

# 一般能力検査(SAT)の追跡研究： その1 変遷と現状 (1961～1977)

原 一 雄  
石 本 菅 生

はしがき

本研究は、国際基督教大学教育研究所 1976 年および 1977 年度研究課題「一般能力検査(SAT)の追跡研究」の成果を基に、現行のSATの特性を検討し、今後の改訂の策定に示唆を求めようとするものである。

## 1. テストの目的

国際基督教大学教養学部1978年度入学試験要領(p.10)には、本学の入学者選抜に用いられる一般能力検査(Scholarstic Aptitude Test, 以下SATと略称)について次のように述べている。

「一般能力検査はICU独特といってもよいもので、この考査は、受験生の知的能力、とくに言語、数および記号を媒介として思考する能力の程度をはかるためのものである。……出題の形式はひじょうに広範囲にわたり、反意語、同意語、ことわざ、三段論法、図形、空間関係、比例、数列から英語にいたるまでと変化にとんでいる。」

このように、SATは本学を受験する高校生たちの基礎的学習能力(適性)の一部を、語意の理解、語意の関係把握(同意・反意)、概念の関係把握、論理(内包・外延)、数的ないし図形的推理などの問題によって測定し、入学後の学力を予測することによって、入学者選抜の重要な資料として参考に供しようとするものである。特にSATが他の人文・社会・自然諸科学領域ごとの学習能力検査と異なる点は、後者が長文の資料を読ませた後に内容の把握力、理解力、応用力など高次の分析的創造的思考能力を時間をかけて試すのに対し、SATでは80分間に100問の類推問題を与

えて、敏速にして正確な判断力と推理力を測ろうと試みている。すなわち、ここでは持久力よりも瞬発的な、ただし冷静にしてかつ柔軟な思考能力を要求するものであると云ってもよいであろう。

## 2. テストの起源と改訂の経過

国際基督教大学教育研究所では開所以来、本学教養学部入試委員会の依頼によって教養学部入学者選抜のための一般能力テストを研究してきたが、1950年後半に至って、それまで実施してきた文部省式進学適性検査に代る新たな検査を、故岡部弥太郎教授と Maurice E. Troyer 教授の指導の下に開発することになった。そこで、アメリカ合衆国の The Psychological Corporation によって刊行されている W. S. Miller 著の Miller Analogies Test (MAT) に構想を得て、日本語および日本文化に適切な内容のSATを1961年に作成し、爾来18年間、毎年改訂を加えてきたのである。\*

ちなみにMATについて簡単に述べると、これは1926年以来、米国内の大学院入学志願者の選抜のために開発されてきたものであり、現在この種のテストとしては Educational Testing Services の Graduate Record Examination (GRE)と並んで盛んに利用され、妥当性、信頼性、予測性において極めて優れたものとして高く評価されている。Columbia Teachers' College, Harvard, Stanford, Minesota 等、志願者にこのテストを要求している一流大学院はかなりの数にのぼっている。なお、このテストは一般には市販されず、指定した施設と専門家以外の実施は許されていない。わが国では本学を含めて3つの機関が委託を受けてMATを施行しているが、各種テストの販売、保管、実施に厳格な米国においても、最も厳しくその内容の機密が保たれているテストの一つである。

前述のように、本学のSATは最初はMATの形式を借りたものの、し

---

\* この作成・改訂にあたっては、年によって異同はあったが、星野 命教授を責任者として教育研究所員の数名が協力してきた。ただし今回の資料分析は原・石本による。

かし、多種多様な思考形式を組み入れることに長年努力してきた。1963年には、1961年版および1962年版の項目分析を行ない、困難度と妥当性を検定した結果、次のような下位検査の組み合わせによる構成とした。

総数 100 問	{	文科系 53問	{	Aタイプ 21問
		理科系 47問		Bタイプ 26問

文科系とは、その内容として「語意の理解」「語意の関係把握」「概念の関係把握」などを含み、理科系Aタイプは「概念の関係把握」の中でも特に理科的な材料を用いたもの、理科系Bタイプは「数的推理」と「論理」などを含んでいた。

1963年以降、これらを幾つかの下位テスト群に分類し、毎年改訂の度に領域名やその中に含まれる項目数の変更を続けてきた。代表的年度の内分けを表示すると表1のようになる。1971年以後には左程大きな変動がなく、現行の14の下位テストに定着することになった。今回、コンピューターで

表1 SAT下位領域と項目数

変数	下位テスト名 (旧分類)	1964	1966	1968	1971
1	計 算 (B)	11	5	10	10
2	数 列 (A <sub>1</sub> )	10	5	10	8
3	数 的 概 念 (H <sub>1</sub> )	6	6	10	8
4	理 科 的 概 念 (H <sub>3</sub> )	9		10	10
5	図 形 (Fg)			10	10
6	論 理 (C)	10	5	6	6
7	反 意 (E)	7		5	6
8	同 意 (F)	9	10	5	4
9	同 原 理 (G)	11	19	5	6
10	ことば・常識 (H <sub>4</sub> , H <sub>5</sub> )	10	25	10	6
11	文 字 列 (A <sub>2</sub> )			5	10
12	乱 字 乱 文 (S)		5	4	4
13	国 語 (D, I, V)	10	20	5	6
14	社会科学的概念 (H <sub>2</sub> )	7		5	6
計		100	100	100	100

資料を分析するに当り、変数名を表1の左端の番号に統一することにした。以下、本論文に表示または図示される変数名は、すべてこの表に準拠することとする。

### 3. 粗点の分布と測定の信頼性

図1は最近4年間の各年度におけるSATの粗点の分布を示すものである。そのいずれもが、ほぼ左右対称のなだらかな単峰型分布を画き、正規分布曲線に極めて近い。これは本検査を作成し始めて以来の積年の努力と経験の賜物である。すなわち、新しい問題を作るに当り、今までにプールされた類似項目の項目分析の資料を参考として、より高い妥当性と、本学を志望する受験者層に見合った困難度を持つテスト項目の作成に非常な労力を払ってきた結果である。

テスト作成者が先ず最初に望むものは高い妥当性である。しかし、このテストの妥当性を検証する外部基準がない現時点においては、下位領域と各項目の face validity に加えて、その構成的内部一致度を求める他に手だてはない。そこで一見無駄に思われるが、毎年度、使用済みの問題を項

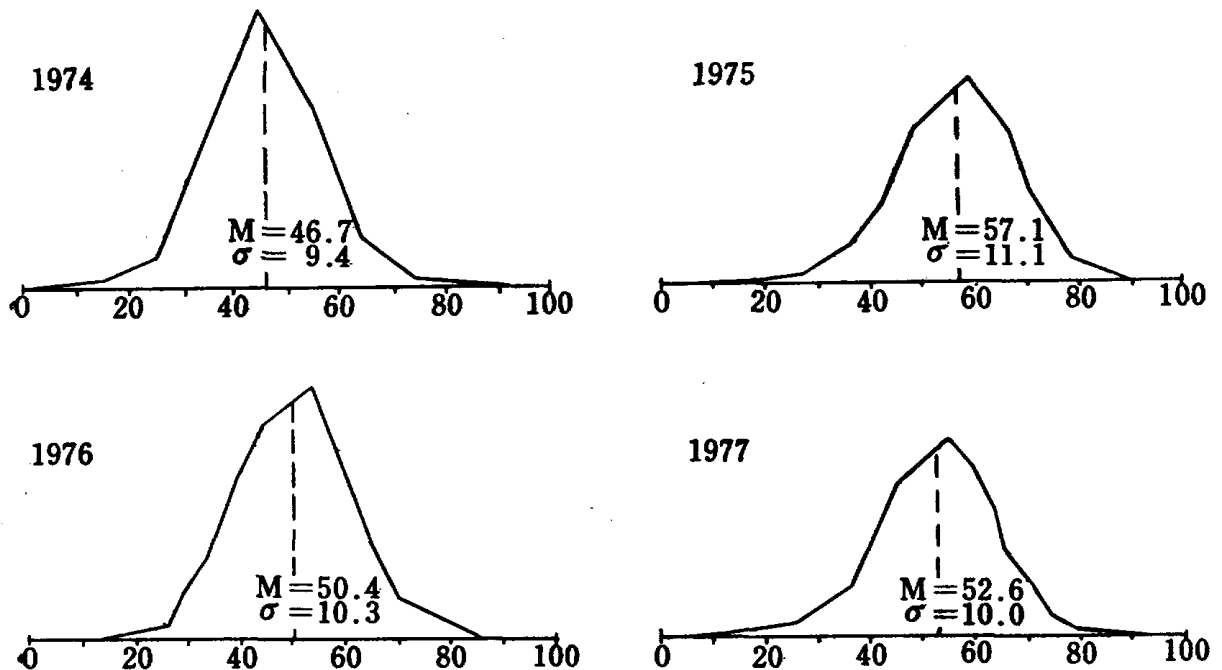


図1 SATの粗点の分布、平均ならびに標準偏差

目分析し、内部妥当性の指標である Flanagan 係数、すなわち、上位27パーセントと下位27パーセントの両群間における弁別度を測定してきた。そして、その論理構造や形式が類似しつつ、且つ Flanagan 係数が、.20以上の弁別力を持つ項目を参考にして、新しい問題の作成を進めてきたのである。

テスト全体として望ましい分布の形は、多分次のようなものであろう。入学者の選抜は高得点者の間で行われるであろうから、弁別を容易にするためには全体として低い得点に skew させたい。しかし偶然性による得点範囲と重なることを極力避けるため、また導入部分において動機づけを高めるためには、容易な項目も多少は用意する必要がある。合計100問、4肢選択の多肢選択法の形式をとるSATにおいては、

$$\text{偶然的得点の平均} : Mc = n \cdot p = 100 \times \frac{1}{4} = 25$$

$$\text{偶然的標準誤差} : \sigma_c = \sqrt{npq} = \sqrt{100 \cdot \frac{1}{4} \cdot \frac{3}{4}} = 4.33$$

$$\text{偶然的得点範囲} : X_{.025} \sim .975 = Mc \pm 1.96\sigma_c = 16 \sim 34$$

(95%の信頼性)

と云うことになる。すなわち、35点以上、100点までに広く分布し、もしその年度の入学者数の総志願者数に対する倍率から予測される足切り点を挟んだ双峰性の分布が得られれば、選抜は非常に容易である。しかし、大勢で、しかも能力の似通った志願者の得点は、多分一ヶ所にかたまり易いであろうから、この最頻値をできる限り足切り点から離したい。これが最頻値を低得点の方向に片寄せたい理由である。

そこで今一度、図1を吟味しよう。全体の平均はいずれも50点前後であり、80点以上の得点が僅少であるところから、これらは受験者にやや難し過ぎたと考えられる。意味のある得点が可能な範囲(35~100)を有効に使用するためには、平均60、標準偏差10あたりが望ましい。上記の観点に立てば、分布の型は1974年度が、また平均と標準偏差については1975年度が最も理想的であったと云えよう。しかも、1976年と1977年度には粗点か

ら標準得点 (Tスコア) への換算がほとんど不要であったと云う利点も無視できない。

入学者選抜試験においては、テストの信頼性が問われることはめったにない。一点の差と云えども及落の判定では重要な差である。しかしながら、評価対象について完璧な妥当性を持つテストがないと同様、信頼性1.00のテストは測定に従事する者の夢であって、現実にもそのような検査はどの科学の分野においても存在しない。そこで、自分の使用する検査用具、すなわちテストの信頼性を知り、測定結果が孕んでいる誤差の巾を考慮して判断を下すと云うことは、良心的科学者が当然持つべき職業的倫理であろう。

SATについては、1978年度の資料から今回初めて信頼性が算出された。受験生の総数3353名から得られたSATの信頼度係数は、Spearman-Brownの修正式を用いた折半法によるものが.813、Kuder-Richardson方式によるものが.811であり、両者は極めて近似した値を示した。

上記の数値を用いて標準測定誤差 (Standard error of measurement,  $\sigma_e$ ) を算定してみよう。もし粗点の標準偏差 ( $\sigma$ ) を10.0と仮定すれば、

$$\sigma_e = \sigma \sqrt{1 - r} = 10 \sqrt{1 - .81} = 4.36$$

である。もし危険率を5パーセントと置けば、

$$1.96\sigma_e = 8.55 \approx 9.00$$

と云う値が得られ、このことはSATのどの得点にも $\pm 9$ の測定誤差を勘案して判断しなければならないということになる。

教育測定に経験のある者は、ここに得られた信頼度係数.81が如何に高い指数であるのかを良く承知しておられよう。それにもかかわらずなお9点前後の誤差が予測されるのであるから、予備調査なしの入試の得点を評価するに当ってはどのような配慮が必要か、今更言を待つまでもない。ましてや多面的評価を目指した複数テストの総合得点を求める場合には、仮に各テストにウエイト付けがなされていないと仮定しても、

$$\sigma_e' = \sqrt{n_1\sigma_{e1}^2 + n_2\sigma_{e2}^2 + \dots}$$

の標準誤差が同時に含まれることを忘れてはなるまい。

#### 4. 因子分析による考察

SATの因子構造を明らかにするために、1974年度以降4年間の入学志願者の資料を用い、まず各年度毎に14の下位テスト間の相関行列を求めた。それらを基にして、次の3つの方法によって因子分析を行った。すなわち、

- (1) 主因子法Ⅰ (対角行列に1.000を挿入)
- (2) 主因子法Ⅱ (対角行列にコミュナリティを挿入)
- (3) Varimax 法

である。

抽出する因子の数は、(1)と(2)の場合には収束の精度を0.0001と定め、また、繰り返しの回数の限度を便宜的に200として計算を実施した。抽出される因子の数は、固有値の近似値にあたる値が負となり計算を進められなくなるまでの繰り返しによって決まるが、(2)の方法による分析では、抽出された因子の数はいずれの資料においても2個であった。

これらの内、(1)の方法は、ある特定の下位テストが他の下位テストとどの因子を共有しているかを個別に検討する場合に便利であり、それぞれの下位テストの問題を作成する人々に多くの示唆を与えることができる。(2)と(3)の方法は、このSATが全体としてどのような主要因子から成り立っているのか、その構造を概観するのに役立つ。そこで今回の研究は、専ら後者の分析方法による結果を基にして行うことにした。

##### 主因子法Ⅱによる分析

結果は図2に年度別に示されている。テストの内容は毎年改訂され、4つの異った志願者群が対象となったにもかかわらず、そこには極めて類似した因子構造の存在することが明らかにされた。すなわち、毎年度、第Ⅰ因子は14のすべての下位テストにほぼ等しく負荷されており、1974年度の下位テスト8を除いて、他のすべての負荷量が.300以上の高い値を示して

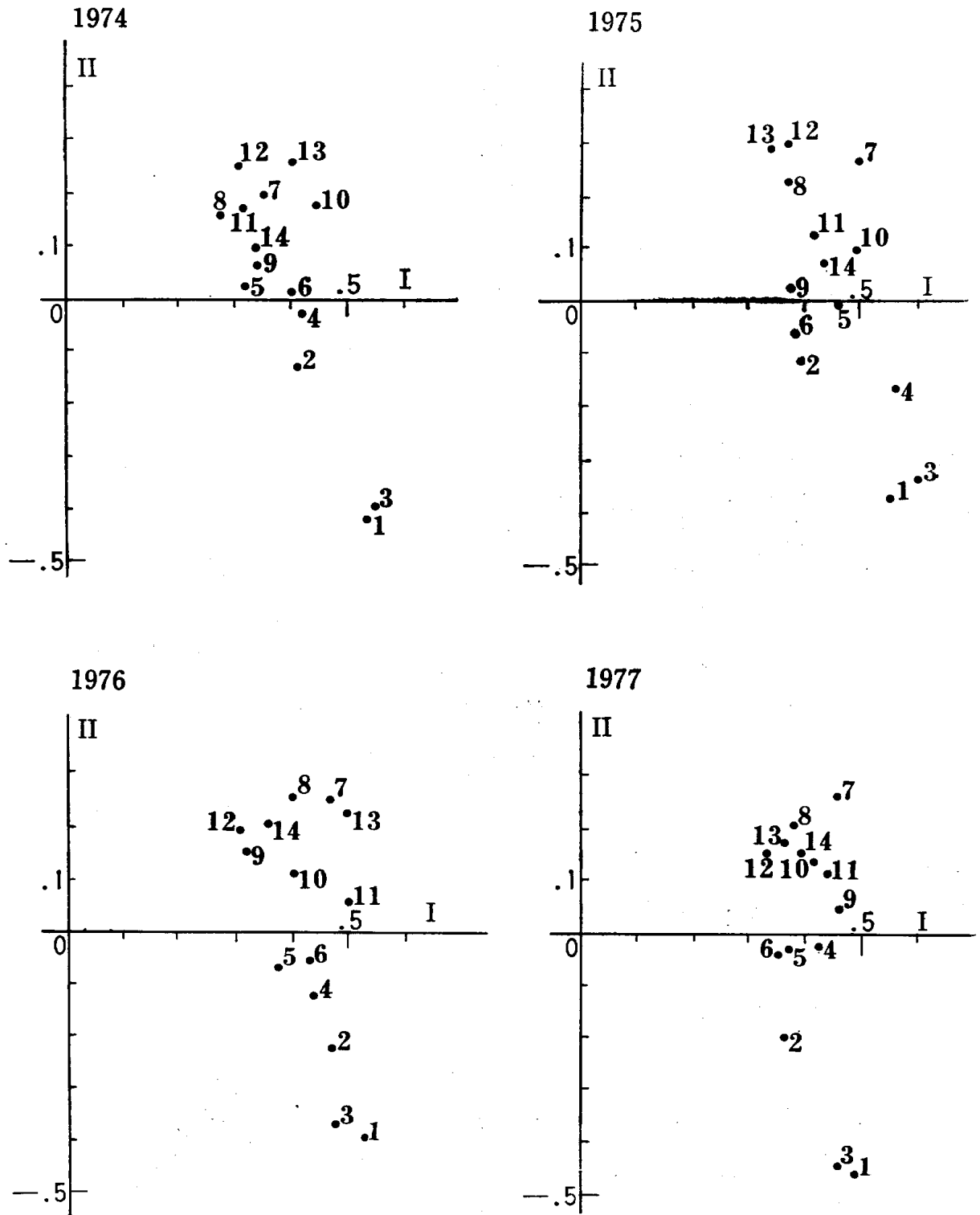


図2 各年度SATの主因子法IIによる因子負荷量  
(数字は表1の変数番号を表わす)



いる。このことから、SATには一般的学習能力ないし適性と称するに相応しい一般因子が包含されているということが明らかに推察される。

第Ⅱ因子は負荷量において第Ⅰ因子にやや劣るものの、これも毎年度、14の下位テストを同じような正、0の近辺、負の3群に分けている。それらの集団に含まれる下位テストの特性から推測すると、第Ⅱ因子は第Ⅰ因子に付随しながら、志願者を更に言語的適性の高い群と非言語的適性の高い群とに二分していると考えられる。すなわち、下位テスト7（反意）、8（同意）、12（乱字乱文）、13（国語）は文科系を得意とする者を、下位テスト1（計算）、2（数列）、3（数的概念）は理科系を得意とする者を弁別するのに役立つであろう。第Ⅱ因子の負荷量が0に近い下位テスト4（理学的概念）、5（図形）、6（論理）、9（同原理）などはいずれも第Ⅰ因子の重要な構成要素であるが、この種の弁別にはあまり寄与をしていないと考えられる。

#### Varimax 法による分析

結果は図3 a～dに示されており、そこから得られる考察は、上に述べた主因子法からの結論とほぼ同じである。すなわち、この方法による回転から得られた第Ⅰ因子は、主因子法の第Ⅱ因子に負の負荷量をもたらした要因に対応しており、毎年度、ほぼ下位テスト1, 3, 2, の順に高い値を示している。第Ⅱ因子は下位テスト7から14までのものであり、主因子法のところで指摘された文科系への適性は、すべてこの中に含まれている。いずれの年度においても、14の下位テストのすべてが Varimax 法の第Ⅰ因子と第Ⅱ因子の作る座標の第1象限内に入ることは、この両者が決して相反する関係にあるものではなく、独立の程度に軽重の差はあっても、互に相補しつつSAT全体、すなわち、主因子法によって抽出された第Ⅰ因子に深く寄与していることを裏付けている。

1975年度以降の資料からは第Ⅲ因子も抽出されており、3ヶ年を通じて

下位テスト2, 5, 6, 11, 12が常に正の負荷量を持つのに対し, 下位テスト3, 4, 9, 14のいずれかは逆に負の方向に負荷されている。すなわち, 数列, 図形, 論理, 文字列, 乱字乱文に対し, 数的概念, 同原理, 社会科学的概念が対比されており, 認知様式の差ないしは課題解決の方略の相違を示していることは, 学習心理学の立場から甚だ興味深い現象と云うことができよう。

1974

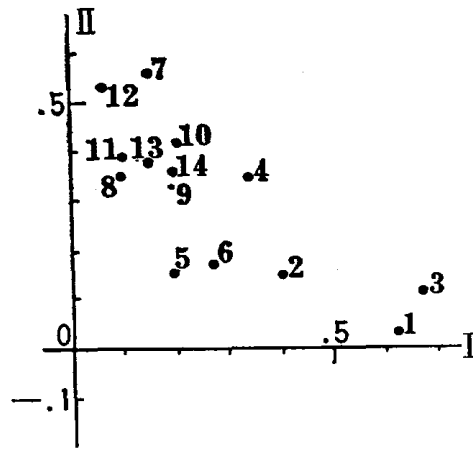


図3a 1974年度SATの Varimax 法による因子負荷量

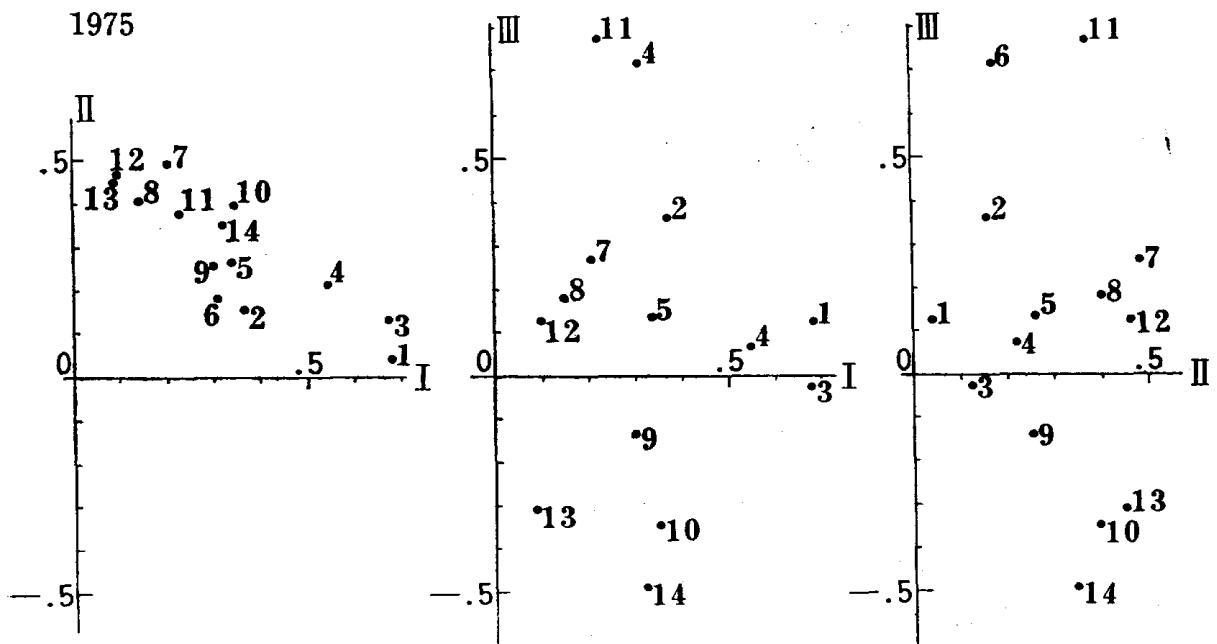


図3b 1975年度SATの Varimax 法による因子負荷量

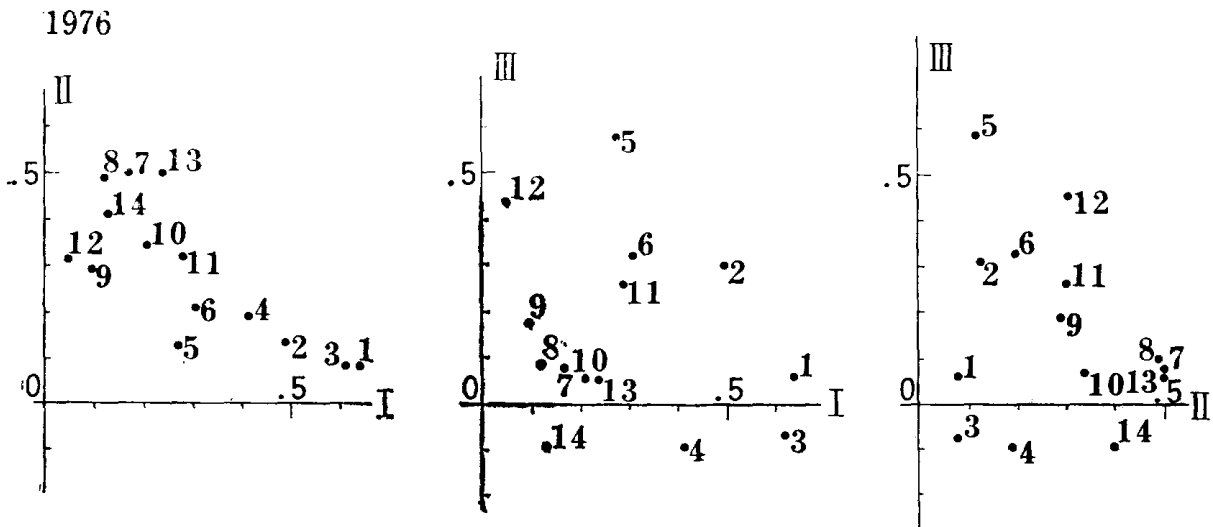


図3c 1976年度SATのVarimax法による因子負荷量

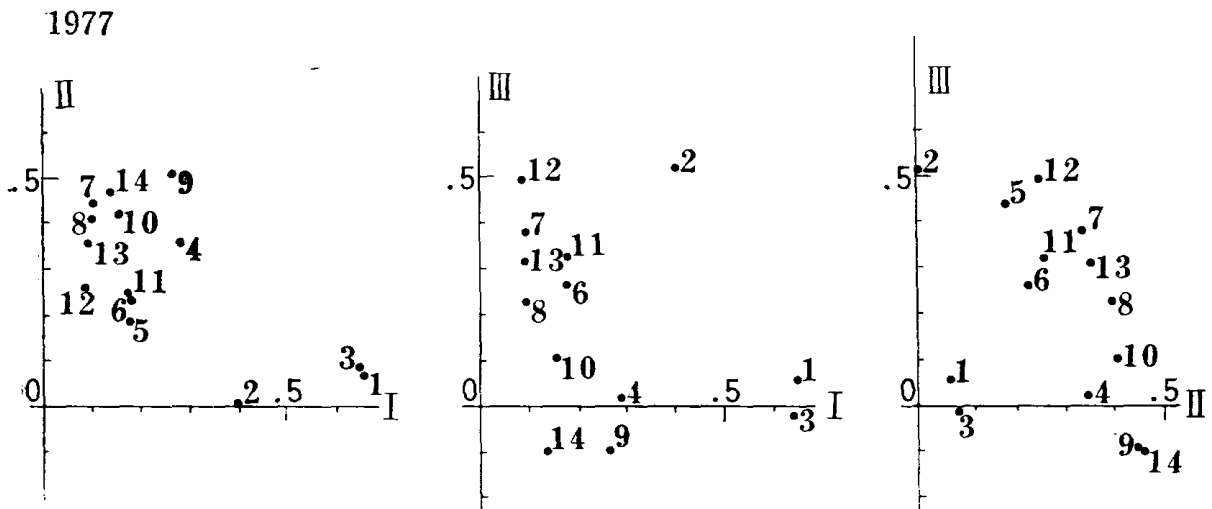


図3d 1977年度SATのVarimax法による因子負荷量

高校調査書，ICU入試，ならびに入学後の学部成績との関係

1974年度入学者311名，1975年度入学者294名，1976年度入学者347名の3群につき，それぞれ，SAT（14領域下位テスト，および総得点），高校調査書の教科成績（国語，社会，数学，理科，外国語，保健体育，美術

表 2 高校調査書成績, SAT成績ならびに入学後の学部成績相互の相関

高校	入 学 部																														
	国	社	数	理	英	体	芸	他	五	全	SAT	FEP	一	般	体	専															
高校	1.00																														
国		0.63	1.00																												
社				0.52	0.54	1.00																									
数					0.52	0.61	0.73	1.00																							
理						0.52	0.48	0.56	0.43	1.00																					
英							0.31	0.28	0.26	0.25	1.00																				
体								0.25	0.22	0.22	0.29	1.00																			
芸									0.25	0.22	0.23	0.29	1.00																		
他										0.27	0.26	0.29	0.21	0.12	1.00																
五											0.81	0.83	0.80	0.83	0.71	0.35	0.28	0.36	1.00												
全												0.81	0.80	0.78	0.80	0.68	0.54	0.39	0.39	0.96	1.00										
入試 SAT													0.00	-0.08	-0.05	-0.06	-0.22	-0.02	0.05	-0.02	-0.09	-0.08	1.00								
学部 FEP														0.23	0.20	0.11	0.01	0.42	0.10	0.02	0.31	0.25	0.24	-0.01	1.00						
一般															0.33	0.35	0.26	0.29	0.23	0.20	0.05	0.40	0.37	0.37	0.09	0.56	1.00				
体																0.19	0.21	0.17	0.23	0.16	0.22	0.15	0.32	0.24	0.26	0.05	0.46	0.54	1.00		
専																	0.33	0.34	0.22	0.24	0.29	0.21	0.10	0.32	0.35	0.36	0.07	0.63	0.69	0.52	1.00



音楽, その他, 5教科平均, 全教科平均), I C U入試 (人文科学・社会科学・自然科学各領域別学習能力, 英語聴解力, 英語読解力, 英語合計), ならびに入学後の教養学部における成績 (Freshman English Program, 一般教育, 体育講義, 専門, ただし1975年度入学者はF E Pと一般教育, 1976年度入学者はF E Pのみ)の計35項目の相関行列を求め, Varimax 法による因子分析を施行した。1974年度入学者の例を示すと, S A T下位テストならびに入試成績を除いた相関行列は表2となり, 全体から15の因子が抽出された。その内, 因子負荷量が.200以上のものを抜萃すると表3のようになる。そこで, これらの表によって, 考察を進めたい。

3学年度を通じて最大の寄与率を示すものは, 高校の5教科平均および全教科平均を軸とし, 入学後の成績, すなわち, F E P, 一般教育, ならびに専修学科の成績に関連する因子である。第2は, 入試の英語考査を軸とし, F E Pと高校の外国語 (英語), ならびに大学の専修科目の成績に関連する因子である。第3は, 主にS A Tに関連する因子であり, 1974年度入学者群では入試の社会科学の学習能力検査と, またここでは示されていないが, 1975年度および1976年度入学者群では自然科学の学習能力検査と因子を共有していることが明らかにされた。

第4番目の因子は, 大学に入学した後の学業成績に影響を及ぼす因子と考えられるが, 高校の成績の中では主要学科目以外のその他と関連があり, 学部成績の中でも体育の負荷が最も重いものである。この因子に関しては, 在学4年間の学業成績が出揃うまで, 追跡調査を続ける要があると共に, 課外の諸活動とも関連させて研究しなければなるまい。

第5番目以下の因子は, 専らS A Tの下位テストの中に見い出されるものであり, それらの解釈はむしろS A Tのみの因子構造の分析でより詳細に検討されるべきであろう。ただし, 1974年度入学者では第8因子として入試の人文科学の学習能力を主要因とし, 僅かばかり社会科学の学習能力にも共通する因子が存在し, また1975年度および1976年度入学者群では, 高校の体育と美術の成績を軸とする因子が見い出されていることは注目に

値しよう。

これらの結果から、ICU入学後の教養学部の成績は、一概に高校の成績のみ、あるいは入試の成績のみから一義的に予測し得るものではなく、入学後に学生が選んだ専門領域、すなわち、専修科目に対する適性と、個人の動機づけ如何によって大いに左右されるものと考えられる。しかしながら、今までの諸研究でも度々指摘されてきたように、今回の調査においてもまた、高校の内申と入試、特に英語の成績は、本学で入学後の学業成績を決定する要因として非常に重要な役割を果しているということが明らかにされたわけである。

おわりに

今回の研究は、本学のSATが入学者選抜試験として評価に必要な条件を、その内容的妥当性、信頼性、ならびに実用性のいずれの点においてもほぼ満足すべき程度に備えていることを明らかにした。従って今後は、ひたすら問題項目の改良と拡充につとめ、なお一層、選抜テストとしての質の向上を図るよう望みたい。

しかしながら、この選抜テストとしての特性は、必ずしも入学後の学業成績の予診ないし診断に必要な要件を同時に兼ね備えていることの保証にはならない。それ故に、かつて入学事務部が行ったように、もしSATの結果を他の入試成績と併せて入学後のガイダンスに利用することを考えるならば、今後更に幾つかの工夫を試みなければならないであろう。例えば、SATの因子構造に従って下位テストの新しい組み合わせを考案し、言語的、数量的、分析的適性を表わす下位得点を求め、それらを入学志願者のプロフィールに記載することが考えられる。もしこれが可能になれば、各領域の学習能力考査の成績を解釈する際に役立つばかりでなく、入学後の専修科目や将来の職業的専門分野の選択にも非常に参考となるであろう。ただし、そのためには、せめて入学者だけについても、外部基準となる他の標準テストを入学後できるだけ早い時期に受けさせ、SATの経験的妥

当性を検証することが是非とも必要となるであろう。

また、専門分野への適性の予診度を高めるためには、卒業前に再度SATを受けさせるとか、あるいは学外の専門家集団に受けてもらうような方法も考えなければならない。この点では特に、在学生のみならず、実社会の諸領域で活躍中の卒業生の協力が望まれる次第である。

最後に、データの分析にはICU計算センターのIBM 370/115を用いたことを記すと共に、本研究にご協力下さった同センターの所員諸氏ならびに今回の研究を助成して下さった大学行政幹部の方々に対して、厚く感謝の意を表したい。



# A Follow-up Study of the ICU Scholarstic Aptitude Test (SAT): I. Its Development and Current Status (1961~1977)

Kazuo Hara  
Sugao Ishimoto

This report is a summary of our joint study under the aforementioned title, conducted at the Institute for Educational Research and Service in ICU during 1976-1978. The study aimed at to self-evaluate our present SAT and to invite criticisms and suggestions from our colleagues for its further improvement.

## 1. Purpose and Characterisitics of SAT.

As being described in the Manual of ICU Admissions Tests, our SAT intends to measure a part of basic learning ability, i. e. scholarstic aptitude, throgh various tests of word and concept comprehensions, understanding of word relations, e. g. synonym and antonym, analogies with mathematical and figural materials, and logical reasonings.

Although the ultimate goal of SAT is same to other learning ability tests for admissions, the former emphasizes more or less a phase of quick judgement and accurate inference, which are supposed to be essential for our clear, flexible and creative thinking in advanced studies.

## 2. Original Saurces and Revising Procedures.

Several members of the IERS were asked to construct a test for the substitution of College Entrance Aptitude Test, which had been administered by the Ministry of Education but became no longer

available in late 1950s. Thus, under the leadership of late Professor Yataro Okabe and Dr. Maurice E. Troyer, and modeling after the Miller Analogies Test of the Psychological Corporation in New York, a new form of SAT has been constructed every year based on the through scrutiny of those questions employed in previous years. Item analysis and factor analysis were two main techniques and the sources of information for revision. The present form consisting of 100 items can be grouped into 14 subtests.

### 3. Distribution of Scores and Reliability of Measurement.

By virtue of item analysis and our effort for standardization, all editions so far constructed have had remarkably similar distributions close to ideal and relatively high reliabilities. Ranges of the standard error of measurement were estimated to assist examiners to evaluate raw scores with necessary caution.

### 4. Factorial Structure.

Correlational matrices calculated from the 14 subtests were factorially analyzed for 4 successive years since 1974. The application of centroid method revealed that all 14 subtests were highly loaded with a general factor. The second factor further differentiated them into verbal and nonverbal aptitudes. Results obtained by varimax method resembled to the above. Interestingly enough, however, the third factor extracted by this method reflected differences of cognitive styles or strategies required for problem solvings.

### 5. Correlations among High School Records, Admissions Test Scores and College Grades.

Correlation Coefficiencies among grades of various subject matters at high school, admissions test scores and grades obtained from the College of Liberal Arts at ICU were factor analyzed by varimax method.

The highest loadings were observed in 5 major subject matters of high school and all college grades. The second factor included high school English, English tests for admissions, Freshman English Program and grades in majoring areas. The third factor related mostly to SAT. The fourth factor seemed to influence college grades, among which physical education had the highest loading.

From these results, it was revealed that the college grades would never be predicted only from high school performances or admissions tests, but students' aptitudes toward their fields of concentration as well as their motivational levels have significant roles to determine their undergraduate academic performances.