

Guttman scale の評価のための一研究

池 田 央

- I. Guttman scale の特徴と問題点
- II. Guttman scale と non-Guttman scale の抽出
- III. Guttman scale と non-Guttman scale の比較
- IV. 要約と結論
- V. References.

近年態度測定や意見調査において、数量化理論をとり入れて個人の態度なり意見なりを一尺度上に表現しようとする試みが多くなされるようになった。その代表的なものとして、Thurstone 法 [47], Likert 法 [36], Guttman 法, Lazarsfeld 法 [33.34.35] 等があげられる。

ここではその一つである Guttman の尺度解析法をとりあげ、それによって作られる scale の長短を実験的に比較することにする。

I. Guttman scale の特徴と問題点

Louis Guttman*が彼の scale 理論を考えはじめたのは1940年代、第二次大戦中であった [18,43]。彼はまず彼のいう scale を定義し、つぎに与えられた一組の test items が彼のいう scale の条件を満足するか否かを検討する。その方法が所謂 Guttman の尺度解析法である。

彼のいう scale と同じような概念をもつ scale は以前にも Bogardus (1925年) の social distance scale, May and Hortshorne (1826年) の scale 等があったが [13], Guttman の scale ほど広く用いられはしなかった。また彼以後にも Loevinger の homogeneous test の概念 (1947

* 現在 The Israel Institute of Applied Social Research および The Hebrew University (Professor of Social and Psychological Measurements として) 在職。

年) [37] が現われたし, Walker (1931年) は unig, および hig の言葉を使って test の反応 pattern を表わそうとしたが [51], それらで狙う基本的理想型は Guttman の perfect scale pattern に相当するものである。しかしこれらも Guttman のそれほど, 理論的にまた実用的に大成したものにはならなかった。

Guttman scale の特徴については文献 [5, 10, 18, 20, 21, 22, 23, 25, 45, 46, 50.] に詳しいが, ここでは, 本論に直接関係の深い事柄だけをとりまとめて略述する。項目のとりあげ方も理論的に分けたものでなく, 叙述に都合のよいように便宜的に分けたに過ぎない。

(1) 目的

Guttman scale の狙いは, ある与えられた集団の各人がある測ろうとする領域に沿って一義的に, また直接的に rank づけしようとするものである。

(2) 測定内容

尺度解析法は態度測定を中心として発達して来たものであるが, その方法は態度に限らず, personality の測定にも, 学力の測定にも, 測定内容の如何にかかわらず適用できる [22], 必要なのは, 形式面からくる要請で, paper と pencil および一定の型式をもった questionnaire である。

(3) 定義

各人を一義的に ranking するような scale の作成が尺度解析法の狙いであるといったが, そこでいう尺度 (scale) とか一義的とかいう意味はつぎの通りである。

いくつかの項目からなる test において, その総得点の大小に応じて, 各人を ranking するものとする。つまり P_i, P_j なる 2 人の人の総得点が T_i, T_j で表わされ, そのとき $T_i > T_j$ が成り立つならば P_i と P_j は P_i の方が高い rank にあるものとする。このときこの test を構成する各項目がどの項目も $I_i \geq I_j^*$ の条件を満していなければ, この ranking は意

*ただしすべての項目が $I_i = I_j$ であってはいけない。そのとき $T_i = T_j$ となるから。

味が不明瞭になってくる。項目のどれかが $I_i < I_j$ と逆の方向に ranking しているのに、全体の得点が $T_i > T_j$ では、 $I_i < I_j$ なる項目はこの ranking に対し、何の働きをしている意義が不明瞭である。

このような項目は Guttman scale の理想型では除かれ、そこでは各項目の持つ方向（大小関係）が全体の得点の方向（大小関係）と平行するようになる。さらにこのような場合には、総得点から逆にその人の各項目の反応をも伺い知ることが可能になる。つまり各人の総得点が判れば、その人が各項目でどう反応したかも知れるようになり、その意味で各項目は総得点の simple function となる。そしてこのような場合、そこで構成された scale は undimensionality を持つものとする。このことが尺度値を記述に用いるとき意味を明瞭化する理由となる。

(4) 累加的性質

このような条件を満たす scale は具体的にはつぎのようにして可能である。いま I_1, I_2, \dots, I_m なる m 個の項目から成る test を考える。各項目の困難度(通過率)を P_1, P_2, \dots, P_m とし、かつ $P_1 > P_2 > P_3 \dots > P_m$ ** であるとする。このとき各人の各項目への反応 pattern が、項目通過の場合を 1, 非通過の場合を 0 で表わすとき、第 1 図のようになれば前記の条件は満足される。

第 1 図 理想的反応 pattern の図

rank	項 目						得点
	I_1	I_2	I_3	I_{m-1}	I_m	
1 位の人	1	1	1	1	1	m
2 位の人	0	1	1	1	1	$m-1$
3 位の人	0	0	1	1	1	$m-2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
m 位の人	0	0	0	0	1	1
$(m+1)$ 位の人	0	0	0	0	0	0
	$P_1 > P_2 > P_3 > \dots > P_{m-1} > P_m$						

* Guttman scale を見出す操作はこれからも分るように自然、項目分析の性格を帯びてくる。

** $P_1 > P_2$ とは P_1 の方がそれだけ通過する人の割合が少いことを示す。

つまり m 個の項目のうち、出来た項目の数を算えてそれを得点とする。その得点の高いものは低いものよりも rank が高いと考える。また各項目を難しさの高いものから低いものへ並べて考える*。このとき項目 I_1 は一ばん難しい項目で得点が m の人 (rank が 1 位の人) でなければ出来ない。得点が $m-1$ の人 (rank が 2 位の人) は一ばん難しい項目は出来ないが、二ばん目に難しい項目以下は全部出来る。得点が $m-2$ の人 (rank 3 位の人) は一ばん難しい項目と、そのつぎの項目は出来ないが、第三番目に難しい項目以下は全部出来る。……

このような場合には総得点が分ればその人は第何ばん目までの項目は出て、第何ばん目以下の項目は出来ないかを完全に予測することが出来る。すなわち、各項目は総得点の simple function であって、この場合総得点 j の人の項目 i の反応 I_{ij} は

$$I_{ij} = \begin{cases} 1 & (\text{もし, } i+j > m \text{ ならば}) \\ 0 & (\text{もし, } i+j \leq m \text{ ならば}) \end{cases}$$

と予測出来る。

ところでこの場合注意しなければならないのは、rank が一つ上るためには一段難しい項目が出来なければならない、しかし、それはその項目より易しい項目はすべて出来て、その上で出来なければならないといという性質である。能力 test であれば、難しい項目が一つ出来て rank が一つ上るためには、それ以前の項目が全部出来る能力の上にプラス α の能力が加ってはじめて rank が上るのである**そのような意味で Guttman scale を構成する項目は累加的性質を持たねばならない。

(5) 再現性の概念

しかし、以上で述べたような各項目の反応が総得点の simple function として一義的に決定され、理想的な累加的反応 pattern を示す場合は実際

* 実際の test 用紙の項目の配列がその順になっていなければならないという意味ではない。

** この性質は能力診断検査または発達段階検査をつくるのに有効である。

にはなかなか得られない。それは理想的 model にしか過ぎない。実際は総得点から各項目の反応を予測してもその何%かは失敗して、100%の成功を示さない。この場合の成功の割合を Guttman は reproducibility (再現性) の概念を用いて表わす。つまり総得点から各項目の反応を再現してみても何%再現し得るか、その割合を coefficient of reproducibility で表わす。そしてこれが高いとき、(普通は 90%以上), その scale は perfect scale の近似として用いられ、そこで測られる対象は unidimensional な次元のもとに測定されると考える。ただこの coefficient of reproducibility は他の多くの coefficient のように 0 から 1 までの間とか、-1 から +1 までの間を変化するのではなく、その下限は項目の困難度に左右される [10]。その意味では Loevinger の index of homogeneity の考え方の方が sensitivity が高い。また Green は [14] において、この Guttman の coefficient of reproducibility の欠点を除く係数として、coefficient of consistency を唱えている。White and Saltz [52] は、他に、Jackson の Plus Percentage Ratio の考え方 [30] なども加えて、これら係数間の比較を行っている。

(6) categorical data の適用

scaling をしていくもとの質問紙への反応が「はい」「いいえ」のいずれかに○をつけるとか、観察法であれば、あらかじめ定められた category のいずれかに○をつけるとかの形である場合、従来の scaling 理論で多くとった立場は、それらの反応に何らかの仮定の上で weight をつけ、その和をある連続体上の一点と考えるとらえて行こうとする立場であった。Guttman の scaling 理論では、そのような categorical な反応は categorical そのままの形で数量化して行こうとする。これは従来のように、時には無理な仮定(たとえば分布の正規性等)を設けて数量化して行くのところが、極めて自然な形で data を扱うことになる。したがってここで得られた scale は Stevens のいう ordinal scale [42] にあたり、

そのことが一方ではある制約を与えることにもなる*

(7) 潜在変量の仮説の否定

各人の反応をそのままの形で生かそうとする Guttman の考えは当然その結果として、表わされた反応がそのまま測りたいものをずばり測っているものと考ええる。Spearman-Thurstone 流の因子分析での factor score とか Lazarsfeld の潜在構造分析法での latent variable とかは、いずれも表わされた data そのものからは直接知ることの出来ない背後にある潜在的なものである。Guttman の考えでは、そのような意味での潜在的構造を仮定しない。よし潜在的なものがあるにしても、それは各項目の反応に直接表現されているものとみる。いわば Spearman-Thurstone, あるいは Lazarsfeld の考えが genotype の研究を目指しているのに対し、Guttman は phenotype の data を尊重しようとする。genotype といっても必ずしも本質的な何かがあるとは限らない。その多くは仮説的なものにとどまっている。Guttman はそれを排除しようとする。しかしそのことのために実際の data を十分説明しきれず、のちに Stouffer らの Harvard 法 [44,49] や Guttman, Foa などの Israel 法 [26,28] へ変化発展することにもなるがここでは扱わない。 ([29] 参照)

(8) single factor の仮定

Guttman の定義する scale によって測定されるものは、一つの一次元的な線上にあって、それは single factor のみに依存すると考える。Spearman-Thurstone 流の因子分析では、各項目はいくつかの common factors と、各項目に独自の unique factor の二つを測っているものと考ええる。しかし Guttman のいう scale ではそのような多数の factors を想

* Guttman scale で表わされた尺度値は質的なものであって、(したがって Guttman は scale value といわず scale type という) これを量的なものと考えて種々の統計量(平均値とか分散とか相関係数等)を求めるのは正しくない、にもかかわらず実際多くそのような処理がなされているのは、質的なものをそのままの形で処理する方法(たとえば non-parametric statistics)の発達が遅れていることにもよるのであろう。

定しない。いくつかの項目が1つの **scale** を構成する以上、そこでは1つの **factor** しか測っていないものとする。しかしその1つの **factor** が Spearman-Thurstone の因子分析の中で表わされる1つの **factor** と同じものであるとは限らない。因子分析により別々の **factor** とみなされるものが Guttman の **scale** 理論では1つの **factor** とみなされるかも知れない。Guttman の **scale** 理論では **phenotype** の **data** を扱う以上、もし **genotype** の因子（もしあるとすれば）が別々のものであっても、**phenotype** 的に差のない項目はいずれも1つの因子を測っているものとして扱われる。したがって同じ **data** も因子分析的に理解するのと、尺度分析的に理解するのとでは結果にくいちがいがでてくる。

(9) 一次元化と多次元化

Guttman の **scale** を実用的にしたのは、一組の与えられた **data** が彼のいう **scale** を構成するか否か検討する方法を同時に考えたからである。その方法としては、

- 1) scalogram board technique [45]
- 2) 紙と鉛筆でできる Cornell technique [21]
- 3) Goodenough の tabulation technique [12]
- 4) Guttman の the least squares method [10, 20, 25]

その他機械を用いた 5) Noland [40], 6) Ford [11], 7) Kahn and Bodine [31] の方法 8) 再現性係数を簡単に見出す Green [14] の方法等があげられるが、そのいずれにも共通する点は、ある与えられた一組の **data** が Guttman のいう **scale** を構成するか否かを検討する点にあるのであって、どういう項目が Guttman **scale** を構成するか明らかにするものではない。

Edwards [5], Edwards and Kilpatrick [7, 8] も指適するように、実際の場合に必要なのは、pretest した数多くの項目の中からどの項目を選べばよいか、その項目選択の方法を示すことである。さらに進んで、あるいくつかの項目の中でどの項目とどの項目とが一つの **scale** を構成し、

また別のどの項目とどの項目とが別の **scale** を構成するか発見する。それによって、いくつかの項目がお互いどのような結びつきをもっているか明らかにし、測定対象の構造を説明することができればぐあいがよい。

Guttman の **original** な尺度解析法はその問題の解決には役立たない。

そのためには2つの方向が考えられる。1つは与えられた資料のうちから、**Guttman scale** を構成すると思われる項目だけを選出し、それと同時に選出された項目群からなる **scale** が、できるだけ **Guttman** のいう理想的な反応 **pattern** をもつ **scale** になるように、積極的に反応 **pattern** を変形して行く方向である。つまり、出来るだけ多くの項目が出来るだけ一次元的なものを測定する項目になるように、項目の一次元化を狙う方向である。その例が、**Haward** 法 [44,49] や **Israel** 法 [26,28] あるいは、**Edwards** の **W-technique** [6] の行き方である。

もう一つは、与えられた資料は与えられたままとし、その資料の中には一次元的な因子のみを想定するのではなくて、多次元的な因子の存在を仮定し、いくつかの次元で資料を説明すると同時に、1つ1つの次元では **Guttman scale** が構成されているようなものを狙うのである。方法においてはかなりの相違がみられるが、その考え方は **Spearman-Thurstone** 流の因子分析法と狙いを同じくするものである。その例として **Coombs and Kao** [3, 4], **Bennett** [1], **Milholland** [39] らの方法があげられるが、まだ十分の実用化をみていない。

II. **Guttman Scale** と **non-Guttman Scale** の抽出

このように **Guttman scale** は種々の問題点を残しながらも、その理論的 **level** の高さと、そこからひきだされるいくつかの興味ある事実 [25,27] に刺激されて、多くの方面に応用され（その例として [41] がある）またその方法もいくつかの方向へ変化発展の途上にある。この節ではまず典型的な **Guttman scale** を実際の **data** の中から取り出す。またそれと比転のために、**Guttman scale** の条件を満たさない **non-Guttman scale** を取り出し、両者を比較することによって、**Guttman scale** が **non-Gutt-**

man scale に比べて、理論的のみならず実際的にも、優れているかどうかをみることにする。

(1) 材 料

Guttman scale 作成の source として、Thurstone の Temperament Schedule 140 項目を用いる。これは personality questionnaires の一種で、被験者は各質問項目で尋ねている事柄について自分を評定し、「はい」「わかりません」「いいえ」のいずれか該当する category に check する。所要時間は教示を含めて約40分前後である。

例, 1. あなたはふつうの人よりそわそわしていますか, 「はい, わかりません, いいえ」 etc.

(2) 被 験 者

〔第1次調査〕 典型的 Guttman scale と, それと比較のための non-Guttman scale 発見のために企てられたもので, 東京都内の公私立大学延4校男子学生約1000名である。そのうち無反応回答を含む資料を除き, 利用した資料は計855名分である。被験者の年齢は18~32才までにわたり, その90%が18~22才, 平均年齢は20才2ヶ月である。調査期間が1955年4月から1956年8月までに行われたものの資料を用いた。

〔第2次調査〕 Guttman scale の信頼性を検討するために企てられたもので, これには第1回と第2回とがある。第1回は1957年4月, 第2回は同年6月, 同一被験者120名(都内某国立大学)に同一 test を施行した。これら被験者は第1次調査の時の被験者とは別人で共通するものはないが, 大学は第1次調査で用いられた大学の中の一校である*。

〔第3次調査〕 Guttman scale の妥当性を検討するために企てられたもので, 第1次調査に用いられたと同じ大学の中から一校を選びその中でお互いどうし personality をよく理解していると思われる group 27名を被験者とした。

* 同一母集団からの異なった sample を得ようとの配慮からである。

調査は1957年11月に行われたが、被験者は第1次調査にも第2次調査にも参加したことの無いものである*。

(3) 四重表の作成

与えられた資料から典型的な Guttman scale を発見するため、Temperament Schedule 7領域^{**}内で各項目間の四重表が作成された。その個数は、1430個におよぶため、資料はすべて punch card に転記され、Remington Rand の counting sorter で分類された。最初は「はい、わかりません。いいえ」の3 categories を各項目が持つため九重表が作成されたが、のち、得点反応に応じて各項目は2分され、四重表が作成された。

(4) 典型的 Guttman scale の抽出

上で求められた四重表を利用して、お互いどうして、以下の基準を満たす項目を拾い上げればそれは Guttman scale を構成する。勿論それは perfect Guttman scale ではない。perfect Guttman scale は実際には存在しないが、その近似として Guttman のみとめる条件—(i) 再現性係数が90%以上であること、(ii) 周辺度数がどの項目も被験者数の20~80%内にあること、(iii) error pattern が random 性をもつこと、[11, 45, 48] を満たすものである。それを通常われわれは Guttman scale と呼んでいる。

さて、四重表が満たす条件とは、項目 i と j との四重表を第2図のようにするとき、

$$\begin{aligned} \text{(i)} \quad & 0.2 \leq \frac{a+c}{N} \leq 0.8 \\ & 0.2 \leq \frac{b+d}{N} \leq 0.8 \\ & 0.2 \leq \frac{a+b}{N} \leq 0.8 \\ & 0.2 \leq \frac{c+d}{N} \leq 0.8 \end{aligned}$$

	i -th item		
	+	-	
j -th	a	b	$a+b$
item	c	d	$c+d$
	$a+c$	$b+d$	N

第2図

* 同一母集団からの異なった sample を得ようとの配慮からである。

** Active, Vigorous, Impulsive, Dominant, Emotionally stable, Sociable, Reflective の7領域である。

$$(ii) \quad a+b \leq a+c \text{ なるとき, } \frac{b}{N} \leq 0.1$$

$$a+b > a+c \text{ なるとき, } \frac{c}{N} \leq 0.1$$

$$(iii) \quad a+b \leq a+c \text{ なるとき, } 2b \leq a \text{ かつ } 2b \leq d,$$

$$a+b > a+c \text{ なるとき, } 2c \leq a \text{ かつ } 2b \leq d,$$

のすべてが成立することである。

この条件が集められた項目間ですべて成立するとき、その項目群は Guttman scale を構成する*。その理由についてはここでは詳述する紙数を持たないので文献〔29, 48〕を参照されたい。

さて、このような基準のもとに集められたいくつかの Guttman scale のうち、ここではさらにつぎの基準により、典型的な Guttman scale を一つ選んだ。それは、

1) 原則として、各項目相互が上にあげた i)–iii) の基準を満足すること。

2) 一つの scale を作るための項目の数を最小限 4 個とする。これは被験者の ranking は五段階法にすることが多く、そのためには項目 4 個を必要とするからである。もう少し目盛の細かい scale を作ろうと思えば、項目の数をふやせばよいが、5 項目以上では 1) の条件を満たすことができず結局 4 項目を選ぶことになった。

3) 出来上った scale で被験者集団が大体等間隔に分れるようにすること、つまり項目の困難度が大体等間隔に並ぶようにすること。

これは前項の条件に関係することであるが五段階に全部の人を分けようと思えば、各 scale の目盛が被験者を 5 つに分けることが出来なくてはならない。

その一つの方法は、被験者を五段階に分けたとき、各段階に入る人数の

* この条件はかなりきびしく、Guttman scale をつくるための十分条件を与える。実際は、これらの条件の一部が満たされなくとも、Guttman scale を作ることはある。

割合が正規分布に近く、7%、24%、38%、24%、7%になるような項目を選ぶことである。しかし Guttman scale は両端に近い人々を sensitive に弁別出来ない欠点を持っている（後述）。したがってこの場合は各人を discriminate する割合が一ぱん大きくなるような基準のもとに項目を選ぶ方が適当と考えられる。Ferguson [9]によれば、それには各段階に入る人数の割合をすべて等しく20%ずつになるようにすればよい。そのための手がかりとしては四重表の周辺度数で positive の反応が30%、45%、55%、70%前後にくるような項目をとるのが、経験上よいようである。

(5) non-Guttman scale の抽出

典型的 Guttman scale の特徴を吟味するには、それと比較できる non-Guttman scale が必要である。その抽出の基準はつぎの通りである。

- 1) 各項目相互間で前項にあげた i)—iii) の基準を満たすものがないこと。
- 2) 項目数は前項で選ばれた Guttman scale の項目数と一致すること。したがってここでは4個である。
- 3) 五段階に分けた時の各段階に入る人数が、大体等しく20%ずつになるようにすること、

つまり 1) 2) 3) 三つの基準のうち 2) 3) は Guttman scale も non-Guttman scale も条件を等しく、1) において、つまり Guttman scale をなすかなさないかの分れ道になる基準では、正反対のものを項目として選ぶことにした。

(6) 選出された項目

このような基準のもとに典型的 Guttman scale を構成する項目として選ばれたのはつぎの4項目で、これにより構成される scale を scale A と名づける。

I (133) いろいろ知らない人たちと会わなければならない仕事が好きですか。（「はい」と答えたものに1を与え、そうでないものに0を与える。以下同様。）

II (49)*たくさんの人と接する仕事が好きですか。

III (78) 知らない人とでもすぐ友だちになれますか。

IV (50) 近所の人としたしくなれますか。

また、上の scale A と対比されるものとして、non-Guttman scale の例として抽出されたのはつぎの 4 項目で、これにより構成される scale を scale B と名づける。

I (18) さわがしい部屋にいてもくつろげますか。

II (103) 期限付きの仕事をしなければならぬと思うといらいらしますか。(「いいえ」と答えたものに 1 を与え、そうでないものに 0 を与える。以下同様)

III (76) 何か心配ごとがあると食欲がなくなりますか。

IV (48) 朝目がさめるまでにだいたい時間がかかりますか。

こうしてつくられた 2 つの scale の特性値を表にかかげると第 1・2 表の通りである。これは Ford [11] の scale analysis の表に多少手を加えて使い易い形にした。これによると scale A では満たされなければならない 4 つの条件をすべて満している。((53), (54), (55) 欄をみよ。) したがって scale A は理想的 Guttman scale に近いものとみることができる。これに反して、scale B では満たされなければならない 4 つの条件のうち 3 つが満たされていない。((53, (55) 欄をみよ。×印のあるのがそれである) したがって scale B は non-Guttman scale の例として扱うことができる。

III. Guttman Scale と non-Guttman Scale の比較

多くの scale 作成者が共通に狙う良き scale の条件は、大きくいって、

* ()内の数字は、もとの Thurstone Temperament Schedule での項目番号を示す。

第1表 Scale Analysis の表

Scale A		Questions		Frequency (Positive)	%	(Positive)	Card Col	Positive Categories		Negative Categories	
		Order	No.					Codes	Wgts	Codes	Wgts
Repro- ducibility	%	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
92.7		I	133	251	29.4	67	Y	8	?+N	0	
Chance Repro- ducibility	%	II	49	357	41.8	55	Y	4	?+N	0	
85.2		III	78	427	49.9	60	Y	2	?+N	0	
Coef. of consisten- cy		IV	50	537	62.0	56	Y	1	?+N	0	
0.51											

Scale Type	Unique Score	Pattern of Re- sponses	Fre- quency Subtotals	Fre- quency Cri- terion I	Errors	Positive Frequencies				Negative Frequencies					
						I	II	III	IV	I	II	III	IV		
(52)	(10)	(51)	(12)	(11)	(53)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	
4	15	++++	19.6 (22.9 %)	139	○	0									
	14	+++ -		19	○	1									1
	13	++ - +		26	○	1								1	
	12	++ --		12	○	2	1/2	1/2						1/2	1/2
3	7	- +++	137 (16.0 %)	79	○	0									
	6	-- ++		67	○	1									1
	5	- + - +		41	○	1		1/2						1/2	
2	3	-- ++	173 (20.2 %)	98	○	0									
	11	+ - ++		32	○	1	1/2					1/2			
	2	-- + -		35	○	1			1/2						1/2
	10	+ - + -		8	○	2	2/3		1/3				1/3		2/3
1	1	---- +	115 (13.5%)	110	○	0									
	9	+ --- +		5	○	1	1								
0	0	-----	234 (27.4 %)	200	○	0									
	8	+ ----		10	○	1	1								
	4	- + --		24	○	1			1						
Total			(23) 855	(22) 855											

Computation of Error

	Total	In Positive Categories				In Negative Categories			
		(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)
Frequency of Error	249	42.3	50.5	20.2	0	0	18.7	52.5	64.8
No. Responses Involved	3420	251	357	427	530	604	498	428	325
Criterion II	(54)	○	○	○	○	○	○	○	○
Total Errors by Question		(42) 42.3	(43) 69.2	(44) 72.7	(45) 64.8				
Percentage of Error	(46) 7.3	(47) 4.9	(48) 8.1	(49) 8.5	(50) 7.6				
Criterion III & IV	(55)	○	○	○	○	○			

第2表 Scale Analysis の表

Scale B		Questions		Frequency (Positive)	%	Card Col.	Positive Categories		Negative Categories	
Repro- ducibility	%	Order	No.				Codes	Wgts	Codes	Wgts
87.5		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Chance Repro- ducibility	85.2	I	18	263	30.8	32	Y	8	? + N	0
Coef. of Consisten- cy	0.14	II	103	353	41.5	45	N	4	Y + ?	0
		III	76	444	51.9	42	N	2	Y + ?	0
		IV	48	554	64.8	38	N	1	Y + ?	0

Scale Type	Unique Score	Pattern of Re- sponses	re quency Subtotals	Fre- quency	Cri- terion I	Errors	Positive Frequencies				Negative Frequencies			
							I	II	III	IV	I	II	III	IV
(52)	(10)	(51)	(12)	(11)	(53)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
4	15	++++		66	○	0								
	14	+++ -	178	20	○	1								1
	13	++ - +	(20.8 %)	38	○	1							1	
	11	+ - + +		42	○	1	1/2				1/2			
	12	+ + - -		12	○	2	1/2	1/2				1/2	1/2	
3	7	- + + +	176	92	○	0								
	6	- + + -	(20.6 %)	29	○	1								1
	5	- + - +		55	×	1		1/2					1/2	
2	3	- - + +	195	103	○	0								
	2	- - + -	(22.8 %)	62	×	1			1/2					1/2
	10	+ - + -		30	○	2	2/3	1/3			1/3		2/3	
1	1	- - - +	158	130	○	0								
	9	+ - - +	(18.5%)	28	○	1	1							
0	0	- - - -	148	80	○	0								
	8	+ - - -	(17.3 %)	27	○	1	1							
	4	- + - -		41	○	1		1						
Total			(23) 855	(22) 855										

Computation of Error

	Total	In Positive Categories				In Negative Categories			
		(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)
Frequency of Error	426	102	74.5	41	0	0	31	34.5	143
No. Responses Involved	3420	263	353	444	554	592	502	411	301
Criterion II	(54)	○	○	○	○	○	○	○	○
Total Errors by Question		(42)	(43)	(44)	(45)				
		102.0	105.5	75.5	143.0				
Percentage of Error	(46)	(47)	(48)	(49)	(50)				
	12.5	11.9	12.3	8.8	16.7				
Criterion III & IV (55)	×	○	○	○	×				

- 1) 信頼性の高いこと。
- 2) 妥当性のあること。
- 3) 一次元性を持つこと。
- 4) 実用性に富むこと。

などがあげられる。Guttman の scale も上にあげたいいくつかの条件を満たそうと試みられたものの一つであるが、実際にそうであろうか。

Guttman scale の方が non-Guttman scale に比べて上にあげた細かい点で本当に優れた test であるといえるかどうか。それを実験的に比較しようとするのが本節の狙いである。

(1) 一次元性

scale A の再現性係数は92.7%であって、scale B の再現性係数 87.5% に比べて高い。したがって scale A では scale type から各項目の反応を予測すれば92.7%の割合でその予測はあたる。しかし scale B でも87.5%はあたるわけでその再現性もそれほど低いとは思われない。したがって一次元の有無を調べるには再現性係数だけでは不十分である。

Festinger [10] はもし周辺分布が一定という条件の下で、各項目が全く独立に反応するときには、各反応 pattern の出現頻数の理論的推定値を計算することができるといっている。それと与えられた反応 pattern とのくいちがいを χ^2 検定か何かで検定し、chance reproducibility によって一次元性をもたないのにもったと考える危険性をのぞこうとしている。

筆者が各項目が独立であると仮定して、その場合に起り得る chance reproducibility の期待値を計算すると scale A で 85.2%, scale B で 85.4%である。これから考えても、全く独立な項目でも再現性係数は高くなり得、再現性係数は一次元性の基準として不十分であることがわかる。ところで実測値とここで計算された期待値とのくいちがいを χ^2 の適合度検定により調べると、scale A も scale B も 1% level で適合性が棄却される。すなわち scale B でも各項目は全く独立ではないのである。しかしそれだからといって scale B を Guttman scale とみなすわけにはい

かない。なぜなら scale A は Ford の 4 つの基準 [11] を満たすのに対し、scale B は満たさないからである。結局 Festinger の方法は Guttman scale を否定するにはよいが、認めるには少し甘い方法だと結論される。

再現性係数に代わるいま一つの方法は Loevinger のあげる等質性指数 (index of homogeneity) である。その考え方は Guttman の考えと相通ずるものがあり [38], しかも彼女のあげる等質性指数は 0 ~ 1 の間を変化する。お互い全く独立な項目では 0 となり、また完全な Guttman scale では 1 となる。前記 scale をこの方法で求めると、近似値ではあるが [15], scale A では $H=0.51$, scale B では $H=0.14$ となる。scale B が等質性に欠ける不完全な scale であることはこれで明瞭である。しかし scale A の指数はこれで非常に高いというわけにはいかない。scale が強度の等質性を要求する場合には等質性指数を measure にする方が適当と思われる。しかしどの値から上を Guttman scale とみなし、どれから下を non-Guttman scale とみなすか基準ははっきりしない。

最近、Guttman の再現性係数の欠点を補うものとして、Green は、均斉性の係数 (coefficient of consistency) をあげている [14]。これによると scale A は 0.51, scale B は 0.14 で等質性指数と同じ値をとっているがこの係数の方が Guttman の scale のために工夫されたものであり、この場合、より適切なものと思われる。この値が 0.50 以上になることがのぞましいが、scale A は辛うじてこの基準を満足している。

このように理想的な scale は一次元的であらねばならぬといっても、一次元的ということは何で評価するかということになると再現性係数、等質性指数、均斉性係数、いずれもその係数や指数のもつ意味にずれがあり、必ずしも同一の結果を得ない。しかしいずれにせよ、一次元性については Guttman scale はそこを目的としたのであるから、non-Guttman scale よりも一次元的であることはたしかである。

(2) 信頼性*

つぎの問題は **Guttman scale** は信頼性も高いかという問題である。信頼性の問題は種々の観点からみることができる [15,16,17]。

A) internal consistency reliability

1) (Kuder-Richardson formula 20 による信頼性)

信頼性を推定する一ばん簡単な方法は折半法による推定である。しかしそれにはあまりにも項目数が少い。ここでは **scale A** と **scale B** の信頼性を比較する意味で少し無理ではあるが **Kuder-Richardson formula 20** [32] により両 **scale** の信頼性を算出した。**scale A**では0.69, **scale B**では0.24で, **scale A**の方がずっと高くなった。

2) (項目数と信頼性) 一般に **test** の信頼性は **test** の長さに依存する。**scale A**のような **Guttman scale** では項目数をそれほどふやすことは不可能であるが, **scale B**のようにある程度等質性を犠牲にすればふやすことが可能である。もし仮に **scale B** の信頼性が4項目で0.24であったとすると, 項目数をふやして **scale A** の0.69までの信頼性に達するには項目の数を何倍すればよいか, **Spearman-Brown** の一般公式 [17] を利用すると,

$$K = \frac{0.69(1-0.24)}{0.24(1-0.69)} \doteq 7.1$$

すなわち **Guttman-scale** のもつ信頼性は **scale B** のような **non-Guttman scale** の実に7倍分の項目の **test** の信頼性に相当する。

3) test-retest reliability

前項の信頼性は主として項目間の **internal consistency** をもとに考えた信頼性である。したがって **internal consistency** の高い **Guttman scale** で信頼性が高く推定されるのは当然ともいえる。**test-retest** 法で信

* 以下の問題を考えるとき, しばしば **Guttman scale** を **interval scale** であるかの如く, **scale type** (質的なもの) を **scale value** (量的なもの) として扱うことが多い。ここでも各 **scale type** は等間隔に並ぶとして扱うので, 多少の誤差はまぬがれない。

信頼性を調べれば違った結果が得られるかも知れない。そこで、第2次調査として都内某国立大学男子学生120名を2ヶ月間隔で retest を行った資料から、信頼性の問題を検討することにした。

1) (項目通過率の変化) 2つの scale A, B を構成するのに各項目の通過率が、それぞれ29%, 42%, 50%, 62%および31%, 42%, 52%, 65%のものをえらんだ。この比率は sample がちがえば当然変化してくるものと考えられる。しかしその変動が小さいものはそれだけ信頼性が高いことになる。仮りに通過率が上のように難しいものから易しいものへといかずに、retest で逆転しているような場合があれば、一度決めた scale type も、sample のちがうたびにちがった scale type を考えなくてはならず困ったことになる。そこで A, B 両 scale の通過率の変動性を調べるために、変動率としてつぎの measure を定義する。

$$D^2 = \sum_{i=1}^K \left(\frac{P_{2i} - P_{1i}}{\sqrt{\frac{P_{1i}(1-P_{1i})}{n}}} \right)^2 + \sum_{i,j=1}^K \left(\frac{P_{2j} - P_{1i}}{\sqrt{\frac{P_{1i}(1-P_{1i})}{n}}} \right)^2 \quad \left(\begin{array}{l} i=1,2,\dots,K \\ j=1,2,\dots,K \end{array} \right)$$

ここで P_{1i} は第1次調査での第 i 番目の項目の通過率、 P_{2i} は第2次調査での第1回目の i 番目の項目の通過率、 P_{2j} は第2次調査で retest した第2回目の第 j 番目の項目の通過率を表わす。 n は被験者数、 K は項目数で、ここでは $n=120$, $K=4$ である。

このとき scale A では $D^2=14.4$ scale B では $D^2=47.3$ となった。これをみても、やはり Guttman scale の方が各項目の困難度変化に安定性がある。それだけ信頼性が高いといえる。

2) (再現性係数) 前項において、項目の通過率の変動率は scale B の方が大きい。項目数が少いため、項目の通過率順位に変動はない。そこで第一次調査で定めた scale type をもとにして各項目の再現性を調べると scale A では第1次調査で92.7%だったのが、第2次調査では1回目92.1%、2回目92.8%になっている。scale B では第1次調査で87.5%だったのが第2次調査では1回目89.0%、2回目90.6%になっている。これで見ると再現性係数の高かったものは別の sample でもやはり高

く出、低かったものはやはり低く出ることがわかる。その意味でも Guttman scale の信頼性は高いといえる。

なお Guttman は categorical data における信頼性を定義し [19],* 再現性係数は各項目の平均信頼性の下限値を示すといっている [24]。それによれば再現性係数の高い Guttman scale はそれだけ彼のいう categorical data の信頼性が高いことになる。

3) (test-retest における各人の scale type の信頼性) 最後の問題は第1回の test で決められた scale type の人が第2回の test でも同じ scale type に属するかどうかの問題である。第2次調査で第1回目に決められた scale type と、第2回目で決められた scale type との相関を被験者 120 名について求めると、scale A では 0.60 scale B では 0.55 を得た。その差はあまり大きくなく、また全体の相関もあまり大きくないが、やはり Guttman scale の方が test-retest reliability が高い。以上の諸点を総合してみても Guttman scale はやはり non-Guttman scale より信頼性の高いすぐれた scale であることがわかる。

(3) 妥当性

妥当性には種々の意味があり、一概に云々することはできない。操作的に言えば、それはある criterion との相関であり、criterion には種々のものをとることができる。したがって一つ一つの criterion について別の妥当性が存在する [17]。

A) internal validity

これは別名 factorial validity といってよいかも知れない。Guttman scale では各項目がある属性 (attribute) について一義的であるような項目のみを集めて来ているので、その意味での妥当性は高くなる** これは因子分析的な factor validity と必ずしも同一ではないが、その意味合いは同じものである。因子分析との関係はここでは述べないが ([29] 参照)

* それは形において Spearman-Brown 公式に似ているが意味合いは異なる。

** 再現性係数が高いことは妥当性が高いための必要条件である。

いまその方法で scale A, scale B の各項目の communality を求めると, scale A では各項目について, $h^2 = .46, .49, .45, .31$ と一番高い communality を示す group に入り, 第1因子の負荷量が高い。つまり single factor を含む。また scale B では各項目について, $h^2 = .34, .22, .14, .13$ でそれほど高くない。 h^2 は factorial validity を示す measure に役立つ [15], この結果からみても Guttman scale の validity は因子分析的意味で non-Guttman scale に比べて高い。

B) external validity

1) (total score との相関) external validity は外部の何かの criterion との相関である。criterion に何をとりかは問題であるが, その一つとしてつぎのように考えることができる。scale A, scale B は Thurstone の Temperament Schedule の sociable area および emotionally stable area からえられたものである。そこで一応 scale A の母集団項目を Thurstone の Temperament Schedule で与えられている sociable area の項目群にとり, また scale B の母集団項目を emotionally stable area の項目群にとる。そうして, それぞれの領域で定められる score をもって scale A および scale B の criterion score のするのである*

こうして, scale A と scale B の妥当性係数を求めると, それぞれ0.75 および0.60を得た。

なおこれらは criterion の score の中に自分自身の score をも含んでいるので, 多少 over-estimate な相関である。これを修正して推定すると [16], scale A では0.74, scale B では0.25となり, scale A が scale B にくらべて断然高い。

* もし, これら sociable area の項目群が本当に scale A の母集団項目ならば, この領域は scale A を内包して一つの大きな Guttman scale を構成するはずであるが実際はそうではない。したがって厳密な意味で sociable area の total score をもって scale A の criterion score とするのはおがしい。ここでは便宜的のものである。

ただこれらの相関は、*criterion* が両者別々のものであり、また各項目と *criterion* との相関もそれぞれ異なるので、一概に *Guttman scale* の方が *validity* が高いとはいえない。しかし理論的に *perfect Guttman scale* では、その *scale score* とある *criterion* との単純相関が、各項目と *criterion* との重相関に等しいので〔23〕、*Guttman scale* での *scale score* と *criterion* との相関、つまり *external validity* はそれだけ *non-Guttman scale* より高くなる。

2) (self-rating と objective rating)

criterion とする変量には、その他つぎのようなものが考えられる。この *scale* を構成する各項目は被験者自身の反応をもとにしたものである。しかし一方各項目についての各人の行動を、他人が評価して、その人の *personality* を記述することも可能である。つまり他人が外からみた *rating* と各人の *self-rating* との一致が一つの妥当性の *measure* となる。この他人の *rating* を *criterion* とするためにつぎの調査を行った。

〔第3次調査〕 被験者はまえにあげたように東京都内某国立大学男子学生27名である。調査対象の27名はお互いによくその *personality* を知っているものである。各人には第1次調査と同じようなやり方で、*Thurstone Temperament Schedule* を自分自身で答えてもらう。つぎに、その27人の中から6人の評定者を選ぶ、その6人は一部の *group* に偏することなく、公平になる様に選ばれる*。被験者甲はたとえば項目「いろいろ知らない人たちと会わなければならない仕事が好きですか」について被評定者乙の行動を考え「はい」「いいえ」のいずれかに判断する。この場合「わかりません」の反応は禁ずる。6人の評定者は各項目について、本人を除く他の26人の *personality* を評定し**「はい」か「いいえ」のいずれかに反応する。そしてその各項目の各評定の後で1つ1つにその評定に対する確

* その基準は筆者の判断による。仲のよい者どうしに *group* を分けて、各 *group* から一人宛選ぶ。

** 本人の *personality* はすでに自己記述で記録されている。

信度を 5 4 3 2 1 の五段階で記入しておく。確信度もっとも大が 5，もっとも小が 1 である。

最後にその評定者が被評定者の personality 全体についてどの位理解しているかの認識度をこれまた五段階法で記入しておく。6 人の評定はそれぞれ独立に行う。最後に 6 人の評定を集め、そのうち各項目で確信度の一番低い者の資料をのぞき、5 人の資料で 3 人以上が「はい」と答えている場合、その被評定者のその項目の objective rating は「はい」とする。また 3 人以上が「いいえ」と答えている場合、その被評定者の objective rating は「いいえ」とする。そうしてさらに 5 人の評定者のうち 3 人が一致しているときその項目の評定の objectivity は 0.4，人が一致しているときの objectivity は 1，5 人が一致しているときの objectivity は 2 と記入する。

6 人の資料から 1 人を除くときに、最低確信度のものが 2 人以上いるときは、評定者の被評定に対する認識度の低いものを捨て、それでも決められない場合は評定者の評定習性（いつも 5 ばかりつけるとか、1 ばかりをつけるとかの全体の偏り）に従って、評定習性の高いものを捨てる。評定者が被評定者になる場合は 5 人分の資料しかないから、その場合は 5 人分の資料全部を利用する。

こうして各人の各項目の objective score（「はい」か「いいえ」に応じて 1 または 0 が与えられる）が決定され、それをその人の criterion measure とする。こうして 27 人の self-rating の反応と objective

第 3 表 self-rating と objectivity-rating の一致度

項 目	scale A				総 計	scale B				総 計
	I	II	III	IV		I	II	III	IV	
self-rating と objective rating の一致(実数と%)	13 (48)	15 (56)	14 (52)	17 (63)	59 (55)	14 (52)	18 (67)	15 (56)	16 (59)	63 (58)
objectivity 平均	1.00	.96	1.22	1.33	1.13	.70	.59	.67	.59	.64

rating によって決定された反応との一致度を各項目についてまとめた表が第3表である。表中 objectivity の平均とは、各項目、各人について決定された objectivity score を27人について平均したものである。

また資料数が少ないので、決定的なことははいえないが、この self-rating および objective rating により決定される scale score の相関を各々求めると scale A では0.01、scale B では0.34となった。

これで見ると scale A と scale B とは殆んど差がなく、むしろ一致の点では scale B の non-Guttman scale の方が勝っていた。それにしても self-rating による scale score と objective rating による scale score とはほとんど無相関で、この方法での Guttman scale の高妥当性の検証は失敗であった。しかし、objective score の平均は scale A の方が高く、したがって objective rating の信頼性では、やはり Guttman scale の方が高いと考えられる。

(4) 項目内容

つぎに Guttman scale と non-Guttman scale の各項目の内容をみる。その一般性はどこまであるか不明であるが、項目内容をみて感ぜられることは、より overt な行動、ないし具体的な行動を聞いている場合の方が Guttman scale を構成し易いのではないかということである。自己記述法は他人が外からみて、なかなか分らないことを記述して貰うのに都合がよいが、そのような内容の項目は自分自身もまた、正確に判断しにくい項目ではなからうか。そういう場合は項目の意味もいろいろに解釈されて等質性が保たれなく、Guttman scale を形成することが少くなるのではないかと思われる。第3次調査の objective rating において、その objectivity を決定したが、各項目の平均値は Guttman scale A の方が non-Guttman scale B よりも高い。これは Guttman scale を構成するような項目は他人からみても判断し易い項目であることを示す。しかしそれにもかかわらず前項でみた客観的判断のくいちがいはどうしたものであろうか。それはまた別に検討を要する問題である*

(5) Guttman scale の問題点

最後に Guttman scale と non-Guttman scale の比較をしていて感ぜられる Guttman scale の主な欠陥をとりまとめてあげる。

1) 項目数をふやせないこと。Guttman scale は多くの項目ではなかなか構成され難い。ここで scale A を発見するために1430個の四重表を作成したが、その中で4個以上の項目からなる Guttman scale はせいぜい数個であった。しかもその scale の score の分布が望ましい形になるのは(ここでは矩形分布)さらに少く、わずかこの例でとりあげた1個のみであった**

もし4個の項目だけでなく、もっと多数の項目からなる Guttman scale を作成しようとなると、もっと困難なことと思われる。もし仮りにそのような scale がつくれたとしても実用的には、せいぜい7個どまりではなからうか。したがってそれで作られる物指としての scale は、きわめて目の粗い物指としてしか使えない。

2) 仮りに多くの項目で Guttman scale が作成され、細かい目盛の物指が出来たとしても、それは sample が変わればすぐその構成を変え、不安定なものであると感ぜられる。**** 第一次調査で設定された Guttman scale が第2次、第3次調査では比較的そのままの形で使えたのは、それが五段階の目盛で、間隔の大きな粗い物指であったからである。

3) Guttman scale は各項目の通過率に強く規定されるといってよい。

* scale A の場合、客観的には、みんな sociable と判断している人のほとんどが、主観的にはそうでなく「人とつき合うのが苦手で嫌いだ、したがって自分は sociable でない」と感じていることによる。これは人間の personality の問題として非常に興味深いことであるが、それは Guttman scale の一般性ではなく、この項目内容に附随する特殊性と想像されるのでここではより立入った考察はしない。

** それととも、10~11頁にあげた3つの基準をすべての項目が満足するものではなかった。6ケの組合せ中1ケは iii) の基準を満たさなかった。

*** 同一母集団からの異なった sample でという意味で、sample の母集団が異質であれば問題は全く別である。

**** その例はたとえば Clark and Kriedt [2] をみよ。

各項目の通過率が高いか、低いか、その両端に近い程形成されやすい。したがって逆にいえば、**Guttman scale** を作るためには、だれもが通過するような項目とか、だれも通過できないような項目を集めるとよい、ということになる。それが極端になれば、どんな場合でも通過率が極端でさえあれば **Guttman scale** を構成することになる。**Guttman** が **reproducibility** の基準とは別に **marginal frequency** の基準を設けた所似はそこにあるわけだが〔20, 45〕, そのことのために **ranking** でいう最上部および最下部に属する人々（最高最低20%内に入る人はだめである）を **sensitive** に弁別出来なくなる。**Guttman scale** を決定するための項目選択の手がかりとしては、項目の通過率が大体30%, 45%, 55%, 70%の5段階か、30%, 40%, 50%, 60%, 70%の6段階かが適当であろうと思われる。さらに細かく分けたとしても25%, 35%, 45%, 55%, 65%, 75%の7段階が限度と思われる。

しかも、この通過率を **sampling fluctuation** によって、順位変化を来たさないだけの安定性のあるものとして推定するためには、どうしても5%以内の精度で、この通過率を推定しなければならない。それには通過率が50%附近の項目をもつ **scale** では400人程度の **sample** 数が必要である。そのため **Guttman** の **Cornell technique**〔21〕や **scalogram board technique**〔45〕による尺度解析では不安である。

4) なおこれは **Guttman scale** としての欠陥ではなく、**Guttman scale** を発見するための **technique** としての問題であるが、**Guttman** の尺度解析法では、与えられた資料が **Guttman scale** をなすかなさないかの判定は比較的すみやかにできるが、与えられた資料群の中から、どの項目をえらび出せば、それが1つの **Guttman scale** を構成するか、判定することは容易でない。すべての項目間相関を求めれば、その中からみつけることは比較的簡単であるが、相関表つまり四重表をつくるまでが大変である。しかし、四重表をつくってから、因子分析を行うときよりは、はるかに操作が簡単ですみやかに行える。それらについては〔29, 48〕を参

照されたい。

IV. 要約と結論

以上 Guttman scale と non-Guttman scale の諸特性を種々の点から考えて来たが、今まで得られた結果を表にまとめて要約にかえる。

第5表 Guttman scale と non-Guttman scale の比較
(×印は Guttman scale の方が不良とでた結果)

要 目		Guttman scale				non-Guttman scale					
項目	項目	I	II	III	IV	I	II	III	IV		
	Positive 反応の割合(%)	29.4	41.8	49.9	62.0	30.8	41.5	51.9	64.8		
scale	scale score	4	3	2	1	0	4	3	2	1	0
	の分布	22.9	16.0	20.3	13.5	27.4	20.8	20.6	22.8	18.5	17.3
	type 類数(%)	0.985				0.996					
	弁別係数 δ	0.985				0.996					
一次元性	再現性係数	92.7				>	87.5				
	chance reproducibility	85.2				=	85.4				
	等質性指数 H	0.51				>	0.14				
	均斉性指数 I	0.51				>	0.14				
信頼性	Kuder-Richardson 20	0.69				>	0.24				
	困難度変動率 D^2	14.4				<	47.3				
	再現性係数 (1回目)	92.1				>	89.0				
	再現性係数 (2回目)	92.8				>	90.6				
	scale type の retest	0.60				>	0.55				
妥当性	internal validity	大					小				
	external validity	大					小				
	total score との相関	0.75				>	0.60				
	total score との相関(修正値)	0.74				>	0.26				
	self-rating と objective rating の一致 (平均)	×		55%		<	58%				
	self-scale type と objective scale type の相関	×		0.01		<	0.34				
	objectivity score の平均	1.13				>	0.64				
	Guttman scale の欠陥	1. 実在性小, 2. 項目少, 3. 目盛の sensitivity 小 4. 両端が弁別できない, 5. 作成の労力大,									

これらの結果から結論的にいえることは、

(1) Guttman scale は一次元性の面から考えて non-Guttman scale よりもすぐれていると考えられる。しかし一次元性を積極的に認めるに十分な measure をいまだわれわれはもたない。

(2) Guttman scale は信頼性の面から考えて、K-R 法、test-retest 法、いずれも場合も non-Guttman scale に比べて優れている。Guttman scale では項目数をふやして信頼性を高めることが容易にできないが、この例では、Guttman scale 4項目は、non-Guttman scale 28項目分に相当することをみた。よって少数項目で信頼性の高い scale を構成するには、Guttman scale は良好である。

(3) Guttman scale は internal validity の面で non-Guttman scale よりも優れている。しかしそれは必ずしも因子分析的な共通因子を1つ含むという意味ではない。両者は操作的に別物をつくる。

(4) external validity つまり、outside criterion との相関においては、理論的には Guttman scale の方が non-Guttman scale より高い。しかし実験的には、それを裏付ける結果と否定する結果とが得られ、簡単に結論できない。実験方法の不備、材料の不適切さ等も考えられる。

(5) 実際的に Guttman scale は項目数をふやせない。したがって、被験者集団を sharp に弁別する scale としては不適當である。せいぜい5～7段階に分類すれば十分のとき、しかも、集団の上下両端を細かく区別する必要のないとき用いられる。

(6) Guttman scale 作成のための労力は通常の item analysis や G-P 分析に比べて大である。しかし、因子分析よりは小である。予備項目は破棄を見込んで十分とる必要（少くも数十個以上）があるが、それで残った Guttman scale の内容が、もとの母集団項目に比べて、どういう意味をもつかは疑問である。

(7) Guttman scale を構成し易い項目は、overt で具体的な行動、ないしは、事実を訊く項目と思われる。しかし、それは確定的ではない。

(8) sample-to-sample stability の高い Guttman scale をつくる

には、項目数を少なくするか、被験者数を多くするかが必要である。4項目の scale をつくるに400人位の被験者が必要と思われる。

(9) 一度 scale を作成してしまったあとの利用面では、Guttman scale が non-Guttman scale よりも、採点その他で不便ということはさほど考えられない。むしろ、既成 scale が、Guttman scale をなすか否かの検討は容易であるから、のちになって現在使用している test が、Guttman scale として、有効であるか否かの check に役立ち、不良 test を誤まって使用しつづける危険性が少い。通常の test ではその検討法をもたない。

(10) ここでは検討しなかったが、Guttman scale は累加的性質を持っているから、発達検査、能力診断検査としての有効な利用が考えられる。

(本学非常勤助手)

V. References

- [1] Bennett, J. F. Multidimensional Unfolding : Determining the Dimensionality of Ranked Preference Data. *Psychomatika*, 1960, 25, 27—43
- [2] Clark, K. E. and Kriedt, P. H. An application of Guttman's new scaling techniques to an attitude questionnaires. *Educ. Psychol. Measmt.*, 1948, 8, 215—223.
- [3] Coombs, C. H. and Kao, R. C. Nonmetric factor analysis. *Engineering Research Bull.*, No.38, 1955.
- [4] Coombs, C. H. and Kao, R. C. On a connection between factor analysis and multidimensional unfolding. *Psychometika*, 1960, 25, 219—231.
- [5] Edwards, A. L. On Guttman's scale analysis. *Educ. Psychol. Measmt.*, 1948, 8, 314—318.
- [6] Edwards, A. L. A technique for increasing the reproducibility of cumulative attitude scales. *J. Appl. Psychol.*, 1956, 40, 263—265.

- [7] Edwards, A. L. and Kilpatrick, F. P. Scale analysis and the measurement of social attitudes. *Psychometrika*, 1948, 13, 99—114.
- [8] Edwards, A. L. and Kilpatrick, F. P. A technique for the construction of attitude scales. *J. Appl. Psychol.*, 1948, 32, 374—384.
- [9] Ferguson, G. A. On the theory of test discrimination. *Psychometrika*, 1949, 14, 61—68.
- [10] Festinger, L. The treatment of qualitative data by scale analysis. *Psychol. Bull.*, 1947, 44, 149—161.
- [11] Ford, R. N. A rapid scoring procedure for scaling attitude questions. In M. W. Riley et al. *Sociological Studies in Scale Analysis*, New Brunswick, Rutgers Univ. Press., 1954.
なお、これは *Publ. Opin. Quart.*, 1950, 14, 507—532. より転載のもの。
- [12] Goodenough, W. H. A technique for scale analysis. *Educ. Psychol. Measmt.*, 1944, 4, 179—190.
- [13] Green, B. F. Attitude measurement. In G. Lindzey (ed.) *Handbook of Social Psychology*, Cambridge, Addison-Wesley Publishing Co., 1954.
- [14] Green, B. F. A method of scalogram analysis using summary statistics. *Psychometrika*. 1956, 21, 79—88.
- [15] Guilford, J. P. *Psychometric Method*. 2nd ed., New York, McGraw-Hill, 1954.
- [16] Guilford, J. P. *Fundamental Statistics in Psychology and Education*. 3rd ed., New York, McGraw-Hill, 1956.
- [17] Gulliksen, H. *Theory of Mental Tests*. New York, John Wiley and Sons, Inc., 1950.

- [18] Guttman, L. A basis for scaling qualitative data. *Amer. Sociol. Rev.*, 1944, 9, 139—150.
- [19] Guttman, L. The test-retest reliability of qualitative data. *Psychometrika*, 1946, 11, 81—95.
- [20] Guttman, L. On Festinger's evaluation of scale analysis. *Psychol. Bull.*, 1947, 44, 451—465.
- [21] Guttman, L. The Cornell technique for scale and intensity analysis. *Educ. Psychol. Measmt.*, 1947, 7, 247—279.
- [22] Guttman, L. The basis for scalogram analysis. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [23] Guttman, L. Relation of scalogram analysis to other techniques. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [24] Guttman, L. Problems of Reliability. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [25] Guttman, L. The principal components of scale analysis. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [26] Guttman, L. Image theory for the structure of quantitative variates. *Psychometrika*, 1953, 18, 277—296.
- [27] Guttman, L. The principal components of scalable attitudes. In P. F. Lazarsfeld (ed.) *Mathematical thinking in the social sciences*. Glencoe, the Free Press, 1954.
- [28] Guttman, L. The Israel alpha technique for scale analysis. In M. W. Riley et al. *Sociological Studies in Scale Analysis*. New Brunswick, Rutgers Univ. Press, 1954.
- [29] 池田 央 Personality Questionnaires における Guttman scale

の研究, 東京大学教育学部修士論文, 1958.

- [30] Jackson, J. M. A simple and more rigorous technique for scale analysis. In a manual of scale analysis Part II. Montreal, McGill Univ. 1949., (mimeograph)
- [31] Kahn, L. H. and Bodine, A. J. Guttman scale analysis by means of I.B.M. equipment. *Educ. Psychol. Measmt.*, 1951. 11, 298—314.
- [32] Kuder, G. F. and Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, 2, 151—160.
- [33] Lazarsfeld, P. F. The logic and mathematical foundation of latent structure analysis. In S. A. Stoffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [34] Lazarsfeld, P. F. The interpretation and computation of some latent structures. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press. 1950.
- [35] Lazarsfeld, P. Conceptual introduction to latent structure analysis. In P. F. Lazarsfeld et al. *Mathematical thinking in the social sciences*. Glencoe, the Free Press, 1954.
- [36] Likert, R. A. A technique for the measurement of attitude scales. *Arch. Psychol.*, N. Y., 1932. No. 140.
- [37] Loevinger, J. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, 61, No. 285.
- [38] Loevinger, J. The technic of homogeneous tests compared with some aspects of "Scale Analysis" and "Factor Analysis". *Psychol. Bull.*, 1948, 45, 507—529.
- [39] Milholland, J. E. *Dimensionality of Response Patterns*. University of Michigan, 1953.

- [40] Noland, E. W. Worker attitude and industrial absenteeism; a statistical approach. *Amer. Social. Rev.*, 1945, 10, 503—510.
- [41] Riley, M. W. et al. *Sociological Studies in Scale Analysis*. New Brunswick, Rutgers Univ. Press., 1945.
- [42] Stevens, S.S. Mathematics, measurement, and psychophysics. In S. S. Stevens (ed.) *Handbook of experimental psychology*. New York, John Wiley and Sons, Inc., 1951.
- [43] Stouffer, S. A. et al. *Measurement and Prediction*. Vol. IV. The Studies in Social Psychology in World War II. Princeton Univ. Press., 1950.
- [44] Stouffer, S. A., Borgatta, E. F., Hays, D. G. and Henry, A. F. A technique for improving cumulative scales. In M. W. Riley et al. *Sociological Studies in Scale Analysis*. New Brunswick, Rutgers Univ. Press, 1944.
なおこれは *Publ. Opin. Quart.* 1952, 16, 273—291. より転載。
- [45] Suchman, E. A. The scalogram board technique. In S. A. Stouffer et al. *Measurement and Prediction*. Princeton Univ. Press., 1950.
- [46] 高木貞二, 心理学における数量化の研究, 東京大学出版会, 1955.
- [47] Thurstone, L. L. and Chave, E. J. *The Measurement of Attitude*. Chicago, Univ. of Chicago Press, 1929.
- [48] Toby, J. and Toby, M. L. A method of selecting dichotomous items by cross-tabulation. In M. W. Riley et al. *Sociological Studies in Scale Analysis*. New Brunswick, Rutgers Univ. Press, 1954.
- [49] Toby, J., Riley, M.W., Fine, P. and Moore, M. Techniques for the improvement of object scales. In M. W. Riley et al. *Sociological Studies in Scale Analysis*. New Brunswick, Rutgers

Univ. Press, 1954.

- [50] Torgerson, W. S. *Theory and Methods of Scaling*. John Wiley and Sons, Inc., 1958.
- [51] Walker, D. A. Answer-pattern and score-scatter in tests and examinations. *Brit. J. Psychol.*, 1931, 22, 73—86 ; 1936, 26, 301—308 ; 1940, 30, 248—260.
- [52] White, B. W. and Saltz, E. Measurement of reproducibility. *Psychol. Bull.*, 1957, 54, 81—99.

A Study for the Evaluation of Guttman's Scale

(English Résumé)

Hiroshi Ikeda

This study is designed for the evaluation of Guttman's scale from the practical point of view. A typical Guttman's and a non-Guttman's scale are constructed from the several items of Thurstone's Temperament Schedule which were administered to 855 male students of 4 colleges in Tokyo. Both scales are compared and evaluated in respects of unidimensionality, reliability, validity and practicability.

The result of the study is summarized as follows :

1. We can conclude that Guttman's scale is superior to non-Guttman's scale in unidimensionality, reliability and internal validity. Both K-R and test-retest reliability of Guttman's scale are higher than those of non-Guttman's. In this experiment we see Guttman's scale is seven times as powerful as non-Guttman's in relation between reliability and the number of test items.
2. Correlation between G's scale and outside criterion is lower than that of non-G's scale, which may mean inadequacy of the experimental design.
3. G's scale is inadequate for the sharp discrimination of the subject group, especially extreme individuals in the group, because it is very difficult to construct G's scale which has a large number of items.
4. Construction of G's scale is more laborious than usual item analysis or G-P analysis, but less than factor analysis.
5. In order to construct G's scale easily, it seems to be wise to provide many question items inquiring overt and concrete behaviors of the subjects.
6. In order to get G's scale which has high sample-to-sample stability, we need either decrease the number of items or increase the number of samples. At least 400 samples are needed to construct G's scale of four items.
7. G's scale, compared with usual scale, is not always inconvenient in practical use, i. e. test administration and scoring. Rather, it has the advantage of checking of applicability of the scale to a new subject group.