

識別性検査 A-1001の「関係判断力・応用力」領域 および「記憶」領域の適応型テスト化の試み

野 口 裕 之

1. はじめに

識別性検査 A-1001は、鉄道事業の運転関係従事員を対象とする一種の知的能力検査で、

- ① 検査項目に鉄道に関係の深い素材を用いる
- ② 検査を実施するのにパーソナルコンピュータを用いる
- ③ 検査を支えるテスト理論に項目反応理論を用いる
- ④ 検査の実施は適応型テスト方式を用いる

という条件の下で、新たに開発するものである。そして、測定領域および問題は、

- I 「知覚の速さ・正確さ」領域
 - ① 「図形・記号の探索」問題
 - ② 「図形・記号の異同弁別」問題
- II 「関係判断力・応用力」領域
 - ③ 「空間図形の推理」問題
 - ④ 「図形・記号系列の作成」問題
- III 「記憶」領域
 - ⑤ 「図形・記号系列の順唱・逆唱」問題

で構成されている。

これまでに、項目プールの作成、予備テスト、項目分析が実施され（野口；1993a,1993b,1994a,1994b）、さらに、項目反応理論（item response theory；IRT）に基づく尺度の構成が試みられた（野口；1995, 1996）。引き続き、本研究では、「II 関係判断力・応用力」領域の「③空間図形の推理」問題および「④図形・記号系列の作成」問題、そして「III 記憶」領域の「⑤図形・記号系列の順唱・逆唱」問題について適応型テスト化の可能性について検討する。

2. 適応型テスト

2. 1. 適応型テストの特徴

受験者個人毎に最適な特性（困難度・識別力等）を持つ項目を選択してテストを編集したならば、全ての受験者に対して効率よく十分精度の高い測定を実施することが可能になる。このように受験者個人毎に最適な項目を選択・編集して実施するテスト方式を「適応型テスト（adaptive testing）」方式と呼ぶ。

適応型テスト方式の利点には

- ① 全ての受験者に対して高い精度の測定が実施可能である
- ② 精度を落とす事なく、受験者一人あたりに実施する項目数を減らす事ができ、その為にテストの実施に要する時間を短縮する事ができる
- ③ 難しい項目が続いて受験者にフラストレーションや不安を起こさせたり、易しすぎる項目が続いて飽きさせたりする事がない
- ④ 受験者の都合に合わせてテストを実施できるなどがある。

2. 2. 適応型テストの方法

適応型テストでは、当該受験者の測定に最適な項目を選択するのに、基本的には「直前に実施した項目に対して受験者が正答した場合には、次にはより難しい項目を実施し、逆に受験者が誤答した場合には、次にはより易しい項目を実施する」という手続きを逐次繰り返すが、実際の測定場面で適応型テストを実施するための具体的方法については測定内容や項目の形式等に応じて様々な種類のものがある。その中で、本研究では、「項目可変型多段階テスト（variable branching multistage test）」と「項目固定型多段階テスト（fixed branching multistage test）」と呼ばれる方法を用いる。

項目可変型多段階テスト：予め項目困難度や識別力の値が推定されている整備された項目プールを用意しておき、その中から逐次最適な項目を直接取り出して受験者に呈

示し、解答させる(図1)。この方法では、受験者が解答した直後瞬時に次に呈示する項目を決定する為の数値計算を行ない、最適項目を決定し、その内容を呈示する必要があるのでコンピュータの利用が不可欠となる。最

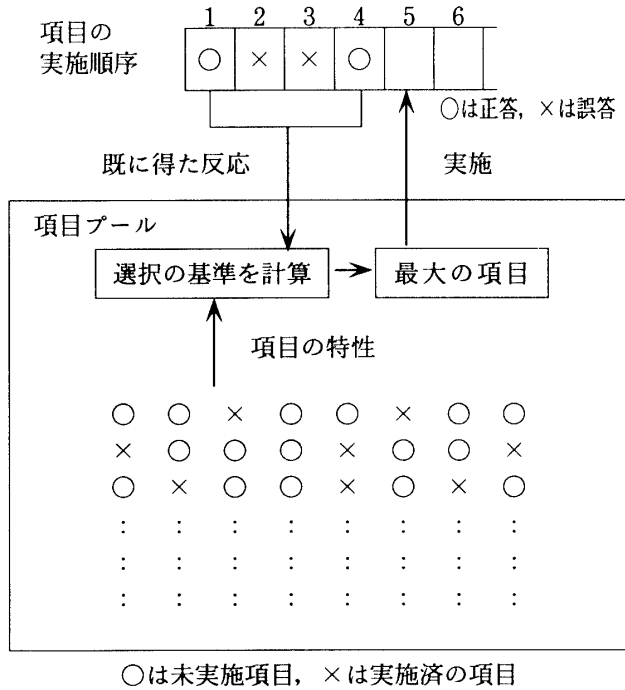


図1 項目可変型多段階テスト

適な項目を決定する基準は、例えば、未実施の項目全てについて、受験者のその時点における推定尺度値に対する項目情報量を計算して、最大の項目情報量を示す項目を次に実施するなどの方法がある。この場合特に「最大情報量方式」と呼ぶ。

項目固定型多段階テスト：予め項目をその特性(特に困難度)に基づいてネットワーク状に配列しておき、受験者の項目に対する反応に応じてネットワーク上の項目を配置された道筋に従って逐次的に呈示し、解答させる。項目ネットワークの配列法によってさらに下位分類することが可能であるが、例えば、多層構造のもの(図2)は初めの数項目についてはツリー状のネットワークが組み立てられており受験者はその解答の正誤に従って枝分かれして進むが、その後は困難度に基づいて設定された層状ネットワークの複数の層の間を受験者は項目に対する解答に基づいて、上下あるいは同じ水準の層へ進む。図2の例では各層毎に未利用の項目が残ることになるが、ネットワーク上の項目を有効に利用するために、各層毎に実施順序を決めておき、当該層に進行する度にその順序に従って受験者に項目を呈示するなどの工夫が加えられることもある。

適応型テストについて詳しくは例えば、野口(1991)を参照されたい。

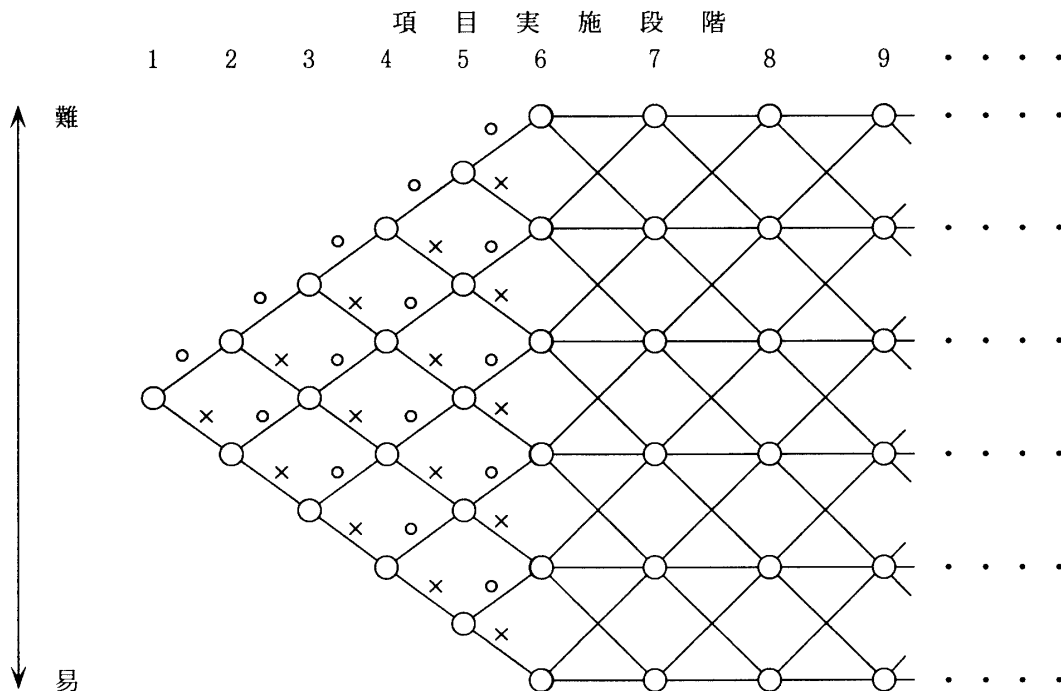


図2 項目固定型多段階テスト

3. 「空間図形の推理問題」の適応型テスト化

3. 1. 問題の概要

本問題では、鉄道線路に用いられる枕木に似た形状の積み木を並べたり積み上げたものを呈示し、受験者はその中にひとつだけ存在する緑色の積み木に接している積み木の数を解答する。各項目毎に、積み木の総数および接する積み木の数が異なり、その結果項目困難度が変化する。全部で30項目から構成され、各項目のパラメタ値は表1に示す通りである。詳しくは、野口、(1993b, 1996)を参照されたい。

表1 「空間図形の推理問題」各項目の項目パラメタ推定値

項目番号	識別力	困難度
1	301	.463
2	302	.375
3	303	.680
4	304	.399
5	305	.883
6	306	.870
7	307	.352
8	308	.833
9	309	.487
10	310	.180
11	311	.729
12	312	.898
13	313	.849
14	314	1.085
15	315	1.247
16	316	1.237
17	317	.941
18	318	.721
19	319	.947
20	320	1.450
21	321	.584
22	322	.543
23	323	1.070
24	324	.738
25	325	.701
26	326	.424
27	327	.766
28	328	1.026
29	329	.467
30	330	.965

3. 2. 適応型テストの構成

本問題では、項目数が全部で30項目と少ないこと、そして、項目が相互に類似しており、特に下位グループを設定する必要もないことから、全30項目から項目プールを構成して、「最大情報量方式」による「項目可変型多段階テスト形式」を採用した。

具体的には以下のような手順で検査が進行するように構成した。すなわち、

①予め指定した5項目（項目番号311, 319, 320, 324, 330）を実施する。

これらの項目は困難度が低いものから高いものまで適当な間隔を持つ様に指定されており、受験者の大雑把な能力推定値を得る。最初の5項目実施後には、受験者の解答、解答の正誤、反応時間が記録される。

②上記の5項目に対する受験者の正誤反応パターンから、推定尺度値を計算し、さらに、残り25項目の推定尺度値に於ける項目情報量を計算する。全ての項目に正答（または誤答）した受験者については、推定尺度値に+3.5（または-3.5）を割り当てる。

③項目情報量の最も大きい項目を、次に実施する。

④この項目の正誤反応を、以前の項目に対する正誤反応パターンに加えて、改めて推定尺度値を計算する。そして、この時点までに実施した項目群のテスト情報量を計算する。全ての項目に正答（または誤答）の受験者については、推定尺度値に+3.5（または-3.5）を割り当てる。

⑤残りの未実施項目について項目情報量を計算して、最大情報量を示す項目を次に実施する。

⑥上の④⑤の手順を繰り返しながら、予め定められたテスト終了基準を満たしたならば、必要な情報を記録して、本検査を終了する。

テスト実施中には、実施した項目の番号、項目情報量、受験者の解答、解答の正誤、反応時間（これはテスト過程の制御には用いない）、各項目実施後の推定尺度値、その時点までに実施した項目によるテスト情報量が記録される。

また、テスト終了時点では、実施項目数、最終的な推定尺度値および実施した項目に基づくテスト情報量が記録される。

3. 3. 方法

テストの実施：本研究では、受験者に対して実際に適応型テスト方式を実施するのではなく、全30項目に対する項目反応パターンが既に得られている受験者のデータを利用する。すなわち、各項目に対する解答および反応時間については既に得られた結果をそのまま用いるが、各

項目の呈示順序については適応型テストを受験した状況をシミュレートする。これにより、受験者があたかも適応型テスト方式で受験し、順次呈示された項目に解答しているような状況を作り出す。本来は全受験者に対して項目の呈示順序が同一であった場合に得られたデータであるため、項目の解答順序が検査結果に影響を与える可能性がある場合にはこのような方法を用いることは適当ではないが、本検査の場合には各項目の独立性が高く、練習問題を最初に実施することもあって解答方法に慣れるのに数項目を要するという可能性は考えられないため、特に問題は無いものと思われる。

受験者：受験者の項目反応には、1996年6月に実施したA大学教育学部の3年生28名分のものを用いた。

テスト終了基準：テストの終了基準については、①実施項目数を予め決めておく、②テスト情報量が予め設定した値を超えた時点でテストを終了する、③テスト実施時間を予め決めておく、④最新の推定値と直前の推定値との差が設定した値以下になった時点で打ち切る、などがあり、これらの中からいずれの基準を採用し具体的な値を設定するのかによって測定効果の評価が変わってくる。本問題の場合、現在のところ基準を唯一に定めるのに十分な情報が得られていないので、「実施項目数」について、10、15、20項目の3条件、「テスト情報量」について、2.00、2.38、3.00の3条件、「テスト実施時間」について、300、450、600秒の3条件、「推定値の差」について、0.05、0.01の2条件を設定し、各テスト終了条件間で比較する。なお、テスト情報量の各条件は、仮に潜在特性尺度の全範囲でテスト情報量が一定であるとした場合に、対応する信頼性係数が0.80、0.85、0.90になるような値にそれぞれ設定した。

テストの評価：個々の結果を評価するための指標として、30項目全てを用いて測定した場合、すなわち全項目テストの結果を基準にして、実施項目数の項目数30に対する比（項目数比）、全項目テストによる推定値との差の絶対値（推定値差）、全項目テストの情報量に対する比（情報量比）、全項目テストの総解答時間に対する比（テスト時間比）を用いる。

3. 4. 適応型テストの効果

適応型テスト方式のシミュレーション結果については、表2に受験者#7のテスト進行過程を、表3に各種終了基準を採用した場合の最終結果を、表4に受験者#26のテスト進行過程を、表5に各種終了基準を採用した場合の最終結果を、そして、表6に全受験者28名分の結果を要約したものを示した。

受験者#7：初期5項目については、2項目に正答し、

推定値が-0.61、テスト情報量が1.744になる。そして適応型テストの部分に入り、最初に項目番号316が呈示され、これに誤答し、推定値が-0.87、テスト情報量が2.015、次に項目番号315に誤答し、推定値が-1.00、テスト情報量が2.161、さらに項目番号314に誤答し、推定値が-1.09、テスト情報量が2.229となり、以下、表2に示すように進行した。各種終了基準を設定した場合の最終結果は表3の通りであるが、

①実施項目数基準については、10項目条件から20項目条件にかけて、推定値差が順に0.40、0.31、0.23と減少し、情報量比が0.705、0.848、0.919と増加している。テスト時間比は順に0.40、0.56、0.70であった。項目数15、すなわち半分の項目数で、テスト情報量が全項目テストの80%を超えた。

②テスト情報量基準については、条件順に項目数比が0.20、0.33、0.53、推定値差が0.22、0.40、0.23、情報量比が0.570、0.705、0.881、テスト時間比が0.25、0.40、0.59となる。約60%の時間、およそ半数の項目で、テスト情報量が80%を超えている。

③テスト時間基準については、300秒の時点で21項目に対して解答し、推定値差が0.14、情報量比が0.942を示しており、解答時間を300秒に設定しても全項目テストとほとんど変わりのない結果が得られている。

④直前の推定値との差を基準とした場合については、基準値が0.05の場合に全項目テストで得られる推定値との差（以下、推定値差と略す）が0.43とやや大きいが、基準値を0.01に採ると推定値差は0.06と小さくなり、情報量比が0.973、テスト時間比が0.86となっている。この場合は、項目数比が0.80であり、実施項目数を20%減らしても測定の正確さにはほとんど影響がない。

受験者#26：初期5項目については、5項目全てに正答し、推定値には3.50が割り当てられ、テスト情報量が0.222になる。そして適応型テストの部分に入り、最初に項目番号327が呈示され、これに誤答し、推定値が1.14、テスト情報量が1.132、次に項目番号323に正答し、推定値が1.28、テスト情報量が1.195、さらに項目番号303に誤答し、推定値が0.90、テスト情報量が1.521となり、以下、表4に示すように進行した。各種終了基準を設定した場合の最終結果は表5の通りであるが、

①実施項目数基準については、10項目条件から20項目条件にかけて、推定値差が順に0.47、0.29、0.14と減少し、情報量比が0.589、0.746、0.884と増加している。テスト時間比は順に0.45、0.62、0.79であった。項目数15、すなわち半分の項目数で、テスト情報量が全項目テストの75%まで到達している。

②テスト情報量基準については、条件順に項目数比が

表2 「空間図形の推理問題」の適応型テスト実施過程（受験者#7の場合）
受験者番号 #7

実施順	項目番号	難易順	正誤	推定値	テスト情報量	累積解答時間(秒)
1	311	4	0			3.1
2	319	24	1			20.0
3	320	9	1			34.8
4	324	27	0			60.3
5	330	15	0	-0.61	1.744	74.2
6	316	11	0	-0.87	2.015	95.3
7	315	17	0	-1.00	2.161	106.6
8	314	20	0	-1.09	2.229	117.6
9	328	19	0	-1.17	2.271	137.2
10	313	5	1	-1.05	2.493	156.5
11	317	13	0	-1.12	2.529	171.9
12	306	10	1	-0.99	2.743	176.3
13	312	18	0	-1.05	2.781	188.8
14	323	25	0	-1.08	2.818	207.7
15	308	23	1	-0.96	2.999	215.4
16	325	12	1	-0.88	3.115	227.4
17	318	22	0	-0.92	3.137	240.3
18	305	2	1	-0.90	3.182	247.1
19	322	21	0	-0.93	3.190	260.2
20	309	14	1	-0.88	3.250	270.6
21	303	29	1	-0.79	3.330	276.6
22	321	26	1	-0.73	3.386	303.5
23	329	7	1	-0.70	3.421	320.1
24	327	30	0	-0.71	3.439	335.1
25	301	6	1	-0.68	3.471	341.9
26	326	16	0	-0.71	3.480	360.4
27	302	8	1	-0.69	3.502	366.3
28	307	3	1	-0.67	3.518	375.7
29	304	1	1	-0.66	3.529	383.1
30	310	29	1	-0.65	3.536	387.7

表3 「空間図形の推理問題」の適応型テスト化の効果（受験者#7の場合）
受験者番号 #7

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	16.	-0.65	3.536	387.70	—	—	—	—
実施項目数									
10	10.	3.	-1.05	2.493	156.47	0.33	0.40	0.705	0.40
15	15.	5.	-0.96	2.999	215.40	0.50	0.31	0.848	0.56
20	20.	8.	-0.88	3.250	270.57	0.67	0.23	0.919	0.70
テスト情報量									
2.00	6.	2.	-0.87	2.015	95.26	0.20	0.22	0.570	0.25
2.38	10.	3.	-1.05	2.493	156.47	0.33	0.40	0.705	0.40
3.00	16.	6.	-0.88	3.115	227.37	0.53	0.23	0.881	0.59
テスト時間									
300秒	21.	9.	-0.79	3.330	276.57	0.70	0.14	0.942	0.71
450秒	30.	16.	-0.65	3.536	387.70	1.00	0.00	1.000	1.00
600秒	30.	16.	-0.65	3.536	387.70	1.00	0.00	1.000	1.00
直前の推定値との差									
.05	14.	4.	-1.08	2.818	207.71	0.47	0.43	0.797	0.54
.01	24.	11.	-0.71	3.439	335.10	0.80	0.06	0.973	0.86

表4 「空間図形の推理問題」の適応型テスト実施過程（受験者#26の場合）

受験者番号 #26

実施順	項目番号	難易順	正誤	推定値	テスト情報量	累積解答時間(秒)
1	311	4	1			29.1
2	319	24	1			122.9
3	320	9	1			139.3
4	324	27	1			177.5
5	330	15	1	3.50	0.222	202.2
6	327	30	0	1.14	1.132	238.9
7	323	25	1	1.28	1.195	308.8
8	303	29	0	0.90	1.521	327.0
9	308	23	1	0.99	1.554	371.5
10	321	26	1	1.14	1.542	420.8
11	318	22	0	0.79	1.817	491.1
12	314	20	1	0.84	1.863	512.1
13	328	19	1	0.89	1.898	540.4
14	312	18	1	0.93	1.925	565.7
15	315	17	1	0.96	1.953	577.2
16	317	13	1	0.99	1.974	600.4
17	325	12	1	1.04	1.976	639.3
18	322	21	1	1.09	1.966	676.9
19	306	10	0	0.79	2.291	716.3
20	316	11	1	0.81	2.314	739.0
21	309	14	1	0.85	2.300	776.5
22	326	16	0	0.75	2.422	792.2
23	313	5	1	0.77	2.428	807.0
24	329	7	1	0.79	2.422	825.6
25	301	6	0	0.69	2.552	830.8
26	302	8	0	0.62	2.644	844.6
27	307	3	1	0.63	2.637	863.7
28	304	1	1	0.64	2.634	882.2
29	305	2	1	0.64	2.639	889.8
30	310	29	1	0.67	2.618	935.0

表5 「空間図形の推理問題」の適応型テスト化の効果（受験者#26の場合）

受験者番号 #26

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	23.	0.67	2.618	935.01	—	—	—	—
実施項目数									
10	10.	8.	1.14	1.542	420.75	0.33	0.47	0.589	0.45
15	15.	12.	0.96	1.953	577.16	0.50	0.29	0.746	0.62
20	20.	16.	0.81	2.314	739.02	0.67	0.14	0.884	0.79
テスト情報量									
2.00	19.	15.	0.79	2.291	716.34	0.63	0.12	0.875	0.77
2.38	22.	17.	0.75	2.422	792.17	0.73	0.08	0.925	0.85
3.00	30.	23.	0.67	2.618	935.01	1.00	0.00	1.000	1.00
テスト時間									
300秒	6.	5.	1.14	1.132	238.88	0.20	0.47	0.432	0.26
450秒	10.	8.	1.14	1.542	420.75	0.33	0.47	0.589	0.45
600秒	15.	12.	0.96	1.953	577.16	0.50	0.29	0.746	0.62
直前の推定値との差									
.05	12.	9.	0.84	1.863	512.12	0.40	0.17	0.712	0.55
.01	27.	20.	0.63	2.637	863.66	0.90	0.04	1.007	0.92

0.63, 0.73, 1.00, 推定値差が0.12, 0.08, 0.00, 情報量比が0.875, 0.925, 1.000, テスト時間比が0.77, 0.85, 1.00となる。受験者#7に比べて適応型テスト化の効果が顕著には現われないが, これは, 3.00条件で項目数比, 情報量比, テスト時間比の全てが1.00であることから, 本受験者を測定するのに適切な項目が項目プール中に不足していたことによるものと思われる。

- ③テスト時間基準については, 600秒の時点で15項目に対して解答し, 推定値差が0.29, 情報量比が0.746であり, まだ十分満足できる測定結果が得られていない。この受験者については解答時間がかかなり長いので, 十分満足できる測定結果が得られていない。
- ④直前の推定値との差を基準とした場合については, 基準値が0.05の場合に全項目テストで得られる推定値との差が0.17とやや大きい, 基準値を0.01に採ると推

定値差は0.04と小さくなり, 情報量比が1.007, テスト時間比が0.92となっている。この場合は, 項目数比が0.90であり, 実施項目数は10%の減少に止まる。

- ⑤本受験者の場合, 受験者#7が全項目テストでテスト情報量が3.536であるのに対して, 全項目テストでもテスト情報量が2.618しか得られず, そのことと解答時間の長さとの適応型テスト方式の評価全体に影響を与えている。

全受験者28名を通して最終結果を見ると,

- ①実施項目数基準については, 10項目条件から20項目条件にかけて, 推定値差が順に平均が0.41, 0.27, 0.17と減少し, 標準偏差も0.55, 0.39, 0.20と減少し, 情報量比が平均で0.71, 0.82, 0.92と増加し, 標準偏差も0.15, 0.13, 0.09と減少している。受験者全体の場合同様受験者#7と同様に項目数15, すなわち半分

表6 「空間図形の推理問題」の適応型テスト化の効果 (要約)
全受験者28名の平均及び標準偏差

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	18.2	-0.09	2.73	469.1	—	—	—	—
	—	7.3	1.21	0.67	167.0	—	—	—	—
実施項目数									
10	10.	5.9	-0.10	1.95	180.1	0.33	0.41	0.71	0.38
	—	2.5	1.56	0.62	76.3	—	0.55	0.15	0.04
15	15.	8.9	-0.11	2.27	265.6	0.50	0.27	0.82	0.56
	—	3.9	1.52	0.73	104.9	—	0.39	0.13	0.04
20	20.	11.4	-0.17	2.54	345.0	0.67	0.17	0.92	0.73
	—	5.0	1.36	0.72	132.4	—	0.20	0.09	0.04
テスト情報量									
2.00	14.1	9.0	-0.09	2.04	256.5	0.47	0.18	0.78	0.51
	8.3	8.0	1.17	0.20	215.6	0.28	0.20	0.16	0.28
2.38	18.1	11.5	-0.10	2.29	316.1	0.60	0.12	0.87	0.64
	8.9	9.2	1.20	0.31	215.0	0.30	0.15	0.12	0.28
3.00	24.2	15.0	-0.10	2.59	402.9	0.81	0.05	0.96	0.83
	7.0	9.1	1.21	0.53	208.6	0.24	0.08	0.06	0.21
テスト時間									
300秒	18.8	10.0	-0.11	2.44	286.0	0.63	0.17	0.88	0.68
	6.6	5.0	1.31	0.83	15.0	0.22	0.27	0.16	0.20
450秒	26.3	15.3	-0.07	2.64	395.7	0.88	0.07	0.96	0.90
	6.0	6.8	1.25	0.76	61.1	0.20	0.14	0.11	0.17
600秒	28.4	16.9	-0.07	2.69	437.1	0.95	0.02	0.98	0.96
	16.9	6.6	1.23	0.72	104.9	0.13	0.05	0.06	0.10
直前の推定値との差									
.05	14.0	8.2	-0.10	2.22	247.0	0.47	0.33	0.80	0.53
	2.6	3.9	1.62	0.80	102.1	0.09	0.59	0.18	0.10
.01	24.1	14.6	-0.04	2.60	396.8	0.80	0.17	0.94	0.85
	5.6	7.0	1.49	0.84	156.2	0.19	0.56	0.19	0.19

目数で、テスト情報量が全項目テストの80%を超えている。

- ②テスト情報量基準については、条件順に項目数比の平均が0.47, 0.60, 0.81, 情報量比の平均が0.78, 0.87, 0.96, テスト時間比の平均が0.51, 0.64, 0.83となる。約60%の時間で項目数もおよそ60%で、テスト情報量が90%に近い値を示している。全体の平均と比較すると、#7の受験者は適応型化によりかなり効率の良い測定を受けたことがうかがわれる。また、推定値差については、平均が順に0.18, 0.12, 0.05, 標準偏差が順に0.20, 0.15, 0.08と減少している。3.00条件の場合はほとんどの受験者が0.1程度の差にとどまっている。
- ③テスト時間基準については、条件順に項目数比の平均が0.63, 0.88, 0.95, 情報量比の平均が0.88, 0.96, 0.98, テスト時間比の平均が0.68, 0.90, 0.96となる。300秒の時点で平均18.8項目に対して解答し、推定値差が平均0.17, 情報量比が0.88を示しており、全項目テストに比較的近い結果が得られている。ただし、この時点では受験者間の差が大きく、例えば推定値差については、標準偏差が0.27あり、受験者によってはいまま少し解答時間が必要なケースが存在する。
- ④直前の推定値との差を基準とした場合については、基準値が0.05の場合に全項目テストで得られる推定値との差が平均で0.33, 標準偏差で0.59と大きな値を示すが、基準値を0.01に採ると推定値差が平均0.17, 標準偏差が0.56となり、平均的にはある程度満足できる値になる。しかしながら標準偏差の値が依然として大きく、中には全項目テストによる推定値とはかなり隔たった推定値でテストが終了するケースもあることを示している。基準値0.01の条件では実施項目数の平均が24.1であるが、これがほぼ同じであるテスト情報量基準の3.00条件では、推定値差の平均が0.05, 標準偏差が0.08であることから、直前の推定値との差をテスト終了基準に用いることにはより慎重な検討が必要である。全体として見ると、既に述べた理由から、直前の推定値との差をテスト終了基準とすることには問題がありそうである。テスト時間基準については1項目あたりの解答時間を制限し、基準時間内に実施できる項目数を増やすなどやり方そのものについて今少し検討する必要がある。実施項目数およびテスト情報量を基準とすることについては、いずれの場合についても適応型テスト化の効果が確認された。それらの中でも、テスト情報量の標準偏差が全体として最も小さい、すなわち、受験者間での測定精度のバラツキが最も小さいテスト情報量基準が実際のテスト場面でも有効に機能する可能性が高いと言える。

4. 「図形・記号系列の作成問題」の適応型テスト化

4.1. 問題の概要

本問題では、ある規則に従って並んだ図形・記号の系列を受験者に呈示し、受験者はその系列の並び方の規則を発見して、呈示された系列の次ぎ及びその次ぎに入るべき図形・記号を解答する。系列の要素となる図形・記号の種類が2, 3そして4種類のもの各10項目、全部で30項目から構成され、各項目のパラメタ値は表7に示す通りである。詳しくは、野口(1993b,1996)を参照されたい。

表7 「図形・記号系列の作成問題」各項目の項目パラメタ推定値

項目番号	識別力	困難度
1	401	1.237
2	402	.694
3	403	.935
4	404	.844
5	405	.652
6	406	.396
7	407	.407
8	408	.657
9	409	.838
10	410	.996
11	411	.660
12	412	.951
13	413	.734
14	414	.658
15	415	.498
16	416	.349
17	417	.545
18	418	.751
19	419	.373
20	420	.665
21	421	.446
22	422	.468
23	423	.670
24	424	.438
25	425	.820
26	426	.951
27	427	.918
28	428	.519
29	429	.751
30	430	.579

4. 2. 適応型テストの構成

本問題についても、「空間図形の推理問題」と同様に、項目数が全部で30項目と少ないこと、そして、項目が相互に類似しており、特に下位グループを設定する必要もないことから、全30項目から項目プールを構成して、「最大情報量方式」による「項目可変型多段階テスト形式」を採用した。

具体的な検査の進行は、

① 予め指定した5項目（項目番号 406, 410, 412, 420, 423）を実施する。

以下、②から⑥まで、「空間図形の推理問題」と全く同じ手順に従う。

テスト実施中には、実施した項目の番号、項目情報量、受験者の解答、解答の正誤、反応時間（これはテスト過程の制御には用いない）、各項目実施後の推定尺度値、その時点までに実施した項目によるテスト情報量が記録される。

また、テスト終了時点では、実施項目数、最終的な推定尺度値および実施した項目に基づくテスト情報量が記録される。

これらについても「空間図形の推理問題」と全く同じである。

4. 3. 方法

テストの実施：本問題についても、受験者に対して実際に適応型テスト方式を実施するのではなく、全30項目に対する項目反応パターンが既に得られている受験者のデータを利用する。すなわち、各項目に対する解答および反応時間については既に得られた結果をそのまま用いるが、各項目の呈示順序については適応型テストを受験した状況をシミュレートする。

受験者：受験者の項目反応には、1996年6月に実施したA大学教育学部の3年生28名分のものを用いた。

テスト終了基準：テストの終了基準についても「空間図形の推理問題」と同様に、「実施項目数」について、10, 15, 20項目の3条件、「テスト情報量」について、2.00, 2.38, 3.00の3条件、「テスト実施時間」について、300, 450, 600秒の3条件、「推定値の差」について、0.05, 0.01の2条件を設定し、各テスト終了条件間で比較する。

テストの評価：個々の結果を評価するための指標として、「空間図形の推理問題」と同じく、全項目テストの結果を基準にして、項目数比、推定値差、情報量比、テスト時間比を用いる。

4. 4. 適応型テストの効果

適応型テスト方式のシミュレーション結果については、表8に受験者#3のテスト進行過程を、表9に各種終了基準を採用した場合の最終結果を、表10に受験者#10のテスト進行過程を、表11に各種終了基準を採用した場合の最終結果を、そして、表12に全受験者28名分の結果を要約したものを示した。

受験者#3：初期5項目については、4項目に正答し、推定値が1.64、テスト情報量が0.795になる。そして適応型テストの部分に入り、最初に項目番号427が呈示され、これに誤答し、推定値が0.77、テスト情報量が1.289、次に項目番号423に誤答し、推定値が0.19、テスト情報量が1.474、さらに項目番号403に正答し、推定値が0.42、テスト情報量が1.628となり、以下、表8に示すように進化した。各種終了基準を設定した場合の最終結果は表9の通りであるが、

① 実施項目数基準については、10項目条件から20項目条件にかけて、推定値差が順に0.10, 0.23, 0.03と振動し、情報量比が0.824, 0.961, 0.984と増加している。テスト時間比は順に0.38, 0.69, 0.83であった。項目数10、すなわち全項目テストの1/3の項目数で、テスト情報量が全項目テストの80%を超えた。

② テスト情報量基準については、条件順に項目数比が0.47, 1.00, 1.00, 推定値差が0.15, 0.00, 0.00, 情報量比が0.946, 1.000, 1.000, テスト時間比が0.61, 1.00, 1.00となる。この受験者を測定するのに項目プール中の全ての項目を用いたとしても十分なテスト情報量が得られないことを示している。これは、適応型テスト方式の問題ではなく、項目プール自体について今後改善の余地のあることを示した結果である。

③ テスト時間基準については、450秒の時点で23項目に対して解答し、推定値差が0.01, 情報量比が0.994を示しており、解答時間を450秒に設定しても全項目テストとほとんど変わりのない結果が得られている。このことは適応型テスト方式自体はうまく機能したということを表している。

④ 直前の推定値との差を基準とした場合については、基準値が0.05の場合に全項目テストで得られる推定値との差が0.21とやや大きい、基準値を0.01に採ると推定値差は0.05と小さくなり、情報量比が0.980, テスト時間比が0.81となっている。この場合は、項目数比が0.63であり、実施項目数を60%程度に減らしても測定の正確さにはほとんど影響がないことが示されている。

受験者#10：初期5項目については、5項目中の2項目に正答し、推定値が-1.52、テスト情報量が0.772に

識別性検査 A-1001の「関係判断力・応用力」領域および「記憶」領域の適応型テスト化の試み

表8 「図形・記号等の系列作成問題」の適応型テスト実施過程（受験者#3の場合）

受験者番号 #03

実施順	項目番号	難易順	正誤	推定値	テスト情報量	累積解答時間(秒)
1	406	12	1			11.8
2	410	24	1			29.1
3	412	5	1			35.6
4	420	28	0			87.3
5	423	16	1	1.64	0.795	103.3
6	427	26	0	0.77	1.289	129.7
7	423	19	0	0.19	1.474	149.8
8	403	20	1	0.42	1.628	160.5
9	408	17	1	0.54	1.707	177.5
10	421	18	1	0.66	1.752	192.9
11	418	27	0	0.55	1.841	238.5
12	414	25	0	0.43	1.919	267.8
13	428	23	0	0.32	1.980	290.3
14	412	21	1	0.41	2.011	307.8
15	415	22	0	0.33	2.043	346.2
16	417	13	1	0.35	2.059	356.1
17	425	14	1	0.37	2.072	367.8
18	422	29	1	0.52	2.064	387.6
19	406	30	0	0.51	2.082	408.7
20	416	11	1	0.53	2.091	417.8
21	409	15	1	0.53	2.102	428.9
22	426	10	1	0.54	2.109	438.5
23	413	7	1	0.55	2.112	447.4
24	429	8	1	0.55	2.115	456.9
25	401	4	1	0.55	2.118	464.1
26	402	1	1	0.56	2.119	472.0
27	407	3	1	0.56	2.122	479.8
28	404	6	1	0.56	2.123	485.7
29	405	2	1	0.56	2.124	494.2
30	410	9	1	0.56	2.125	502.6

表9 「図形・記号等の系列作成問題」の適応型テスト化の効果（受験者#3の場合）

受験者番号 #03

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	22.	0.56	2.125	502.61	—	—	—	—
実施項目数									
10	10.	7.	0.66	1.752	192.86	0.33	0.10	0.824	0.38
15	15.	8.	0.33	2.043	346.19	0.50	0.23	0.961	0.69
20	20.	12.	0.53	2.091	417.81	0.67	0.03	0.984	0.83
テスト情報量									
2.00	14.	8.	0.41	2.011	307.76	0.47	0.15	0.946	0.61
2.38	30.	22.	0.56	2.125	502.61	1.00	0.00	1.000	1.00
3.00	30.	22.	0.56	2.125	502.61	1.00	0.00	1.000	1.00
テスト時間									
300秒	13.	7.	0.32	1.980	290.25	0.43	0.24	0.932	0.58
450秒	23.	15.	0.55	2.112	447.38	0.77	0.01	0.994	0.89
600秒	30.	22.	0.56	2.125	502.61	1.0	0.00	1.000	1.00
直前の推定値との差									
.05	16.	9.	0.35	2.059	356.10	0.53	0.21	0.969	0.71
.01	19.	11.	0.51	2.082	408.69	0.63	0.05	0.980	0.81

表10 「図形・記号等の系列作成問題」の適応型テスト実施過程（受験者#10の場合）
受験者番号 #10

実施順	項目番号	難易順	正誤	推定値	テスト情報量	累積解答時間(秒)
1	406	12	1			12.9
2	410	24	0			39.5
3	412	5	1			48.0
4	420	28	0			79.1
5	423	16	0	-1.52	0.772	110.4
6	404	15	1	-1.08	1.009	124.9
7	427	19	1	-0.26	1.317	136.3
8	409	20	0	-0.54	1.438	166.3
9	418	17	0	-0.76	1.508	207.1
10	408	18	0	-0.91	1.550	244.7
11	411	13	1	-0.83	1.636	259.4
12	428	23	0	-0.91	1.659	284.6
13	417	25	0	-0.97	1.679	309.6
14	415	11	0	-1.20	1.656	330.8
15	425	8	1	-1.15	1.711	340.3
16	401	9	1	-1.13	1.752	347.9
17	413	7	1	-1.10	1.791	363.5
18	422	21	0	-1.16	1.818	402.7
19	407	14	1	-1.09	1.855	416.5
20	421	10	1	-1.04	1.886	430.0
21	426	26	0	-1.06	1.911	453.8
22	403	6	1	-1.04	1.937	459.8
23	419	22	0	-1.10	1.953	488.0
24	414	4	1	-1.08	1.976	497.5
25	430	27	1	-0.85	2.026	516.7
26	402	3	1	-0.85	2.039	523.6
27	405	2	1	-0.84	2.048	532.9
28	424	29	0	-0.85	2.054	655.4
29	416	1	1	-0.84	2.061	671.3
30	429	30	0	-0.84	2.064	761.3

表11 「図形・記号等の系列作成問題」の適応型テスト化の効果（受験者#10の場合）
受験者番号 #10

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	16.	-0.84	2.064	761.30	—	—	—	—
実施項目数									
10	10.	4.	-0.91	1.550	244.70	0.33	0.07	0.751	0.32
15	15.	6.	-1.15	1.711	340.34	0.50	0.31	0.829	0.45
20	20.	10.	-1.04	1.886	429.99	0.67	0.20	0.914	0.56
テスト情報量									
2.00	25.	13.	-0.85	2.026	516.72	0.83	0.01	0.982	0.68
2.38	30.	16.	-0.84	2.064	761.30	1.00	0.00	1.000	1.00
3.00	30.	16.	-0.84	2.064	761.30	1.00	0.00	1.000	1.00
テスト時間									
300秒	12.	5.	-0.91	1.659	284.56	0.40	0.07	0.804	0.37
450秒	20.	10.	-1.04	1.886	429.99	0.67	0.20	0.914	0.56
600秒	27.	15.	-0.84	2.048	532.89	0.90	0.00	0.992	0.70
直前の推定値との差									
.05	16.	7.	-1.13	1.752	347.89	0.53	0.29	0.849	0.46
.01	26.	14.	-0.85	2.039	523.56	0.87	0.01	0.988	0.69

識別性検査 A-1001の「関係判断力・応用力」領域および「記憶」領域の適応型テスト化の試み

なる。そして適応型テストの部分に入り、最初に項目番号404が呈示され、これに正答し、推定値が-1.08、テスト情報量が1.009、次に項目番号427に正答し、推定値が-0.26、テスト情報量が1.317、さらに項目番号409に誤答し、推定値が-0.54、テスト情報量が1.438となり、以下、表10に示すように進行した。各種終了基準を設定した場合の最終結果は表11の通りであるが、

- ①実施項目数基準については、10項目条件から20項目条件にかけて、推定値差が順に0.07, 0.31, 0.20と振動し、情報量比が0.751, 0.829, 0.914と増加している。テスト時間比は順に0.32, 0.45, 0.56であった。項目数15, すなわち半分の項目数で、テスト情報量が全項目テストの80%を超えている。
- ②テスト情報量基準については、条件順に項目数比が0.83, 1.00, 1.00, 推定値差が0.01, 0.00, 0.00, 情報量比が0.982, 1.000, 1.000, テスト時間比が0.68,

1.00, 1.00となる。2.00条件で既に項目数が25に達しており、本受験者の場合も受験者#3と同じく、項目プール中の全ての項目を用いたとしても十分なテスト情報量が得られないことが示された。

- ③テスト時間基準については、300秒の時点で12項目に対して解答し、推定値差が0.07, 情報量比が0.804であり、比較的早い時点である程度満足できる測定結果が得られている。その後時間の経過による測定精度の向上は、実施項目数が増加するにもかかわらず、あまり顕著なものではない。これも項目プール中に含まれる適切な項目が不足していることの現われと言える。
- ④直前の推定値との差を基準とした場合については、基準値が0.05の場合に全項目テストで得られる推定値との差が0.29とやや大きいが、基準値を0.01に採ると推定値差は0.01と小さくなり、情報量比が0.988, テスト時間比が0.69となっている。この場合は、実施項目

表12 「図形・記号等の系列作成問題」の適応型テスト化の効果（要約）
全受験者28名の平均及び標準偏差

	項目数	正答数	推定値	テスト情報量	テスト時間	項目数比	推定値差	情報量比	テスト時間比
全項目テスト	30.	20.3	0.23	2.03	617.0	—	—	—	—
	—	3.5	0.98	0.16	151.4	—	—	—	—
実施項目数									
10	10.	5.8	0.24	1.57	218.9	0.33	0.36	0.78	0.37
	—	1.2	1.20	0.17	46.9	—	0.27	0.06	0.08
15	15.	8.6	0.24	1.82	340.0	0.50	0.18	0.90	0.56
	—	1.7	1.07	0.18	73.2	—	0.14	0.05	0.09
20	20.	11.9	0.21	1.95	455.5	0.67	0.08	0.96	0.75
	—	1.8	1.01	0.15	95.1	—	0.09	0.03	0.10
テスト情報量									
2.00	21.9	13.7	0.22	1.96	452.1	0.73	0.08	0.97	0.75
	6.1	6.5	0.99	0.12	119.8	0.20	0.09	0.03	0.17
2.38	30.0	20.3	0.23	2.03	617.0	1.00	0.00	1.00	1.00
	—	3.5	0.98	0.16	151.4	—	—	—	—
3.00	30.0	20.3	0.23	2.03	617.0	1.00	0.00	1.00	1.00
	—	3.5	0.98	0.16	151.4	—	—	—	—
テスト時間									
300秒	13.2	8.0	0.30	1.72	279.0	0.44	0.29	0.85	0.49
	2.6	2.7	1.15	0.18	24.8	0.09	0.24	0.09	0.15
450秒	20.8	13.4	0.21	1.91	413.7	0.70	0.08	0.95	0.72
	5.8	6.4	1.02	0.16	34.1	0.19	0.10	0.06	0.20
600秒	26.4	17.5	0.22	2.00	531.7	0.88	0.02	0.99	0.89
	4.3	5.5	0.98	0.16	77.0	0.14	0.05	0.02	0.13
直前の推定値との差									
.05	15.9	9.0	0.24	1.84	378.2	0.53	0.17	0.91	0.63
	2.1	2.2	1.03	0.18	94.4	0.07	0.12	0.06	0.12
.01	22.0	13.0	0.24	1.99	508.1	0.73	0.05	0.98	0.82
	3.6	2.1	0.99	0.15	147.2	0.12	0.08	0.02	0.09

数が26, すなわち項目数比が0.87であり, 実施項目数は約10%の減少に止まる。

全受験者28名を通して最終結果を見ると,

- ①実施項目数基準については, 10項目条件から20項目条件にかけて, 推定値差が順に平均が0.36, 0.18, 0.08と減少し, 標準偏差も0.27, 0.14, 0.09と減少し, 情報量比が平均で0.78, 0.90, 0.96と増加し, 標準偏差は 0.06, 0.05, 0.03と減少している。受験者全体として, 項目数15, すなわち半分の項目数で, テスト情報量が全項目テストの90%に達している。ただし, テスト情報量の平均値そのものは1.82と決して高くない。
- ②テスト情報量基準については, 条件順に項目数比の平均が0.73, 1.00, 1.00, 情報量比の平均が0.97, 1.00, 1.00, テスト時間比の平均が0.75, 1.00, 1.00となる。受験者# 3および#10のいずれの場合でも見られたことであるが, 受験者全体としても項目プール中の全ての項目を用いたとしても十分なテスト情報量が得られないことが示された。
- ③テスト時間基準については, 条件順に項目数比の平均が0.44, 0.70, 0.88, 情報量比の平均が0.85, 0.95, 0.99, テスト時間比の平均が0.49, 0.72, 0.89となる。450秒の時点で平均20.8項目に対して解答し, 推定値差が平均0.08, 情報量比が0.95を示しており, 全項目テストにほとんど近い結果が得られている。
- ④直前の推定値との差を基準とした場合については, 基準値が0.05の場合に全項目テストで得られる推定値との差が平均で0.17, 標準偏差で0.12とやや大きな値を示すが, 基準値を0.01に採ると推定値差が平均0.05, 標準偏差が0.08となり, 平均的にはある程度満足できる値になる。しかしながらこの場合に解答項目数が, 平均22.0, 標準偏差3.6と比較的大きく, この後で適切な項目がもはや項目プール中に残っておらず, そのため推定尺度値に変化が生じなかっただけとも考えられる。

全体として見ると, 既に述べた理由から, 直前の推定値との差をテスト終了基準とすることには問題がありそうである。この点は「空間図形の推理問題」と同様である。テスト情報量基準については, 本問題の場合項目プールそのものが十分に整備されたものとは言えず, そのため, 十分な評価を下すことができない。実施項目数およびテスト時間については, 項目プールを所与とした条件の下ではその効果が確認された。いずれにせよ, 本問題については, 適応型テスト方式が相対的には良く機能することが確認されたが, 基準にとった全項目テスト自体が十分な精度を持つとは言えず, さらに項目プールの整

備を行なってから, 再び適応型テスト方式の有効性について検討する必要がある。

表13 「図形・記号等系列の記憶問題」各項目の項目パラメタ推定値

項目番号	識別力	困難度
1	501	.517
2	502	1.265
3	503	.591
4	504	.724
5	505	.476
6	506	.426
7	507	.594
8	508	1.075
9	509	.599
10	510	.490
11	511	.428
12	512	.854
13	513	.551
14	514	.480
15	515	.513
16	516	.692
17	517	.585
18	518	.954
19	519	.666
20	520	.492
21	521	1.040
22	522	.872
23	523	.539
24	524	.928
25	525	.437
26	526	.991
27	527	.809
28	528	.733
29	529	.567
30	530	.640
31	531	.970
32	532	.680
33	533	.552
34	534	.686
35	535	.627
36	536	.981
37	537	1.084
38	538	.577
39	539	.496
40	540	.662

5. 「図形・記号等系列の記憶問題」の適応型テスト化

5. 1. 問題の概要

本問題は、順唱課題と逆唱課題の二つの部分から構成される。受験者は、画面中に順次呈示された複数の図形・記号等（いずれも呈示時間は1.0秒）について、その系列を呈示順に記憶し、順唱課題では図形・記号等の系列をそのままの順序で再生し、逆唱課題では図形・記号等の系列を逆順で再生して解答する。図形・記号等には具体的には、数字、色、時刻表記号、駅名、列車種別、の5種類を用いる。呈示する図形・記号等の数を、順唱課題では4、5、6、7と、逆唱課題では3、4、5、6と変えることによって項目の困難度を予め変化させてある。項目数は、順唱課題が20、逆唱課題が20の、全部で40項目から構成され、各項目のパラメタ値は表13に示す通りである。詳しくは、野口（1993b, 1996）を参照されたい。

5. 2. 適応型テストの構成

本問題では、項目が数字・色・時刻表記号・駅名・列車種別、の5種類の材料から構成されており、適応型テスト化するにあたり、ひとりの受験者について呈示される項目の材料の種類が偏らないように配慮する必要がある。既に、野口（1993b）で本問題に含まれる項目の1次元性が確認されているが、材料と受験者との間に何らかの相性が存在する可能性を全く否定することもできないので、全ての材料にわたって解答させるように配慮する必要がある。このため、本問題については「項目固定多段階適応型テスト方式」を採用した。

項目の層状ネットワークは1系列あたりに含まれる図形・記号等の数および 図形・記号等の種類に基づいて、

表14のような層状ネットワークを設定した。表中の数字は項目番号を表わし、水準は下の層が易しく、上の層が難しくなるように配置されている。

具体的には以下のような手順で検査が進行する。すなわち、

- ①順唱課題について、「数字」の4項目を下から順に実施する。
- ②次に、「色」に移るが、この時、先ず「数字」で正答した項目の中で一番上の水準のものと同じ水準にある項目を実施する。この項目に正答ならば一つ上の水準の、誤答ならば一つ下の水準の項目を次に実施する。さらに次の項目についても、基本的にはこの手順を繰り返すが、一つ上（あるいは下）の項目が既に実施済の場合が出てくる。その場合は、「色」項目を終了して「時刻表」項目に移る。また、一番上の水準で正答、または、一番下の水準で誤答の場合も、「色」項目を終了して、「時刻表」項目に移る。
- ③「時刻表」項目は、「色」項目の最後の項目が、正答ならばそれより一つ上の水準の、誤答ならばそれより一つ下の水準の項目から実施する。その後の手順は②の「色」項目と同様にする。
- ④さらに、「駅名」、「列車種別」項目も同様に実施する。「列車種別」項目で、一つ上（あるいは下）の項目が既に実施済の場合や、一番上の水準で正答、または、一番下の水準で誤答の場合には、順唱課題を終了する。
- ⑤逆唱課題についても、「数字」の4項目を下から順に実施する。そして、以下②から④の手順を繰り返す。
- ⑥必要な情報を記録して、本検査を終了する。

テスト実施中には、実施した項目の番号、受験者の解答の正誤、反応時間が逐次記録され、最終項目実施後には、実施項目数を順唱課題・逆唱課題別およびそれらの合計、推定尺度値、テスト情報量が計算され記録される。

表14 「図形・記号等系列の記憶問題」の項目ネットワーク

■ 順唱課題						
	記号数	数字	色	時刻表	駅名	列車種別
水4	7	504	508	512	516	520
3	6	503	507	511	515	519
2	5	502	506	510	514	518
準1	4	501	505	509	513	517
■ 逆唱課題						
	記号数	数字	色	時刻表	駅名	列車種別
水4	6	524	528	532	536	540
3	5	523	527	531	535	539
2	4	522	526	530	534	538
準1	3	521	525	529	533	537

5. 3. 方法

テストの実施：本問題についても、受験者に対して実際に適応型テスト方式を実施するのではなく、全40項目に対する項目反応パターンが既に得られている受験者のデータを利用する。すなわち、各項目に対する解答および反応時間については既に得られた結果をそのまま用いるが、各項目の呈示順序については適応型テストを受験した状況をシミュレートする。

受験者：受験者の項目反応には、1996年6月に実施したA大学教育学部の3年生28名分のものを用いた。

テスト終了基準：本問題では、項目ネットワークに従って進行し、最終的に次に実施すべき項目が尽きた段階でテストを終了する。

テストの評価：個々の結果を評価するための指標として、「空間図形の推理問題」および「図形・記号系列の作成問題」と同じく、項目数比、推定値差、情報量比、テスト時間比を用いる。

5. 4. 適応型テストの効果

適応型テスト方式のシミュレーション結果については、図3および図4に受験者#13および#28の進行状況を、表15に各受験者の実施結果を全受験者分について、表16にそれを要約したものを示した。

受験者#13：図3に示すように、順唱課題の「数字」4項目(501, 502, 503, 504)に解答し、それら全てに正答したため、次の「色」で最も難しい水準4の項目(508)に進むが、これに誤答し、1つ下の水準3の項目(507)に正答し、「時刻表記号」で最も難しい水準4の項目(512)に進むが誤答、1つ下の水準3の項目(511)に正答し、「駅名」の水準4の項目(516)に進むが誤答、1つ下の水準3の項目(515)に正答し、「列車種別」の水準4の項目(520)に進み正答、逆唱課題に進み、「数字」の4項目(521, 522, 523, 524)中、水準1, 2, 4の3項目に正答し、「色」の水準4の項目(528)に進み正答、「時刻表記号」の水準4の項目(532)に進み誤答、水準3, 2の項目(531, 530)に連続して誤答、水準1の項目(529)に正答して、「駅名」の水準2の項目(534)に進み正答、水準3の項目に正答(535)、水準4の項目(536)に誤答、「列車種別」の水準3の項目(539)に進み正答、水準4の項目(540)に正答してテストが終了した。最終的に25項目に解答して、17項目に正答、推定尺度値は0.30、テスト情報量は2.30、そして総解答時間は319.20秒であった。この受験者に対して全ての項目を順に実施した場合の結果は、推定尺度値が0.28、テスト情報量が2.79で総解答時間が449.02秒であった。全項目

テストと比べると、項目数比が0.63、推定尺度値の差の絶対値が0.02、テスト情報量の比が0.82、総解答時間比が0.71であった。

受験者#28：図4に示すように、順唱課題の「数字」4項目(501, 502, 503, 504)に解答し、そのうち水準1と2の2項目に正答したため、次の「色」で水準2の項目(506)に進むが、これに誤答し、1つ下の水準1の項目(505)にも誤答したため、「時刻表記号」では最も易しい水準1の項目(509)に進むが誤答、「駅名」の水準1の項目(513)に進み正答、1つ上の水準2の項目(514)にも正答し、さらに水準3の項目(515)に進む。これに誤答して、「列車種別」の水準2の項目(518)に進み正答、1つ上の水準3の項目(519)に誤答、逆唱課題に進み、「数字」の4項目(521, 522, 523, 524)中、水準1, 2, 3の3項目に正答し、「色」の水準3の項目(527)に進み誤答、1つ下の水準2の項目(526)に誤答、さらに1つ下の水準1の項目(525)に正答、「時刻表記号」の水準2の項目(530)に進み正答、水準3の項目(531)に誤答、「駅名」の水準2の項目(534)に進み正答、水準3の項目(535)に誤答、「列車種別」の水準2の項目(538)に進み誤答、水準1の項目(537)に誤答してテストが終了した。最終的に25項目に解答して、11項目に正答、推定尺度値は-1.56、テスト情報量は2.68、そして総解答時間は308.40秒であった。この受験者に対して全ての項目を順に実施した場合の結果は、推定尺度値が-1.65、テスト情報量が2.95で総解答時間が440.67秒であった。全項目テストと比べると、項目数比が0.63、推定尺度値の差の絶対値が0.09、テスト情報量の比が0.91、総解答時間比が0.70であった。

全ての受験者について各受験者毎に、全ての項目を順に実施した場合(全項目テストと表示)の正答項目数、推定尺度値、テスト情報量、総解答時間と適応型テスト方式をシミュレートした場合の実施項目数、正答項目数、推定尺度値、テスト情報量、総解答時間、そして、両テスト間の実施項目数比(適応型/全項目、以下同じ)、推定尺度値の差の絶対値(適応型テスト-全項目テスト)、テスト情報量比、総解答時間比について表15に示した。ここで、テスト情報量は当該受験者の推定尺度値でそのテストが示す情報量を、総解答時間は各項目に対する解答時間の総和を表わしている。また、表15の結果を要約したものが表16である。

実施項目数については、全項目テストでは当然のことながら全受験者が40項目であるのに対して、適応型テストでは、平均24.7、標準偏差3.15である。受験者個人毎に求めた実施項目数比については、平均0.62、標準偏差0.08である。いずれの受験者についてもおよそ60%程度

識別性検査 A-1001の「関係判断力・応用力」領域および「記憶」領域の適応型テスト化の試み

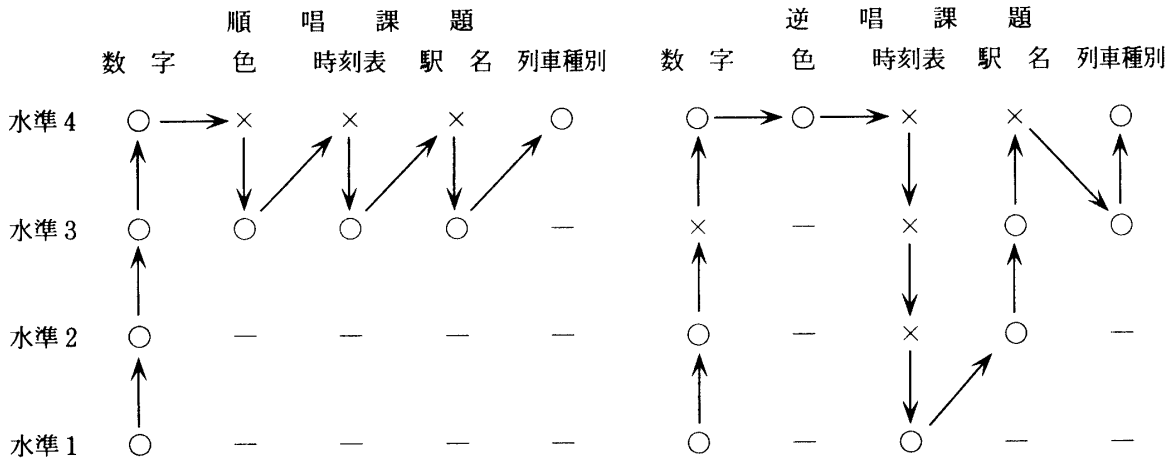


図3 受験者#13の適応型テスト進行状況

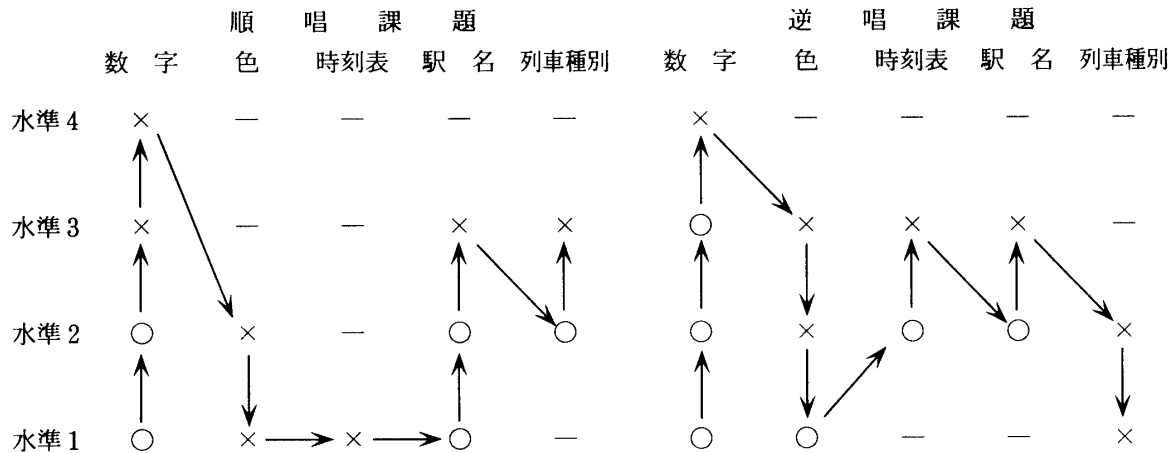


図4 受験者#28の適応型テスト進行状況

の項目を適応型テストで実施しており、実施項目数は40%減らすことができた。

正答項目数については、全項目テストでは平均26.2、標準偏差5.54であるのに対して、適応型テストでは、平均15.9、標準偏差1.92である。実施項目数が両テスト間で異なるので直接比較することはできないが、各テストの実施項目数を考慮しても適応型テストの正答項目数の標準偏差が全項目テストに比べて小さくなっている。このことは、受験者の能力水準の違いにもかかわらず受験者間で正答した項目数に大きな違いのないことを示しており、言い換えると、受験者間で解答した項目の困難度水準が異なるにもかかわらず正答項目数に違いがないということで、適応型テスト方式による受験者への項目の割り当てが適切に行なえたことを示している。

推定尺度値については、全項目テストでは平均0.06、標準偏差0.89であるのに対して、適応型テストでは平均0.07、標準偏差0.96であり、両テストの推定尺度値間の

相関は0.947であった。推定尺度値の差の絶対値については、平均0.24、標準偏差0.27であり、最大値が0.62、最小値が0.00、そして、差の絶対値が0.1以下の受験者が28名中10名であった。また平均2乗誤差は0.093である。基準が各受験者の推定尺度値であり、真の尺度値ではないという限界はあるが、真の尺度値が平均0.0、標準偏差1.0の分布を示す集団で個人毎の推定尺度値の分布の標準偏差が0.3（分散に直すと、0.09）である場合に、そのテストの信頼性係数がおよそ0.9になることに照らして評価すると、受験者全体として実施項目数の減少が尺度値の推定に特に望ましくない影響を与えているとは考えられない。

テスト情報量については、全項目テストでは平均2.77、標準偏差0.29であるのに対して、適応型テストでは平均2.29、標準偏差0.34であった。受験者個人毎に求めたテスト情報量比については、平均0.82、標準偏差0.06である。適応型テスト化によるテスト情報量の減少は避けら

表15 「図形・記号等系列の記憶問題」の適応型テスト化の効果

受験者番号	記憶問題：全項目テスト					記憶問題：適応型テスト					適応型テストの効果			
	項目数	正答数	推定値	情報量	総解答時間	項目数	正答数	推定値	情報量	総解答時間	項目比	推定値差	情報量比	解答時間比
1	40	28	0.13	2.84	391.89	24	18	0.72	2.11	224.67	0.60	0.59	0.74	0.57
2	40	31	0.65	2.65	411.59	23	16	0.28	2.13	244.33	0.57	0.37	0.80	0.59
3	40	33	0.89	2.55	415.91	21	16	0.94	2.13	241.90	0.52	0.05	0.84	0.58
4	40	28	0.20	2.82	402.11	25	17	0.17	2.28	273.84	0.63	0.03	0.81	0.68
5	40	34	1.27	2.37	367.00	23	17	0.75	2.04	247.72	0.57	0.52	0.86	0.67
6	40	25	-0.13	2.92	503.34	22	14	0.02	2.20	332.03	0.55	0.15	0.75	0.66
7	40	23	-0.37	2.97	400.27	25	16	-0.08	2.33	265.97	0.63	0.29	0.78	0.66
8	40	23	-0.38	2.97	354.72	24	12	-1.00	2.46	243.18	0.60	0.62	0.83	0.69
9	40	17	-1.38	3.00	631.73	28	14	-1.38	2.57	460.77	0.70	0.00	0.86	0.73
10	40	27	0.12	2.84	628.90	31	20	0.04	2.66	500.22	0.77	0.08	0.94	0.80
11	40	33	1.12	2.44	310.34	22	17	1.06	2.03	190.56	0.55	0.06	0.83	0.61
12	40	18	-1.21	3.01	411.21	29	16	-1.01	2.69	291.81	0.72	0.20	0.89	0.71
13	40	29	0.28	2.79	449.02	25	17	0.30	2.30	319.20	0.63	0.02	0.82	0.71
14	40	25	-0.23	2.94	529.64	26	16	-0.38	2.42	354.25	0.65	0.15	0.82	0.67
15	40	29	0.45	2.73	372.38	24	17	0.85	2.13	245.02	0.60	0.40	0.78	0.66
16	40	26	0.08	2.86	447.19	23	15	0.46	2.27	293.57	0.57	0.38	0.79	0.66
17	40	26	0.13	2.84	493.96	26	16	-0.13	2.34	353.42	0.65	0.26	0.82	0.72
18	40	27	0.16	2.83	483.15	26	16	0.09	2.40	366.59	0.65	0.07	0.85	0.76
19	40	30	0.65	2.65	446.85	23	18	1.04	1.96	279.46	0.57	0.39	0.74	0.63
20	40	26	0.18	2.83	445.95	26	18	0.58	2.36	333.56	0.65	0.40	0.83	0.75
21	40	32	1.16	2.42	320.44	19	15	1.21	1.73	160.96	0.47	0.05	0.71	0.50
22	40	21	-0.66	3.01	429.20	30	16	-0.88	2.71	304.53	0.75	0.22	0.90	0.71
23	40	17	-1.32	3.01	322.56	27	15	-1.12	2.57	206.50	0.67	0.20	0.85	0.64
24	40	37	2.41	1.63	357.51	16	15	2.72	1.12	169.16	0.40	0.31	0.69	0.47
25	40	23	-0.67	3.01	390.22	26	13	-1.04	2.61	248.86	0.65	0.37	0.87	0.64
26	40	29	0.29	2.79	443.63	26	18	0.30	2.39	325.73	0.65	0.01	0.86	0.73
27	40	23	-0.42	2.97	404.14	27	15	-0.93	2.50	268.71	0.67	0.51	0.84	0.66
28	40	14	-1.65	2.95	440.67	25	11	-1.56	2.68	308.40	0.63	0.09	0.91	0.70

表16 「図形・記号等系列の記憶問題」の適応型テスト化の効果(要約)

	全項目テスト		適応型テスト	
	平均	標準偏差	平均	標準偏差
項目数	40.0	0.00	24.7	3.15
正答数	26.2	5.54	15.9	1.92
推定尺度値	0.06	0.89	0.07	0.96
テスト情報量	2.77	0.29	2.29	0.34
総解答時間(秒)	428.77	78.47	287.68	76.85
	平均	標準偏差		
項目数比	0.62	0.08		
推定尺度値差	0.24	0.27		
テスト情報量比	0.82	0.06		
総解答時間比	0.66	0.07		
平均2乗誤差	0.093			

れないが、それでも全項目テストのおよそ80%の水準を保ち、しかも、受験者毎に実施項目が異なるにもかかわらずテスト情報量比の標準偏差が極めて小さく、適応型テスト化することによって特に測定精度の面から不利益を被る受験者はいないことが示されている。

総解答時間（秒）については、全項目テストでは平均428.77、標準偏差78.47であるのに対して、適応型テストでは平均287.68、標準偏差76.85であった。受験者個人毎に求めた総解答時間比については、平均0.66、標準偏差0.07である。総解答時間の減少効果に受験者間の違いはほとんど見られず、およそ70%弱に短縮している。

この適応型テスト方式は全項目テストに比べて、実施項目数が60%程度に押さえられ、総解答時間がおよそ70%弱に短縮できるにもかかわらず、テスト情報量が80%程度の水準を保ち、推定尺度値にも特に望ましくない影響が出ている様子の見られないことから、実用水準で十分に機能するものと思われる。

6. まとめ

「空間図形の推理問題」では、項目可変型の適応型テスト方式を適用したところ、実施項目数基準の場合、およそ半分の項目数でテスト情報量が全項目テストの80%を超え、テスト情報量基準の場合、約60%の時間および項目数でテスト情報量が全項目テストの90%を超えるなど、いずれの場合も適応型テスト化の有効性が確認された。中でも、受験者間での測定精度のバラツキが最も小さいテスト情報量基準が実際のテスト場面でも有効に機能する可能性が高いと言える。

「図形・記号系列の作成問題」についても、項目可変型の適応型テスト方式を適用したところ、実施項目数基準の場合、およそ半分の項目数でテスト情報量が全項目テストの90%を超え、テスト情報量基準の場合、約60%の時間および項目数でテスト情報量が全項目テストの90%を超えるなど、「空間図形の推理問題」と同様にいずれの場合も適応型テスト化の有効性が確認された。ただし、本問題の場合は、評価の基準となる全項目テストのテスト情報量自体が十分高い値を示しているとは言えず、そのため、十分な評価を下すことができないが、項目プールを所与とした条件の下ではその効果が確認された。すなわち、本問題については、適応型テスト方式が相対的には良く機能することが確認されたが、基準にとった全項目テスト自体が十分な精度を持つとは言えず、従って、項目プールそのものが十分に整備されたものとは言えず、さらに項目プールの整備を行ってから、再び適応型テスト方式の有効性について検討する必要がある。

項目可変型の適応型テスト方式に共通して検討すべき

問題点もある。

第一に、直前の推定値との差をテスト終了基準とすることには問題がありそうである。これは、その時点における推定値と直前の推定値との差の絶対値が設定された基準内に入ったとしても、そこでテストを終了させずに続行させた場合に基準値を超える推定値の変化の見られるケースがあるからである。たまたま局所的に推定値の変化が小さい場合テストを終了させてしまうことになる。第二に、1項目あたりの解答時間を制限することについても検討する必要がある。本研究では受験者が「次に進む」ことをコンピュータに対して反応しない限り次の項目を呈示することのないようにしていた（「図形・記号系列の作成問題」について、解答時間が20秒を超える場合は、例え正しい解答であっても誤答として取り扱ったが、画面に呈示された項目はそのままにしていた）が、テスト時間基準を採用する場合には基準として設定された時間内に実施できる項目数がある程度確保する必要がある。従って、1項目あたりの解答時間を制限し、基準時間内に実施できる項目数を増やすなどやり方そのものについて今少し検討する必要がある。

「図形・記号等系列の記憶問題」については項目固定型の適応型テスト方式を採用したところ、実施項目数は40%減らすことができ、受験者間で解答した項目の困難度水準が異なるにもかかわらず正答項目数に違いがなく、受験者への項目の割り当てが適切に行なえた。また、テスト情報量は、全項目テストのおよそ80%の水準を保ち、総解答時間（秒）は、受験者ほぼ全員について、およそ70%弱に短縮した。これらのことから、本問題について適応型テスト方式は有効であり、実用水準で十分に機能するものと思われる。

これらの結果を解釈するにあたって共通する問題点は、それぞれの結果を評価するにあたり、全項目テストのテスト情報量および全項目テストに基づく推定尺度値を用いているということである。前者については、「図形・記号等系列の作成問題」のように全項目テスト自体が必ずしも十分な精度を示しているとは言えないような場合には、適応型テスト方式の効果についても相対的な評価にとどまり、直接実用水準に結びつく評価とはならないし、後者については、あくまで全項目テストでの推定値であって、それ自身が真の尺度値とは異なっており、評価基準として十分な条件を満たすか否かは全項目テストの特性に依存するということである。

このような解釈上配慮しなければならない問題点があるが、「空間図形の推理問題」と「図形・記号等系列の記憶問題」については、受験者の疲労効果などにも配慮するならば、少なくとも全項目テストの代替テストとし

て実用に耐えるものと判断できる。

本研究の結果を踏まえて、

- ① 今回は、受験者ひとりひとりについてテスト実施過程を追跡し検討する目的から受験者集団をある大学のひとつのクラスに限定したが、受験者数が28名と少なく、今後はさらに受験者の範囲を広げて多人数を用いて結果を検討する。
 - ② さらに、シミュレーションによる検討だけではなく、実際に適応型テスト方式で受験者集団に実施し、受験者がどのような反応を示すかについて、テスト実施過程を観察したり、アンケート調査を実施する。
 - ③ 「図形・記号系列の作成問題」については項目プールをさらに整備する。
- などが今後の課題として残された。

文 献

- 野口 裕之 1991 項目反応理論に基づくテストの作成法, 芝 祐順 編著 項目反応理論——基礎と応用——, 東京大学出版会, 51-86.
- 野口 裕之 1993a コンピュータを用いた適応型知能検査の開発 (1), 東日本旅客鉄道安全研究所委託研究報告書.
- 野口 裕之 1993b パーソナル・コンピュータをベ-

スとする識別性検査 A-1001用項目プールの作成, 名古屋大学教育学部紀要——教育心理学科——, 40, 141-158.

野口 裕之 1994a コンピュータを用いた適応型知能検査の開発 (2), 東日本旅客鉄道安全研究所委託研究報告書.

野口 裕之 1994b パーソナル・コンピュータをベースにした識別性検査に対する受験者の反応, 名古屋大学教育学部紀要——教育心理学科——, 41, 25-38.

野口 裕之 1995 識別性検査 A-1001の「知覚の速さ・正確さ」領域の IRT 尺度化, 名古屋大学教育学部紀要——教育心理学科——, 42, 59-71.

野口 裕之 1996 識別性検査 A-1001の「関係判断力・応用力」領域および「記憶」領域の IRT 尺度化, 名古屋大学教育学部紀要——教育心理学科——, 43, 37-49.

附 記

本研究を実施するにあたって、ご協力を賜りました受験者の皆様方、そして色々なご配慮をいただきました岡山大学教育学部の 藤森 進 助教授に深く感謝申し上げます。

(1997年9月16日 受稿)

ABSTRACT

An Investigation of the Measurement Efficiency of Computerized Adaptive Version of Adult Intelligence Test : A-1001

NOGUCHI Hiroyuki

The purpose of this paper is to study the applicability of adaptive testing procedure for Spatial reasoning problems (30 items), Figure series completion problems (30 items) and Memory problems (40 items) in computer based adult intelligence test : A-1001. Variable branching procedure was used for the former two problems and fixed branching procedure was used for the latest problem. This study was “real data” simulation ones using the responses of 28 examinees to the 30, 30 and 40 items of A-1001 respectively. The adaptive testing procedure was simulated on item responses of the examinees who had taken conventional version of A-1001. All of 28 examinees’ testing process were examined in detail.

It was shown that adaptive version was effective for all of three types of problems in A-1001, for example, adaptive version was shown to be shorter than conventional version with little or no loss in the amount of test information. But results of this study are limited to simulational testing situation and do not include an evaluation of the possible psychological or motivational effects of adapting testing version.

Key words : computer based test, adaptive testing, item response theory, adult intelligence test, computer simulation study