さらにアンケートでそれを検証した例<sup>4)</sup>である。

この論文では、計算機マニュアルのわかりやすさを構成すると思われる（形態素解析などにより）自動計測可能な要因（基礎指標）を仮説的に取り上げ、「わかりやすさ」を評定する第一のアンケートを行った。この結果に対して因子分析を用いて、4つの要因（「簡潔さ」、「読みやすさ」、「理解しやすさ」、「親しみやすさ」）を取り出した。次にこれらの要因を用いて「わかりやすさ」の計量モデルを構成した。第二のアンケートでは、評定者をテクニカルライターと工学系の大学院生の二群に分けて評定し、重回帰分析により異なる計量モデルを得た。これらのモデルができたことにより、マニュアルの各作成段階で、機械的に「わかりやすさ」とそのプロフィールが得られる。

本論文の評価手法として参考となるとされた点は、評価シート記入者によると「綿密な論理展開であり、特に問題なし。仮説を設定し、統計学的に結論を導き出し、その妥当性を統計学的に評価する論理展開は、お手本となる」とした。

分科会のミーティングにおいても、因子分析から重回帰分析までのモデル導出に至る各種統計表の提示は見習うべき点であるとされた。

### 3. 統計手法を使うときの留意点

上に紹介した二つの論文はともに、顧客満足度や「わかりやすさ」といった心理量を扱っている。だからこそ慎重に、仮説を提示し、アンケート内容あるいは因子分析の経過を示し、統計処理の結果を示して、仮説の検証とその間の論理のトレーサビリティを確保しようとしたと考えられる。情報システムを定量的に評価しようとするとき、何をもちいて有効性を論じるかについては、さまざまな考え方がありうるが、「良さ」という主観を計測するにおいては、特にその指標値を明確に定義し、実測して、統計値を提示し、有意性を明示<sup>c)</sup>する必要がある。

ここでは、こうした観点から情報システムの有効性評価において統計的手法を用いる際の留意点をいくつか述べる。

c たとえば、分散分析では分析表を示す代わりに、F値と自由度を表記する簡易的な形式が一般的である。「F(1/7)=27.17\*\*」とは、その要因の自由度が1、誤差の自由度が7、F値が27.17であり、それは危険率0.01 (\*\*: p<0.01) で有意であることを述べている。

### 3.1 実験計画を立てる

統計的手法を使う場合に留意しなければならないことは、実験計画を立てることである。研究目的に合わせてどのような仮説を立て、何を比較し、どのようなデータを収集し、分析目的とデータの性質に合わせてふさわしい方法をあらかじめ決めておくことが肝要である。

特に情報システムの評価では、得られるデータが少数であることが多く、同じ条件下での繰り返しの観察が難しいことから、そのための綿密な実験計画が必須となる。

### 3.2 被験者（参加者）の統制

情報システムの有効性は相対的に表現される。多くの場合、同じ課題を異なる条件下で実施しその経過や結果を比較するが、その比較の対象は、提案手法の適用前のシステム（前後比較）であったり、他の類似システム（クラス比較）であったり、別途定められている基準値（基準値比較）であったりする<sup>d</sup>。

前後比較およびクラス比較において、当該システムと対照システムのそれぞれの条件下で、被験者に同じ課題を与えて使用のパフォーマンスを見る実験を行うことがある。ここで被験者の採用に関して気をつけるべきことがある。

- 狭い範囲で被験者を募らない
- 一人の被験者を何度も採用しない
- 被験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被験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<sup>e</sup>になってしまうおそれがある。

また、被験者を公募する場合でも、一人の被験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。被験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被験者を当てるか、学習が成立しにくい作業であれば、同じ被験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」<sup>f</sup>する。

被験者は一律ではない。たとえば、当該システムに対してノービスな被験者と熟練した被験者が存在しうる。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。被験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、被験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。

d これらをコントロール（統制）群と呼ぶことがある。

e 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。

f Counter-balance. たとえば、同じ課題をA, Bの条件で行う場合、被験者を実験順序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。

### 3.3 心理量の扱い

心理量の扱いにも注意を要する。情報システムの評価に関わる心理量は、調査票（アンケート）などを用いて得られる調査対象者の“感覚”の表現（rating）である。ここにはいくつかの問題がある。

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<sup>5)</sup>。それによって調査票の内容も変わってくる。

二つ目は個人差の問題である。感覚には個人差があるし、同じ質問に対しても理解の幅や深さも異なる。これを、予備実験を繰り返して、質問の仕方、ワーディングを改善する。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。

三つ目の問題は、順序尺度の問題である。“感覚”は、「とても○○だ」、「やや○○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。そのため数量化と呼ばれる一連の手法がある。順序尺度を間隔尺度としてみなす場合もあるが、十分な論拠を示す必要がある。

### 3.4 少数データの扱い

少数のデータを取り扱う場合に、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ていることである。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こと。

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、できるだけ正規分布に近づける努力をする必要がある。三つ目が、異常値の取り扱いである。これについては、棄却検定法<sup>7)</sup>などを用いることで、“良心的に”異常値を捨てる。

### 3.5 統計手法

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。

表1 比較でよく使う統計的手法

目的	統計手法	前提
度数の差の検定	直接確率計算 <sup>6)</sup>	
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 分析でよく使う統計的手法

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	

このうち、あまり知られていない「直接確率計算」と「共分散構造分析」について簡単に触れておく。ただし、こうした統計処理を行う前に、クロス集計表を作るなどの基本操作を忘れてはならない。

### 3.5.1 直接確率計算

直接確率計算<sup>6)</sup>は、2群間の1条件または2条件で発生する事象の度数(観測数)の偏りが偶然ではないことを検定する方法である。経験的に、カイ二乗検定ではデータが少ないために検出されにくい分布の差が、度数の差では検出されることがある。

検定の原理は、一つの条件について、二つの観測値が得られたとして、それが偶然であるといえる確率はどのくらいかを見るものである。たとえば、千秋楽に7勝7敗の関取が勝って「給金を直す」確率はどのくらい偶然かを見る。以下は文献6)の例である。ある年の一場所、千秋楽を7勝7敗で迎えた関取は8人いた。結果は2人が負け、6人が勝った。これが偶然であると言える確率は、「早見表・直接確率計算1×2」の度数N=8(2対6)の項を見ると片側検定で $p=0.145$ (14.5%)とある。これは、5%水準よりも大きいので、有意であるとは言えない。つまり、まだ偶然と言い切れる。ちなみに1対7だとしたら、 $p=0.035$ で有意である。

二つの条件でも直接確率比較ができる。たとえば、ある中学校の話<sup>g</sup>。授業開始のチャイムが鳴って2分以内で着席できない生徒は、300人中52人いた。そこで「朝の読書会」を導入した。半年後、同様にカウントして着席できない生徒は36人いた。朝の読書会の効果はあったと言えるだろうか。導入以前(条件1)では着席できない生徒と着席できていた生徒の数は、52対248、導入後(条件2)では36対269として直接確率を計算する。この結果は片側検定で $p=0.0415$ で有意である。十分効果があったと言える。ちなみに、これをカイ二乗検定してみると、 $\chi^2(1)=3.409$ 、 $.05 < p < .10$ でこちらは有意傾向にとどまった。

g この例は、文献6)のものから、直接確率比較とカイ二乗検定で違いが出るように数値を変更した。

### 3.5.2 共分散構造分析

アンケート調査においては、必ずしも分析の対象となる変数が観測数ではない場合も少なくない。たとえば、広告の印象がよければ購買意向が高くなると仮定した場合、広告の印象が独立変数、購買意向が従属変数に当たる。しかし、この場合の独立変数は直接観測できないものがある。このようなものを潜在変数(構成概念)因果モデルという。

重回帰分析では「Aの原因としてBとCの2つがある」という図式を適用して、【想定したモデル内において】原因や結果が何であるかを考えることができるが、さらに複雑な要因をもった、たとえば「Aの原因としてBとCの2つがある。BはさらにDとEに影響する」といった関係を検証する場合に有効なのが共分散構造分析<sup>エラー! 参照元が見つかりません。</sup>である。

共分散構造分析は、従来の因子分析に対して提唱された「確証的因子分析」が発展したものである。因子分析は、観測変数のデータから、その背後にある因子という「内生変数」について調べるものであるが、因子分析によって抽出された因子群を潜在変数としてモデル化することができる。

近年では、共分散構造だけではなく潜在変数の平均構造を解析するモデルも開発されたことから、構造方程式モデル(SEM: Structural Equation Models)と呼ばれることが多い。これにより、「Aの原因としてBとCの2つがある。BはさらにDとEに影響する」影響を与えられる度合いには男女間で差があるといった分析が可能になる。

## 4. 今後の活動

有効性評価手法分科会の今後の活動について述べる。

### 4.1 アーキテクチャや設計の良さに関して

情報システムのアーキテクチャや設計そのものの良さを評価する必要はないだろうか。情報システムは組織の中で長期間にわたって徐々に修正、拡張されていくという立場からは、それを容易にするアーキテクチャの有効性を論ずる必要があると考える。

情報システムの有効性は、実際に運用に供されて本当の評価が得られるというのは正しい。しかし、それには相当の時間がかかる。まして、企業情報システムのアーキテクチャの有効性評価については、5~10年の時間を要する。研究として、評価のために複数の実装が必須であるとするのでは、あまりに費用や時間がかかりすぎる。また、たとえ設計は良くても、実装がまずかったために、十分な評価が得られなかったという誤判定もあり得る。実装や運用によらずにそれらの良さを判定する方法はないだろうか。たとえば、概念モデルや業務シナリオによる設計の表現を評価できないか検討したい。

## 4.2 質的アプローチに関して

統計的手法は、安定した環境で一定量のデータを必要とする。当該システムに、前もって目的に沿った効果の指標を採取できる仕組みを組み込むことで、評価が容易になるように持って行くことができる。しかし、事業に対する情報システムのインパクトをどのように評価するかについては、一定量のデータを集めて統計的に判断することはできない。これが、情報システムの個別一回性の問題である。

個別一回性の強い対象に対する評価指標として、看護学をオリジネータとする質的研究法の Grounded Theory Approach (GTA)<sup>8)10)11)</sup>が、心理学や教育学などの他分野でも一定の評価を得るまでになっている。先端的な情報システムでは、多くの被験者を集めて使用実験をして評定値を得るという方法はそぐわない。少数の先端的ユーザに対する半構成的インタビューから、その人たちが認識する情報システムのモデルを示すことによって、新しい価値を見いだしていくことが、その情報システムの進化につながることもある。そこで、有効性評価手法分科会としては、GTAの情報システム評価への応用について、見解を示したいと考えている。

GTAについては、いまのところ調査研究の準備中である。欧米では、予稿レベルではあるが、情報システムの評価にGTAを使用する試みがすでに報告されている<sup>12)13)14)</sup>。これらの文献調査をするとともに、GTAの具体的な適用方法について、指針を示すことができるか検討を進める。

## 4.3 調査票の事例集に関して

有効性評価の調査票については、可能ならば、事例集を作成することも重要な貢献と考えている。これは、いくつかのケースを想定してあらかじめ調査項目を示すと同時に、統計的手法も併せて示すものである。調査項目のワーディングは、調査結果にも大きく影響するため、特に慎重に行う必要がある。たとえば、社会調査法<sup>15)</sup>などを参考に、曖昧な言葉、難しい言葉、ステレオタイプによる偏り、二重質問、誘導質問などを避ける、また、個人質問と一般質問、事実判断、価値判断を区別して偽答を導かないなどの工夫が必要である。分科会参加メンバの業績などを提供してもらい、調査項目を決めワーディングを検討して下案を作成する。これを実際の研究などで使用してもらいながら、質問項目とワーディングを段階的に洗練していく。一定の精度で結果が得られると認められた段階で、調査票に名称をつけて公表する。

今後の研究において情報システムの評価を調査票で行う場合には、この調査票を使用してもらい、論文および研究報告には調査項目を掲示するのではなく、調査表の名称を示すだけでよいものとする。もちろん、調査票で改善を要するものがあれば、それを調査票に定期的に反映することで、ますます調査票の精度を高めていく。

## 5. まとめ

以上、「有効性評価手法分科会」設立の経緯と活動目標について述べ、その活動である情報システムの有効性評価の文献調査の成果を報告した。

まず、有効性評価方略の分析結果を述べた。ここでは、実環境またはそれに近い状態で動作させた情報システムを、時間などの物理量やアンケートによる心理量を用いて、対照システムと比較して有効性を述べるという方略がとられていた。ただし、多くは有効性を客観的に説明する証拠に乏しい。文献調査で参考になると挙げられた2論文は、この点、アンケート内容の掲示、統計結果の明示がなされており、有効性の論理がトレースできた。

次に、有効性を客観的に説明する証拠としての統計処理方法について紹介した。統計処理は単に統計処理ツールにデータを流し込めばいいというものではない。実験計画を立て、被験者を統制し、良い調査票を作り、少数データとしての配慮をしたうえで、適正な統計方法を使う必要がある。

最後に今後の取り組みについて述べた。情報システムの有効性を設計レベルで評価する方法の検討、ごく少数の意見に基づく質的研究の方法としてのGTAの可能性の検討、調査票の事例集の作成について期待を述べた。各方面からのコメントを歓迎する。

## 参考文献

- 1) 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.
- 2) 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.
- 3) 小原孝一郎: 「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.
- 4) 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.
- 5) 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.
- 6) 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>)
- 7) 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>)
- 8) 狩野 裕: 「共分散構造分析とソフトウェア」, BASIC 数学連載, 現代数学社, 1996.2~1997.3.
- 9) 戈木クレイグヒル滋子: 「質的研究方法ゼミナール-グラウンデッドセオリーアプローチを学ぶ」, 医学書院, 2005.
- 10) 木村康仁: 「ライブ講義 M-GTA-実践的質的研究法」, 弘文堂, 2007.

- 11) 西條剛央：「ライブ講義：質的研究とは何か（SCQRM ベーシック編）」，新曜社，2007.
- 12) Goede, R., and De Villiers, C.: “The Applicability of Grounded Theory as Research Methodology in studies on the use of Methodologies in IS Practices,” *Proc. of SAICSIT 2003*, ACM, pp. 208 –217, 2003.
- 13) Fitzgerald, S., Simon, B., and Thomas, L. “Strategies that Students Use to Trace Code: An Analysis Based in Grounded Theory,” *Proc. of ICER '05*, ACM, pp. 69-80, 2005.
- 14) Matavire, R., and Brown, I. “Investigating the Use of “Grounded Theory” in Information Systems Research,” *Proc. of SAICSIT 2008*, ACM, pp. 139-147, 2008.
- 15) 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.