

インクブロットテストの Form Definiteness 評定の尺度化

東京大学教育情報科学研究室 松 井 仁

Scaling technique on Form Definiteness of the Inkblot test

Hitoshi MATSUI

— Abstract —

In this research, a new scaling procedure was applied in order to measure personality traits (θ) on Form Definiteness of the Inkblot test. This procedure was based on the item response theory and the graded response model was applied. The criterion in terms of Form Definiteness in this research was referred to the Holtzman Inkblot Technique. The scores of Form Definiteness were originally scored on a 3-point scale. 43 homogenous items were selected by the principal factor method and the item parameters were calculated by the heuristic method. And the estimates of θ were calculated by the maximum likelihood method. Then the checking model assumption and expected model features was discussed.

目 次

1. 問 題
2. 調査と被験者
3. 評定カテゴリー
4. 項目分析
 - 4-1 主因子解による等質な項目の抽出
 - 4-2 α 係数の算出
 - 4-3 項目の反応分布
5. 項目反応モデルの導入
 - 5-1 モデルの仮定
 - 5-2 項目特性関数
 - 5-3 項目パラメタの推定
 - 5-4 モデルの当てはまり
6. まとめ
7. 参考文献

1. 問 題

ロールシャッハテストに代表されるインクブロットテストは、伝統的に心理診断・治療の分野で大きな役割を果たしてきている。しかし、このテストの尺度化に関しての信頼性・妥当性の問題は、他の人格検査と比べて未解決な面が多い。特に、尺度としての信頼性の検討について言えば、評定カテゴリーについて各図版で測定しているものが本当に被験者の同じ潜在特性を測定しているものなのか・どれだけ安定した測定なのか...といった項目の等質性からの議論は、国内においては皆無である。

議論の伸展しない原因として様々なものが考えられるが、能力テストの構造と比較してみるとインクブロットテストの持つ次の特徴があげられよう。それはインクブロットテストでの刺激図版を1項目としてとらえる（以下、本論ではインクブロットテストの刺激図版を項目と呼ぶことにする）と、インクブロットテストの中の代表ともいえるロールシャッハテストが、“たった”10項目から構成されているテストであるという様に、項目数が非常に少ないことがあげられる。能力テストの枠組内で発展してきた信頼性に関する議論は、基本的に十分な項目数と十分な被験者数が得られた上での尺度化を前提としている側面がある。このようなことから研究者は、最初からインクブロットテストを項目の等質性の議論の対象外として来たと考えられる。

また、インクブロットテストのスコアの得点化について考えると次のような問題も指摘できよう。このテストのスコアの集計法は各項目の各反応カテゴリーに対する正反応数（または得点）の単純加算にその基礎をおいている。この集計法は、古典的テストモデルにその発想の基礎があるといえる。一般に古典的テストモデルでの得点は、尺度に含まれる項目数・項目の困難度に依存する。よって、現状では異なる項目（刺激図版）を利用して再テストを実施してもその尺度値を比較することは不可能であり、心理治療の効果を測定する目的で再テストする場合も同じ項目を使わざるをえない。その場合、インクブロットテストでのこのような再テストは学習効果の影響で治療効果を把握する目的としては問題が生じる。理

想としては、異なる項目での再テストが構成されて、治療前のテストとその尺度値の比較が可能であることが望まれる。しかし、このことは古典的テストモデルの枠内で作成された尺度では不可能である。

この様な状況の中で、インクプロットテストの尺度値についての研究方法としては次の2つの方向性があるだろう。一つは、既存のインクプロットテストなどの実施法に見合う信頼性の評価方法と集計方法を開拓することである。もう一つは、実施法は伝統的なものから離れても、“インクのしみ”という刺激を使って研究者が被験者から得たい情報をより効率的に得る実施・尺度化の手法を、既存の集計法を応用することによって開拓することであろう。

以前に筆者は、筋肉運動反応（通常のロールシャッハテストで言えば、人間運動反応（M）と動物反応（FM）を合わせた評定カテゴリー）の尺度化について検討した（松井，1991）。その手続き上、筋肉運動反応の評定カテゴリーに対して等質な項目の抽出を試みた。そして、項目反応モデルを適用した尺度化を抽出した項目を利用して試みた。この研究で利用したデータは、集団式で得られたものであるため、通常のロールシャッハテストの様に面接法で得られたものと質が異なる。また、反応を得る際の教示も1項目について複数の反応を許す通常の手法と異なり、「1項目につき1つずつの反応する」と総反応数を制限する形で実施した。よって、この一連の研究で得られた尺度値の特性や、項目（図版）の特性についての議論を通常のロールシャッハテスト等に一般化することはできない。しかし、“インクのしみ”という刺激から得られる個人の特性の測定と尺度化の可能性については検討できるであろう。このような研究の方向性を確認した上で、筋肉運動反応に引き続いて本論ではインクプロットテストの形態水準に関する尺度化の可能性について検討する。

2. 調査と被験者

ここで利用するデータは、松井（1991）で利用した集団式の質問紙形式によって得られたものと同じものを利用する。

質問紙として、ロールシャッハテストの10項目とHIT（Holtzman Inkblot Techniqueの略、Holtzmanら1961）からの50項目の合計60項目を130mm×90mmに縮小印刷した冊子を作成した。教示は、各項目に対して「何に似ていると思うか」と言う形で行われ、被験者は1項目につき1つずつ自由記述で反応する。このような方法で

実施したところ、60項目に対しておよそ1時間で回答を得ることができた。被験者は、18才から25才までの大学生・専門学校生であった。調査の結果、時間内に最後の項目まで達しなかったものを除いて402人のデータを得た。

3. 評定カテゴリー

本論では、インクプロットテストを「被験者が、あいまいな“インクのしみ”の刺激（項目）からどれだけ具体的に/明細化の度合いが高い内容の反応を創り出せるのか？」という課題を暗黙のうちに課しているものと考えた。そして、この課題に対処する能力を測定する道具としてこのテストを捉え直すことにした。

インクプロットテストの課題に対して、項目（刺激図版）と完全に一致する反応内容はおそらく「（インクのしみ）であろう。しかし、このような反応内容に対する検査者の評価は否定的なものであり、被験者も通常このような反応は示さない。インクプロットテスト場面においては、被験者はより想像力を働かせて何か具体的なものを反応しようと努力しているものと考えられる。この様に考えると、項目に対して被験者が反応を示さないことは、「被験者にとって、その項目は具体的に/明細化の度合いが高い内容の反応を創り出すには難しすぎたため、反応を創出できなかった」と解釈できよう。逆に、被験者のより高い能力の反映として、具体的に形の定義され・より高度な明細化が期待される反応内容の創出が見られるものと考えた。このようにして、被験者の反応内容に対して表1の基準にしたがって3段階評定を実施し、その評定値 y_i （以下本論では、項目 i に対する評定値を y_i で示す）を本論での Form Definiteness 評定得点と定義する。

この y_i は、HITの評定カテゴリーとの対応で言うところのようになる。 $y_i = 0$ は、Rejectionに当てはまる被験者の反応に当てはまる。 $y_i = 1$ は、Holtzmanでの反応内容を構成する概念の具体性を5段階で評価する評定カテゴリー Form Definiteness（以下 Holtzmanのものについては、略してFDとして本論の評定カテゴリーと区別する）のFD=0からFD=2までの評定値と一致するものである。これは、反応内容が非形態もしくは漠然とした形態認知しか期待できないものである。漠然とした形態認知とは、図版と反応内容との形態の適合の努力をさほど被験者がしなくとも、大まかにその反応内容が当てはまってしまうような認知を示す。これは、Klopperの言う漠然反応・片口で言えば許容水準（-+）レベルまでの

反応と対応している。具体的には、「火」「雲」「獣」といった反応内容があげられる。 $y_i = 2$ は、 $y_i = 1$ の反応内容からより明細化の進んだ認知が期待されるものである。Holtzmanの評定で言えば、FD=3からFD=4の評定と一致するものである。また、片口では良好水準(+1)以上が得られると期待される反応と対応する。具体例としては、「ドレスを着て踊っている少女」「エッフェル塔」などがあげられる。

この様に Form Definiteness 評定は、ロールシャッハテストの評定カテゴリの中で形態水準の評定と関係があるものと捉えられよう。ただし、本研究で得たデータは集団式の実施法で得たものであるため、通常の間接法で行われる「質問段階」がない。つまり、被験者の反応と図版との適合性については十分な情報を得ることができない。よって、この研究の結果をそのまま一般の形態水準評定の問題に置き換えることはできない。ただ、いずれの立場の形態水準の評定でも、まず反応内容の具体性/明細化の可能性によって大まかな水準が決められ(第1段階とする)、その後図版と反応の対応で評定の修正が加えられる手続き(第2段階とする)をとっている。例えば HIT では、この2つの段階をFD評定とFA(Form Appropriate)評定に分化させているが、他の方法では分化していない。この様な中で、本論での Form Definiteness 評定は、形態水準の評価過程の第1段階に関係する

尺度化と位置付けて議論を進めることは可能であろう。

4. 項目分析

この尺度化に利用する項目を抽出するために以下の手続きをとった。

4-1 主因子解による等質な項目の抽出

等質な項目を抽出する手がかりを得るために、項目間の相関係数を求め、主因子解を求めた。主因子解を求める時の共通性の推定値として、各項目の重相関係数の二乗を当てた。この主因子解によって得られた固有値は、第1因子から順に 7.38, 1.23, 1.00, ...であった。各項目の共通性の推定値の合計は 15.32であったため、この第1因子だけで共通性の総和の48%を説明していることが示された。本論では後に述べるように項目の等質性を前提とした尺度化のモデルを適用するため、この中で項目の因子負荷量が 0.30以上の43項目を尺度化の対象項目として選択した。選択された項目の因子負荷量は表2の r_i に示した通りである。

また、解釈可能な因子軸の存在の有無を確認するために、念のため2因子群・3因子群についてバリマックス回転した結果を見てみたが、その項目群の特徴に特に共通するものは得られなかった。

表1：評定カテゴリ

本論文での評定値 y_i	各評定値での被験者の状態 (能力)	HIT での評定	具体例
$y_i = 0$	具体的な反応内容を創出するには被験者にとって項目が困難なため何等反応を創り出すことができなかった。	Reject	
$y_i = 1$	反応は創り出せたが、具体性の高い内容に至ることができなかった。従って、図版との対応が不明確であっても、まずまずその内容に当てはまることが予想される内容の反応が創り出された。	FD=0~2	バクテリア・地獄・抽象画 獣・眼・原爆雲 人・蛇・きのこ 等
$y_i = 2$	具体性の高い内容を持つ反応を創り出すことに成功した。	FD=3・4	踊っている少女・メリーゴーランド等

表 2：項目パラメータ一覧

項目通し 番 号	オリジナル 項 目	因子負 荷量 r_i	$P_{y_i=1}$	$P_{y_i=2}$	項目パラメタ			最適化スコア			順序の 整合性
					a_i	$b_{y_i=1}$	$b_{y_i=2}$	$y_i=0$	$y_i=1$	$y_i=2$	
H02	A36	0.34	0.17	0.79	0.36	-2.74	2.41	1.78	-0.22	-0.83	ok
H03	A40	0.33	0.21	0.90	0.35	-2.36	3.96	1.70	-0.38	-1.05	ok
H05	A08	0.35	0.21	0.86	0.38	-2.19	3.05	1.75	-0.48	-0.47	
H10	A17	0.40	0.35	0.76	0.44	-0.93	1.81	1.31	-0.71	-0.72	ok
H12	A18	0.33	0.12	0.69	0.35	-3.40	1.52	1.94	-0.16	-0.50	ok
H14	A23	0.35	0.42	0.88	0.38	-0.56	3.38	1.24	-0.91	-0.85	
H15	A24	0.35	0.38	0.86	0.38	-0.80	3.09	1.24	-0.84	-0.58	
H18*	R06	0.32	0.22	0.80	0.34	-2.33	2.62	1.65	-0.53	-0.33	
H19	A28	0.34	0.22	0.72	0.37	-2.20	1.68	1.73	-0.41	-0.63	ok
H20	A31	0.41	0.29	0.89	0.45	-1.30	2.99	1.51	-0.60	-0.78	ok
H21	B09	0.39	0.22	0.80	0.42	-1.94	2.20	1.84	-0.52	-0.57	ok
H22	A03	0.31	0.23	0.88	0.33	-2.31	3.82	1.52	-0.38	-0.90	ok
H23	A06	0.41	0.36	0.80	0.45	-0.84	2.07	1.41	-0.87	-0.66	
H24*	R04	0.39	0.21	0.88	0.43	-2.01	3.04	1.94	-0.44	-0.98	ok
H26	A04	0.35	0.15	0.86	0.38	-2.84	3.08	2.48	-0.35	-0.97	ok
H27	B43	0.38	0.21	0.82	0.41	-2.06	2.45	1.70	-0.34	-0.92	ok
H28	B15	0.47	0.35	0.86	0.54	-0.75	2.25	1.49	-0.74	-1.15	ok
H31	A45	0.38	0.12	0.72	0.41	-2.97	1.57	2.37	-0.20	-0.64	ok
H32	B03	0.37	0.24	0.87	0.40	-1.86	3.13	1.51	-0.47	-0.57	ok
H34	B04	0.36	0.25	0.72	0.39	-1.83	1.62	1.46	-0.46	-0.53	ok
H35	R03	0.30	0.02	0.81	0.32	-6.42	2.92	3.91	0.06	-0.81	ok
H36	B06	0.40	0.33	0.91	0.43	-1.04	3.47	1.34	-6.61	-0.21	ok
H37	B11	0.46	0.42	0.91	0.52	-0.41	2.96	1.38	-0.95	-1.37	ok
H38	B08	0.45	0.26	0.88	0.51	-1.35	2.64	1.99	-0.74	-0.65	
H39	A30	0.35	0.21	0.97	0.38	-2.25	5.60	1.95	-0.46	-2.43	ok
H42*	R07	0.41	0.16	0.64	0.46	-2.35	0.90	2.45	-0.50	-0.42	
H43	B12	0.42	0.23	0.93	0.46	-1.72	3.66	1.92	-0.49	-1.63	ok
H44	B14	0.39	0.18	0.68	0.42	-2.32	1.21	1.89	-0.25	-0.67	ok
H45	B13	0.35	0.24	0.95	0.37	-1.99	4.81	1.57	-0.44	-1.32	ok
H46	A16	0.31	0.12	0.90	0.32	-3.74	4.23	2.11	-0.21	-0.94	ok
H47	B16	0.36	0.37	0.92	0.38	-0.89	4.02	1.25	-0.70	-1.09	ok
H48*	R08	0.42	0.32	0.79	0.47	-1.07	1.91	1.73	-1.03	-0.36	
H49	B18	0.44	0.39	0.90	0.49	-0.61	2.88	1.38	-0.90	-0.87	
H51	B22	0.31	0.19	0.81	0.32	-2.70	2.82	1.80	-0.66	-0.25	
H52	B26	0.35	0.17	0.64	0.37	-2.61	1.04	1.88	-0.19	-0.69	ok
H53	B27	0.43	0.23	0.64	0.48	-1.67	0.85	1.85	-0.30	-0.85	ok
H54*	R09	0.40	0.38	0.87	0.44	-0.74	2.88	1.29	-0.76	-0.95	ok
H55	B21	0.48	0.29	0.83	0.55	-1.12	2.00	1.85	-0.66	-1.11	ok
H56	B35	0.41	0.39	0.82	0.45	-0.66	2.22	1.31	-0.85	-0.82	ok
H57	B40	0.39	0.23	0.80	0.42	-1.86	2.14	1.86	-0.48	-0.78	ok
H58	B42	0.31	0.18	0.98	0.32	-2.84	7.02	1.91	-0.43	-0.78	ok
H59	A39	0.45	0.33	0.84	0.50	-0.95	2.20	1.61	-0.82	-0.81	
H60	R10	0.43	0.25	0.86	0.47	-1.55	2.53	1.99	-0.65	-0.74	ok

(注) オリジナル項目は、“A”，“B”のつくものが HIT のそれぞれ Form A, Form B からの引用，
“R”のつくものがロールシャッハテストからの引用である。

4-2 α 係数の算出

主因子解によって抽出された43項目の等質性を確認するために、その指標として α 係数を算出した。その結果、0.87と言う値が得られた。これより、この43項目でまずまずの等質性を持っていると考えることができる。よって、一応この評定カテゴリーの1因子性を認めた上で以下の尺度化を進める。

4-3 項目の反応分布

今回の評定カテゴリーに対する各項目の反応分布を示したのが表2の $p_{y_i=1}$, $p_{y_i=2}$ である。

$p_{y_i=1}$ は、 $y_i=1$ に満たない、つまり $y_i=0$ の評定カテゴリーに分類される反応をした被験者の比率である。この値が、低すぎる項目は反応を創出することが容易すぎる項目であるといえる。実際にこれが5%を下回る項目は、H35の1項目だけであった。 $p_{y_i=2}$ は、 $y_i=2$ に満たない反応を示した。つまり $y=0$ と $y=1$ のレベルの反応を示した被験者の累積比率である。 $p_{y_i=1}$ の場合と逆に、この値が高すぎる項目は、 $y_i=2$ レベルの具体性・明細化を創出することが困難な項目と考えられる。実際にこれが95%を上回った項目は、H39, H58の2項目であった。全体としては、 $p_{y_i=2}$ が高い方に偏り気味ではあるが、 $p_{y_i=1}$, $p_{y_i=2}$ 共に極端に偏った散らばりを見せていないため、Form Definiteness 評定に関する項目として前項で選ばれた43項目すべてを採用することにした。

5. 項目反応モデルの導入

前項で抽出された43項目に対する被験者の反応パターンから、項目の識別力・困難度の特性を考慮した尺度値を得るために項目反応モデルを導入する。ただし、ここで扱うデータは、項目反応モデルで通常利用される二値データ（以下、二値データによる項目反応モデルを binary model と称する）とは異なり、段階づけられたものである。よって、Samejima (1969) に示される多段階モデルでの項目反応モデル（以下、これを graded response model と称する）を利用することにする。以下、Samejima に従って項目パラメタ・被験者の特性パラメタを簡便法 (heuristic method) によって算出する手続きを紹介する。

5-1 モデルの仮定

項目 i に対して、「具体的/明細化の度合いが高い反応を持つ反応」を創出させる被験者の心的機能の強弱を示す

連続量を x_i とする。すると、graded response model での項目反応モデルは、以下の5つの仮定の上に構成されることになる。

- 仮定① x_i は、個人の特性値 θ を除けば互いに独立になっている量である（局所独立の仮定）。
- 仮定② ことなる項目 i, j についての x_i, x_j は、二変量正規分布に従う。
- 仮定③ θ と各項目の x_i は、相関係数が ρ_i の二変量正規分布に従う。
- 仮定④ 項目 i についての評定値 y_i （本論のデータは $y_i=0,1,2$ ）は、測定対象となっている心的機能を示す量 x_i の大小関係と対応している。
- 仮定⑤ 項目 i については、その中のどの評定値 y_i についても識別力は等しいとする。

これらの仮定の内、仮定①～仮定③は、binary model の場合と同じく heuristic method で解く場合の項目反応モデル共通の共通の仮定である。仮定④・仮定⑤は y_i と x_i の順序的な整合性を保つための graded response model 固有の仮定である。

ここで、個人の特性を示すパラメタ θ は、各項目に共通して測定されている潜在的な変数であると解釈できる。よって、仮定③での ρ_i は、この潜在的な変数と各項目で測定している心的機能の間の相関係数である。よって、先の主因子解で得られた各項目の因子負荷量 r_i に当てることにする。

5-2 項目特性関数

binary model の場合と同様に、項目 i について y_i をとる確率を θ の関数として $P_{y_i}(\theta)$ と設定する。これを graded response model での項目特性関数と呼ぶことにする。また、この関数を設定する手続きの都合上、項目 i について y_i 以上値を得る確率を θ の関数として $P_{y_i^*}(\theta)$ で示すことにする。

次に、項目 i の y_i の値に対応させて、評定値ごとの困難度 b_{y_i} を設定する。本論で扱うデータで y_i が実際にとる値は、0～2の整数値である。そして、もし前述の仮定④が成立するならば、被験者は必ず最低の評価値 $y_i=0$ を得ることになる。つまり、 $y_i=0$ という値を得ることの困難度は、設定された θ の範囲内で最も低い値になるはずである。ここで、 θ については正規分布が仮定されており、その範囲は実数全領域である。よって、 $b_{y_i=0}=-\infty$ と示せる。また、実際の評定値は、 $y_i=2$ までなので $y_i=3$ という値を被験者がとることはあり得ない。つまり、後の計算の便宜上 $y_i=3$ を仮定してもこれについて強いて言えば「 $y_i=3$ を得ることは無限に困難である」と解

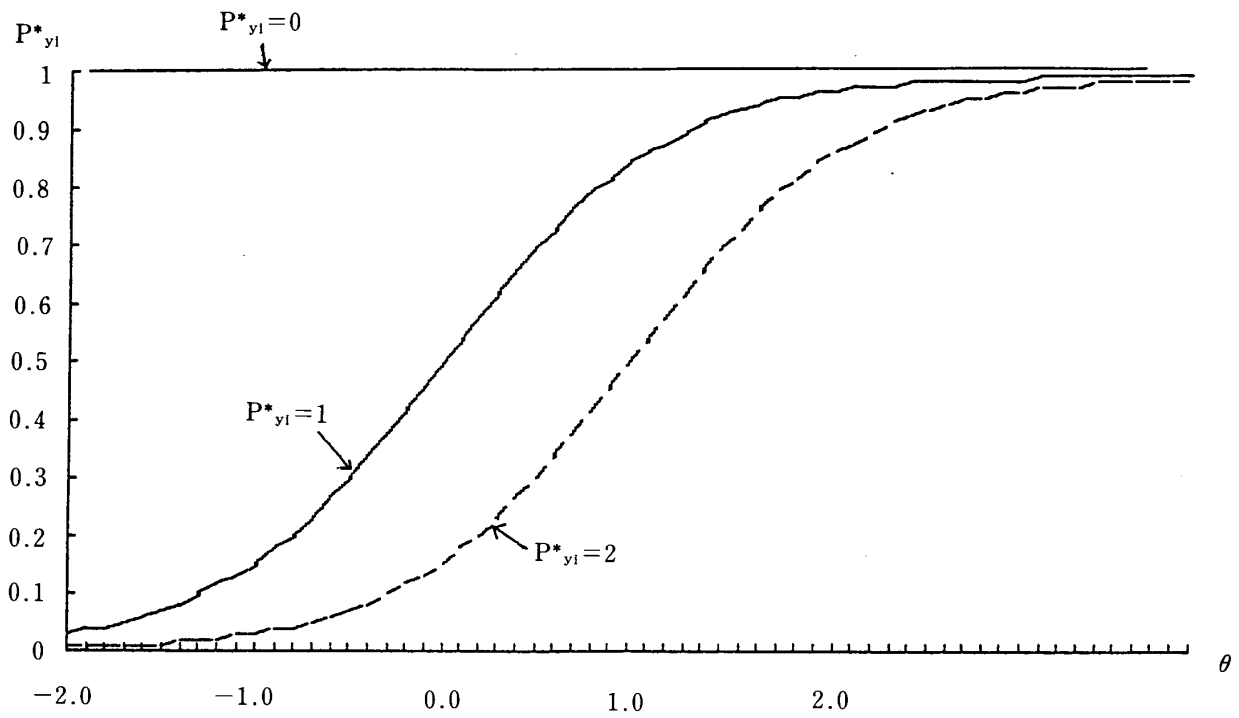


図 1 : P^*_{y1} の例

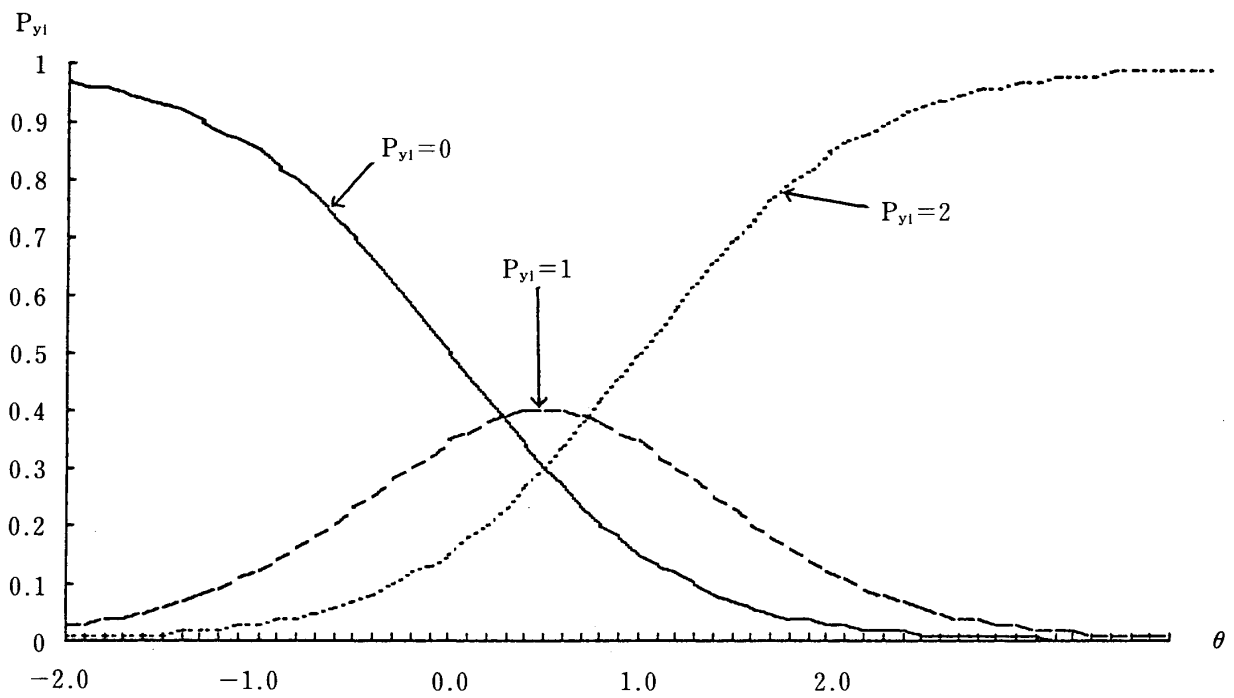


図 2 : 項目特性関数 P_{y1}

積できるため、 $b_{y_i=3} = \infty$ と書ける。この様に考えると項目 i の書く評定値 y_i の困難度には次の関係が成立する。

$$-\infty = b_{y_i=0} < b_{y_i=1} < b_{y_i=2} < b_{y_i=3} = \infty$$

ここで、各パラメタの推定値を解析的に得るために、正規分布関数をロジスティック関数に近似して扱う。 a_i を項目 i の各評定値に共通な項目識別力であるとする。すると、項目 i について評定値 y_i 以上の値をとる確率 $P_{y_i}^*(\theta)$ は、仮定⑤より共通の項目識別力 a_i を用いて、2パラメタ・ロジスティック関数として次の様に書ける。

$$P_{y_i}^*(\theta) = [1 + \exp \{-1.7a_i (\theta - b_{y_i})\}]^{-1}$$

但し $y_i = 0, 1, 2, 3$

この関数は θ についての単純増加関数である。ただし、 b_{y_i} の値の条件により、 $y_i = 0, 3$ の場合に限りこの関数の値は θ の全領域にわたって次の定数で書ける。

$$P_{y_i=0}^*(\theta) = 1, \quad P_{y_i=3}^*(\theta) = 0$$

ここで、仮定④に示した様に、評定値の順序性が保たれているならば、どの θ に関しても y_i 以上の評定値を得ることよりも $y_i + 1$ 以上の測定値を得ることが困難であるため、その確率はより小さいはずである。よって、 θ の全領域について次の関係式が成り立つ。

$$P_{y_i}^*(\theta) > P_{y_i+1}^*(\theta)$$

この様な性質を持つ $P_{y_i}^*(\theta)$ を利用すると、 y_i の値そのものをとる確率である項目特性関数 $P_{y_i}(\theta)$ は、項目 i に対する被験者の反応が y_i に達したが $y_i + 1$ には達することができなかった確率と解釈して次の様に書けるだろう。

$$P_{y_i}(\theta) = P_{y_i}^*(\theta) - P_{y_i+1}^*(\theta)$$

但し $y_i = 0, 1, 2$

この様にして得られる $P_{y_i}^*(\theta)$ 及び項目特性関数を、 $a_i = 1.0, b_{y_i=1} = 0.0, b_{y_i=2} = 1.0$ の場合を例として図示すると、図1・図2の様になる。

5-3 項目パラメタの推定

このモデルでは各項目 i について、 $a_i, b_{y_i=1}, b_{y_i=2}$ の3つの項目パラメタが必要である。これらは、標準正規分

布関数の逆関数 $\Phi^{-1}(p)$ を利用すると、それぞれ次のように書ける。

$$a_i = \frac{\rho_i}{\sqrt{1 - \rho_i^2}}, \quad b_{y_i} = \frac{\Phi^{-1}(p_{y_i})}{\rho_i}$$

但し $y_i = 1, 2$

ただし、 ρ_i は前述の様に各項目の因子負荷量 r_i をあてるため、 p_{y_i} と r_i これらより全ての項目パラメタの値が計算できる。その結果が、表2の項目パラメタ欄に示した値である。

5-4 被験者の特性値 θ の推定

特性値 θ の被験者から、43項目に対する反応パターン $\{y_1, \dots, y_{43}\}$ が得られる確率を $P_g(\theta)$ とする。仮定より、 θ が条件付られた状況のもとで各項目の $P_{y_i}(\theta)$ は相互に独立であるため、 $P_g(\theta)$ は次の様に書ける。

$$P_g(\theta) = \prod_{i=1}^{43} P_{y_i}(\theta)$$

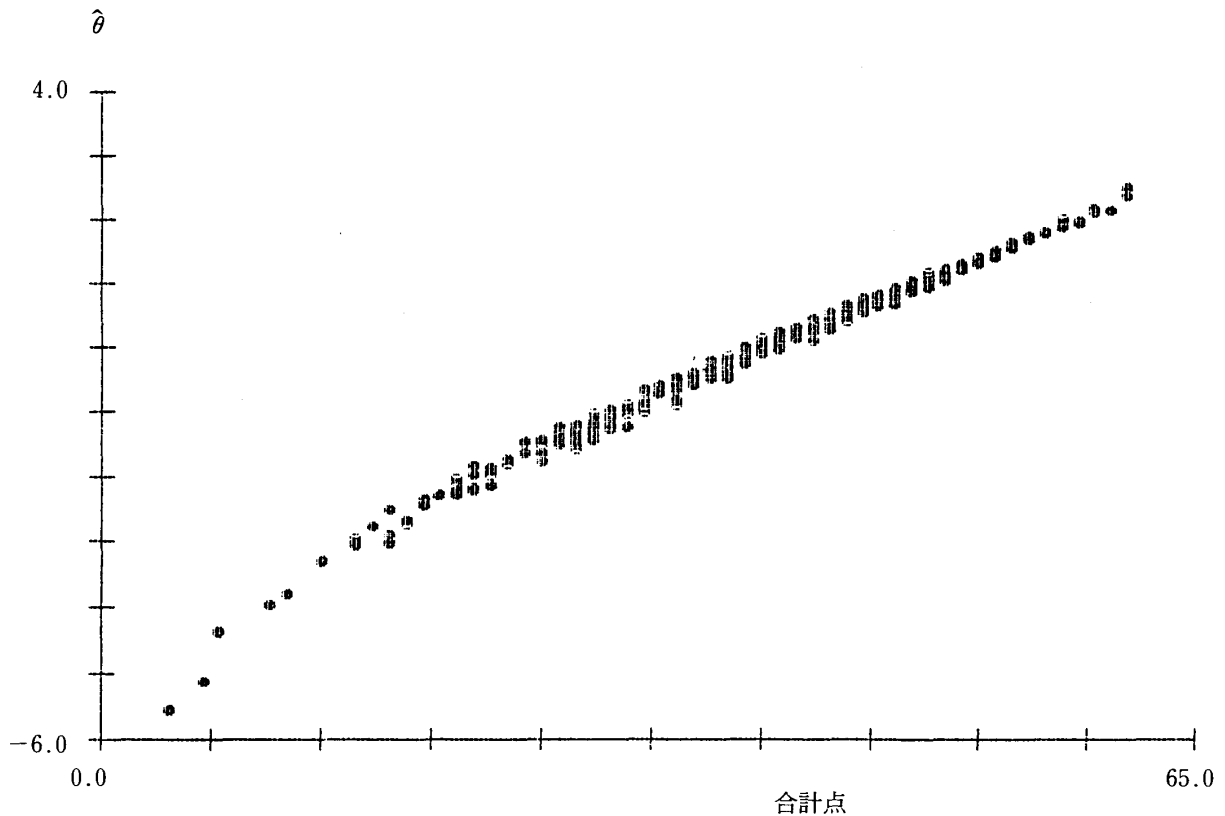
反応パターン $\{y_1, \dots, y_{43}\}$ が得られたとき、 $P_g(\theta)$ を θ の関数と見れば θ に関する尤度関数 $L(\theta)$ と考えることができる。よって、 $L(\theta)$ を最大にする θ の値を被験者の特性の推定値 $\hat{\theta}$ とする。実際の計算は、対数尤度を利用し、その導関数が0になる θ の値をニュートンラフソン法で数値計算した。

ここで利用した項目は、主因子解より等質な項目と見なされるものである。よって、項目の測定値 y_i の合計点と θ の推定値の大小関係はほぼ一致しているはずである。これを確かめるために、縦軸に $\hat{\theta}$ 、横軸に合計点をとって、その分布を図示したのが図3である。これを見ると、明らかに $\hat{\theta}$ と合計点には対応関係があり、これまでの項目得点の合計による集計にかわって、この推定値を被験者の指標として使用しても問題ないことが解る。

5-5 モデルの当てはまり

ここでは、①今回利用したデータが、どれほど項目反応モデルの仮定を満たしているか？②項目反応モデルを導入した利点がどれ程生かされるのか？という2点に絞って検討を加える。

仮定①については、この43項目は主因子解の第1因子に因子負荷量が大きいものとして得られたものである。このことは、第1因子という潜在的な特性以外では各項

図3：推定値 θ と合計点

目は独自のものを測定しているものと解釈できよう。また、 α 係数の値からも、これらの項目の等質性は問題ないであろう。よって、これらのことからこの43項目については局所独立の仮定は満たされていると見なして差し支えないであろう。

仮定②・仮定③については、これらを直接確かめる方法は現実的にはない。

仮定④については、数量化Ⅲ類を用いて、43項目の各評定カテゴリーに与えられる最適スコア値の大小と評定値の順序の整合性を検討した。この手続きで得られた最適スコア値は、表2に示した通りである。この表では、最適スコア値と評定値 y_i の大小関係は逆転しているが、43項目中31項目がその順序の対応について一致している。一致しなかった項目に関しても、 $y_{i=1}$ と $y_{i=2}$ の最適スコア値の順序が入れ替わっているだけであった。よって、全体としては、仮定④についてまずまず満たされているデータであったように考えられる。

仮定⑤についても、直接これを確かめる方法はない。

ただ、もしこれが現実のデータと大きく食い違っている仮定であれば、項目特性関数の形に問題が生じることになるであろう。特に、識別力を示すパラメタ a_i は、項目特性関数の項で示した $P_{y_{i=1}}(\theta)$ 、 $P_{y_{i=2}}(\theta)$ のそれぞれ $\theta = b_{y_{i=1}}$ 、 $\theta = b_{y_{i=2}}$ での曲線の傾きを決める値である。よって、この仮定に対しては、項目毎に項目特性関数を形作る関数 $P_{y_{i=1}}(\theta)$ 、 $P_{y_{i=2}}(\theta)$ の理論的な形と、 θ のレベル毎での実際の反応頻度を図示したものを比較してその一致を検討を加える。ただし、全ての項目についてここで図示することは紙面の都合上不可能なので、例として項目 H53 について図4に示す。これを見ると分布の両端ではずれが見られるが、 $\theta = 0$ を中心とした最も被験者が密な θ の領域では理論的な関数曲線と実際の観測値頻度が良く一致していることがわかる。このことより、少なくとも項目 H53 についてこの仮定は良く当てはまっており、項目特性関数の種類の選択も問題が無かったと言えよう。ただし全ての項目にわたって、必ずしもこの様な一致が一様に見られたわけではない。項目

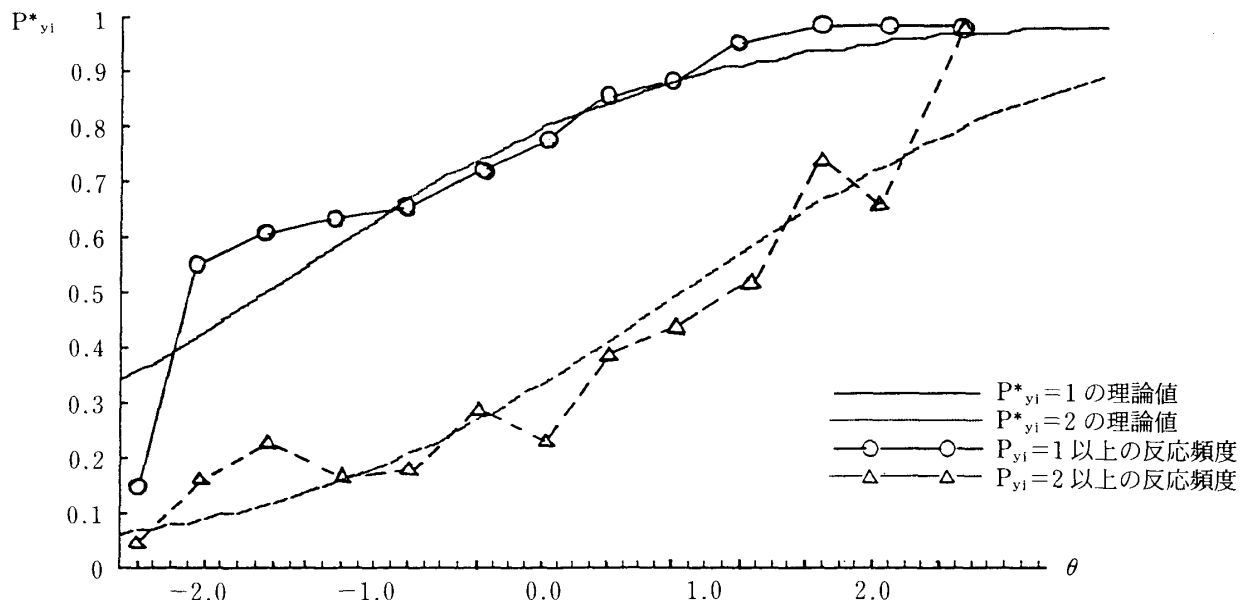


図4： $P^*_{y_i}$ の理論値と観測値の比較

項目グループA (21項目)

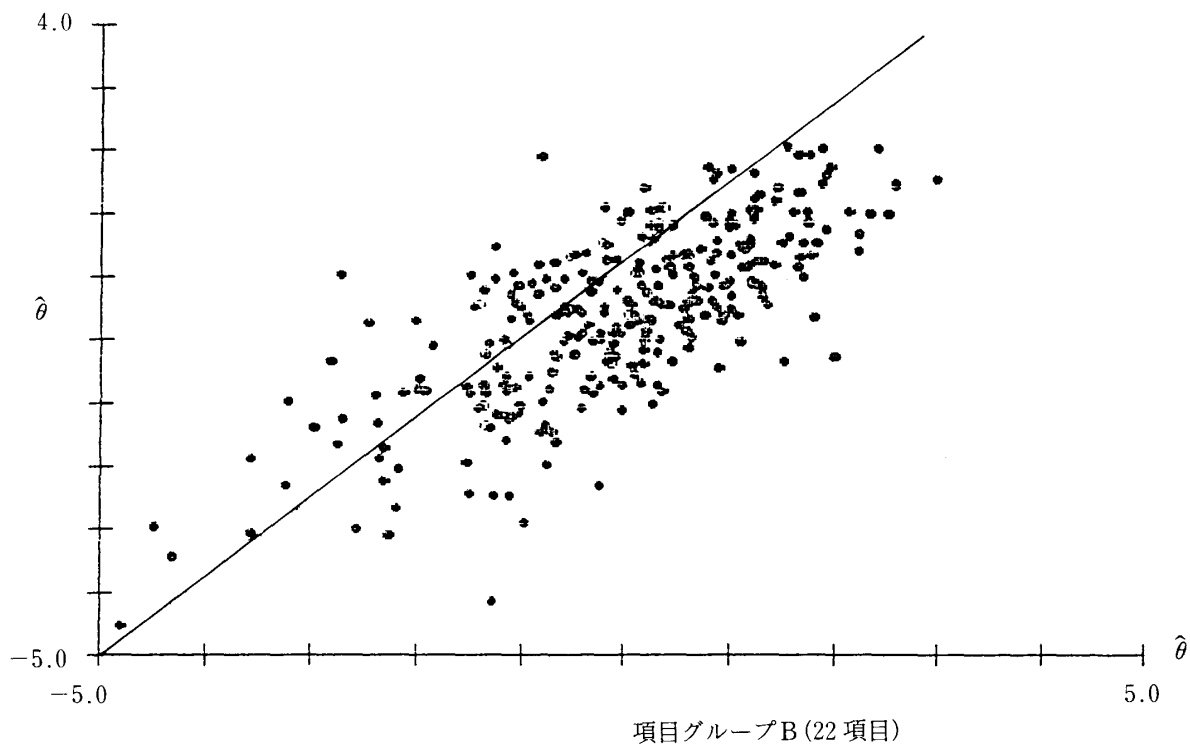


図5： 2つの項目グループでの推定値

によっては、この仮定や項目特性関数の選択自体を検討する必要を感じさせるものもあった。この問題の検討は今後の課題といえる。

項目反応モデルを導入した利点に関しては、異なる項目を利用して被験者の尺度値を得ても、同一の同レベルの尺度値が得られることを確認する。これが確認できれば、異なる項目を利用しても同一尺度軸上で被験者の特性を比較できることへの可能性が開けることになろう。ただし、ここでは新たなデータを得ることができなかった。よって、項目パラメタを得たデータについて項目を任意に2つのグループに分け、それぞれの項目グループに対する被験者の反応パターンから推定値の一致について検討した。各項目グループで得られた推定値の同時分布を示したのが、図5であり、推定値間の相関係数は、0.74であった。項目を2つのグループに分けることにより項目数が減少し、各グループでの推定値の精度は低下する。よって、直線的に一致するとまではいかないが、大まかな尺度値のレベル毎に見ればこれらは良く対応しているといえよう。よって、この尺度化は異なる項目グループを使った再テストの結果の比較等に応用の期待が持てるものと言える。

6. まとめ

本論では、集団式で実施したインクプロットテストについて Form Definiteness 評定の視点から3段階評定を実施した。そしてこの評定に関して等質な項目群が抽出され、その反応パターンについて項目反応モデルの graded response model を適応して尺度化を試みた。その結果、この評定においてモデルに比較的当てはまりのよい尺度化が可能であることが確認できた。

今後の課題としては、この評定に関して等質な項目をより多くプールすることが必要である。また、この尺度を利用する上では、ここで定義した Form Definiteness 評定と他の人格特性の概念との関係を検討する中で妥当性の議論を深めていく必要がある。

7. 参考文献

Holtzman, W.H., Thorpe, J.S., Swartz, J.D. and Herron E.W. 1961
Inkblot perception and personality. Published for the Hogg Foundation for Mental Health.

松井 仁 1991
ロールシャッハテスト出の信頼性について～その2～

東京大学教育学部心理教育相談室紀要 第13集。

Samejima, F. 1969
Estimation of latent ability using a response pattern of graded scores.
Psychometric Monographs. No.17.

Samejima, F. 1988
Comprehensive latent trait theory.
Behaviormetrika, vol 24, pp1-24.

(指導教官 渡部 洋 助教授)